John Billingsley · Peter Brett *Editors*

Mechatronics and Machine Vision in Practice 3

Springer

Mechatronics and Machine Vision in Practice 3

John Billingsley · Peter Brett
Editors

Mechatronics and Machine Vision in Practice 3

Springer

Editors
John Billingsley
School of Mechanical and Electrical Engineering
University of Southern Queensland
Darling Heights, Toowoomba, QLD
Australia

Peter Brett
National Centre for Engineering in Agriculture
University of Southern Queensland
Darling Heights, Toowoomba, QLD
Australia

ISBN 978-3-319-76946-2 ISBN 978-3-319-76947-9 (eBook)
https://doi.org/10.1007/978-3-319-76947-9

Library of Congress Control Number: 2018934398

Printed on acid-free paper

This Springer imprint is published by the registered company Springer International Publishing AG part of Springer Nature
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

This book is based on selected papers from the conferences on Mechatronics and Machine Vision in Practice, held in Manila and Toowoomba in 2014 and 2015. Each paper passed a rigorous peer-reviewing process for acceptance.

In keeping with the previous books in this series, each chapter presents the work of researchers who are seeking practical solutions for application in real working environments. Collectively the book is a rich resource embracing topics that range from the industrial sectors of agriculture to mining, sport and medicine.

The field of mechatronics has advanced in step with the great leaps in computational performance and software sophistication. This has enabled microprocessors to exert their influence on the real world through applying intelligent control to physical devices. Sensing is a key factor in this. Different sensing techniques are reported here, including tactile and vision, enabling the control of more complex mechanisms and soft-bodied robots.

Working in the natural environment, the challenge of automatic recognition and the measurement of target parameters stems from variable presentation or behaviour and also variable operating conditions. The examples presented in agriculture and medicine are a good illustration and indicate successful techniques, whether automatically determining crop health and yield, recognising human motion or controlling surgical devices towards a deep tissue target.

In contrast with previous books on mechatronics and machine vision in practice, a significant number of chapters focus on systems designed for human interaction and deciphering human motion. Examples illustrate assistive actuation of hip joints, the augmentation of touch sense in artificial hand prostheses and helping stroke survivors in repetitive motion therapy. Interactive mechatronics and the experience of developing machine interfaces have enabled an examination of how we use mechatronics in the service of training, and even to consider why computer games perhaps appear to capture attention more readily than a human instructor!

Mechatronics continues to be an exciting and developing field. It is now an essential part of our world and living experience. This and the previous books illustrate the journey in developing the use of mechatronics so far. We anticipate

that you will find the chapters here an equal source of inspiration for new devices to solve the challenges of new applications and of course as a resource for teaching and inspiring the new generation of mechatronics engineers.

Toowoomba, QLD, Australia

John Billingsley
Peter Brett

Contents

Part I
Mechatronics Serving Complex Applications in Practice

3D Shape Sensing of Elongated Objects Using Fibre Bragg Gratings

Karsten Hoehn and Andrew Olsson

Abstract Navigating in GPS deprived areas can be a challenging task, especially when the environment doesn't offer sufficient visual clues for stereo camera matching. This paper describes the development of a 3D shape sensor based on optical fibres and the resulting "Sensor Wand". The intension was to develop a Wand that could be easily deployed in small holes like drill holes for mining or geo-technical applications and determine their 3D shape, which is a vital information for Drill and Blast Engineers in mining, or a geo-technical assessment in civil applications. Initially the expected response of the Wand was modelled for a number of configurations. A Wand was then manufactured to verify the simulated results. The mechanical manufacturing proved itself to be significantly more complex than originally anticipated. As a result, the sensor accuracy was below expectation. Improvements to the manufacturing process were identified to increase the performance of the next generation of the Sensor Wand.

Keywords 3D shape sensing · Fibre optic sensing · Fibre bragg grating FBG

1 Introduction

Meaningful navigation requires the knowledge of the current location and pose. In many applications this information can be derived from Global Positioning Systems, Inertial Navigation Units, (stereo) cameras, or a combination of these systems. However, there are applications where it is neither practical nor even

K. Hoehn (✉) · A. Olsson
CSIRO Mineral Resources Business Unit, Queensland Centre for Advanced Technologies, Pullenvale, QLD, Australia
e-mail: Karsten.Hoehn@csiro.au

A. Olsson
e-mail: Andrew.Olsson@csiro.au

J. Billingsley and P. Brett (eds.), *Mechatronics and Machine Vision in Practice 3*,
https://doi.org/10.1007/978-3-319-76947-9_1

possible to utilise these traditional localisation systems. Such examples include medical, underwater and downhole sensing in the oil, gas and mining industries.

The Commonwealth Scientific and Industrial Research Organisation (CSIRO) through its Minerals Resources Business Unit developed a 3D shape sensor based on Fibre Bragg Gratings (FBGs) embedded in optical fibres that can be used in these challenging environments.

2 Sensing Concept

An FBG is a wavelength selective reflector, with the nominal reflected wavelength depending on the grating constant. Additionally, the reflected wavelength is subject to external influences like temperature and strain as they alter the grating constant. An FBG can therefore be looked at as an optical strain gauge. We have incorporated a number of these FBG sensors a deflection sensor wand. Based on the knowledge of the various strains along the length of the wand the 3D shape of the sensor wand can be reconstructed in real-time (Fig. 1).

There are some commercial and pre-commercial systems available to measure the 3D shape using fibre optics, such as

- the Luna Fibre Optic Shape Sensor [1], which has an extremely high spatial resolution as it uses Rayleigh backscatter to determine the strain along the length of the sensor. Because of this sensing principle it has a limited sensor length of 1.5 m and is very expensive; and

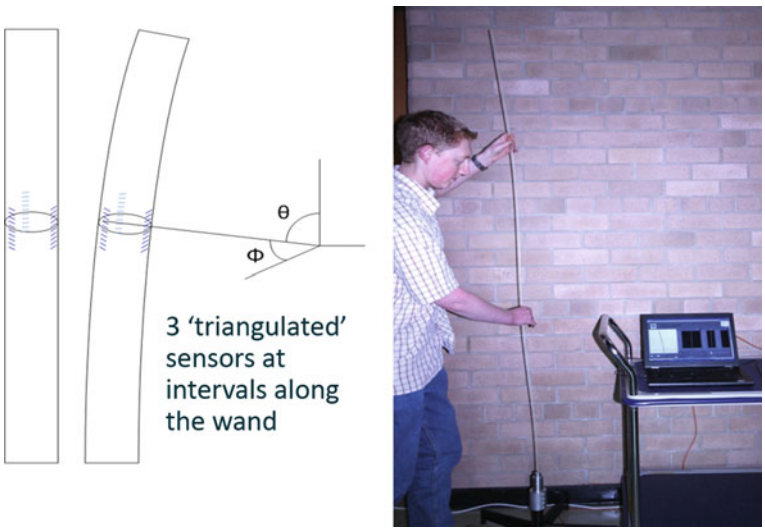


Fig. 1 Principle of "triangulation" and "deflection"

- the Fraunhofer Fibre Optical 3D Shape Sensor (FibreNavi) [2], which requires a highly sophisticated optical setup to burn the FBGs into the fibre.

The CSIRO design uses a simple configuration utilising four optical fibres with FBGs embedded in a linear arrangement. These fibres are available off the shelf and then were embedded in a glass-fibre composite tube.

3 Simulations

3.1 Static Model and Simulations

An initial static model was set up to investigate the sensor performance. It is capable of computing a straight, an arced, or a sinusoidal sensor layout, both in the x and y plane (the depth is given by the z value). Additional measurement noise can be added in the form of normal distributed random error with a configurable standard deviation. An example calculation with blue for the theoretical curve and red for the reconstructed path using 1 pm signal noise is shown on Fig. 2. The plot on the right shows the error versus measurement length or depth.

This figure shows that the reconstruction fits very well initially, but drifts at the "turning points". The reconstruction algorithm assumes a constant curvature between sample points, however, this is not a valid assumption for a sinusoidal or helical path.

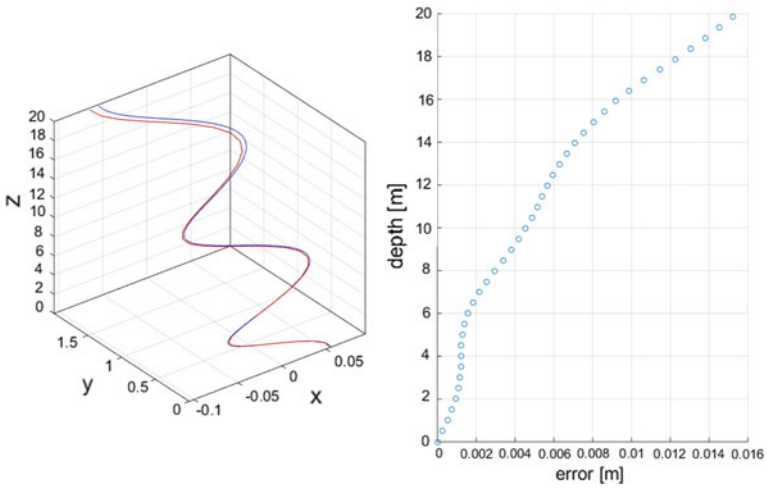


Fig. 2 Simulated shape reconstruction for an arced (R = 100 m) Sensor Wand, which is also buckled into a helix (pitch = 8 m), with 0.5 m spacing between stages and 1 pm normally distributed measurement error

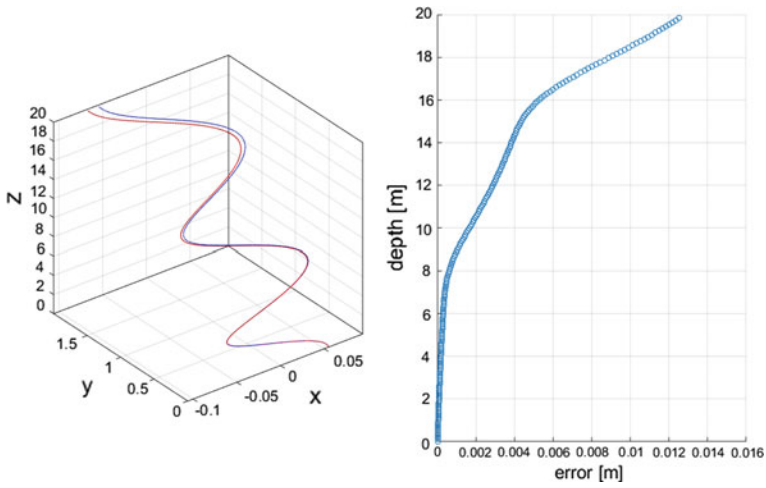


Fig. 3 Simulated shape reconstruction for an arced (R = 100 m) sensor wand, which is also buckled into a helix (pitch = 8 m), with 0.1 m spacing between stages and 1 pm normally distributed measurement error

The same calculation can be performed with a spacing of 0.1 m between the stages and the results are shown in Fig. 3. This finer spacing reduced the total error at the end of the path by 20%.

3.2 Dynamic Model and Simulations

It is worth noting that the above simulations use a static model that can be seen as a "snapshot". In reality the sensor will progress into the hole and so it is possible to measure the defection at a given depth multiple times, which will reduce all random and not systematic errors. Therefore a dynamic model was developed to simulate the progression of the Sensor Wand through the hole and compare the new readings with the old measurements. Averaging of the repeated deflection angle measurements was used to combine the consecutive readings.

A series of images generated by a simulation of a shorter wand going through a deeper hole is shown in Fig. 4. These graphs show the theoretical path section in blue, the reconstructed sample points as crosses and the fitted path as a dotted, red line. The Deflection Sensor length used in this computation is 2 m, the hole depth is 20 m and no signal noise was added. Longer sensor configurations can be computed, and the dimension are only reduced for illustration purposes.

The Sensor Wand measures deflection, which is a relative parameter, therefore the algorithm needs a start point and direction for the absolute 3D shape

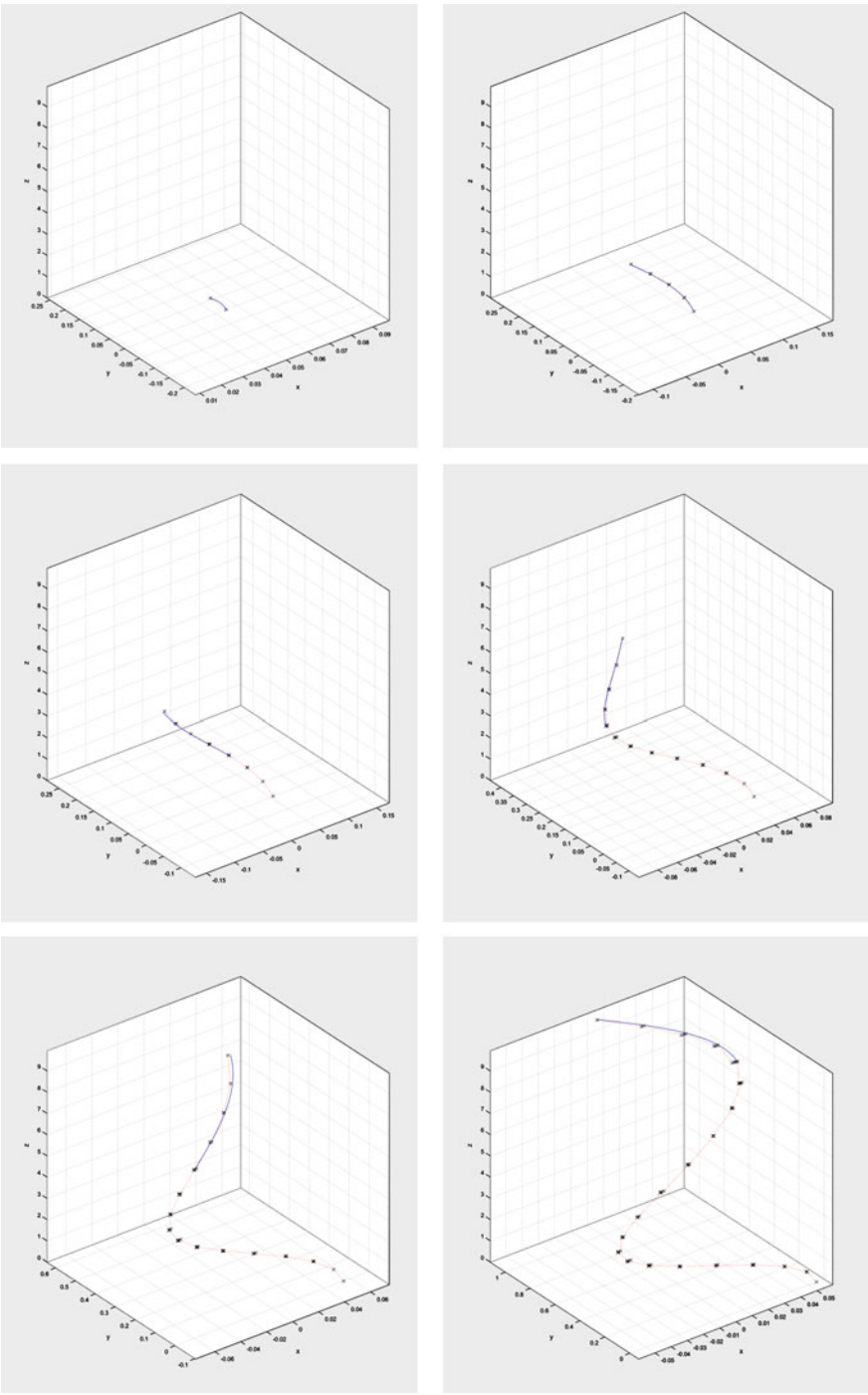


Fig. 4 Progression of the sensor wand through a deeper hole

determination. These values can be assumed to be known at the beginning of the hole. Once fully submerged the algorithm uses its previous estimated path to determine these values.

3.3 Initial Discussion

The Sensor Wand model is designed to simulate the result one would get by inserting the wand into a hole. It computes the reconstructed 3D shape of the hole and the deviation of the theoretical path as a function of the following input parameters:

- Depth of hole
- Theoretical or design hole shape
- Measurement interval
- Measurement noise
- Distance between FBG stages
- Number of FBGs per stage
- Sensor geometry
- Interrogator resolution

The model has been used with various input parameters to identify the influence of particular parameters and their error contribution. We observe that when the interrogator resolution is reduced the error will grow, and that once it reaches 100 pm, the interrogator resolution becomes the dominant error contribution factor. Reducing the Sensor Wand length also increases the error. Reducing the distance between the FBG stages on the other hand reduces the error as demonstrated in Figs. 2 and 3. If the interrogator resolution is increased again to a value of 1 pm, which is achievable with state-of-the-art interrogator systems, the resulting error returns back to the original size of around 5–6 cm.

The heavy oversampling, which happens when a point in the hole is measured again by a consecutive FBG stage, is responsible for the good performance even for a short Sensor Wand.

Unfortunately the fabricated Sensor Wand, which was constructed during this project showed a number of imperfections. Therefore it was not possible to verify the simulated data with the measurements. However, the measurement results show qualitatively the expected behaviour.

4 Sensor Wand Performance

4.1 Testing

The Sensor Wand was manufactured and tested on a surveyed rail track at the Queensland Centre for Advanced Technologies. Some images from our field-test studies are shown in Fig. 5. The image on the right shows the calibration jig, which is used to calibrate the "zero-strain" wavelength values of the non-deflected FBGs. Furthermore, the jig contains an encoder wheel to determine the penetration depth of the Sensor Wand. The testing was conducted on different sections of the track with curvatures between 30 and 200 m radius. Additionally, the sensor was tested on a straight section of the track.

4.2 Algorithm

The wavelength and peak level values for each FBG were recorded during the testing and stored in a hierarchical data format grouped by penetration depth. A recording of 18.5 m depth in 25 cm steps resulted in a data file containing roughly 2.5 million values. The data processing was conducted in Matlab. The first task after reading in the data was to apply a filter to remove double peaks and outliers.

Fig. 5 Sensor wand testing on surveyed track

Furthermore, the filter verifies that the peaks are within the expected wavelength range. Values which differ by more than ±1 nm from the factory calibration, corresponding to roughly 0.5 mℰ, are tagged to be ignored during processing. Reasons for incorrect wavelength values could be glitches in the interrogator, resulting in some peaks not being recorded, or temporary "over-stretching" of some FBGs during manual handling.

The next step is the data pre-processing. During this step the algorithm searches for the FBGs that are already in the hole and computes the mean values from the recorded data sets. At the same time, the on-the-fly calibration values are generated for the FBGs currently in the calibration jig.

Finally, the deflection algorithm uses the mean FBG wavelength and the calibration values to compute the 3D shape, which accumulates up as the Sensor Wand progresses deeper into the hole.

4.3 Initial Results

The 3D reconstruction of a straight track section from a manually recorded data set is shown in Fig. 6. Obviously the reconstructed path is not straight and shows a significant deflection in the "y" direction.

To investigate the reason for the deflection of the reconstructed path, the calibration values and the final wavelength values for the fully penetrated Sensor Wand were compared. Theoretically, both values should be equal as no deflection force is applied during the on-the-fly calibration, and when fully submerged in the hole.

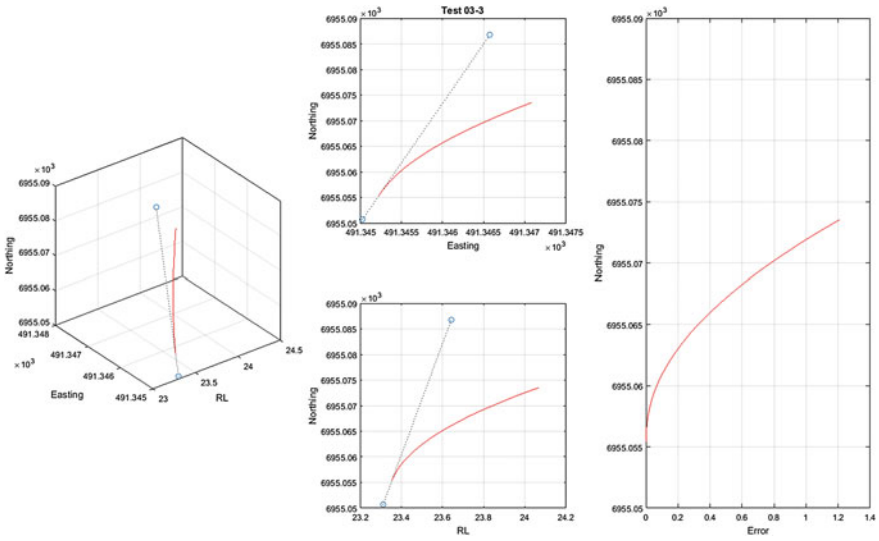


Fig. 6 3D reconstruction of test 3 (straight section, manual recording)

The wavelength difference between the straight Sensor Wand and the calibration values are plotted in Fig. 7. The results vary substantially from zero, indicating that the Sensor Wand is NOT force free when laid out straight. The 3D reconstruction of the above data set was repeated, but instead of the "on-the-fly" calibration, the "forced-straight" values were used to in the deflection algorithm. To avoid a circular reference, the forced-straight values from a repeated measurement series were taken. The result of this reconstruction is shows in Fig. 8, an almost perfect straight

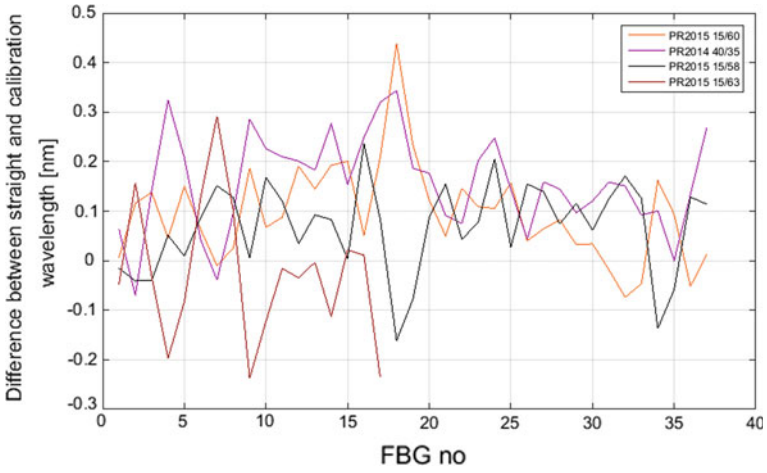


Fig. 7 Comparison between the straight and the calibration wavelength

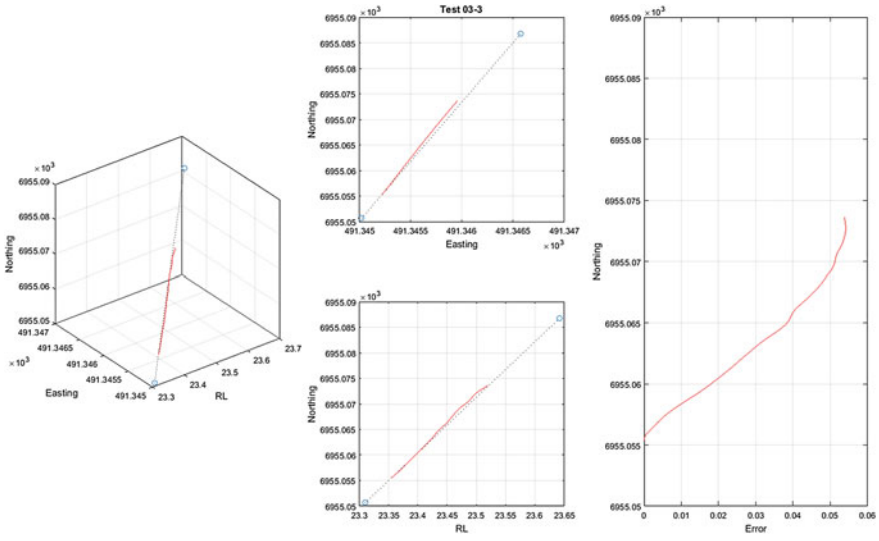


Fig. 8 3D reconstruction of test 3 (straight section, manual recording) using the forced-straight calibration

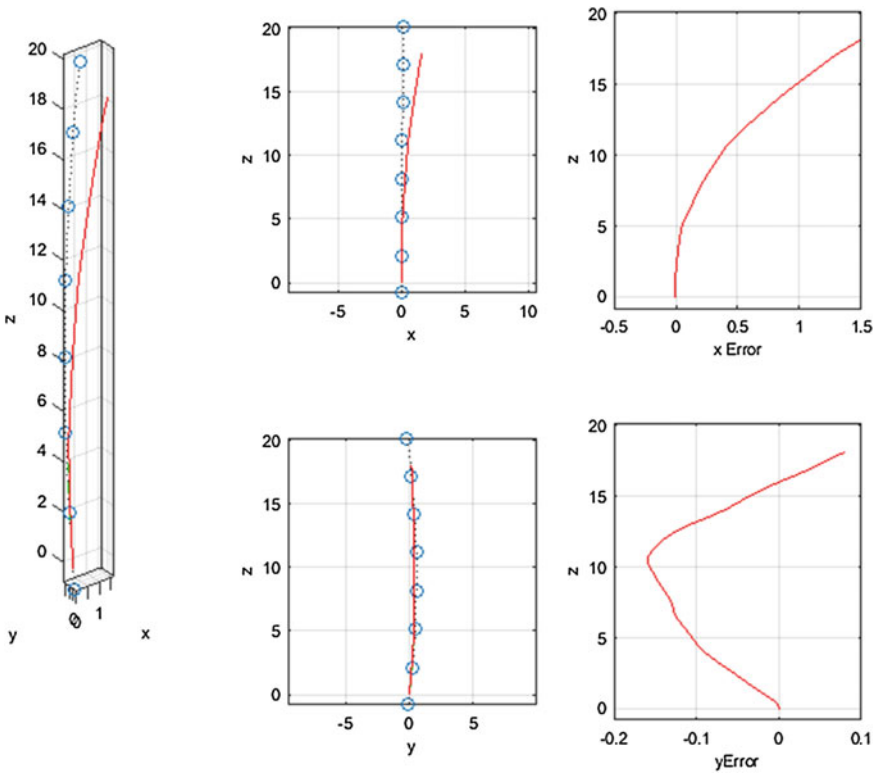


Fig. 9 3D reconstruction of a bend track section using the forced-straight calibration

line at the beginning with a slight curve at the top. The maximum deflection, or error in this case is less than 6 cm over 18.5 m.

The forced-straight calibration can also be applied to the other recordings. The reconstruction of a bent track part is shown in Fig. 9. This figure shows a good correlation in the y plane, however the reconstructed shape still deviates significantly in "x". This deviation may be affected by the fibre misalignment issues which occurred during the manufacturing as the FBGs were not aligned radially, two of the four fibres had a constant offset, and the entire Sensor Wand contained a twist. Additionally, the FBGs were not aligned longitudinally, but showed an offset of several millimetres. Therefore the algorithmic assumption that all FBGs are within one cross-section plane was not fulfilled. Consequently we enhanced the reconstruction model to correct for the misalignment.

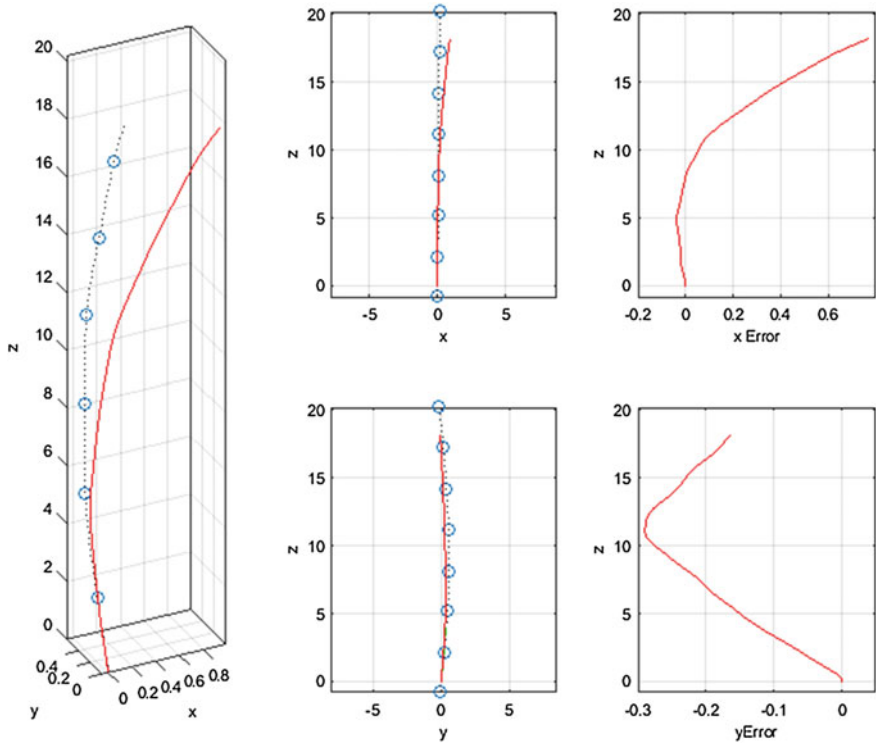


Fig. 10 3D reconstruction of test in Fig. 9 using the enhanced model

The enhanced model can be used to reconstruct the test data from Fig. 9. The result is shown in Fig. 10. The resulting error is reduced significantly, however, still not to expectations. The enhancement uses only a first order approximation of the misalignment and future Sensor Wands will need to be better characterised to determine these misalignments better, in case they can't be avoided during production.

5 Discussion

All results were heavily influenced by the fibre misalignment and this artefact dominates the plots. Given the limitations of the constructed Sensor Wand these results are very promising nevertheless. The algorithms developed as part of this project managed to reconstruct the 3D shape of the Sensor Wand and reduced some of the errors introduced by the artefacts in the wand. However, the measurement accuracy did not meet the expectations from the simulations due to the wand artefacts.

Future Deflection Sensor Wand developments should take potential residual stresses in the composite material into account; and if possible, optimise the design

to reduce these stresses. Furthermore, it is important to characterise the actual location of the FBG along the tube so that the compensation algorithm can use the correct position offsets rather than a first order estimate.

References

1. Rice, T. 2008. NASA-inspired shape-sensing fibers enable minimally invasive surgery. *NASA Tech Briefs* 32 (2): 12.
2. Waltermann, C., et al. 2014. Femtosecond laser aided processing of optical sensor fibers for 3D medical navigation and tracking (FiberNavi). In *23rd international conference on optical fibre sensors*, ed. J.M. LopezHiguera, et al.

A Drilling Technology Sensing Interaction with Cut Mediums to Discriminate Behaviour and Properties of Mediums Ahead on the Tool Path

P. N. Brett, X. Du and S. Wilde

Abstract An innovative tactile approach to automatically detect mediums and underlying structures ahead on a cutting tool path has been implemented in surgery. In this paper a description of a mechatronic approach to surgery is described and shows the merits of the approach. This is supplemented by evidence of suitability of the method to a wider range of mediums and applications outside of surgical application.

Keywords Smart drilling robot · Interfaces · Surgical · Ceramic
Medium guided

1 Introduction

An innovative tactile method to automatically discriminate mediums and structures ahead on a cutting tool trajectory has been demonstrated successfully in surgery [1]. The method enables preservation of fine tissue structures and simultaneous determination of the state of the process, medium and tool performance. Most important, this is used to achieve high tissue preservation and low tissue trauma in surgery [2–4]. The principles enable extension to other application processes and mediums outside of surgery. This paper reports performance in metallic, ceramic and rock mediums showing potential for other application sectors, and to date has relied on measurands coupled within the cut medium. In this approach transients reflect the

P. N. Brett (✉)
University of Southern Queensland, Queensland, QLD 4350, Australia
e-mail: Peter.Brett@usq.edu.au

X. Du
Brunel University, London, UK

S. Wilde
MoD, London, UK

J. Billingsley and P. Brett (eds.), *Mechatronics and Machine Vision in Practice 3*,
https://doi.org/10.1007/978-3-319-76947-9_2

state and behavior of the cut medium and to properties ahead on the drilling trajectory. Much of the work shows that by using the coupled transient between torque and feed force the characteristic changes can be used to discriminate a variety of conditions simultaneously, reliably and in real-time.

2 Application to Surgery

The surgical robotic micro-drill is well known in Otology where it has been applied to expose delicate tissue structures in the cochlea hearing organ prior to placement of a cochlear electrode (cochlear implant). In this application the device is used to expose the delicate endosteal membrane underlying the bony capsule (shell) of the cochlea to a precise diameter without penetrating the membrane. This it achieves, with significant variability of tissue hardness, tissue thickness and orientation of the plane of the tissue interface to the axis of the drilling trajectory. The process is robust, increasing tissue preservation over conventional approaches, and reducing trauma to the tissues and hearing organ. A standard drill bit is used in the process.

Originally the drill was supported on a flexi-arm to aid alignment under the surgical microscope used in the procedure Fig. 1. An operating microscope is required, as visual guidance to locate the line of drilling trajectory at this small scale is beyond unaided visual feedback to the surgeon. The drill bit used is a 0.8 mm diameter burr and the positional tolerance with respect to tissues is measured within a few microns to achieve the required outcome.

The innovative tactile sensing scheme utilises information within the coupled transients of feed force and drill torque. The information is extracted by the drilling system to describe the state of the process relative to deforming tissues, the state of the tissue and drill bit. The detection of tissues and structures ahead are also automatically discriminated in a similar approach, using the same coupled transients.

Fig. 1 The surgical micro-drill on an arm

Figure 2 shows a typical drilling transient when drilling the bone tissue of the capsule of the cochlea in a laboratory (Fig. 3), without involuntary disturbances induced by operator or patient. Both feed force and drilling torque transients are shown. Both measurands rise and fluctuate simultaneously as a result of their coupling through the tissue. Before approaching the interface at the membrane, the bone tissue becomes more flexible, and there is a reduction in force. The tissue interface is detected by the simultaneous rise in torque with a corresponding progressive reduction in feed force. Further information on the residual tissue is found from the changes during progression of the transients.

More complex are the transients in theatre of Fig. 4. Here further disturbances are evident and are associated with involuntary motion of the patient, general disturbances in the operating room, and by the operator. Yet the robotic device is able to discriminate these and is able to determine the risk presented to the process. The resulting exposed window onto the endosteum is visible at 0.6 mm in diameter,

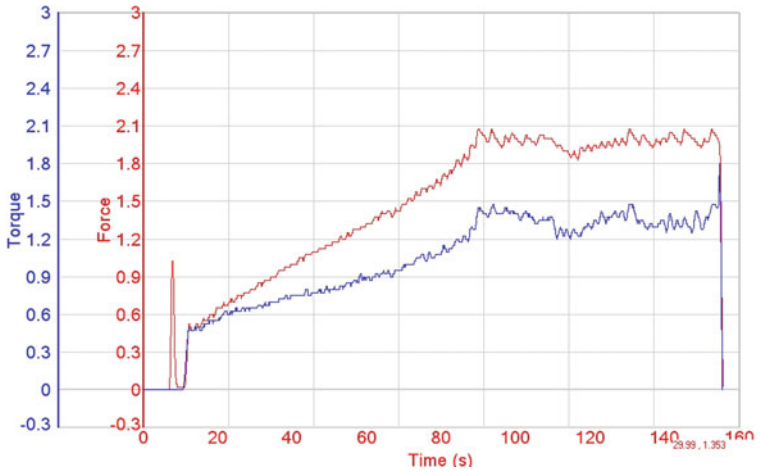
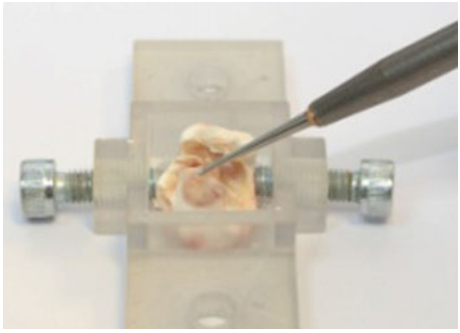


Fig. 2 Drilling transients for cochleostomy in the laboratory (arm fixed drill)

Fig. 3 Drill bit in tissues



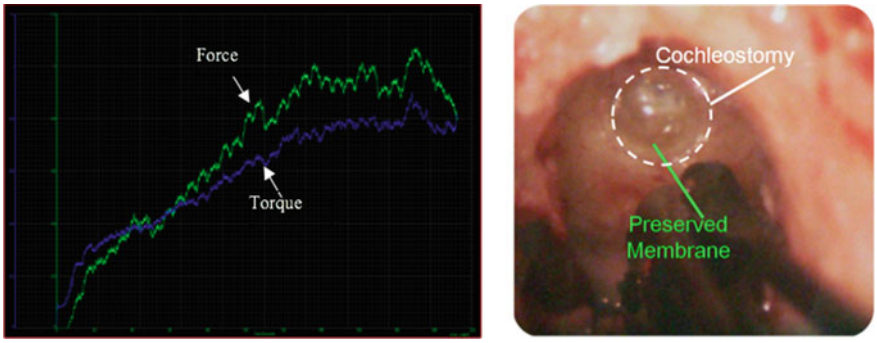


Fig. 4 Drilling transients in the operating room and resulting exposed membrane in cochleostomy

and is intact. For the record, the approach preserves the membrane without fail, which is a significant advantage over conventional drilling, and the amplitude of disturbances induced in the cochlea approach only 1% of conventional surgical drilling, producing a substantially atraumatic approach that is important in ear surgery. The production of a cochleostomy using conventional tools is beyond the threshold of human sensitivity and dexterity and therefore difficult to achieve reliably with the underlying membrane intact. The tissues are delicate and the ear is sensitive to interaction that is not smooth in nature. This tissue guided method in surgical robotics shows a smooth and sensitive approach to drilling.

To explore further, recently the drill has been designed into a hand guided form. Here the machine is taking account of normal disturbances induced by both the operator and patient and enables the operator to apply the drill such that is interacts smoothly throughout the drilling process. The same tactile sensing scheme is used to simultaneously discriminate effects while discriminating important phenomena critical to success of the operating process. This is used to control the drill automatically while guided by the surgeon (Fig. 5).

Fig. 5 The hand guided drill

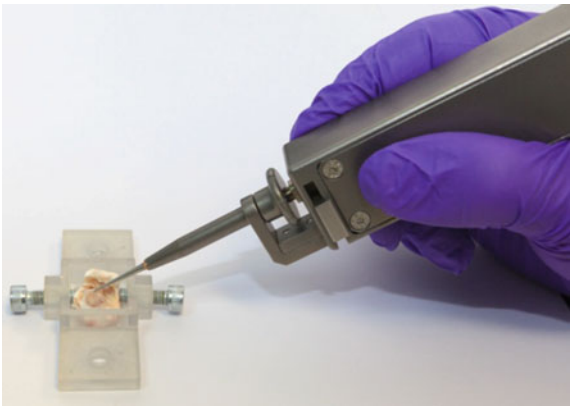


Fig. 6 Left to right: drilling in cranial and nasal tissues respectively

In this form the device is appropriate to the process of drilling, as in many procedures a readily deployed tool is efficient. It both enhances and reinforces surgical skill as the operator is performing the procedure in the same manner as with a conventional drilling tool. This arrangement imposes low setup and training overhead that has significant advantage over extensive, complex robot systems for example.

The success in cochleostomy has driven investigations in other procedures. Figure 6 shows similar results to expose membranes beneath cranial and nasal tissues respectively.

3 Drilling in Other Applications

The advantage of being able to detect differences in the approaching medium, or an underlying structure are fundamental to the identification of reaching a target without encroaching into the underlying medium. This is important to critical processes where exposure, or penetration, of the underlying medium or strata will introduce a hazard, undesirable or uncontrollable event, or may cause undue wear of the cutting surfaces should the medium be harder than can be tolerated. Mining, drilling of oil or gas wells for energy are such applications. The coupled force and torque scheme can detect the state of the drill bit and deploying mechanism too.

Further, it has sensitivity to underlying structures that can also be isolated in a medium, if in close proximity. In manufacturing the process can be used to control the quality of finish to a hole by avoiding unacceptable break-out in metallic and composite materials during the drilling process.

3.1 Metallic Mediums

Investigations have demonstrated drilling in metallic mediums and have shown similar characteristics to that of drilling in bone tissues (Fig. 3). The approach to sensing can be applied to control the quality of the hole produced on drill exit.

3.2 Ceramic Mediums

Further studies in ceramic materials have shown the method able to discriminate interfaces before they are encountered and the type of ceramic material while drilling.

Using the sensing technique, the automatic drilling approach was applied to drilling a composite structure with stratified layers of ceramic material with differing properties. In Fig. 7, the section through the composite structure is shown. Moving from left to right, on the left is a soft porous ceramic followed by a harder fibrous membrane. This is subsequently followed by a more dense ceramic containing harder ceramic particles. The automatic approach is able to discriminate the types of medium, the material interfaces. Finally it avoids penetration of the final interface and places the drill tip a designated distance in close proximity to the surface. The tactile approach is able to discriminate types of medium as porous or particulate, and to discriminate the presence of hard particles in close proximity to the drill path. The example of Fig. 7 shows torque (blue) force (red) plots. As expected, these reflect behaviour in signals one would expect moving both from hard to soft and from soft to hard mediums at material interfaces and local discontinuities, such as particles or cavities present in the mediums. Using the coupled transient behaviour, the drilling system is able to automatically describe the changing cutting environment for the application.

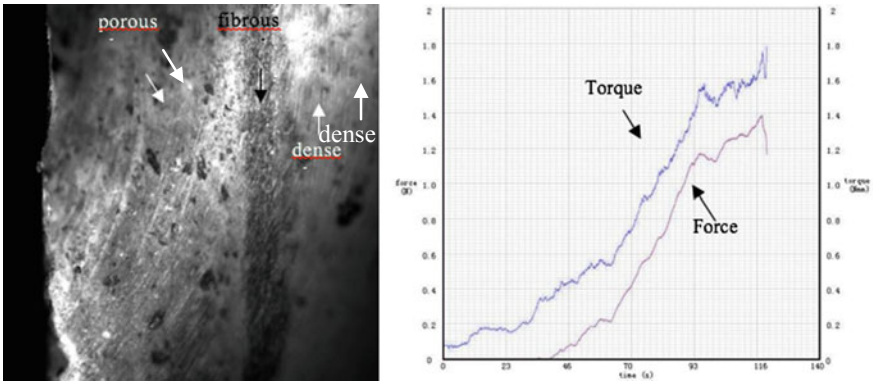


Fig. 7 Detecting interfaces in ceramic materials using coupled force and torque sensing transients

4 Conclusion

The robotic tissue guided micro-drilling method applied to surgery is able to discern tissue interfaces ahead on a drill path. This enables tools to cut up to delicate tissue interfaces without penetration. The process can be deployed to advantage by maximising tissue preservation and minimizing tissue trauma during cutting in surgery. This mechatronic solution in drilling is a relatively simple method to apply, achieves excellent results, and has been demonstrated in the operating theatre. The preparation of a cochleostomy, a critical part of the cochlear implantation process, is a challenging surgical process requiring preservation of tissues and minimal disturbance trauma in the delicate tissues of the hearing organ. The required skill is at the threshold of human dexterity and sensitivity, and the process is achieved reliably and consistently in a robotic solution, even when hand guided.

Application to other mediums such as metals and ceramics, bordering on the range of mediums with the hardness of rock shows equal success to detect and avoid, or control, penetration as required.

References

1. Taylor, R.P., X. Du, D.W. Proops, A.P. Reid, C. Coulson, and P.N. Brett. 2008. A sensory-guided surgical micro-drill. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science* 224 (7): 1531–1537.
2. James, C., K. Albegger, R. Battmer, S. Burdo, N. Deggouj, O. Deguine, et al. 2005. Preservation of residual hearing with cochlear implantation: How and why. *Acta Oto-Laryngologica* 125 (5): 481–491.

3. Zou, J., P. Bretlau, I. Pyykkö, J. Starck, and E. Toppila. 2001. Sensorineural hearing loss after vibration: An animal model for evaluating prevention and treatment of inner ear hearing loss. *Acta Oto-Laryngologica* 121 (2): 143–148.
4. Coulson, C.J., M. Zoka Assadi, R.P. Taylor, X. Du, P.N. Brett, A.P. Reid, and D.W. Proops. 2013. A smart micro-drill for cochleostomy formation: A comparison of cochlear disturbances with manual drilling and a human trial. *Cochlear Implants International* 14 (2): 98–106.

Schematic Design of an X-ray Based Robotic Welding Inspection System

Jinxiao Liu, Jie Li and Xingsong Wang

Abstract For inspecting welding seams of large-scale equipment such as storage tanks and spherical tanks, automated mobile robotic inspecting system is more effective compared with manual operations. A wall-climbing and inspection robot needs not only stable climbing ability, but also high positioning accuracy. In this paper, a flat-panel X-ray inspection based wall-climbing robotic system is developed for intelligent detecting of welding seams. The robot system consists of two Mecanum wheels based measuring cars climbing on both side of the tank wall, each of which is equipped with either a digital flat-panel or an X-ray emitter. On each car, a permanent magnet adsorption mechanism is employed to let it absorbed and climbing on the tank still wall, and a visual path tracking module is used for tracking the welding lines to be detected. To let the flat-panel X-ray work properly, two laser tracking system are applied to ensure each of two cars on the two sides of the tank wall move synchronous exactly with a limited tolerance. Some initial experiment was conducted and reported.

Keywords Robotic welding inspection · Visual path tracking · Double side positioning · Flat-panel X-ray

1 Introduction

Recently,with a widely application of the welding technology in various industrial fields such as electricity, bridge, aviation, automotive and marine, weld quality testing becomes very important [1]. Manual non-destructive testing (NDT) needs inspectors with high level of experience, and some detecting environment is harmful to humans [2]. Therefore, in order to improve the reliability and efficiency of detection and reduce detecting costs, automated non-destructive detecting technology has become an important research topic.

J. Liu · J. Li · X. Wang (✉)
School of Mechanical Engineering, Southeast University, Nanjing, China
e-mail: xswang@seu.edu.cn

J. Billingsley and P. Brett (eds.), *Mechatronics and Machine Vision in Practice 3*,
https://doi.org/10.1007/978-3-319-76947-9_3

X-ray inspection has many advantages in terms of testing welds, radiography can more visually display the size and shape of internal defects about the work piece and easily determine the nature of the defect [3], the negatives of radiography can be saved to future research. The basic principle of ray detection is as following: when uniform intensity beam is irradiated through the object, if the object exists the defective or structural differences, it will change the radiation attenuation of the object, so that the transmitted beam intensity of different parts is different. Using a certain detector to detect the transmitted radiation intensity can determine internal defects and material distribution of the object [4].

For the large equipment such as pipes and tanks, automated climbing robot can effectively replace the manual, it achieve fast and efficient testing [5]. In recent years, climbing machine has develop very rapidly, robot application with testing equipment has been widely used [6]. The magnetic adhesion system often be used to climb the iron wall, it depends on the magnetic force between the robot and the climbed structures. It is used only for the ferromagnetic material structures. Some climbing robots made use of this adhesion system as Rvc robot [7], shown in Fig. 1. An intelligent pipeline inspection tool with an instrumented vehicle was introduced into a pipeline travels through its extension, driven by the fluid itself and could perform various functions [8] (Fig. 2). There is a Weld line detection with tracking system [9] (Fig. 3). The wall-climbing robot for detecting not only needs the ability to climb walls, and also required precise position control, tracking and positioning.

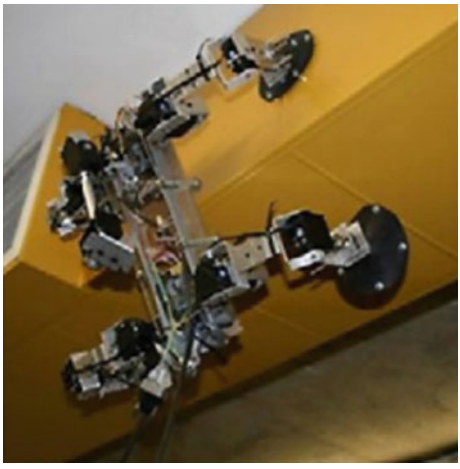


Fig. 1 Reconfigurable vertical climber using magnetic pads

Fig. 2 Smart tool inspection ITION and scraper

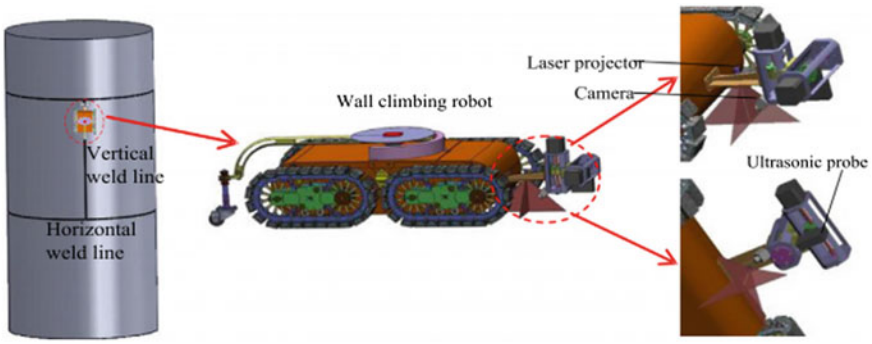


Fig. 3 Weld line detection and tracking system

2 Configuration the X-ray Based Welding Inspecting Robots

In this paper the welding seam inspection robot includes two Mecanum wheels vehicles, a permanent magnet adsorption mechanism, an X-ray machine, a tracking module, a simultaneous position module, and a motion control module, see Fig. 4. Mecanum wheels vehicles have omnidirectional movement capability that can adapt to the complex environment, with a full range of movement, provide precise location for detection [10]. Permanent Magnetic mechanism is fixed in the vicinity of the wheels of vehicles, the equipment in the vicinity of the wheels of the robot can provide enough suction force and stronger obstacle ability.

Detecting equipment includes a main body of x-ray machines and a DR board used for it [11]. The body of x-ray machines is mounted in the major vehicles (Car1), DR board is installed in the minor vehicles (Car2). Two cars do synchronized motion on both sides of the iron wall and make preparations for detecting properly. The tracking module includes a tracking camera and a tracking board. Function of the tracking module is to track and determine the location of the

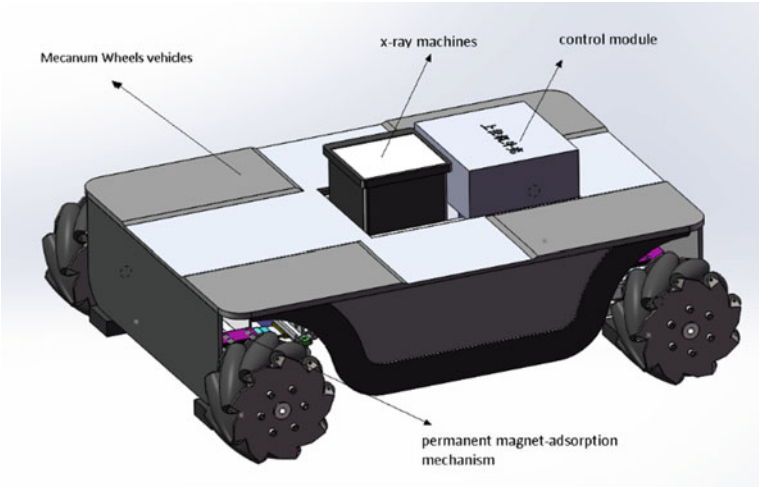


Fig. 4 A Mecanum wheels robot with digital flat-panel X-ray detector

welding line, the data of the tracking module will be sent to the motion control module which handled the data to ensure the vehicle running along the welding line. The simultaneous position module is composed of laser rangefinders. Two vehicles are equipped with positioning systems to determine the coordinates of each sides. After coordinate transformation, we get the actual displacement deviation of the two vehicles, then the motion control module correct the position of the two vehicles to ensure synchronous motion of the two vehicles, see Fig. 5.

The main purpose of weld inspection robot is used for inspecting welding seam to large-scale equipment such as storage tanks and spherical tanks, double vehicles synchronous is innovative used to detect, the weld inspection is more quickly and efficiently, and this can save a lot of manpower and resources. X-ray machine can

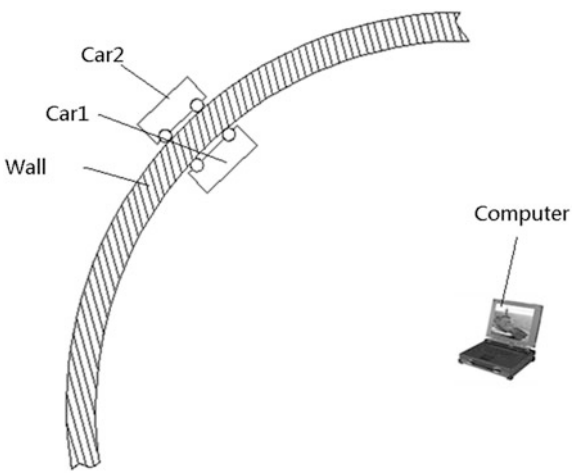


Fig. 5 Principle of welding inspection robot system

clearly show welding seam defects [12], robot can achieve real-time detection through wireless transmitting with a computer. Robots key techniques contains the magnetic adsorption system, the tracking method and the simultaneous position method.

3 Visual Tracking of Welding Lines

With the development of intelligent direction of the mobile robot, the control problem [13] has been paid more and more attention by the people. Path tracing is one of the core problem of motion control for mobile robot. Under a guiding way, how to make a robot tracking pre-specified movement path is very important [14]. Visual navigation captures images of the surroundings through the camera mounted on the robot and uses image processing techniques to extract useful information and guided the robot motion [15, 16]. The Visual navigation has a wide range of signal detection and more complete information. Visual navigation will be a major future development direction of mobile robot navigation.

However, this method is easily influenced by lighting conditions, and the image processing cycle is longer and has large amount of computation. Single visual guidance is frequent in the application of visual navigation, which is done by a camera to photograph the forward path case [14]. According to the deflected distance and angle information of the path, robot can real-time adjust the position and direction. Under the single visual guidance mode, the data inputted to controller is less and the control method is relatively simple, it is suitable for small mobile robots. Tracking module includes a tracking camera and a tracking board, see Fig. 6.

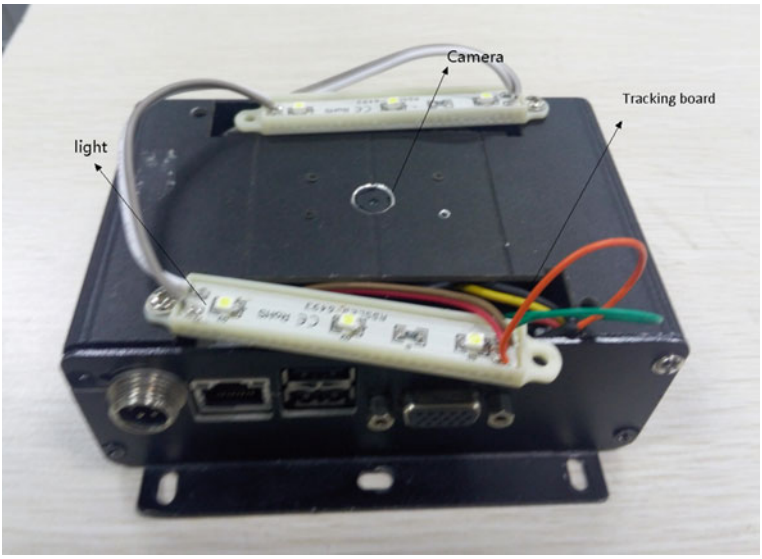


Fig. 6 Tracking module

The purpose of visual guidance is separating the reference path (guide line) from the image, and getting the deflected distance and angle of the robot relative to guide line. Due to the presence of noise about reflective light and shake, images will be processed in order to remove the interference and extract useful information [17].

The process of image acquisition and processing is divided into the following steps: image acquisition, image preprocessing, light regulation, pattern matching, path fitting and calculation of the deflected distance and angle. Of which, the image preprocessing includes median filtering and thresholding and removing noise. Not all of the image acquisition must adjust the light, only the camera equipped with a flash and adjustable parameters will go through this step. In the case of lighting conditions change has resulted path cannot be recognized, it can re-shoot automatically adjusting the exposure time and intensity according the environmental conditions [17].

This ability to automatically adjusting makes the robot less dependent on light conditions and adaptability better, and easier identify path. Pattern matching is to identify the path markers, such as turn signs, parking signs. Fitting path separation method is using the image to obtain border points of guide line and fitted the centerline path, then according to pixel distance and geometry calculate the deflected distance and angle.

Center line of the image and the deflected distance and angle of the guide line has been defined as shown in Fig. 7. Red dot represented the center of the image, declination of angle and distance of the guide line are represented by d and α. When the guide line is in the right of center, $d > 0$; on the contrary, $d < 0$. When the angle from the vertical center line of the image to guide line is counterclockwise, $\alpha > 0$; on the contrary, $\alpha < 0$.

The path tracking control objective of Mecanum wheels robot is: robot advance along the guide line and the camera lens is centered on the center line of the guide line. Tracking process requires robots running smoothly, change of speed and direction is coherent and steady. Fuzzy control mainly changes the linear velocity and angular velocity to adjust the robot pose, the linear velocity determine the forward speed of the robot and the angular velocity determine steering of the robot.

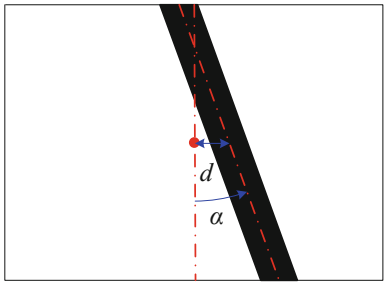


Fig. 7 Deflected distance and angle of the center line of path

4 Synchronous Positioning and Control

The laser rangefinder belong to the external sensor. And similar with sonar, it is active to transmit probe signal and receive the reflected signal to calculate the distance to the target by the interval [18]. The difference is its accuracy and speed greatly improved.

Simultaneous position module is composed by laser range-finders, it also is a laser tracker see Fig. 8. Using laser position system to locate the robot is divided into steps: establishment of the coordinate system, the coordinate transformation, the error elimination. Two laser ranging and positioning systems are arranged on both sides of the iron wall. The system includes a laser range-finder and Pan/Tilt/Zoom (PTZ) which can freely rotate the head according to the actual position of the each vehicle to ensure the laser-finder tracking the vehicle and measure the distance [19].

According to the data measured we establish the coordinate system I (coordinate system of Car1) and the coordinate system II (coordinate system of Car2) to calculate and determine the coordinates of the two vehicles. According to difference of settings about two coordinate systems, we can convert data in the coordinate system II to the coordinate system I. Then, We get relative coordinates of two vehicles in the coordinate system I and compared the coordinates to obtain the difference of coordinates between the two vehicles. Next, we adjust the body position by the motion control system to make them in the same position on the wall. In this process, the coordinate data exists errors, so we had a step to eliminate errors. The main source of error is coordinate conversion and wall thickness, we will get more accurate data through calculating and processing.

Fig. 8 The laser tracker

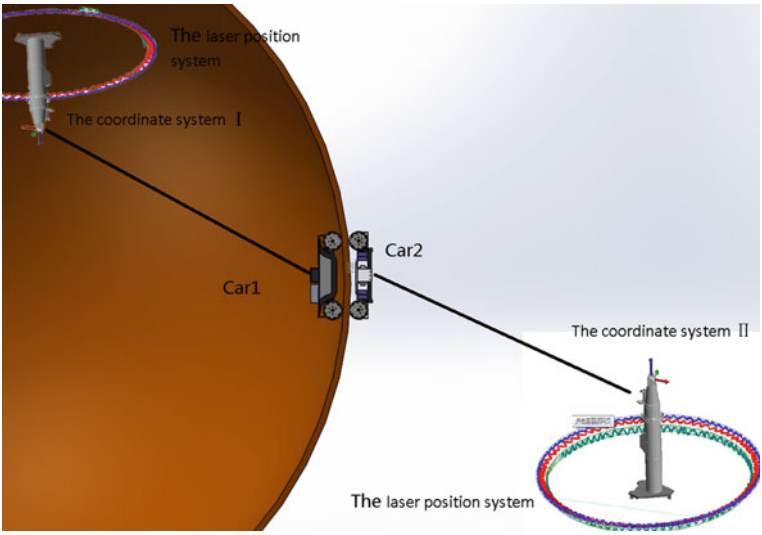


Fig. 9 Simultaneous position system

We have also designed an alternative synchronization method, the main principle is as follows: The x-ray machine imaging is be used to determine the positional deviation of the two vehicles, then robot controls the movement to eliminate the position deviation [20]. Ray source is mounted on the Car1 and DR board is mounted on the Car2, x-ray detector image shows position of center-ray sources in the DR board, we can obtain deviation of the two center through calculating. Every once inspection get a set of images, the robot repairs vehicle location according to the image information to guarantee ray projection in the center position of DR board and two vehicles at the same location, see Fig. 9. This process needs real-time image processing.

By comparison of the two methods, laser positioning is established by an external coordinate system to determine the location of the two vehicles, ray-image position correct position error of center point of the two vehicles. Laser range-finder clearly determine the coordinates of the two vehicles, coordinate conversion and processing is more complex and need to consider effect of wall thickness. The x-ray image is used to correct the position of the vehicle to ensure synchronization of two vehicles but the location is not very high precision.

5 Some Initial Experiments

Key technology of X-ray weld inspection robot is tracking the welding line and double vehicles synchronous positioning. Robot motion was smooth and reliable in experiment, see Fig. 10. Visual tracking module for weld identification determines

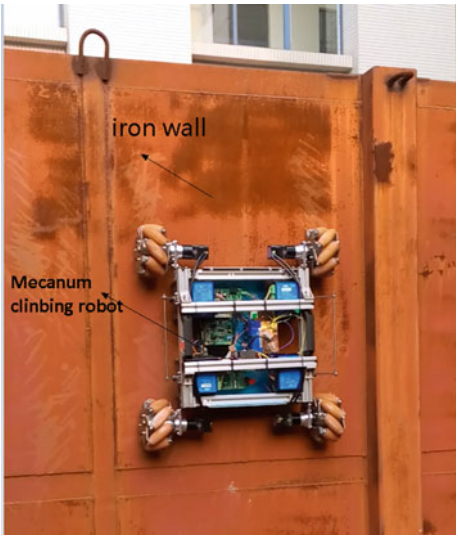


Fig. 10 Robot in experiment

the deflected distance and angle of the vehicle body relative to the welding seam to ensure the body just above the weld and the welding seam in the range of ray detector [21]. Double vehicles simultaneous position aim is to achieve normal work of digital flat-panel x-ray detector for weld, it is the key to ensure accurate inspection.

The key technology of the visual tracking is recognizing and processing the iron wall welding seam to determine the guide line [22], and then we can get the deflected distance and angle of the vehicle body relative to the weld, this was a prerequisite for weld inspection. This process requires the image preprocessing and removing interference and noise to obtain useful information, including median filtering, thresholding, removing noise. Median filtering is a nonlinear signal processing method that can overcome blurred image details caused by the linear filter and is most effective to filter out interference of noise pulse and image scanning [23].

Threshold segmentation is based on a certain threshold to divide the path out from the background image for edge detection. The basic idea of thresholding is to determine a threshold value, then compare each pixel value with the threshold value. According to the comparison result, pixels are divided into foreground (path identifier lines) and the background and the effect depends on the choice of threshold. In order to adapt to different lighting conditions, we used the optimal threshold method dynamically to set threshold and make the threshold increased with enhance of the light and threshold reduced when the light reduced. For complex background larger discrete interference point in the complex background, we use noise removal method to eliminate its impact on the guide line enhance the reliability and stability of the image recognition.

The key technologies of simultaneous position is to determine the relative position of the two vehicles, the laser position system establishes two coordinate

systems, and determines the coordinates of the two vehicles in the corresponding coordinate system based on vehicle distance data. Through coordinate transformation, we can get relative coordinates of two vehicles in the coordinate system I and obtain the difference of coordinates between the two vehicles by comparing the coordinates. The control system correct position of two vehicles to ensure synchronous operation of the two vehicles. Laser range-finder also need to track the vehicle's location, so as to ensure the accuracy of the measured distance. Laser range-finder was installed on the PTZ with 360 rotating to ensure accurate tracking. We will use appropriate algorithms to reduce the corresponding error.

Robot motion control included the following, tracking information is transmitted to the vehicle body motion control system, robot control the vehicle to travel along the welding seam by tracking algorithm, the control system of two vehicles are connected via wireless communication to ensure accurate motion of two vehicles; The data measured by laser position system is converted to the displacement of two vehicles, Car2 keeps synchronous operation with Car1 by wireless communication. After the two-vehicle position kept pace, robot can inspect the weld, image by ray detection is transferred wirelessly to work computer to achieve remotely view [24].

6 Initial Conclusions and Discussions

In the investigating and analysis of the request from the industries, it is very difficult and important to exactly find and track the existed welding line, which is still a challenge problem for the image processing researchers because of the uncertainty of the appearance due to long time using. The other key problem may be the algorithm development to find defects through the digital image of the flat-panel of X-ray. It is occasionally that, in the authors' point of view, the two key problems are all related to image processing.

Acknowledgements This partly supported by the National General Administration of Quality Supervision, *Public welfare of industry quality and inspection, Special research funding*, project no. 201410028.

References

1. Vilaça, Pedro, and Wayne Thomas. 2012. *Friction stir welding technology. Structural connections for lightweight metallic structures*, 85–124. Berlin: Springer.
2. Sbihli, Scott Leo, Jason Howard Messinger, Francois Xavier De Fromont, and Robert Carroll Ward. 2015. Reference speed measurement for a non-destructive testing system. U.S. Patent 9,003,880, issued 14 April 2015.
3. Tamizharasi, G., and S. Kathiresan. 2014. Methodologies to detect defects in thin layered material. *Middle-East Journal of Scientific Research* 2373–2378.

4. Hurley, K., B.L. Dingus, R. Mukherjee, P. Sreekumar, C. Kouveliotou, C. Meegan, and G. J. Fishman. 1994. Detection of a γ-ray burst of very long duration and very high energy. *Nature* 372 (6507): 652–654.
5. Zhang, Houxiang, Jianwei Zhang, Guanghua Zong, Wei Wang, and Rong Liu. 2006. A real pneumatic climbing robot for glass-wall cleaning. *IEEE Robotics & Automation Magazine* 13 (1): 32–41.
6. Utz, Hans, Stefan Sablatnög, Stefan Enderle, and Gerhard Kraetzschmar. 2002. Miro-middleware for mobile robot applications. *IEEE Transactions on Robotics and Automation* 18 (4): 493–497.
7. Peters, G., D. Pagano, D.K. Liu, and K. Waldron. 2010. A prototype climbing robot for inspection of complex ferrous structures. In *Proceedings the 13th international conference on climbing and walking robots and the support technologies for mobile machines*, Nagoya, Japan, 150–156.
8. Arizmendi, C.J., W.L. Garcia, and M.A. Quintero. 2015. Automatic welding detection by an intelligent tool pipe inspection. *Journal of Physics. Conference Series* 628 (1): 012082 (IOP Publishing).
9. Zhang, Liguo, Wei Ke, Qixiang Ye, and Jianbin Jiao. 2014. A novel laser vision sensor for weld line detection on wall-climbing robot. *Optics & Laser Technology* 60: 69–79.
10. Borgolte, Ulrich, R. Joelper, H. Hoyer, Helmut Heck, W. Humann, J. Nedza, I. Craig, R. Valleggi, and A. Sabatini. 1995. Intelligent control of a semiautonomous omnidirectional wheelchair. In *Proceedings of the 3rd International Symposium on Intelligent Robotic Systems* 95: 113–120.
11. Glockmann, Walter, and Thomas Herwig. 1989. X-ray scanner for detecting plastic articles. U.S. Patent 4,884,289, issued November 28 1989.
12. Mirapeix, J., P.B. García-Allende, A. Cobo, O.M. Conde, and J.M. López-Higuera. 2007. Real-time arc-welding defect detection and classification with principal component analysis and artificial neural networks. *NDT & E International* 40 (4): 315–323.
13. Astudillo, Leslie, Oscar Castillo, Patricia Melin, Arnulfo Alanis Garza, Jose Soria, and Luis T. Aguilar. 2006. Intelligent control of an autonomous mobile robot using type-2 fuzzy logic. *Engineering Letters* 13 (2): 93–97.
14. Maalouf, Elie, Maarouf Saad, and Hamadou Saliah. 2006. A higher level path tracking controller for a four-wheel differentially steered mobile robot. *Robotics and Autonomous Systems* 54 (1): 23–33.
15. Ackerman, Christopher, and Laurent Itti. 2005. Robot steering with spectral image information. *IEEE Transactions on Robotics* 21 (2): 247–251.
16. Krupa, Alexandre, Gabor Fichtinger, and Gregory D. Hager. 2007. Full motion tracking in ultrasound using image speckle information and visual servoing. In *2007 IEEE international conference on robotics and automation*, 2458–2464.
17. Shan, Qi, Jiaya Jia, and Aseem Agarwala. 2008. High-quality motion deblurring from a single image. *ACM Transactions on Graphics* (*TOG*) 27 (3): 73.
18. Ye, Cang. 2008. Mixed pixels removal of a laser rangefinder for mobile robot 3-d terrain mapping. In *IEEE international conference on information and automation, 2008. ICIA 2008*, 1153–1158.
19. Derenick, Jason, Christopher Thorne, and John Spletzer. 2005. On the deployment of a hybrid free-space optic/radio frequency (FSO/RF) mobile ad-hoc network. In *2005 IEEE/RSJ international conference on intelligent robots and systems, 2005* (*IROS 2005*), 3990–3996.
20. Weitkamp, Timm, Ana Diaz, Christian David, Franz Pfeiffer, Marco Stampanoni, Peter Cloetens, and Eric Ziegler. 2005. X-ray phase imaging with a grating interferometer. *Optics Express* 13 (16): 6296–6304.

21. Xu, Peiquan, Xu Guoxiang, Xinhua Tang, and Shun Yao. 2008. A visual seam tracking system for robotic arc welding. *The International Journal of Advanced Manufacturing Technology* 37 (1–2): 70–75.
22. Imbert, Christophe Claude, and Jinchi Zhang. 2013. Weld seam tracking system using phased array ultrasonic devices. U.S. Patent 8,365,602, issued February 5 2013.
23. Perreault, Simon, and Patrick Hébert. 2007. Median filtering in constant time. *IEEE Transactions on Image Processing* 16 (9): 2389–2394.
24. Hohl, Lukas, Ricardo Tellez, Olivier Michel, and Auke Jan Ijspeert. 2006. Aibo and Webots: Simulation, wireless remote control and controller transfer. *Robotics and Autonomous Systems* 54 (6): 472–485.

Low Cost Cascade Controllers for Non Linear Hydraulically Driven Joints

Peter Milani

Abstract The usefulness of low cost cascade controllers for manipulation of hydraulically driven rotary joints was examined by measuring the controller's effectiveness in practice using single joint control and describes the implementation across seven joints in a serial manipulator. The ideal hydraulic dynamics is examined and how non-linear kinematics of a cylinder driven joint, valve hysteresis and asymmetrical volumes of a single ended cylinder affect accurate control of the joint using a single control layer. We linearize the control problem by implementing a feed forward velocity with error loop to handle valve actuation and achieve linear joint velocity. This lower control was connected to more complex higher controllers via a standard interface implemented with ROS Control. The overall system allows the implementation of complex rate and trajectory controllers to manoeuvre the tool position of the manipulator.

Keywords Control · Manipulators · Hydraulics

1 Introduction

Large scale manipulators are common in industry and the opportunity to automate them is increasing. However they are mostly still operated by a human operator closing the loop to ensure the correct control of the tool end of the manipulator. Automation of hydraulic large scale manipulators has been done to achieve a supervisory control of forestry equipment [1]. Research has also been undertaken to improve positioning of tooling in large scale manipulators by studying the important factors in positioning accuracy [2]. The research underpinning this study was performed in an attempt to develop a seven axis, hydraulically actuated manipulator capable of carrying loads in excess of 200 kg at a full extension of 2 m and portable enough to be transported across a mine site and setup within minutes.

P. Milani (✉)
University of Southern Queensland, Toowoomba, QLD, Australia
e-mail: petermilani80@gmail.com

J. Billingsley and P. Brett (eds.), *Mechatronics and Machine Vision in Practice 3*,
https://doi.org/10.1007/978-3-319-76947-9_4

Our initial goal was teleoperation, using the Jacobian to convert an operator's task-space velocity commands into the joint space of the manipulator. A user interface that could receive commands in all six degrees of freedom of the operator's taskspace, x, y, z and roll, pitch, yaw was developed. However, our two methods for implementing the conversion relied on a highly linear implementation of the joint space control in the manipulator. This led to the development of our own low cost controllers for joint position and velocity.

If such joints are linearised, then a large range of automation options become available. These include the implementation of trajectory controllers, leveraging six dimensional path planning. As a basis for implementation, using Robot Operating System makes manipulator path generators such as MoveIT! [3] a realistic option for sensor integration and control of the arm.

2 The Problem

Due to the torque requirements, the simplest hydraulic implementation was to actuate each rotary joint with a hydraulic cylinder in the form outlined in Fig. 1. It is shown in the position of zero joint angle. As the cylinder retracts, the joint will move with a negative angular velocity. The x axis is to the right, the y axis is upward, and the joint follows a right handed convention. Thus, if the joint angle, ϕ, is 0 at full cylinder extension, then as the cylinder retracts, the joint moves with a negative angular velocity. The relationship of the joint's angular position and velocity to the cylinder's rod position and velocity is the following non-linear relationship based on the cosine rule:

$$\theta = \cos^{-1}\left(\frac{(x+l)^2 - p^2 + q^2}{-2pq}\right) \tag{1}$$

$$\phi = \beta + \theta + \alpha - \pi \tag{2}$$

where x_c is the cylinder's extension, l is the minimum length of the cylinder, q is the distance from the joint of the insertion cylinder end, and p is the distance from the joint of the insertion of the rod end. ϕ is the joint angular position and is related to θ, the angle of the cosine rule by the summation given in (2).

Applying a constant linear velocity in the cylinder will result in different angular velocities depending on the position of the joint. This is demonstrated in Fig. 2 where the change in joint angle for a constant cylinder extension is plotted against the joint angle in accordance with (1) and (2). This shows that at the limits of the joint, the change in joint angle to the change in cylinder length is at its maximum and varies throughout the stroke in a non-linear fashion.

The dynamics of a hydraulic cylinder can be modelled as an integrator for a given valve setting. Merritt [4] has studied the dynamics of a variety of hydraulic

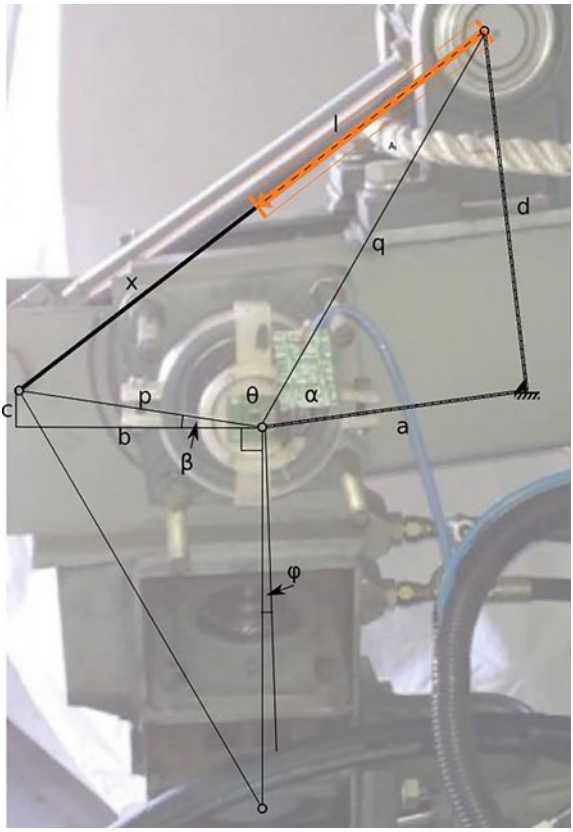


Fig. 1 Experimental apparatus with free body diagram superimposed on image

actuators and valves. However for our analysis, we simplify [4] to consider that the load pressures on the hydraulic actuators are significantly lower than the supply pressure or that they are appropriately compensated, so that we can ensure that the various relationships between valving and load pressures do not feature prominently in this analysis. Nevertheless there are sufficient, hydraulic nonlinearities to affect light load control of the hydraulic joint. In the experimental apparatus, a Hydraforce SP08-47C, 4 way, 3 position, closed proportional cartridge valve were used in series with a counterbalance valve cylinder. These valves were subject to *deadband, asymmetry and hysteresis* which all were non-linear features present by design. Also present are *valve* and *cylinder stiction* which can cause sudden stopping and breakaway of the spool or piston when actuated. An additional feature that is not present is the compression and expansion of fluid lines from the valve to the cylinder. This is assumed to be minimal as the apparatus' lines are steel pipe.

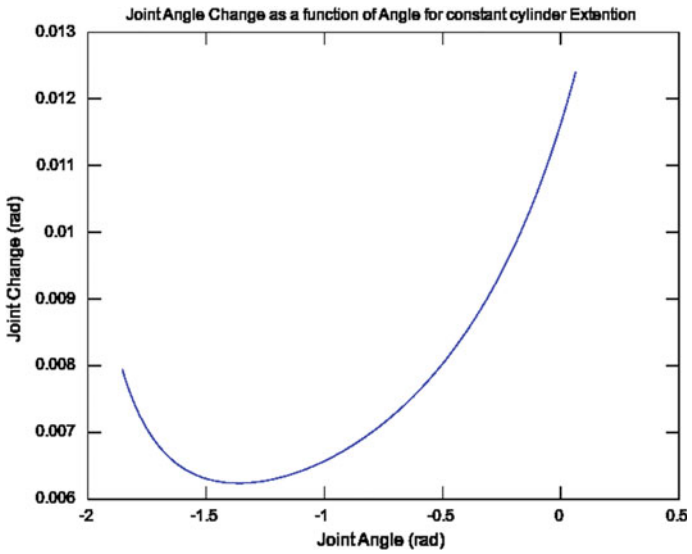


Fig. 2 The geometry of the Joint causes a non-linear change in joint angle for a constant extension rate. The effect is plotted against joint angle to illustrate the acceleration that occurs to the joint in certain parts of the movement

2.1 Valve Nonlinearities

Valve nonlinearities include deadband, symmetry and hysteresis. Deadband occurs in closed centre valves by design through overlapping of porting within the valve. This also ensures sufficient mechanical sealing of the valve and a minimal signal to actuate the valve and avoid false actuation by stray currents [5]. The proportional valves were designed to be significantly symmetrical, that is flow to Port A is similar to flow to Port B, however, there is minute asymmetry of the order of 1–2%, due to the valve design. Hysteresis in valves due to friction within the valve can be up to 7% of the valve spool location [5].

There are a number of common methods for reducing the effect of these non-linearities such as tuning the input signal to minimise deadband. An oscillating signal on top of the control signal called *dithering* improves the linearity of the valve performance, reduces hysteresis and sudden *breakaway* of the valve, by keeping the valve constantly in motion reducing static friction. An alternate option is the selection of more precise, specifically designed valves that have limited hysteresis and asymmetry or employ closed loop control of the valve spool, with additional cost.

The factors present in the valve can be represented as below:

$$\dot{m}_v = \begin{matrix} K_i(1+N(0,0.03))i & i > i_{\text{db}} \\ 0 & 0<i<i_{db} \end{matrix} \tag{3}$$

where $\dot{m}_v$ is the fluid flow through the valve, i is the applied current to the coils of the valve which, when above the deadband i_{db}, results in flow. Hysteresis $N(0, 0.03)$ is modelled as a random variable taken from the normal distribution about zero and a standard deviation of 3%. K_i is a constant affected by asymmetry of the valve, thus has different values depending on direction of actuation:

$$K_i = \begin{matrix} K_{P\gg A} & i<0 \\ K_{P\gg B} & i>0 \end{matrix} \tag{4}$$

where $K_{P\gg B}$ and $K_{P\gg A}$ are proportional gains when flow is to port B and port A respectively.

2.2 Cylinder Nonlinearities

Cylinders are subject to many of the same types of non-linear phenomena such as static friction and asymmetry due to cylinder ratios, common in single ended cylinders, where the volume of the cylinder rod reduces the flow required to move the cylinder in contraction compared to extension. The static friction also changes depending on use due to the effects of lubrication and oil temperature. The relationship between the cylinder extension velocity $\dot{x}_c$ and the flow rate $\dot{m}_c$ is:

$$\dot{m}_c = \begin{matrix} \dot{x}_c A_p & P \gg A \\ \dot{x}_c(A_p - A_r) & P \gg B \end{matrix} \tag{5}$$

where A_p and A_r are the areas of the piston and rod respectively. These linearly affect extension velocity.

2.3 Valve Cylinder System

Discounting any leaks in the system it is usual to assume that (3) and (5) are equal. The pure integrator of the ideal system is apparent when operating above the dead band threshold:

$$\begin{aligned} \frac{X_c(s)}{I(s)} &= \frac{K_{P\gg A}(1+N(0,0.03))}{s(A_p - A_r)} P \gg A \\ \frac{X_c(s)}{I(s)} &= \frac{K_{P\gg B}(1+N(0,0.03))}{sA_p} P \gg B \end{aligned} \tag{6}$$

The differing and noisy gains as part of this integrator require more than fixed parameters for an open loop control mainly due to the effects of hysteresis and which randomly affects the gain of the valve and cylinder ratio which provides a constant bias to the velocity. This, combined with the effect of joint geometry, gives rise to the effects shown in the experimental data in Fig. 3.

Figure 3 shows the effect of a constant valve setting on velocity against joint angle. The data points at −1.95 and 0.05 rad represent the joint limits. The forward (red) and backward (blue) rotation of the joint are shown and a fourth order polynomial has been fit to the data to demonstrate the general trend. The effect of cylinder ratios is demonstrated with the difference in trend lines in both directions. The difference between velocities between −1.5 and −0.2 rad, described by the trend lines is 17.5% greater in the backward direction. This corresponds directly to the cylinder ratio whose volume is 18% less in the backward direction thus commensurately quicker for a given flow. The increased velocity per unit joint change at the joint limits is similar to the theoretical shape given by the joint geometry.

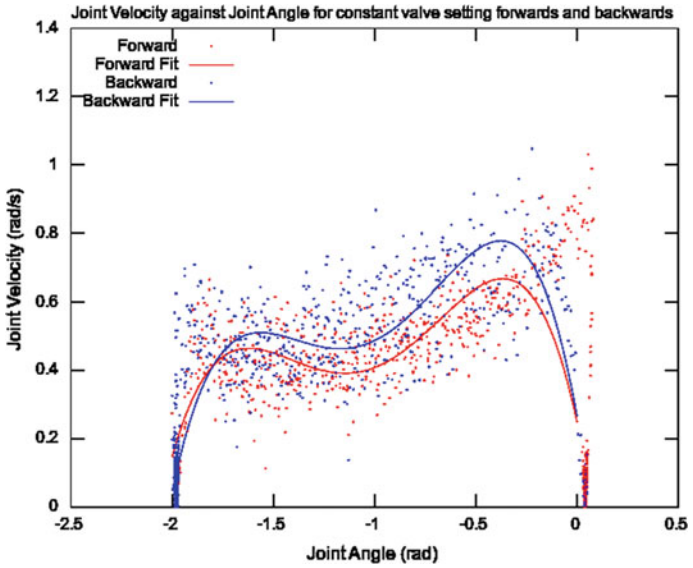


Fig. 3 Plot of experimental data showing the effect of nonlinearities on joint velocity for a constant valve setting. A 4th order polynomial has been fitted to both forward and backward movements

3 Cascade Controllers

Our implementation of the cascade controller utilised two main levels, a lower controller, responsible for joint velocity control in a hard real-time context and a higher level control which implements a common hardware interface and interface with the ROS middleware.

3.1 Lower Controller

We implemented a cascade controller which aimed to linearise the joint velocity by closed loop velocity control. This was implemented with a PI controller to eliminate the effects of hysteresis, cylinder ratios, as well as the effects of the geometry of the joint. In order to investigate the possibility of doing this cheaply, low cost components were used.

A 10 bit absolute magnetic encoder, based on the Austria Microsystems AS5040, was utilised at a cost of around $40 per board. There was some noise in the LSB due to incorrect alignment, as a result the velocity estimation is low pass filtered in software. The velocity control was implemented on an ATMEL ATMega328P microcontroller packaged in an Arduino Uno for $30 per board. The software utilised a custom state machine implemented in AVR-C for direct manipulation of registers. A PWM signal was output from the microcontroller to the Joystick input of a Hydraforce Proportional valve controller for current control of the A and B solenoids of the valve. It utilised 50% for all off, >60% for flow to A port and <40% for flow to B port. The min current of the valve was tuned to ensure that minimal flow occurred at those transition points.

As there were seven joints on this manipulator, the joint control was split between two microcontrollers. Each microcontroller controlled up to 4 joints with the encoder information travelling over an SPI bus implemented on a reconfigured RS-232 port. The Chip Select lines for this SPI Bus were implemented manually through two digital outputs through a 74HC139 dual 2–4 line decoder. Use of the 74HC139 saved two pins on the microcontroller which was nearly fully utilised. The PWM signal was generated utilising the Arduino Analog Write libraries which are based on hardware PWM generation. The control period of the lower controller was 1370 ms.

Software control of each of these joints relied upon a simple state machine given in Table 1.

Table 1 State machine for lower controllers

State 1	Event	State 2	Action
Init	config_msg_recieved	Pause	None
Pause	comd_recieved	Run	None
Run	write_timeout	Pause	None

The lower controller utilised a Feedforward plus PI control of the remaining error. PID control alone works poorly due to the incompressibility of hydraulic fluid making the system very stiff. As a result, any negative valve setting will instantly reverse the direction of flow, causing hammering and poor velocity outcomes. Therefore, a signal proportional to the velocity is fed forward taking into account the cylinder ratio. Added to this is a PI signal based on the remaining velocity error, due to other nonlinearities and geometry of the joint. The signal is not allowed at any stage to reverse the direction of flow, at minimum flow it closes the valves. The feedforward control is shown in Fig. 4.

3.2 *Upper Controller*

The upper controller interfaced with the Robot Operating System (ROS) middleware. The Upper Controller was implemented using the *ros_control* package [6], which allows controllers to be implemented as dynamic libraries and loaded at runtime. As part of this, the lower controller is exposed in a standard format. In this case the lower controller was exposed as a velocity interface, allowing velocity references to be set and joint position and velocity to be queried. Other parameters, mainly those associated with the lower controllers (PI gains, joint limits etc.), were set at run time as robot parameters in the ROS Master parameter server.

The upper controller and the ROS Master were implemented on a Beaglebone, a 1 GHz ARM computer running Ubuntu Precise 14.04 with ROS Indigo. The Beaglebone is a low cost computer at around $50 which exposes a number of hardware interfaces, multiplexed across the available output pins in two headers. The SPI1 peripheral was enabled via the device tree overlay on the 4.1.+ kernel and accessed by the spidev library, packaged with Ubuntu [7].

Communication to the lower controllers was via the second SPI interface available on the Beaglebone, which was capable of implementing communication to two slaves (our two lower controller microcontrollers) in hardware. The SPI bit rate was implemented at 200 Kb/s, limited to minimise the impact of interrupt

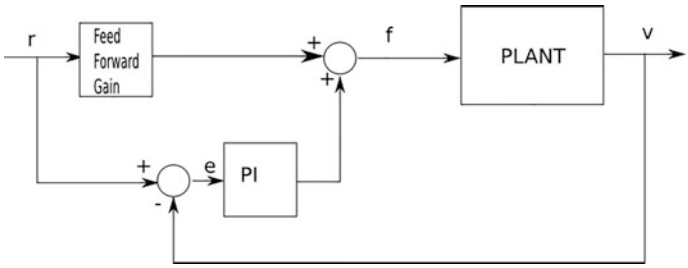


Fig. 4 Velocity control implemented as a feed forward gain moderated by the proportional and integral of the error

servicing on the control loop. All communication between the upper and lower controllers was interrupt driven. Three message types for *configuration, reading,* and *writing* were implemented as a union of a C++ struct object and an eight bit data stream. The C structure made generating the data stream straightforward as both shared the same memory locations. The data stream was sent over SPI and reassembled in the microcontroller. An 8-bit checksum was incorporated for error checking.

The link between the upper and lower controller was synchronous, with periodic reading of jointstate from lower controller, performing an update on the upper controller and then writing the necessary commands to the lower controller. There was only one plugin implemented, it was capable of receiving both joint position reference commands for joint position control, or desired joint velocity commands. The update step executed the statemachine in Table 2. For the joint position control, the update step calculated the necessary command via a Proportional Integral Derivative (PID) summation. In the velocity_forwarding, the latest value received from the command topic was applied as commands with no additional processing.

4 Experimental Results

The following results demonstrate the linearising effect of the Feedforward velocity plus PI error controller. Note that the controller does not show any of the effects demonstrated in Figs. 2 and 3. The Feedforward gain is zero and the Error Proportional gain is 100. No gain was applied to the integral of the error. The particle cloud consisted of three runs in forward velocity and two runs with backwards velocity.

The response in Fig. 5 contrasts with that of Figs. 2 and 3 in that the non-linear effects due to joint geometry are not present. The use of the feedforward gain has provided sufficient drive to the joint and a proportional gain on the error was sufficient to remove errors in velocity due to geometry. The noise in the solution is due to noise in the velocity estimation as the control period had some noise to it, and the joint velocity was insufficient to elicit a change in encoder value despite having 1024 pulses per revolution. This velocity control has a linearising effect on the joint response and allows other more standard controllers to be utilised in the upper control layers.

Table 2 State machine for the upper controller

State 1	Event	State 2	Action
Paused	fwd_comd_recieved	Velocity_Forwarding	doVelocityControl
Paused	position_comd_recieved	Position_Control	doPositionControl
Paused	nothing	Paused	doPaused
Velocity_Forwarding	fwd_comd_timeout	Paused	doPaused
Position_Control	position_comd_timeout	Paused	doPaused

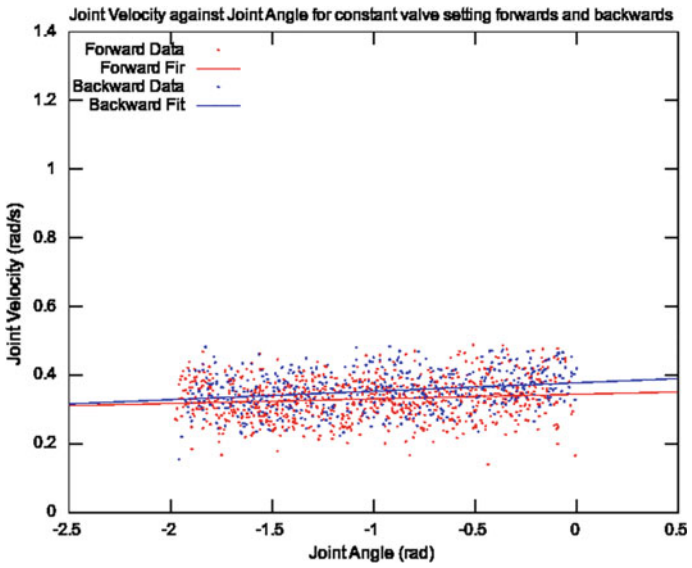


Fig. 5 Joint velocity plotted as a function of joint angle

5 Conclusion

The issue of servo joint control for hydraulic joints is not a new one. The dynamics of hydraulic joints is well established however, automation of large scale hydraulic equipment with multiple joints is still not widely available. This work aimed to linearise the highly nonlinear response for control input that is present in cylinder driven joints, in addition to utilising low cost components. The experimental results demonstrate the controller's effectiveness achieving velocity control of the joint and presenting that control as a velocity interface. As this has been implemented across multiple joints of a serial manipulator, each joint exposed as a linearized velocity controller, it allows the possibility for trajectory controller and manipulator automation to be realised.

Acknowledgements This research was in part funded by Australian Coal Associates Research Program (ACARP) Project C12027.

References

1. Hansson, A., and M. Servin. 2010. Semi-autonomous shared control of large scale manipulator arms. *Control Engineering Practice* 18: 1069–1076.
2. Cho, Y.K., C.T. Haas, and S.V. Sreenivasan. 2004. Position modeling for automated construction manipulators. *Journal of Construction Engineering and Management*, 50–58.

3. Sucan, Ioan A., and Sachin Chitta. “MoveIt!” [Online]. Available: http://moveit.ros.org.
4. Merrit, Herbert E. 1967. *Hydraulic control systems*. USA: Wiley.
5. SP08-47C. Spool, 4 Way, 3-Position, Closed Centre, Hydraforce, 2.110.2.
6. Bohren, J., D. Coleman, A. Tssoukian, and W. Meussen. ROS control, an overview. In *ROSCON 2014*.
7. SPIdev, linux-sunxi.org/SPIdev.

State of the Art and Key Technology of Soft-Bodied Robots

Zhisheng Zhang, Qi Zhang, Fang Jia and Yifan Zhou

Abstract As a new type of robot, soft-bodied robots have wide potential applications in the fields of exploration and exploitation, rescue, medical service. This paper presents a review on the current research status, which, on the basis of functional characteristics, can be classified as mobile robots, including peristaltic motion, rolling motion, omega-type motion and jumping motion, and operation robots. Moreover, the key technologies of soft-bodied robots, mainly including actuators, active materials, physical modeling, control strategies and manufacture processes, are summarized. Finally, some bottlenecks and research directions on soft-bodied robots in future are discussed as well.

Keywords Soft-bodied robots · Motion characteristics · Drive modes
Physical modeling · Control strategies · Manufacture processes

1 Introduction

The robots have obtained wide application in many fields which range from traditional industry to health care, education, entertainment, exploration and biological engineering. Based on the underlying materials, robots can be classified as rigid robots and soft-bodied robots. Rigid robots mainly refer to traditional rigid robots, discrete hyper redundant robots and hard continuum robots [1].

Z. Zhang · Q. Zhang · F. Jia (✉) · Y. Zhou
School of Mechanical Engineering, Southeast University, Nanjing 211189, China
e-mail: jfang@seu.edu.cn

Z. Zhang
e-mail: oldbc@seu.edu.cn

Q. Zhang
e-mail: 220140251@seu.edu.cn

Y. Zhou
e-mail: yifan.zhou@seu.edu.cn

J. Billingsley and P. Brett (eds.), *Mechatronics and Machine Vision in Practice 3*,
https://doi.org/10.1007/978-3-319-76947-9_5

Based on physiological structure and kinematics characteristics of mammals, the bodies of traditional rigid robots have multiple flexible joints connected by stiff links. And these links are driven by motor, hydraulic and other actuators. Besides, the traditional rigid robots can quickly and accurately performing a wide variety of tasks. The rigid robots have the advantages of simple control, accurate position, high stabilization and reliability. However, the traditional rigid robots have the following disadvantages: the flexibility of motion is limited, the adaptability to environment is bad, the movement in narrow space is restricted, and the robot cannot through opening smaller than their nominal dimensions [2]. The drawback of traditional rigid robots limits its applications in the fields of dynamic, unknown and unstructured complex environment, such as military, disaster relief and scientific exploration. In order to improve flexibility of traditional rigid robots, researchers developed discrete hyper redundant robots by increasing motion pairs and degrees of freedom (DOF) of traditional rigid robots.

Hard continuum robots are further developed depending on discrete hyper redundant robots. It can work in unstructured environments and provide the continuous deformation ability [3]. Recently, researchers developed a variety of hard continuum robots. Rezaei et al. designed a snake robot based on snake muscular system [4]. Based on underwater propulsion principles of fish, Fan et al. presented a fish-imitated underwater robot [5]. He et al. studied a mimic octopus arms robot comprising several disks and four NiTi alloy wires. It can simulate a bending motion by driving every wire independently [6].

As a new type of robot, soft-bodied robots are inspired by biology in nature. These robots can be used for delicate tasks in unstructured environments with following advantages. (1) Owing to an infinite number of DOF and continuous deformation ability, they can change shape and size when pressurized. (2) They have distributed and continuum deformation, thus the robot tip can attain every point in three-dimensional workspace. (3) By using large strain deformation, they can conform to obstacles. Meanwhile, they can generate little resistance to compressive forces and therefore can squeeze through openings that are smaller than nominal dimensions. (4) Based on flexibility of soft-bodied robots, they can carry soft and fragile payloads without causing damage.

This paper presents a review on the current research status and key technologies of soft-bodied robots. Firstly, based on functional characteristics, the soft-bodied robots are classified. Then, the details of the key technology, mainly including actuators, active materials, physical modeling, control strategies and manufacture processes, are discussed. Finally, a whole lot of problems and the trends in soft-bodied robots technology are investigated.

2 Review of Soft-Bodied Robots

Soft-bodied robots are inherently compliant and express large strains in normal operation. According to organizational structures and kinematic characteristics of soft-bodied organisms in nature, the robots have high environment adaptability and wide potential applications in fields, such as military, exploration and exploitation, rescue and medical service. There are a variety of classification methods about the soft-bodied robots. From the point of morphology and functionality of soft structures, the soft-bodied robots can be divided into the static water bone structure, the muscle hydrostatic skeleton structure and other structure types [2]. From the point of power supply ways, the soft-bodied robots are grouped into the cable driven soft-bodied robots and the wireless driven soft-bodied robots [7]. From the point of functional characteristics, there are two categories of the soft-bodied robots: the mobile robots and the operation robots [8]. This section describes some of the most interesting examples of these soft-bodied robots.

2.1 Mobile Robots

The mobile robots are capable of moving on unknown terrain. By analyzing the motion characteristics, the mobile robots can be divided into peristaltic motion, omega-type motion, rolling motion and jumping motion.

Peristaltic Motion. Peristalsis is the main moving mode for small mollusks, such as worms, earthworms and tobacco hornworms. In addition, the cephalopods (such as octopus, squid and starfish) can also creep with the soft tentacles. According to the peristaltic motion mode of soft-bodied organisms in nature, a lot of bionic peristaltic robots are developed. The robots are composed of many segments which can be operated normally while some of the units are out of order. Hence, they have strong adaptability to the environment.

Based on the mechanics of caterpillars, Kate et al. developed a soft-bodied robot named SoftBot [9]. The robot is divided into three main components: the body wall (elastic silicone rubber), the actuators and the control system (nervous system). The soft-bodied robots can move forward by controlling the shape memory alloy (SMA).

According to simple structures and motion mechanisms of the Caenorhabditis elegans, Yuk et al. designed a small crawling robot [10]. The robot with several body segments is designed to generate a sinusoidal undulating motion. The overall design of the robot includes two key features: (1) the actuators are designed using a thermal SMA because it is reasonably comparable to the actual worm's wall muscles; (2) The simple binary motion control mechanism is inspired by muscular activations of C. elegans [10]. As shown in Fig. 1, the motion modes of the robot are comparable to those of a real worm.

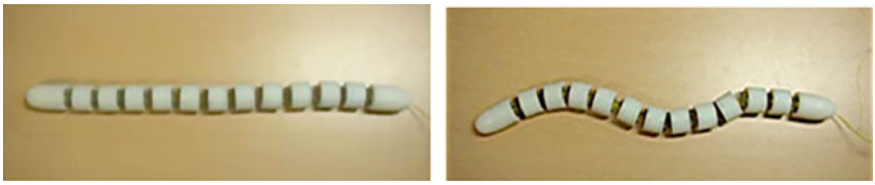


Fig. 1 Images of the motion modes of the robot [10]

Shepherd et al. devised a multi-gait soft robot that can crawl and creep by using pneumatic actuators [11]. Soft lithography is used to fabricate this pneumatically actuated robot, and the design of pneumatic channel is based on the pneu-net (PN) architecture. The PNs are a series of chambers embed in a layer of extensible elastomer and bond to an inextensible layer, and these chambers inflate like balloons during actuation [11]. The structure of a multi-gait soft robot is shown in Fig. 2a. This robot is quadrupedal, and sensors are not needed. It only needs five actuators and a simple pneumatic valve control system that operates at low pressure. Figure 2b shows the actuation sequence for the PNs that generate the creep locomotion. The robot is actuated by pressurizing the PNs in sequence and can navigate through an obstacle.

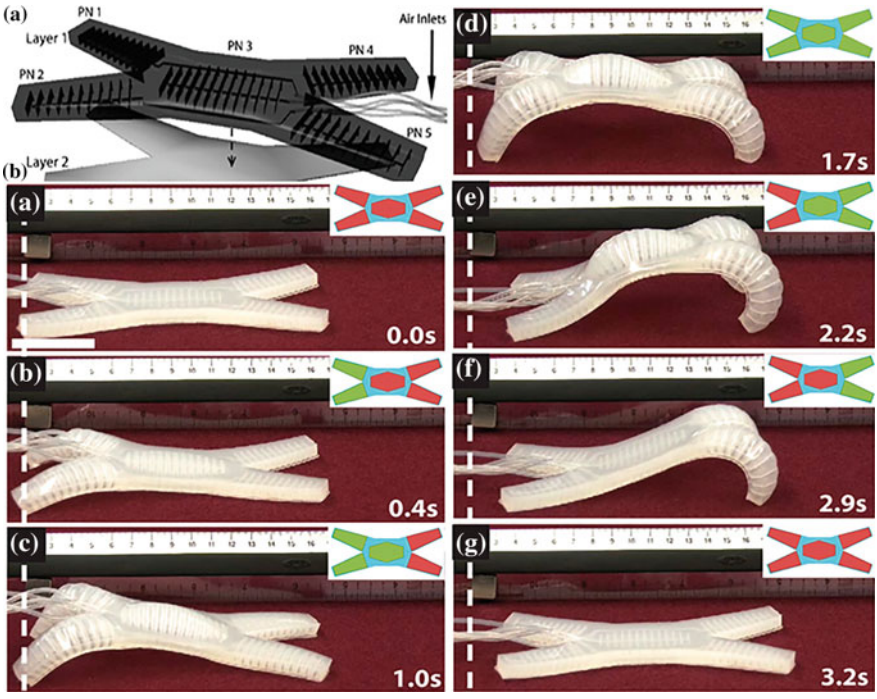


Fig. 2 The structure and creeping motion of the multi-gait soft robot [11]

Besides the above-mentioned robots, there are many soft-bodied robots with crawl motion ability. Based on motion principle of earthworms, Ma et al. proposed a micro bionic and peristaltic robot which can simulate the walking motion of earthworms [12]. Li developed an earthworm-like robot which coves hole under the earth [13]. Liu et al. studied a biomimetic earthworm robot which is actuated by using springs made of SMA [14]. Fu et al. devised a guiding robot for active catheters based on SMA [15]. On the foundation of the motion characteristics of mollusks, Du et al. designed a novel soft robot which can roll, crawl and omega-type motion [16].

Omega-type Motion. Omega-type motion is a special form of peristalsis and peculiar to inchworms locomotion in biological world. An inchworm moves forward by alternating two kinds of stroke motion: the longitudinal extension motion and omega shape bending motion (see Fig. 3) [17]. Most soft-bodied robots designed so far are based on the motion of inchworm. Not only can they creep through a small space, but also pass through a high barrier with body suspension [18].

Inspired by inchworm motion, Menciassi et al. presented a microendoscopy device [19]. An inchworm locomotion device is basically made up of two types of actuators: the extensor and the clamper. The extensor produces a positive displacement, while the clamper is used to securely clamp the device. The simplest device consists of two clampers in its end and one extensor in its middle [19]. Maeda et al. designed a bionic inchworm robot [20–23] which is made of polymer gel. The periodic self-oscillating motion of the gel is produced by dissipating chemical energy of the oscillatory Belouzov–Zhabotinsky (B–Z) reaction [20] (details given in Sect. 3.1).

Koh et al. developed a soft-bodied robot named Omegabot [17] based on the motion of inchworms, which has an omega shape bending motion between extension motions. The robot is fabricated using smart composite microstructures (SCM) and actuated using SMA coil spring actuators. It consists of multiple segments where the first and last segments touch the ground and the middle part does not need to touch the ground. Combined with silicon polymers attached at the outer tip of the first segment and the inner tip of the last segment, the segments at the

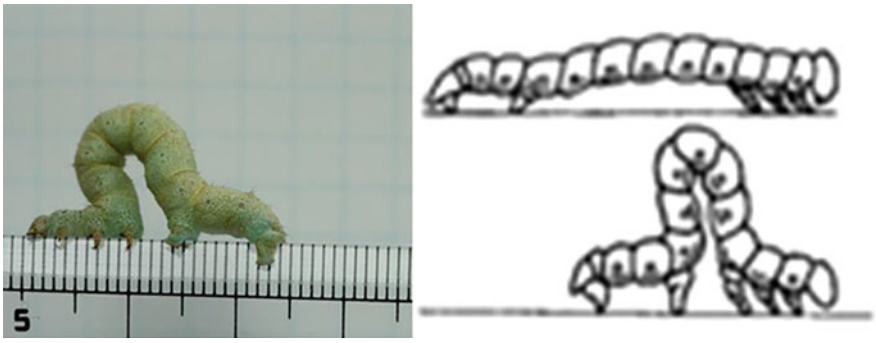


Fig. 3 Omega-type motion of an inchworm [17]

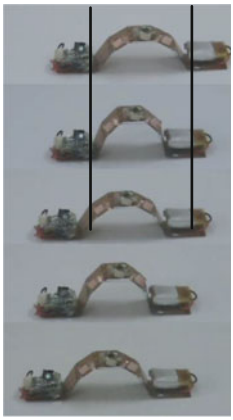


Fig. 4 The crawling motion of the Omegabot [17]

middle create an omega shape as the actuator is activated, and extend as the actuator is deactivated. The robot moves forward by alternating the two motions. Figure 4 shows the crawling motion of the Omegabot.

Rolling Motion. Compared with the peristalsis motion and omega-type motion, a rolling motion has a fast speed and high efficiency. There are two categories of rolling motion: the active rolling motion and the passive rolling motion [18]. The active motion relies on internal energy. For example, caterpillar muscle tissue includes longitudinal muscle and oblique muscle. The rolling energy of caterpillar is generated by adjusting their muscle tension. Moreover, the passive motion relies on external energy, such as wind and gravity.

Kimura et al. proposed an innovative flexible robot with a new hydrostatic skeleton driving mechanism [24]. The main components of the robot are a hermetically-sealed outer covered with looped structure and flexible crawlers with hydrostatic skeleton (HS). The HS crawlers mainly consist of fluid chambers and pressure tubes. By pressurizing the pressure tubes, the angle between two adjacent fluid chambers changes, then the robot is actuated rolling forward.

Inspired by physiological structures of caterpillars, Lin et al. devised a soft-bodied robot named GoQBot [25]. The robot has a soft silicone body of 10 cm long and two major actuation units (i.e. embedded SMA drive coil): the anterior flexor and posterior flexor. Those units can be controlled independently. Through the SMA drives, the robot can quickly become circular and has capability of rolling motions, the rolling speed can be up to 200 cm/s [25].

According to hydrostatic bone structures of sea cucumber, Zheng et al. developed a pneumatic hydrostatic soft robot [26]. As shown in Fig. 5a, the robot consists of air bags, actuated airbags and sensors (installed on the actuated airbags). The actuated airbags are connected to a compressor and a pump via a valve. The internal pressures of the actuated airbags are individually controllable. The air bags straddle the two adjacent actuated airbags. Figure 5 shows the motion of the pneumatic hydrostatic soft robot. By analyzing actuated airbags charging and

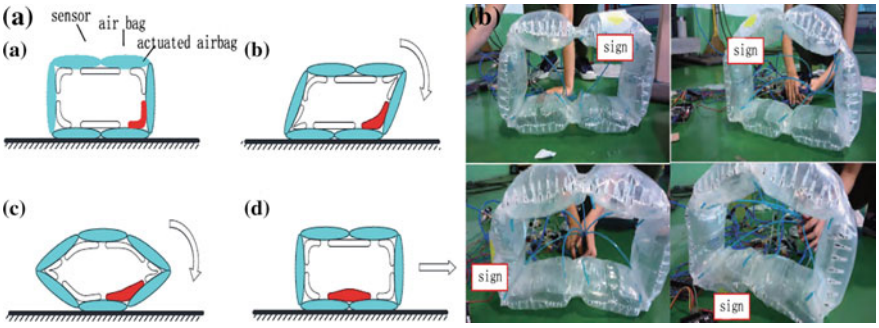


Fig. 5 The structure and rolling motion of the pneumatic hydrostatic soft robot [26]

discharging processes, a mathematical model of driving force is set up. The robot can be actuated by pressurizing the actuated airbags in sequence.

Jumping Motion. In addition to the above motions, the soft-bodied robots can also achieve jumping motion. Shepherd et al. developed a tripedal robot which is capable of jumping by using an explosive combustion of hydrocarbons triggered by an electrical spark [27] (details given in Sect. 3.1).

Sugiyama et al. designed a deformable robot in a wheel shape [28]. The robot consists of a circular elastic shell with a set of soft actuators inside and eight SMA coils inside the robot can be used to store this elastic energy. The Robots will produce deformation by controlling the SMA actuators, and the deformation allows the elastic potential energy to be stored. If rapidly release, it can generate large force to make the robot jump.

By using multi-material three-dimensional (3D) printing, Bartlett et al. built a combustion-powered robot whose body transition is from a rigid core to a soft exterior [29]. Powered by combustion of butane and oxygen, this robot is able to perform untethered jumping. The robot consists of the main explosive actuator and three pneumatic legs (see Fig. 6a). A rigid core module that contains power and

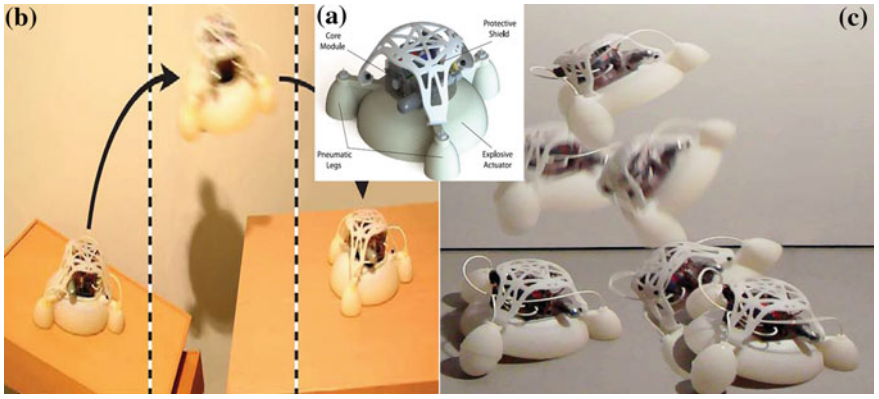


Fig. 6 The structure and jumping motion of the functionally graded soft robot [29]

control components sits atop the main body. It is protected by a semi-soft shield. To initiate a jump, the robot inflates a subset of its legs to tilt the body in the intended direction. Upon combustion, the bottom hemispheroid balloons out, pushing against ground and propelling the robot into the air. Figure 6b performs a targeted jump off of an angled surface onto a table. The directional jump of this robot is shown in Fig. 6c. This robot autonomously jumps 0.76 m (six body heights) high and the directional jumping is up to 0.15 m (0.5 body lengths, 20% of jump height) [29].

2.2 *Operation Robots*

Robots for gripping are mostly inspired by mollusks. The operation robots have the potential to grasp objects in pose-invariant and shape-invariant conditions and can thus adapt to the environment relying on compliance. By analyzing the muscle geometry and biomechanical properties [30–33] of octopus, an European octopus project team presented artificial muscles of octopus tentacles' structure [34] and put forward some corresponding control methods [35, 36].

Inspired by structure of octopus arms, several operation soft robots are developed. Martinez et al. proposed robotic tentacles with three-dimensional mobility shown in Fig. 7, capable of large deformation and with camouflage ability [37]. For the robotic tentacles, the pneumatic networks are based on three indistinguishable channels along the longitudinal direction of the tentacles. By pressurizing one of its channels, the 3D tentacles can achieve progressive bending. The motion is based on the selection of distribution, configuration and size of an embedded pneumatic network.

Most operation soft robots developed so far are based on mimic octopus arms, and few examples of other kinds of operation soft robots exist. Xu et al., for example, devised a peristaltic and soft-bodied xy-sorting table by using the pneumatic actuators [38, 39]. As shown in Fig. 8, the table consists of the flexible silicone layer (A) with the inflatable cells (B) on the rigid ground (C), a camera (E) and a robotic arm (F). The sorting table is actuated by the peristaltic actuation. The principle of the peristaltic actuation is to produce different wave shapes in the flexible layer of the table, which can drive objects on its surface [38]. Since the

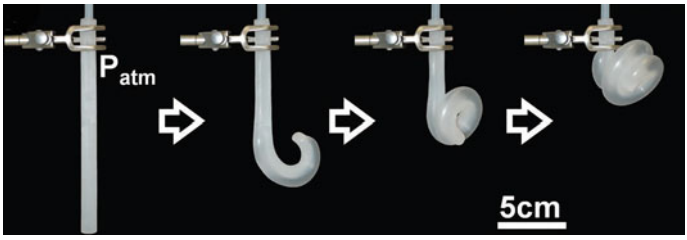


Fig. 7 The motion of a tentacle [37]

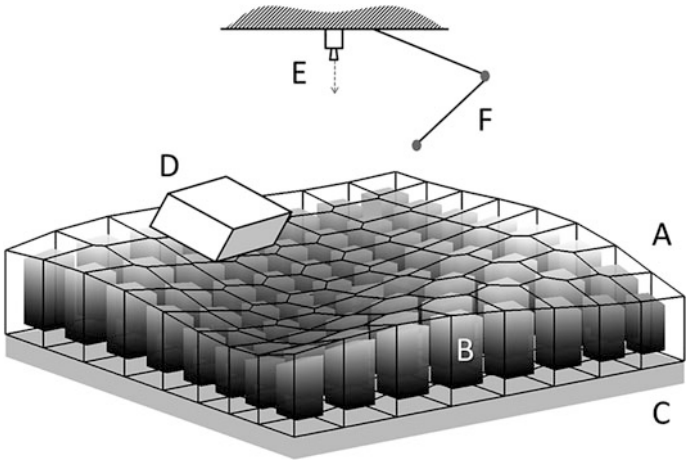


Fig. 8 The structure of a peristaltic xy-sorting table [38]

slippery, delicate organs are difficult to handle and they are easily damaged by rigid robotic grippers. Thus, the sorting table is used to solve the problems of sorting ovine offal in industrial meat processing.

3 Key Technology of Soft-Bodied Robots

The research on soft-bodied robots involves the biology, chemistry, machinery, electronics, material, control, and many other fields. The current situation of key technologies, including actuators, active materials, physical modeling, control strategies and manufacture processes, and some corresponding application cases in different aspects, are introduced in this section.

3.1 Actuators

Actuators can be divided into four categories: shape memory alloy actuator, electro-active polymers actuator (EAP), pneumatic/hydraulic actuator and chemical actuator. In this section, the features of those actuators are described, individually.

Shape Memory Alloy Actuator. The shape-memory alloy can remember its original shape. That is to say when it is deformed, it will recover to its pre-deformed shape after being heated [40]. SMA actuation is based upon a solid-state phase transition controlled by temperature changes. SMA has a unique property of its pseudo elastic behavior under high temperature (austenitic state), this kind of alloy can carry recoverable strain several times more than general metal.

One of the advantages to use shape-memory alloys is the high level of recoverable plastic strain that can be induced. However, if the SMA is used in an environment where the ambient temperature is uncontrolled, permanent damage on SMA actuator may occur.

On the foundation of constant volume design, Menciassi et al. designed an artificial earthworm [41]. Inspired by the contraction and expansion of earthworm link fluctuations, the robot has capability of peristalsis. The robot has four modules which are driven independently. Each module is actuated by one or more SMA springs. The robot is covered by a shaped silicone material which can be used as a platform to be inserted tiny legs for obtaining differential friction conditions.

Seok et al. proposed a soft-bodied robot named Meshworm [42, 43], the design principle is based on the unique antagonistic arrangement of radial/circular and longitudinal muscle groups of Oligochaeta. The robot includes an elastic fiber mesh tube and NiTi coil actuators. The mesh tube structure is composed of multiple spiral polymeric fibers woven into a tube shape. According to antagonistic arrangement between radial/circular muscle and longitudinal muscle, the robot is capable of achieving peristalsis.

By using the SMA actuators, Mao et al. developed a starfish-like soft robot [44]. Soft body allows the robot to achieve a high mobility via shape changes, withstand external damages, and adapt to different terrains through passive or active deformation. As shown in Fig. 9a, the SMA actuators are embedded into the five flexible rays. The diaphragm seal is applied to prevent the system from leaking water. Through this model, the SMA spring can be precisely designed and controlled according to the displacement and force requirements. Multi-gait locomotion can be achieved by applying appropriate control patterns and the related locomotion strategy of the soft robot. As shown in Fig. 9b, the starfish-like soft robot moves with a creeping or analogous peristaltic motion on a clammy rough terrain.

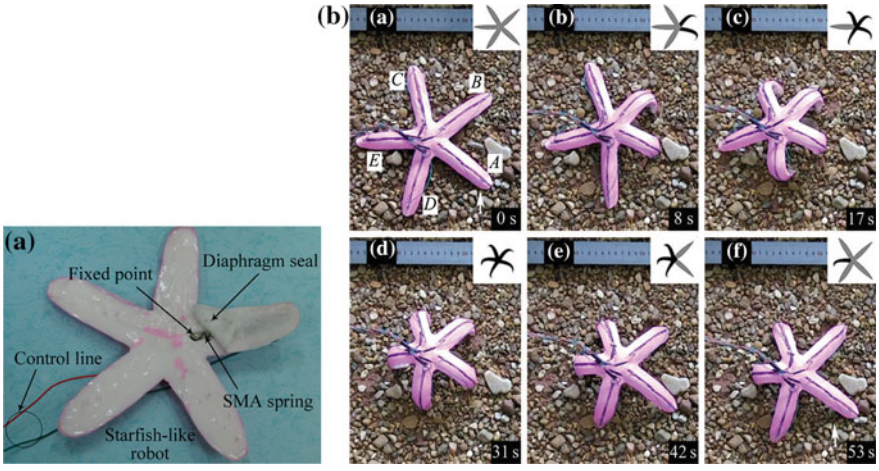


Fig. 9 The structure and a sequence of snapshots of starfish-like soft robot [44]

Electroactive Polymers Actuator Electroactive polymers exhibit changes in size or shape when stimulated by an electric field. When an electrical field is applied, a force is applied to each partial charge and causes rotation of the whole polymer unit. This rotation causes electrostrictive strain and deformation of the polymer. EAP materials can be easily manufactured into various shapes due to the ease in processing many polymeric materials, which makes them very versatile [45].

Li et al. devised a tortoise-like soft-bodied robots that can crawl and swim underwater actuated by ion conductive polymer film (ICPF) [46]. It has quick response and can be driven by low voltage (1.5 V) in the water or wet conditions. The robot's limbs are all made of ICPF. According to the phenomenon that thin film cathode generates bend when ions move and swap in the water under the action of electric field force, the robots can complete the task flexibly, such as creeping and walking.

Otake et al. developed a starfish-like soft robot which actuated by EAP [47, 48]. The electro-active polymer systems of the starfish-like soft-bodied robots can be divided into two parts: the polymer and an electric field. The robots are made of electro-active polymer gel in constant electric fields. According to penetration of the surfactant solution, a beam of gel develops wave forms in uniform electric fields. As shown in Fig. 10, spatially varying electric fields generated by linear array, or a matrix arrangement of electrodes can bring out interesting motions of the gel, such as a starfish-shaped gel robot that turns over by deforming its whole body [47].

Chi et al. designed a novel smart structure of soft robot that is driven by dielectric elastomer [49]. The soft robot uses a flexible elastomer as the body and uses dielectric elastomer actuator as the muscle. The structure of the soft robot is shown in Fig. 11. The smart structure of the robot is influenced by pre-stretch, structural stiffness and voltage. The robot can slowly crawl on the horizontal plane

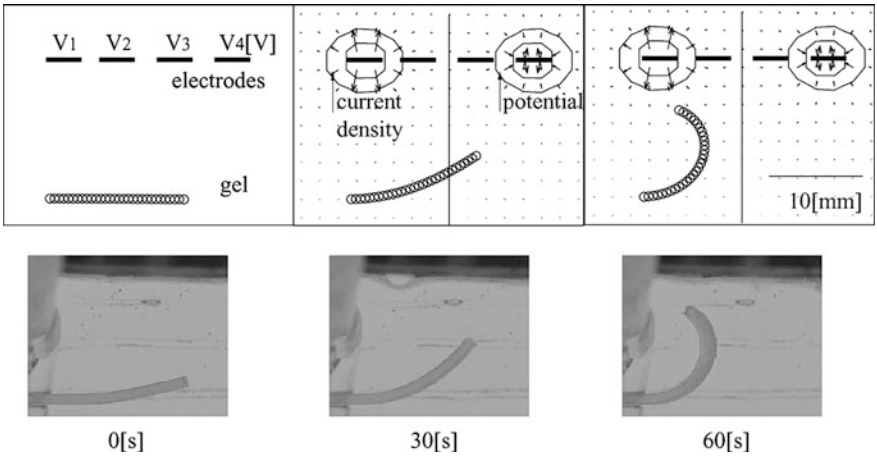


Fig. 10 Transformation of the beam-shaped gel [47]

Fig. 11 Novel dielectric elastomer structure of soft robot [49]

when its arch upward and giving it an alternating high voltage. The speed is about 1 cm per second under the alternating voltage of 8 kV and 5 Hz.

Pneumatic Actuator. The pneumatic actuator has been developed as a mature and widely used technology, which has the following advantages: (1) compressed air is easily generated; (2) it is environmentally benign; (3) it is lightweight; (4) it is essentially in viscid and thus allows rapid motion.

Based on the 'Jamming' principle, Steltz et al. presented a new paradigm in soft robots [50]. Jamming is the mechanism by which particulate material can transit between a liquid-like and a solid-like state with only a small change in confining volume [50]. The robot consists of some cells and an internal actuator. The cellular compartments (cells) around its outer perimeter are filled with a jamming (granular). Similarly, the center of the robot is filled with a fluid. Through deformation of the unjammed cells, the cells volume will expand and the shape of the robot will change.

By using the pneumatic actuators, Fei et al. designed a modular soft robot [51]. The robot consists of three deformable spherical cells as shown in Fig. 12. According to the deflating and inflating of the spherical cells, the size of each spherical cell will be changed, then enabling the soft robot to move forward. The motion speed is influenced by the speed of the spherical cell's deflating and inflating. When the robot moves, its quality and friction with the ground will be changed by inflating and deflating the spherical cells. The robot's actual moving distance is less than that of the theoretical analysis, due to the friction force and the inflation rate, and the robot can move forward 12 cm in one circular.

The new design of PN is described by Mosadegh et al. [52]. This PN reduces the amount of gas needed for inflation, thus increases its speed of actuation. The actuator consists of an extensible top layer, and an inextensible bottom layer reinforced with embedded paper as shown in Fig. 13a. The extensible layer of the PN contains gaps between the inside walls of each chamber, so that the two inside walls are thinner, and have greater surface area than the other exterior walls. As shown in Fig. 13b, the actuator can bend from a linear to a quasi-circular shape rapidly in 50 ms.

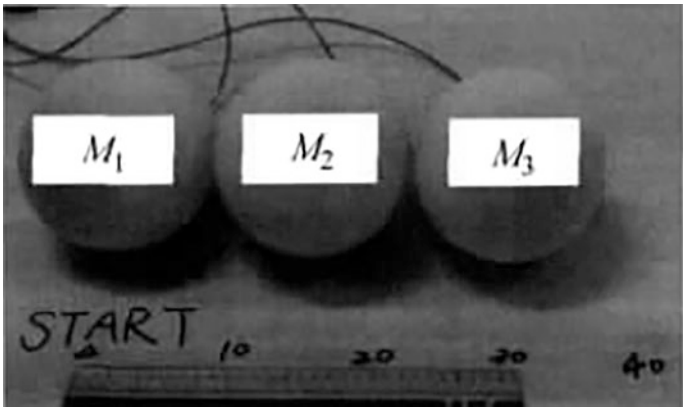


Fig. 12 The structure of the modular soft robot [51]

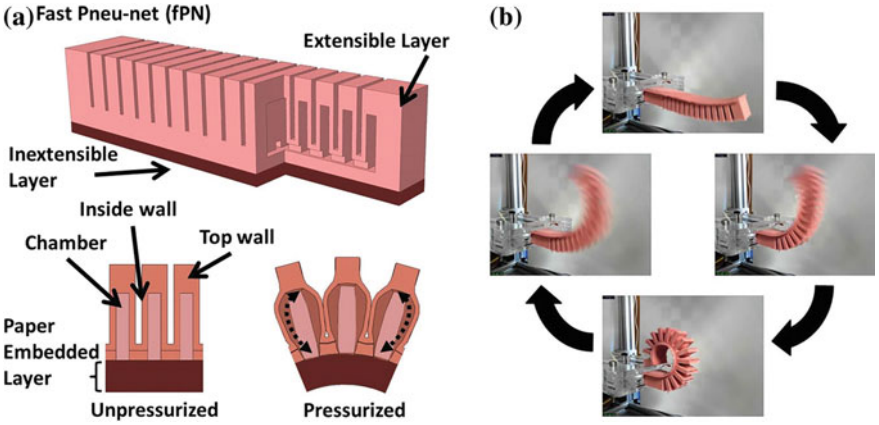


Fig. 13 The structure and motion of the PN actuator [52]

Chemical Actuator. The chemical drive mode refers to converting chemical energy into mechanical energy with chemical reaction, and thus actuating the robots motion. At present, one of the typical chemical reactions is the B–Z chemical oscillating reaction, and the other is the combustion reaction. The overall process of the B–Z reaction is the oxidation of an organic substrate, such as malonic acid (MA) or citric acid, by an oxidizing agent (bromated ion) in the presence of a strong acid and a metal catalyst [20].

Maeda et al. developed a bionic inchworm robot [20–23] which is made of polymer gels. The polymer gels have cyclic reaction networks as in a metabolic process itself. With a cyclic reaction, the polymer gel swells and shrinks autonomously [20]. The gel membrane is immerses into an aqueous solution containing malonic acid, sodium bromate and nitric acid. One end of the gel strip is

sandwiches in the incision of the silicone rubber. By dissipating chemical energy of the oscillatory B–Z reaction, the periodic self-oscillating motion of the gel is produced.

Shepherd et al. devised a tripedal robot which is capable of jumping motion using an explosive combustion of hydrocarbons triggered by an electrical spark [27]. As shown in Fig. 14, this robot incorporated a passive valving system, which exhausts the product gases automatically and actuate the same PN repeatedly. Explosive chemical reactions producing pulses of high-temperature gas for PN actuation provides simple, rapid, co-located power generation, and motion in soft robots [27]. The jumping height will be increased, the energy efficiency will be improved, and the jump of the robot can be driven directly by controlling the timing of the sparks.

Lee et al. presented light-driven hydro-gel actuators that can rapidly undergo nonlinear motions by combining reduced graphene oxide (RGO) nanosheets and elastin-like polypeptides (ELPs) [53]. RGO efficiently generates heat by absorbing NIR light. ELPs can possess tunable thermal responsiveness in aqueous solutions. As shown in Fig. 15a, the fingers of a hand-shaped hydrogel bending and unbending in response to the location of a NIR laser spot. Figure 15b shows the bending

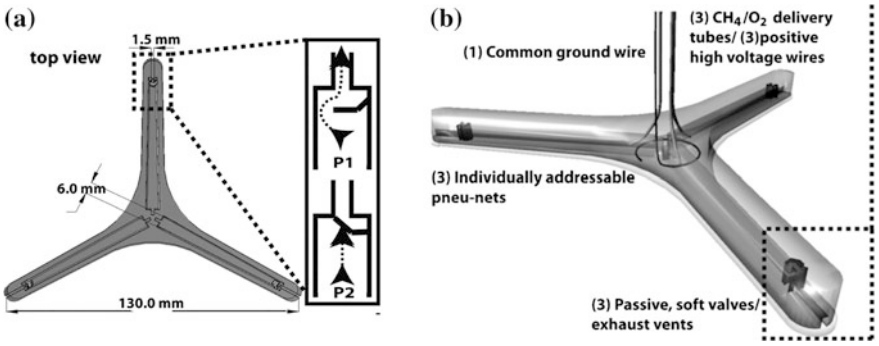


Fig. 14 The structure and control of the tripedal robot [27]

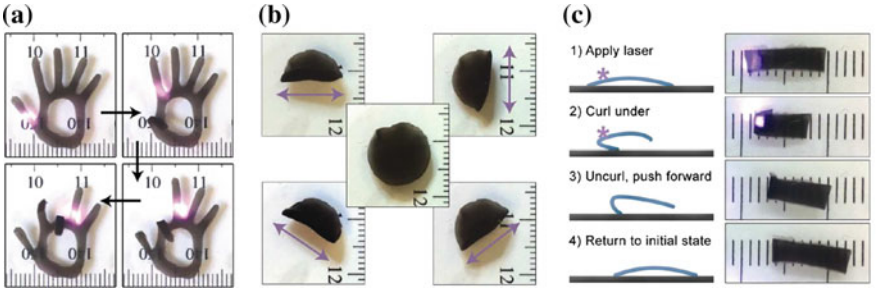


Fig. 15 Controlled motion of hydrogel actuators [53]

directions of robot can also be manipulated. Figure 15c shows a light-driven crawler. The laser is applied so as to induce gel curling. Subsequent uncurling during recovery is removed pushes the gel forward (1 mm tick marks) [53].

3.2 Active Materials

Soft-bodied robots greatly depend on the novel active materials. The currently used material which can be used to fabricate soft-bodied robots includes silicone rubber, polymer gel, acrylonitrile butadiene styrene (ABS) plastic, SMA and EAP.

Among them, silicone elastomers are adopted primarily and the most typical castable silicone rubbers are Ecoflex, Polydimethylsiloxane (PDMS), Dragonskin 20 and VTV800. Ecoflexs (Ecoflex 00-30 or Ecoflex 00-50) are mostly used for actuating layer because of its highly extensibility under low stresses, and PDMS (Sylgard 184) is also often used for strain limiting layer due to its relatively inextensibility under the stress induced by pressurization. PDMS and Ecoflex bond well with each other to form composite structures. For example, multi-gait soft robot [11] and robotic tentacles [37] are both fabricated by using these tow silicone elastomers. Similarly, Dragonskin 20 and VTV800 are elastic over a large strain, resistant to tearing, and can be prepared with a wide range of cured stiffness. For instance, GoQBot [25] is fabricated by using these tow silicone elastomers.

Polymer gel is another kind of frequently used materials, and utilizing chemical reaction to drive the robot. For example, actuator in literature [53] is fabricated by light-driven hydro-gel, similar to the optical response process of plant cells, material bends to the light side. ABS plastics are often used to make body transitions from a rigid core to a soft exterior. Meanwhile, SMA and EAP are often used to make actuators of soft-bodied robots.

3.3 Physical Modeling

Currently, the method of physical modeling of soft-bodied robots mainly includes two categories: experimental analysis and theoretical analysis [2].

The experimental analysis is the method of studying the physiological structure and kinematics characteristics of soft-bodied organisms, analyzing and building the mechanics and kinematics model of biological, and then setting up the physical model of the soft-bodied robots. Many researchers have developed models by using this method. Griethuijsen et al. used this method to measure the kinematics parameters of caterpillar crawling in horizontal and vertical condition [54]. Based on electromyography of caterpillar muscle, Woods et al. concluded the dynamic properties of caterpillar muscle curve [55]. Through measurement of the motion inspired by escape behaviors in caterpillars, Lin et al. demonstrated how to replicate it in a soft robot and provided details about dynamic analysis of the motion [25].

What's more, Mazzolai et al. presented stress-strain curve of the octopus tentacle by dissecting and analyzing its muscle structure [31] and Yekutieli et al. used three-dimensional reconstructions to analyze the octopus movements [56]. On foundation of sea cucumber's hydrostatic bone structure, Zheng et al. set up a mathematical model of driving force [26].

The theoretical analysis refers to improving the existing analysis method of the rigid robots, and then building the physical model of the soft-bodied robots. By using this method, Yiang et al. provided a finite element simulation scheme for biological muscular hydrostats [57]. Renda et al. presented a dynamic model of a continuum soft robot arm, which drives by cables based upon a rigorous geometrically exact approach [58]. What's more, He et al. used this method to establish constitutive equation of SMA [59]. To analyze the thermo-mechanical property of the SMA spring under martensitic (reverse) transformation, Mao et al. developed a deformation model of the SMA spring wire with a bias spring load [44]. Meanwhile, Bar-Cohen et al. used this method to establish constitutive equation of EPA [60]. Additionally, Kuksenok et al. modeled the photoinduced reconfiguration and directed motion of polymer gels by studying the behavior of the SP-modified gels [61].

3.4 Control Strategies

The traditional control of the rigid robot uses sensors to control of each DOF by independent feedback. The soft robots theoretically have infinite DOF, but the number of sensors and actuators is finite in any practical soft robot. At present, more mature control technology is the distributed hierarchical control and the low-level control.

The distributed hierarchical control theory is inspired by octopus arms. An octopus uses a distributed and hierarchical control system to control its multiple flexible arms. The distribution of functionalities exists between the octopus central nervous system (CNS) and the peripheral nervous system (PNS) [62]. Pfeifer et al. proposed a soft-bodied robot's design principles for intelligent systems which can enhance performance through interaction with the environment [63, 64]. Li et al. presented a sensorimotor control strategy of octopus crawling behavior [35]. The control process of this strategy could be summarized as follows: (a) the octopus visual system preprocesses visual input which covers 360° of the environment, and sends the visual information to the CNS; (b) The octopus CNS determines crawling direction and which arm(s) to use, and then initiation signals to the PNS; (c) According to the signals sent by the CNS, the PNS controls individual arms by using local embedded motor programs to achieve the crawling behavior [35]. Branson et al. discussed an algorithm based on octopus neurophysiology that utilizes division of functionality to achieve simple and robust control of continuum arm. The controller is composed of a single CNS communicating with PNS elements located in each arm [65].

The low-level control for soft robots is pressure control using pressure transducers. Most soft robots use open-loop valve sequencing to control body-segment actuation. The valve is turned on for some time to pressurize the actuator and then turned off to either hold or deflate the actuator. Many soft-robotic systems use this control method. Based on experience, Shepherd et al. showed that the multi-gait soft robot's motion can be controlled by pressurizing the PNs in sequence [11]. On the basis of this method, Zheng et al. used the time-stamped back propagation (BP) neural network to predict delay time for increasing the stability of internal model feedback control [26]. Mao et al. used this approach to control the SMA spring coils in an open-loop sequential control [44].

3.5 Manufacture Processes

The manufacture processes for soft-bodied robots include soft robot body fabrication, actuators for soft robots and stretchable electronics [66].

The soft robot body fabrication mainly includes shape deposition manufacturing (SDM), smart composite microstructures (SCM), soft lithography and other methods. SDM is a solid freeform fabrication process, which systematically combines material deposition with material removal processes [67]. With this process, rigid and flexible materials can be composed into one part. iSprawl [68] robots and Stickybot [69] robots are both developed using shape deposition manufacturing which enable the integrated use of soft and compliant parts. SCM is a manufacturing process which uses micro laser machining to cut out a flat polymer film into the required shape, to create an integrated structure. This process is especially favorable for small-scale biomimetic robots [66]. Wood et al. built a micro crawling robot using SCM [70]. Soft lithography is a group of non-photolithographic techniques that uses organic (soft) materials to transfer patterns to the substrate material [71]. Multi-gait soft robot [11], tripedal robot [27] and the fast pneu-net actuator [52] are fabricated using soft lithography. In addition, there are many robot parts fabricated from ABS plastic molds constructed with a 3D printer, such as GoQBot [25], robotic tentacles [37], the combustion-powered robot [29] and the starfish-like soft robot [44].

The actuator manufacturing for soft robot includes SMA embedding process and piezo composite actuator manufacturing [66]. The SMA embedding batch process mainly includes depositing and structuring the SU-8 socket, preparing the SMA foil, connecting the SMA foil and structuring the top protection layer, wetting chemical etching of the SMA foil, embedding the SMA with SU-8, structuring the SU-8 body and finishing the actuator element by lift-off [72]. Since the piezo actuator has extremely small deformation in applications, how to enhance the deformation of this actuator has become the focus of piezo composite actuator manufacturing. Wood et al. designed a composite type of piezo actuator by SCM process [73].

Stretchable electronics is a technology for building electronic circuits by depositing stretchable electronic devices and circuits onto stretchable substrates or embedding them completely in a stretchable material such as silicones or polyurethanes [74]. By using new conducting material or novel line structure, electrical and mechanical properties are less degraded under mechanical deformation, researchers developed many elastic electronics. The stretchable electronics for soft-bodied robot includes mesh structure, vertical ripple structure, planar horseshoe pattern and spontaneous wrinkles [73].

4 Future Research

The technology of soft-bodied robots has developed greatly and rapidly in recent years. Compared with the mollusks, there is a world of difference in the motion mechanism between the soft-bodied robots and real soft-bodied organisms. In the future, several actions of them need to be taken as follows.

(1) *Novel material for soft-bodied robots*. The currently used materials are nonlinear and difficult to be measured accurately. The novel active materials that have different mechanical properties under the different directions and ranges of pressure are supposed to be studied. Firstly, the material which is used to fabricated soft actuators or sensors should be researched and developed. This kind of material can be both deformed as the deformation of the soft-bodied robots, and used normally while the soft-bodied robots are under extreme distortion. Some of the materials can be used to fabricate actuators which can be embedded in the interior of the soft-bodied robots. Similarly, the sensors which can be attached on the surface of the soft-bodied robots can also be fabricated by such materials. In addition, the novel material can be mixed with the material which is used to make soft-bodied robots should be researched and developed. Meanwhile, the material has the function of sensing and detecting. By using this material, we can manufacture the sensor detection unit, a widespread thing in all parts of the body, which is like the nerve cell of the mollusks, and the position of the body can be measured accurately. Finally, we concluded that the research and development of materials with the characteristics of rigidity and flexibility are of great significance to the continued advancement of soft-bodied robots. These materials can produce larger deformation, and sharp objects could not easily damage it.

(2) *Modeling for soft-bodied robots*. The soft-bodied robots could produce nonlinear deformation under the function of forces and the structure of them is continuous, so it is difficult to measure their shape exactly. What's more, accurately simulating the operation of soft-bodied robots involves analysis of solid and fluid mechanics, kinematics, electro-mechanics, thermodynamics and chemical kinetics, so it is hard to set up accurate physical model of soft-bodied robots. To achieve accurate models of soft-bodied robots, the developing of the simulation platform is needed. The platform can be used for simulating the shape and position of soft-bodied robots as well as the expected function of it.

(3) *Control strategy for soft-bodied robots.* Soft-bodied robots theoretically have infinite DOF, but the number of sensors and actuators in any practical soft robot is finite. It is a challenge for us to decide what to measure and how to use the measurements to control the mobile of soft-bodied robots. Closed-loop control system of the soft-bodied robots which combines the environment feedback with the robot feedback can be introduced in the future. This system can realize accurate control of soft-bodied robots. By combining distributed neural control with classical control, the realization of accurate control of soft-bodied robots is a good direction to control algorithm for future research [8].

(4) *Manufacture technology for soft-bodied robots*. Tradition rigid connectors and electrodes cannot be used with soft structures, so new ways to connect actuators both mechanically and electrically must be developed [1]. During the process of manufacturing robot with silica gel and other materials, soft-bodied robots will deform to a large extent due to machining errors and uneven distribution of the material. Therefore, there is a strong need for more focused aspects in the method of manufacturing as well as the processing technology for soft-bodied robots.

(5) *Human-computer interaction techniques for soft-bodied robots*. The structures and motions are quite different between soft robots and humans, and it will lead to a great deal of inconvenience to human operation. Hence, due attention has to be paid to the design of hardware and software, the research of operator feedback and input devices for soft robots.

(6) *Potential application for soft-bodied robots*. The soft-bodied robots have a great flexibility and an infinite number of DOF as well as continuous deformation ability. It can be arbitrarily changed within a wide range of its shape and size. The rigid robots have the advantage of simple control, accurate position, high stabilization and reliability. By combing the soft-bodied robots with the rigid robots, the function which each robot cannot achieve alone can be achieved through coordination and cooperation. For example, snail is composed of rigid shell and internal structure which is flexible. The protection is provided by rigid shell and the forward motion is realized through internal structure. Additionally, the soft-bodied robots that are composed of rigid internal structure and soft external structure can also be designed. Rigid structure can be used to provide support force and the external structural can protect objects.

5 Conclusion

This paper provides a comprehensive overview and classification of the soft-bodied robots. Based on the functional characteristics, the soft-bodied robots can be divided into two categories: the mobile robots and the operation robots. Meanwhile, many corresponding robots are presented. Moreover, the main technologies of soft-bodied robots, mainly including actuators, active materials, physical modeling, control strategies and manufacture processes, are summarized. There are four types of the actuators: SMA, EAP, pneumatic/hydraulic actuator and chemical actuator.

The features of these actuators as well as the details of other technologies are introduced. In addition, in terms of the existing problems, the improved advices and directions for future research are proposed in this paper.

Acknowledgements This partly supported by the Natural Science Foundation of China, project No. 51275090.

References

1. Trivedi, D., C.D. Rahn, W.M. Kier, et al. 2008. Soft robotics: Biological inspiration, state of the art, and future research. *Applied Bionics and Biomechanics* 5 (3): 99–117.
2. Cao, Y.J., J.Z. Shang, K.S. Liang, et al. 2012. Review of soft-bodied robots. *Chinese Journal of Mechanical Engineering* 48 (3): 25–31.
3. Sun, L.N., H.Y. Hu, M.T. Li, et al. 2010. A review on continuum robot. *Robot* 32 (5): 688–694.
4. Rezaei, S.M., F. Barazandeh, M.S. Haidarzadeh, et al. 2010. The effect of snake muscular system on actuators' torque. *Journal of Intelligent and Robotic Systems: Theory and Application* 59 (34): 299–318.
5. Fan, B., H. Jiao, N. Jia, et al. 2014. Design and analysis of the fish-imitated underwater robot. *Machine Design Research* 30 (2): 242–247.
6. He, B., Z. Wang, Q. Li, et al. 2013. An analytic method for the kinematics and dynamics of a multiple-backbone continuum robot. *International Journal of Advanced Robotic Systems* 10 (1): 257–271.
7. You, X.D., X.B. Song, and F. Chen. 2014. Research on the classification and processing manufacturing of soft robots. *Processing Automation Instrumentation* 35 (8): 5–9.
8. He, B., Z.P. Wang, H.F. Tang, et al. 2014. Review of soft robot. *Journal of Tongji University (Natural Science)* 42 (10): 1596–1603.
9. Kate, M., G. Bettencourt, J. Marquis, et al. 2008. *SoftBot: A soft-material flexible robot based on caterpillar biomechanics*. Medford, MA: Tufts University.
10. Yuk, H., D. Kim, H. Lee, et al. 2011. Shape memory alloy-based small crawling robots inspired by C. elegans. *Bioinspiration & Biomimetics* 6 (4): 1001–1004.
11. Shepherd, R.F., I. Filip, C. Wonjae, et al. 2011. Multigait soft robot. *Proceedings of the National Academy of Sciences* 108 (51): 20400–20403.
12. Ma, J.X., M.D. Li, Z.J. Bao, et al. 1999. Micro peristaltic robot simulating earthworm and its control system. *Journal Shanghai Jiaotong University* 33 (7): 855.
13. Li, Y. 2004. *The research on an earthworm-like robot which cove hole under the earth*. Shanxi: Northwestern Polytechnical University.
14. Liu, W.T., X.S. Fang, and Y.Q. Chen. 2005. Realizing of SMA actuators for biomimetic earthworm. *Chinese Journal of Sensors and Actuators* 18 (3): 623.
15. Fu, Y.L., X.L. Li, and Z.G. Liang. 2008. Design of guiding robot for active catheter based on shape memory alloy. *Chinese Journal of Mechanical Engineering* 44 (9): 76–82.
16. Du, Y., M. Xu, E. Dong, et al. 2011. A novel soft robot with three locomotion modes. In *2011 IEEE International Conference on Robotics and Biomimetics*, *ROBIO 2011*, 98–103. Phuket: IEEE Press.
17. Koh, J.S., and K.J. Cho. 2009. Omegabot: Biomimetic inchworm robot using SMA coil actuator and smart composite microstructures (SCM). In *Proceedings of the 2009 IEEE international conference on robotics and biomimetics*, *ROBIO 2009*, 1154–1159. Guilin: IEEE Press.
18. Du, Y. 2013. *The research on the deformable soft robot with multi-locomotion modes*. Anhui: University of Science and Technology of China.

19. Menciassi, A., and P. Dario. 2003. Bio-inspired solutions for locomotion in the gastrointestinal tract background and perspectives. *Philosophical Transactions of the Royal Society of London A: Mathematical, Physical and Engineering Sciences* 361 (1811): 2287–2298.
20. Maeda, S. 2008. *A study on self-oscillating gel actuator for chemical robotics*. Tokyo: Waseda University.
21. Hara, Y., S. Maeda, R. Yoshida et al. 2009. Development of novel self-oscillating molecular robot fueled by organic acid. In *2009 IEEE/RSJ international conference on intelligent robots and systems*, *IROS 2009*, 2223–2227. St. Louis, MO: IEEE Press.
22. Nakamaru, S., S. Maeda, Y. Hara, et al. 2009. Development of novel self-oscillating gel actuator for achievement of chemical robot. In *2009 IEEE/RSJ international conference on intelligent robots and systems*, *IROS 2009*, 4319–4324. St. Louis, MO: IEEE Press.
23. Nakagawa, H., Y. Hara, S. Maeda, et al. 2010. A novel design of nanofibrous gel actuator by electrospinning. In *2010 10th IEEE conference on nanotechnology*, *NANO 2010*, 1135–1138. Ilsan, Gyeonggi-Do: IEEE Press.
24. Kimura, H., F. Kajimura, D. Maruyama, et al. 2006. Flexible hermetically sealed mobile robot for narrow spaces using hydrostatic skeleton driving mechanism. In *2006 IEEE/RSJ international conference on intelligent robots and systems*, *IROS 2006*, 4006–4011. Beijing: IEEE Press.
25. Lin, H.T., G.G. Leisk, and B. Trimmer. 2011. GoQBot: A caterpillar inspired soft-bodied rolling robot. *Bioinspiration & Biomimetics* 6 (2): 26007–26020(14).
26. Zheng, J., X. Song, Z. Jiang, et al. 2014. The driving force mechanism and control strategy of a pneumatic hydrostatic soft robot. *Robot* 36 (5): 513–518.
27. Shepherd, R.F., A.A. Stokes, F. Jacob, et al. 2013. Using explosions to power a soft robot. *Angewandte Chemie* 125 (10): 2964.
28. Sugiyama, Y., and S. Hirai. 2006. Crawling and jumping by a deformable robot. *The International Journal of Robotics Research* 25 (5–6): 603–620.
29. Bartlett, N.W., M.T. Tolley, J.T.B. Overvelde, et al. 2015. A 3D-printed, functionally graded softrobot powered by combustion. *Science* 349 (6244): 161–165.
30. Matzner, H., and Y.B. Gutfreund. 2000. Neuromuscular system of the flexible arm of the octopus physiological characterization. *Journal of Neurophysiology* 83 (3): 1315–1328.
31. Rokni, D., and B. Hochner. 2002. Ionic currents underlying fast action potentials in the obliquely striated muscle cells of the octopus arm. *Journal of Neurophysiology* 88 (6): 3386–3397.
32. Laschi, C., M. Cianchetti, B. Mazzolai, et al. 2012. Soft robot arm inspired by the octopus. *Advanced Robotics* 26 (7): 709–727.
33. Zheng, T., D.T. Branson, E. Guglielmino, et al. 2013. Model validation of an octopus inspired continuum robotic arm for use in underwater environments. *Journal of Mechanisms and Robotics* 5 (2): 1567–1576.
34. Cianchetti, M., V. Mattoli, B. Mazzolai, et al. 2002. A new design methodology of electrostrictive actuators for bio-inspired robotics. *Sensors and Actuators B: Chemical* 142 (11): 288–297.
35. Li, T., K. Nakajima, M. Calisti, et al. 2012. Octopus inspired sensorimotor control of a multi-arm soft robot. In *2012 9th IEEE international conference on mechatronics and automation*, *ICMA 2012*, 948–955. Chengdu: IEEE Press.
36. Kuwabara, J., K. Nakajima, R. Kang, et al. 2012. Timing-based control via echo state network for soft robotic arm. In *2012 international joint conference on neural networks*, *IJCNN 2012*, 1–8. Brisbane, QLD: IEEE Press.
37. Martinez, R.V., J.L. Branch, C.R. Fish, et al. 2013. Robotic tentacles with three-dimensional mobility based on flexible elastomers. *Advanced Materials* 25: 205–212.
38. Stommel, M., and W.L. Xu. 2015. Optimal, efficient sequential control of a soft-bodied, peristaltic sorting table. *IEEE Transactions on Automation Science and Engineering* 12: 1–10.
39. Stommel, M., W.L. Xu. 2015. Qualitative control of soft robotic peristaltic sorting tables. In *2015 IEEE international conference on mechatronics*, *ICM 2015*, 82–87. Nagoya: IEEE Press.

40. Shape-memory alloy. https://en.wikipedia.org/wiki/Shape-memory_alloy.
41. Menciassi, A., S. Gorini, G. Pernorio, et al. 2004. A SMA actuated artificial earthworm. In *Proceedings 2004 IEEE international conference on robotics and automation*, 3282–3287. New Orleans, LA: IEEE Press.
42. Seok, S., C.D. Onal, R. Wood, et al. 2010. Peristaltic locomotion with antagonistic actuators in soft robotics. In *IEEE international conference on robotics and automation*, *ICRA 2010*, 1228–1233. Anchorage, AK: IEEE Press.
43. Seok, S., C.D. Onal, K.J. Chok, et al. 2012. Meshworm: A peristaltic soft robot with antagonistic nickel titanium coil actuators. *IEEE/ASME Transactions on Mechatronics* 18 (5): 1485.
44. Mao, S., E. Dong, H. Jin, et al. 2014. Gait study and pattern generation of a starfish-like soft robot with flexible rays actuated by SMAs. *Journal of Bionic Engineering* 11 (3): 400–411.
45. Electroactive polymers. https://en.wikipedia.org/wiki/Electroactive_polymers.
46. Li, D.S., and L. Nie. 2007. Analysis on dynamics for flexible leg of ICPF actuated tortoise-like micro-robot. In *Proceedings of the 2007 IEEE international conference on mechatronics and automation*, *ICMA 2007*, 1102–1107. Harbin: IEEE Press.
47. Otake, M., Y. Kagami, M. Inaba, et al. 2002. Motion design of a starfish-shaped gel robot made of electro-active polymer gel. *Robotics and Autonomous Systems* 40 (02): 185–191.
48. Otake, M., Y. Kagami, and H. Inoue. 2004. Pattern formation theory for electroactive polymer gel robots. In *2004 IEEE international conference on robotics and automation*, 2782–2787. New Orleans, LA: IEEE Press.
49. Li, C., Y.H. Xie, X.Q. Huang, et al. 2013. Novel dielectric elastomer structure of soft robot. In *Electroactive polymer actuators and devices (EAPAD) 2015*, 1–5. San Diego, CA.
50. Steltz, E., A. Mozeika, et al. 2009. Jamming skin enabled locomotion. In *2009 IEEE/RSJ international conference on intelligent robots and systems*, *IROS 2009*, 5672–5677. St. Louis, MO: IEEE Press.
51. Fei, Y.Q., H.Y. Lv, and X.Y. Shen. 2013. Moving mode of modular soft robot. *Journal Shanghai Jiaotong University* 47 (12): 1870–1873.
52. Mosadegh, B., P. Polygerinos, C. Keplinger, et al. 2014. Pneumatic networks for soft robotics that actuate rapidly. *Advanced Functional Materials* 24 (15): 2163–2170.
53. Wang, E., M.S. Desai, and S.W. Lee. 2013. Light-controlled graphene-elastin composite hydrogel actuators. *Nano Letters* 13 (6): 2826–2830.
54. Van Griethuijsen, L.I., and B.A. Trimmer. 2009. Kinematics of horizontal and vertical caterpillar crawling. *Journal of Experimental Biology* 212 (10): 1455–1462.
55. Woods, W., and S.B. Fusillo. 2008. Dynamic properties of a locomotory muscle of the tobacco hornworm Manduca Sexta during strain cycling and simulated natural crawling. *Journal of Experimental Biology* 211 (6): 873–882.
56. Yoram, Y., M. Rea, H. Binyamin, et al. 2007. Analyzing octopus movements using three-dimensional reconstruction. *Journal of Neurophysiology* 98 (3): 1775–1790.
57. Liang, Y., R.M. Mcmeeking, and A.G. Evans. 2006. A finite element simulation scheme for biological muscular hydrostats. *Journal of Theoretical Biology* 242(1): 142–150 (2006).
58. Renda, F., M. Giorelli, M. Calisti, et al. 2014. Dynamic model of a multi-bending soft robot arm driven by cables. *IEEE Transactions on Robotics* 30 (5): 1109–1122.
59. He, G.P., and Z. Lu. 2005. Self-reconfiguration of under actuated redundant manipulators with optimizing the flexibility ellipsoid. *Chinese Journal of Mechanical Engineering* 18 (1): 92–97.
60. Bar-Cohen, Y. 2004. *Electroactive polymer (EAP) actuators as artificial muscles, reality, potential, and challenges*. Scitech Book News.
61. Kuksenok, O., and A.C. Balazs. 2013. Modeling the photoinduced reconfiguration and directed motion of polymer gels. *Advanced Functional Materials* 23 (36): 4601–4610.
62. Letizia, Z., S. German, A. Claudio, et al. 2009. Nonsomatotopic organization of the higher motor centers in octopus. *Current Biology* 19 (19): 1632–1636.
63. Pfeifer, R., F. Iida, and J. Bongard. 2005. New robotics: Design principles for intelligent systems. *Artificial Life* 11 (1): 1–2.

64. Pfeifer, R., M. Lungarella, and F. Iida. 2007. Self-organization, embodiment, and biologically inspired robotics. *Science* 318 (5853): 1088–1093.
65. Branson, D.T., R. Kang, E. Guglielmono, et al. 2012. Control architecture for robots with continuum arms inspired by octopus vulgaris neurophysiology. In *2012 IEEE international conference on robotics and automation*, *ICRA 2012*, 5283–5288. Saint Paul: IEEE Press.
66. Cho, K.J., J.S. Koh, S. Kim, et al. 2009. Review of manufacturing processes for soft biomimetic robots. *International Journal of Precision Engineering and Manufacturing* 10 (3): 171–181.
67. Shape deposition manufacturing. http://www.cs.cmu.edu/∼rapidproto/students.99/myh/project2/SDM.html.
68. McClung, A.J., J.G. Cham, and M.R. Cutkosky. 2004. Rapid maneuvering of a biologically inspired hexapedal robot. In *2004 ASME international mechanical engineering congress and exposition*, *IMECE*, 1195–1202. Anaheim, CA: IEEE Press.
69. Kim, S., M. Spenko, S. Trujillo, et al. 2007. Whole body adhesion: Hierarchical, directional and distributed control of adhesive forces for a climbing robot. In *2007 IEEE international conference on robotics and automation*, *ICRA'07*, 1268–1273. Rome: IEEE Press.
70. Wood, R.J., S. Avadhanula, R., Sahai, et al. 2008. Microrobot design using fiber reinforced composites. *Journal of Mechanical Design, Transactions of the ASME* 130(5): 680–682.
71. Soft lithography. http://nanotechnology.wikia.com/wiki/Soft_Lithography.
72. Leester-Schädel, M., B. Hoxhold, C. Lesche, et al. 2008. Micro actuators on the basis of thin SMA foils. *Microsystem Technologies* 14 (4): 697–704.
73. Wood, R.J. 2008. The first takeoff of a biologically inspired at-scale robotic insect. *IEEE Transactions on Robotics* 24 (2): 341–347.
74. Stretchable electronics. https://en.wikipedia.org/wiki/Stretchable_electronics.

Part II
Advanced Machine Vision Applications in Practice

Single Player Tracking in Multiple Sports Videos

Carlos Anthony B. Petilla, Gary Daniel G. Yap, Nathaniel Y. Zheng, Patrick Laurence L. Yuson and Joel P. Ilao

Abstract Performance analysis for basketball development programs are based on the athletes' movement patterns, playing position on an area, and ball acquisition. The objective of the system is to automate the performance analysis by providing raw statistical data based on the player's behaviors produced through tracking the player inside the basketball court. Using visual features described by SURF, unoccluded players are localized using a tracking by detection approach with an observed accuracy of 40%. Player positions are also adjusted to compensate for distortions, which improves player localization by 0.11%, on average.

Keywords Sports · Tracking · Feature detection · Location estimation

1 Introduction

Today, the most prominent applications of computer vision are object detection and tracking. These applications could be observed in the work of Yang, Chen et al. who implemented a multi-camera system [1] that detects and tracks individuals within a specified area. Aside from the multi-camera system, the work of Zhixing and Bhanu uses a single camera system [2] that detects and tracks individuals in a

C. A. B. Petilla · G. D. G. Yap · N. Y. Zheng · P. L. L. Yuson · J. P. Ilao (✉)
College of Computer Studies, De La Salle University, 2401 Taft Avenue, Manila 1004, Manila, Philippines
e-mail: joel.ilao@delasalle.ph

C. A. B. Petilla
e-mail: carlos_petilla@dlsu.edu.ph

G. D. G. Yap
e-mail: gary_yap@dlsu.edu.ph

N. Y. Zheng
e-mail: nathaniel_zheng@dlsu.edu.ph

P. L. L. Yuson
e-mail: patrick_yuson@dlsu.edu.ph

J. Billingsley and P. Brett (eds.), *Mechatronics and Machine Vision in Practice 3*,
https://doi.org/10.1007/978-3-319-76947-9_6

simulated crowd setting. Similar computer vision techniques are applied in a variety of applications, such as analysis of sports videos to extract information on player behavior patterns. This eliminates the constraints of manual analysis of player movement patterns inside a playing field during a game. Information from player movement patterns and player positions are useful for developing effective training programs.

Applications for sports analysis provide a more consistent and concise set of performance metrics such as player movements and player positions. SportVU [3] is a real-time sports analysis system developed by STATS for gathering basketball analytics in the National Basketball Association (NBA). The system uses six high-definition cameras to record the position of the referee, players, and the ball. SportVU analyzes the data to calculate each player's speed, traversed distance, ball possession, and position. Another similar application, SAGIT [4] is a tracking and measuring system for calculating the covered distance and the average velocity of basketball players based on their movements. SAGIT enables an automated acquisition of tracking information from the video recordings of games. To capture the whole playing field and compensate for the unpredictable events that happen during a game, all the systems mentioned for sports analysis requires high end cameras to produce high definition videos positioned at elevated locations to capture an unobstructed view of the playing field for easier tracking of each player.

Developing a tracking and detection system capable of utilizing ordinary recorded videos from the same event captured on different cameras from different vantage points eliminates the requirement for a rigid setup as mentioned above. This chapter will focus on the implementation of the core module of the developed tracking system, the *Player Tracking module*, which involves object tracking and localization.

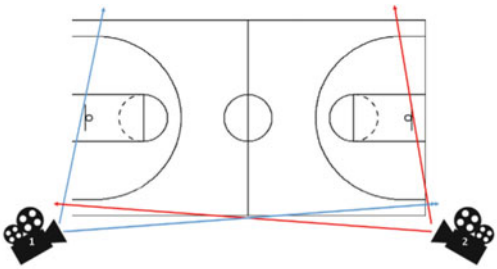
2 System Architecture and Implementation

2.1 System Setup

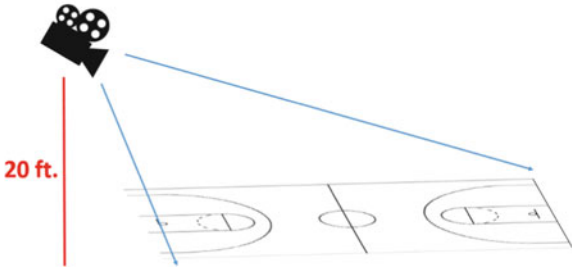
Two cameras were positioned 20 ft from the ground from two corners of the court as shown in Fig. 1. The 20 ft elevation is sufficient to capture an unobstructed view of the entire basketball court. The recorded videos were in .MOV video file format with a frame resolution of 720 × 1280.

2.2 Architectural Design

Figure 2 shows the modular architecture of the entire system. The modules comprising the entire system covers the ff. functionalities: acquiring multiple videos of



a) Two cameras located at the corners of the basketball court.



b) Cameras are elevated 20ft from the ground.

Fig. 1 Camera position and elevation for the preliminary data gathering.

the same sports event, tracking a single player throughout the duration of the video sequences, estimating the position of the player relative to the court, synchronizing all the players' location data from each corresponding video, and translating the synchronized data into player performance metrics such as player tracks, field coverage, speed and movement profiles.

3 Player Tracking

The Player Tracking module handles the object tracking and location estimation process of the system as shown in Fig. 3. After converting the videos into a set of video frames, a single player is manually selected and is tracked throughout the duration of each video. During tracking, the player's location inside the basketball court is recorded.

Some conditions need to be observed during player selection before a player can be successfully tracked: he/she must not be occluded by other foreground objects or any background structures, and the size and shape of region of interest corresponding to the player must be large enough to contain enough visual features.

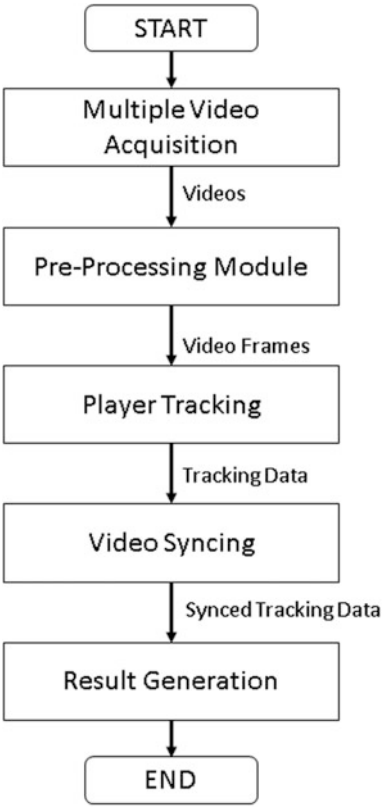


Fig. 2 System block diagram

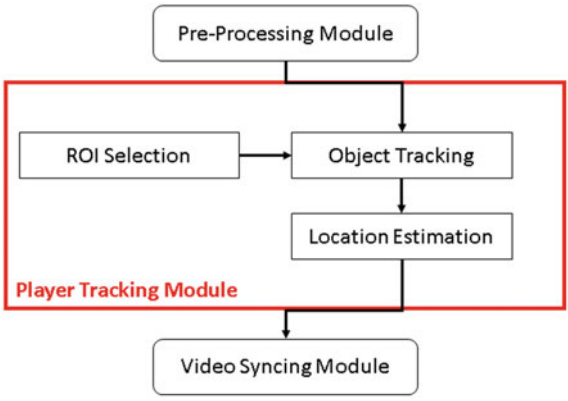


Fig. 3 Player tracking block diagram

3.1 Object Tracking

The player is tracked based on visual features described by SURF (Speeded Up Robust Features) [5]. SURF is an improved feature descriptor inspired by Scale Invariant Feature Transform (SIFT) which uses corner and blob detection. It is scale and rotation invariant, and uses approximate values for the system to process faster with nearly identical results. With the approximate values, SURF describes feature points based on its repeatability, distinctiveness, and robustness.

As shown in Fig. 4, tracking a selected region in the image frame containing a moving player is achieved by matching its corresponding SURF features in each frame. The point feature matching algorithm uses Symmetric Nearest Neighbor (SNN) to identify matching features. SNN compares the feature vector for each of the target's SURF point from the previous frame against the detected SURF points on the next frame. A match is considered if two feature vectors are similar and the distance between the features, calculated through the Sum of Absolute Difference (SAD) of the features pixel location, satisfies a given similarity threshold. The amount of visual features detected in an image is directly proportional to its size as illustrated in Fig. 5 where the selected target decreases in size as it moves farther away from the camera as the video progresses.

In an event of an occlusion, a motion estimator predicts the succeeding position of the target based on its previous location. The selected target is considered occluded if other foreground objects, represented as blobs, overlaps with the blob representing the selected target, as shown in Fig. 6.

The succeeding positions of the occluded target are determined based on prediction of the implemented motion estimator, a Kalman Filter. Assuming that the target is moving with constant velocity, the Kalman filter is modelled using the state equations provided in Eqs. 1, 2, 3 and 4.

$$px_k = px_{k-1} + vx_{k-1}dt \tag{1}$$

$$py_k = py_{k-1} + vy_{k-1}dt \tag{2}$$

$$vx_k = vx_{k-1} \tag{3}$$

$$vy_k = vy_{k-1} \tag{4}$$

where x_k contains the target's 2D coordinates, (p_x, p_y), and the velocity in each direction, (v_x, v_y).

Simultaneously, the player's real-world court position is localized thru location estimation.

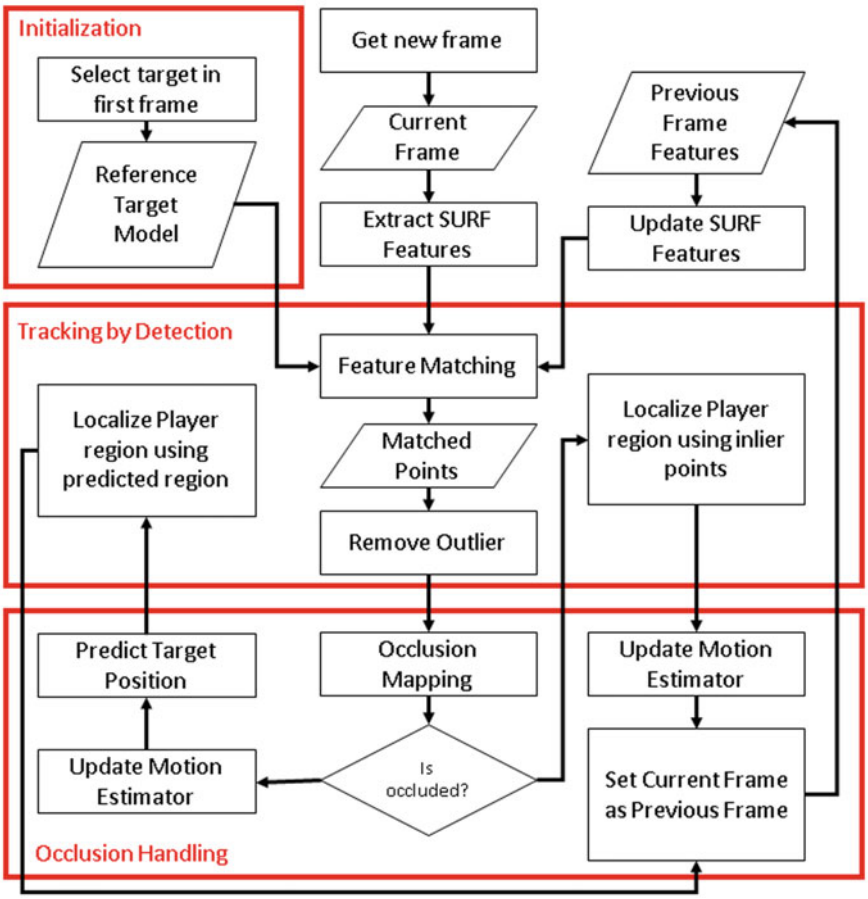


Fig. 4 Object tracking block diagram

3.2 *Location Estimation*

Figure 7 shows the implementation of the *Location Estimation sub-module*. The location estimation is performed to transform the location of a player from a 3D perspective to a 2D perspective. Given the court region extracted from an image frame showing perspective distortion, which is a part of **A**, and the dimension size of the basketball court layout image, **B**, the projective transformation matrix, λ, Eq. 6, is determined to transform the location of the player.

$$\boldsymbol{A}\lambda = \boldsymbol{B} \tag{5}$$

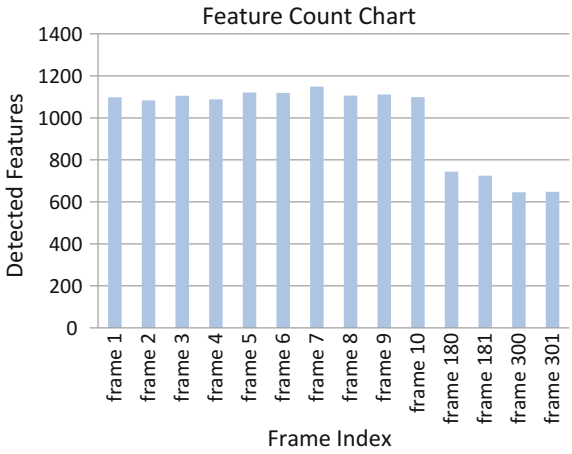


Fig. 5 Amount of features detected on a target moving further away from the camera

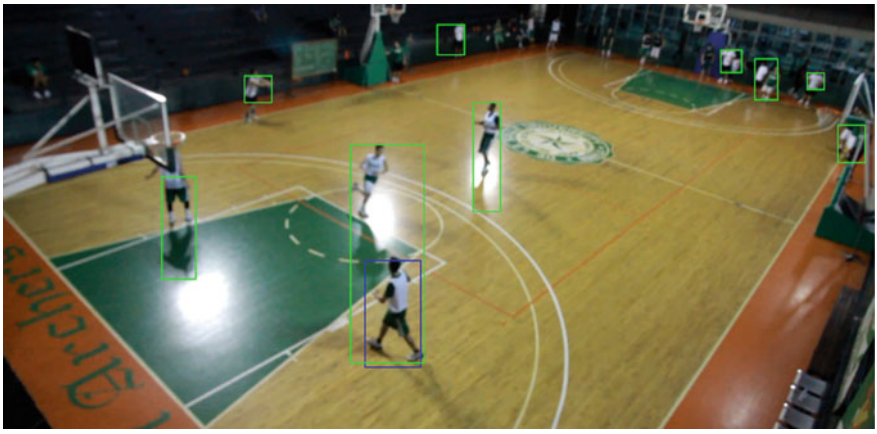


Fig. 6 Bounding boxes from blobs representing the detected foreground object. The green box represent the foreground objects while the blue box represents the target

$$\lambda = \left(\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{A}\right)^{-1}\boldsymbol{A}^{\boldsymbol{T}}\boldsymbol{B} \tag{6}$$

Player location information is converted from pixel units to physical units of length using a *Spatial calibration factor*. Given the length and width measurements of the basketball court layout image and the basketball court, the conversion factor is computed, using Eq. 7.

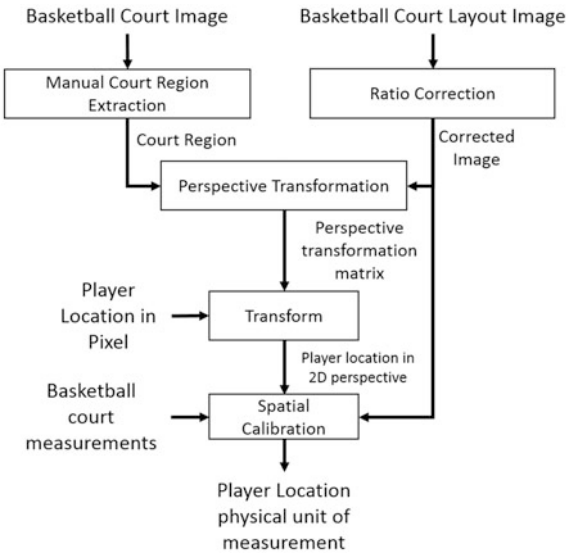


Fig. 7 Location estimation block diagram

$$Spatial\ calibration\ factor\ = \frac{physical\ measurement\ unit}{pixel} \tag{7}$$

4 Results and Analysis

4.1 Object Tracking

To quantify the tracking performance, the location and shape of the boundix boxes of tracked player locations were considered. A set of ground truth data containing ideal locations of bounding boxes for a single player, for all frames of a reference video, was manually built. The ground truth bounding boxes were compared with the system tracker's bounding box by measuring the *Spatial Overlap*, which associates the predefined ground truth bounding box to the system tracker's bounding box, as seen in Eq. 8.

$$Spatial\ Overlap\ = \frac{area(GT_{BOX} \cap ST_{BOX})}{area(GT_{BOX} \cup ST_{BOX})} \tag{8}$$

The spatial overlap will be the basis for calculating the metrics for evaluating the performance of the object tracker. These metrics are: (1) *True Positives*, (2) *False Positives*, (3) *Closeness of Tracks*, and (4) *Track Matching Error*.

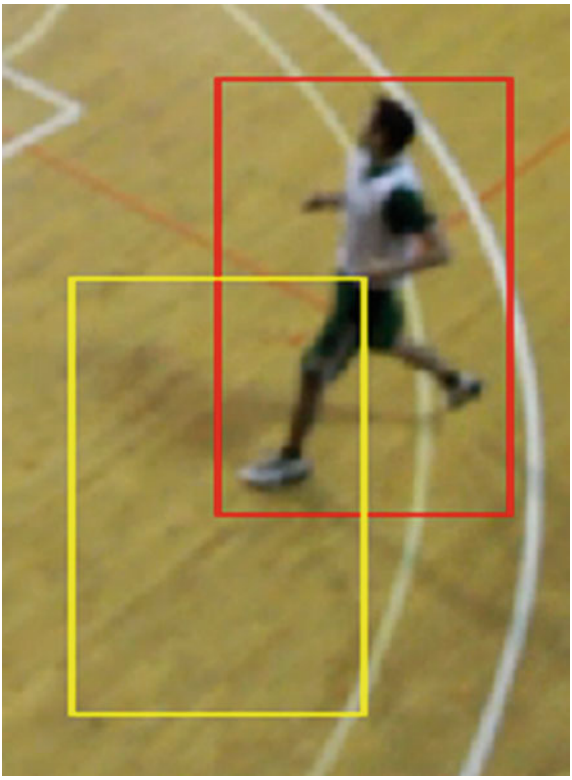


Fig. 8 The spatial overlap at 15% between the ground truth, red box, and tracking bounding boxes, yellow box

True positives and false positives provide an overview of the algorithm performance. A track is considered a true positive if the corresponding spatial overlap is greater or equal to an accuracy threshold of 15%. Anything below the specified threshold is considered a false positive. Based on [6], the threshold of 15% was chosen as different values of the accuracy threshold do not significantly affect the evaluation results. A 15% threshold is sufficient enough to not miss any possible associations for both bounding boxes as illustrated in Fig. 8.

The Closeness of Tracks and Track Matching Error indicate the accuracy of estimating the position of the selected target. The Closeness of Tracks is calculated using Eq. 9

$$Closeness\,of\,Tracks = \frac{\sum_{t=1}^{n} L(a_t) \times mean(a_t)}{\sum_{t=1}^{n} L(a_t)} \tag{9}$$

$$a_t = \{SO(GT_1, ST_1), \ldots, SO(GT_n, ST_n)\} \tag{10}$$

$$L(.) = Number\ of\ Frames \tag{11}$$

where a_t is a vector containing the spatial overlap values at time t and $L(.)$ denotes the number of frames for that specific time frame. The Track Matching Error is calculated through Eq. 12

$$Track\ Matching\ Error = \frac{\sum_{t=1}^{n} L(d_t) \times mean(d_t)}{\sum_{t=1}^{n} L(d_t)} \tag{12}$$

$$d_t = \sqrt{(x_{GT}, y_{ST}), (x_{GT}, y_{ST})'} \tag{13}$$

where d_t is the Euclidean distance between the centroids of the ground truth bounding box and the system tracker's bounding box.

Based on the evaluation metric presented, Table 1 shows the performance of the object tracker for each selected basketball video for each metric. An accuracy score of 60.7 and 64.3% was recorded from the basketball videos wherein the selected target traversed a straight path unoccluded. The object tracker was able to track the selected target throughout the entire video series as represented by the 100% true positive scores. For the 2nd target under the 1st video, although the selected target also traversed a straight path unoccluded, the poor performance of the object tracker originated from the selection of the target. As shown in Fig. 9, the selection of the target included the features from other foreground objects, which affected the tracking and matching process.

The performance of the object tracker for the 1st video under the 1st target is close to the results from the 3rd video since both trackers were lost during tracking. For the 1st video, the tracker's performance degraded due to the size change of the target as it moves further away from the camera. The amount of features matched dropped below the required amount of features for matching (3 SURF Points). For

Table 1 Evaluation metric results for each selected basketball videos and their corresponding targets

Metric	Video 1		Video 2		Video 3
	Target 1	Target 2	Target 1	Target 2	Target 1
Average spatial overlap	0.346	0.023	0.607	0.643	0.353
True positives	90.9%	7.38%	100%	100%	50.8%
False positives	9.1%	92.62%	0%	0%	59.2%
Average closeness of tracks	0.397	0.044	0.614	0.616	0.549
Weighted standard deviation of CT	0.121	0.108	0.05	0.166	0.28
Average track distance error	13.05	117.65	11.06	17.15	88.55
Weighted standard deviation of DE	6.26	41.65	4.77	13.58	119.86

Fig. 9 Selection of the 2nd target from basketball video 1

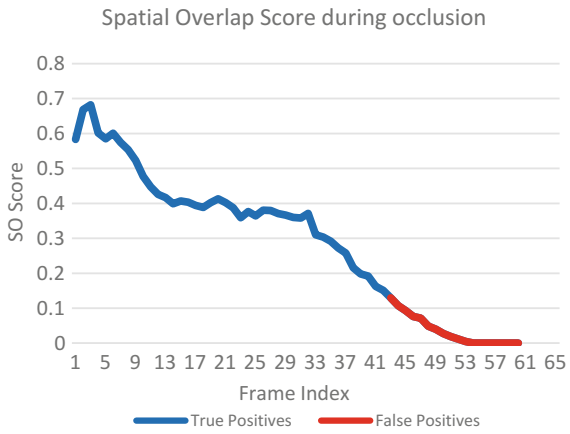


Fig. 10 Spatial overlap during occlusion

the 3rd video, the errors in tracking were caused by the dismal performance of the occlusion handler.

Based on the mentioned videos, the occlusion events only last for 30–60 frames, in which only an average of 60% from the entire duration can be considered as true positives, as seen in Fig. 10. The red line represents the amount of false positives during the event of occlusion.

The poor performance of the occlusion hander during object tracking is caused by the choice of motion models implemented for the motion estimator Kalman Filter. The motion estimation process predicts the succeeding position of the target

based on its previous location. This estimation technique does not satisfactorily handle scenarios wherein the tracked object's movement is erratic, such as in the Basketball player tracking example.

4.2 Location Estimation

In order to test the accuracy of the location estimation sub-module, two kinds of tests were performed: the small scale and the large scale location estimation tests.

For the small scale location estimation test, a flat checkerboard with 22 by 22 square cells, each measuring 2.54 cm on each side, was used. The analysis on the accuracy is performed by providing three sets of selected locations: horizontal, vertical, and diagonal locations, shown in Fig. 11. For the large scale location estimation test, the half court of the hybrid basketball court with two badminton court layouts was used, as shown in Fig. 12. The analysis on the accuracy is performed by providing the selected locations shown in Fig. 13.

For each test, two kinds of selected locations are compared: the ideal locations, which are generated based on the measurement specifications of the checkerboard and the badminton court layout, respectively, and the measured locations, which are

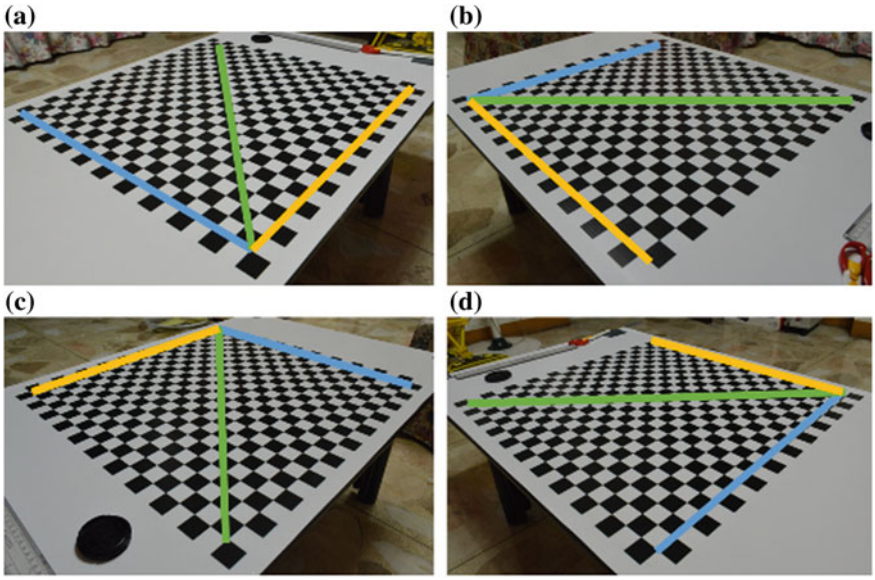


Fig. 11 Checkerboard in vantage points A, B, C, and D with diagonal (green), horizontal (yellow), and vertical (blue) set of locations

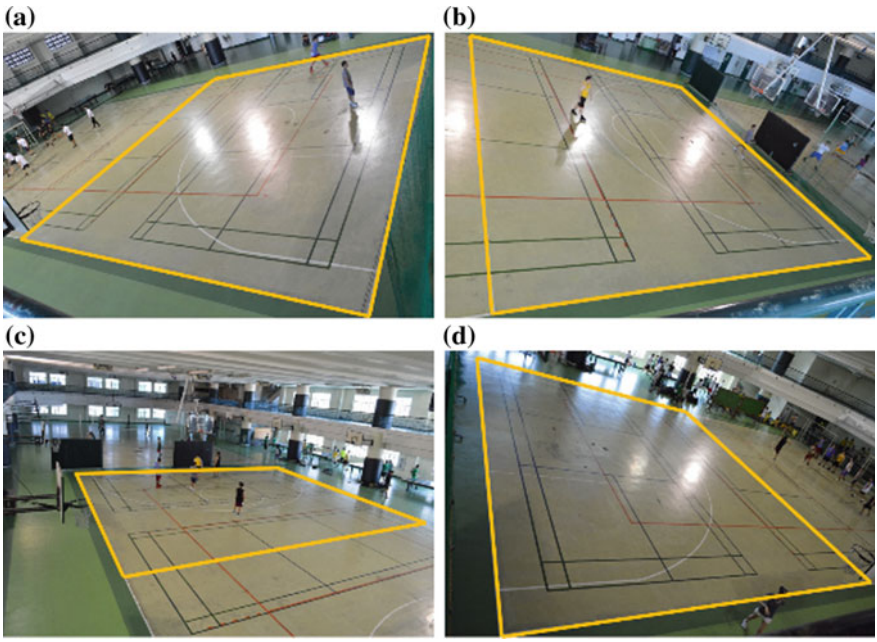


Fig. 12 Half hybrid court with court region reference viewed in vantage points A, B, C, and D

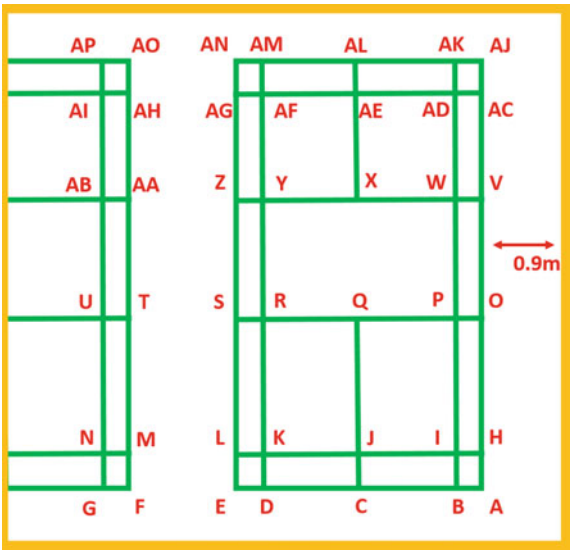


Fig. 13 Half hybrid court region with selected locations

obtained by manual plotting of selected locations. The comparisons were repeated in four different vantage points as shown in Figs. 11 and 12.

After performing the tests, the accuracy of the small scale test is 99.82%, derived from the maximum distance error and the diagonal length of the checkerboard of 79.0263 cm. On the other hand, the accuracy of the large scale test is 99.62% relative to the diagonal length of 19.6026 m, shown in Table 2. Based from the result, the small and large scales have 99% accuracy, and thus, it could be said that the location estimation is invariant to different scales.

Although the accuracy is satisfactory, large distance errors occur at random regardless of the camera location. To reduce the distance error values, the Least Squared Error adjustment algorithm is used. The use of Eqs. 14 and 15 as weights to the ideal locations of the tests in Eqs. 16 and 17 reduces the distance between the ideal and measured locations, and therefore, reduces the large distance error values as well.

$$f_x(\alpha, \beta) = \frac{x_{error} - \bar{x}_{error}}{\hat{f}(x, y)} \tag{14}$$

$$f_y(\alpha, \beta) = \frac{y_{error} - \bar{y}_{error}}{\hat{f}(x, y)} \tag{15}$$

$$X_{real} = f_x(\alpha, \beta)\hat{x} + \bar{x}_{error} + \hat{x} \tag{16}$$

$$Y_{real} = f_y(\alpha, \beta)\hat{y} + \bar{y}_{error} + \hat{y} \tag{17}$$

$$X_{adjusted} = x_{actual} - X_{real} \tag{18}$$

$$Y_{adjusted} = Y_{actual} - Y_{real} \tag{19}$$

As shown in from Table 2, the accuracy of the small scale test increased to 99.88% while the accuracy of the large scale test increased to 99.78%.

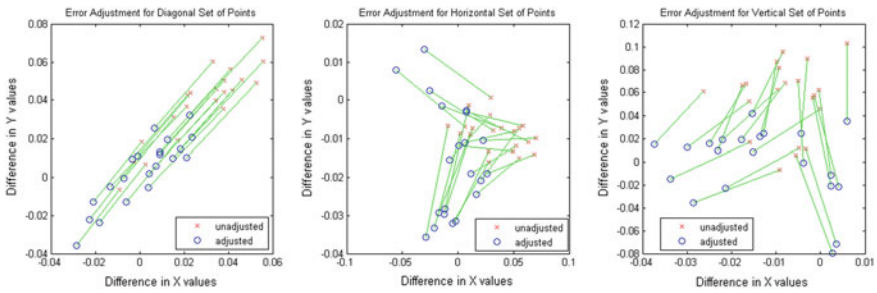


Fig. 14 Unadjusted (red) and adjusted (blue) position error values of the diagonal, horizontal, and vertical sets of selected locations in vantage point A in a scatter plot

Table 2 Average distance errors, maximum distance errors, and the percentage accuracy of the small and large scale tests results

Vantage point	Checkerboard (cm)				Half hybrid court (m)			
	Average distance errors		Maximum distance errors		Average distance errors		Maximum distance errors	
	Unadjusted	Adjusted	Unadjusted	Adjusted	Unadjusted	Adjusted	Unadjusted	Adjusted
A	0.0483	0.0266	0.1034	0.0795	0.0301	0.0168	0.0588	0.0373
B	0.0706	0.0493	0.1284	0.0913	0.0365	0.0214	0.0627	0.0423
C	0.0427	0.0410	0.0976	0.0897	0.0386	0.0246	0.0697	0.0273
D	0.0596	0.0293	0.1444	0.0653	0.0325	0.0204	0.0750	0.0300
	Percentage accuracy		99.82%	99.88%	Percentage accuracy		99.62%	99.78%

Table 3 Maximum distance errors of camera 1 and 2 including the percentage accuracy

Maximum distance errors (cm)	Camera 1	% Accuracy	Camera 2	% Accuracy
Distorted	0.3032	98.45	0.2053	98.95
Undistorted	0.0857	99.56	0.0692	99.65

Corresponding to the distance errors, using Eqs. 18 and 19, the scatter plot in Fig. 14 shows the improved distance errors.

During the tests, the errors may also be attributed to lens distortion effects on the resulting image, and inaccurate calibration parameters caused by human errors when extracting the court region and plotting of locations, due to light glares in the image, inconsistent badminton court layout, and uneven court flooring.

Because of the lens distortion problem, an additional test was conducted to determine the effect of lens distortion on the accuracy. Given a new set of ideal and measured locations, the accuracy using images not corrected for distortion, is around 98%. After undistorting the lens, the accuracy increased to 99%, shown in Table 3; which indicates that lens distortion correction improves the accuracy of tracking.

5 Conclusion

We have described a system that can provide performance analysis of a single selected player in sports videos. For this study, the sports videos used were basketball training videos recorded from two different vantage points, capturing an unobstructed view of the basketball court.

The developed system first requires selection of a player who will be tracked throughout the entire video. The system can successfully initialize the tracking process if certain conditions were met upon the player selection phase.

The Tracking process adapts the tracking by detection approach, which utilizes the visual features in an image. The evaluation of the performance of the object tracking is based on the spatial overlap between a set of ground truth bounding box and a set of system tracker's bounding box. Using the spatial overlap as the scoring metric on videos with corresponding manually built ground truth data, the performance of the object tracker is 40% for the basketball videos where the selected targets are unoccluded. For situations wherein the target is occluded, the tracker is unsuccessful since the implemented motion estimator, the Kalman Filter, can only predict the succeeding positions of the target based on its previous location. During the occlusion handling process, the tracker drifts away from the target within a time frame of 60 frames or less. From the beginning of the occlusion until the tracker is lost, only 60% of the predicted locations can be considered as correct detections.

The performance analysis of the selected player is based on the translated locations of the player from 3D to the 2D perspective on the basketball layout or template. To decrease the large distance errors between ideal and measured locations, location adjustment was performed to compensate the effects of the radial distortion caused by the lens of the recording camera. The location adjustment improved the accuracy of location estimation an average of 0.11%.

Acknowledgements We would like to extend our thanks De La Salle University Men's Basketball Team Coach Marco Januz "Juno" Sauler for sharing his expertise and insights in the field of basketball and the De La Salle Green Archers Basketball team for allowing us to record their training sessions. We would also like to extend our heartfelt gratitude to Mr. Carlo Ochotorena and Mrs. Cecile Ochotorena for sharing their knowledge and resources in the undertaking of the recording of the training sessions. This study would not have been made possible without the generous contributions of the aforementioned individuals.

References

1. Yang, T., F. Chen, D. Kimber, and J. Vaughan. 2007. *Robust people detection and tracking in a multi-camera indoor visual surveillance system.* Beijing. https://doi.org/10.1109/icme.2007.4284740.
2. Zhixing, J., and B. Bhanu. 2012. Single camera multi-person tracking based on crowd simulation. Pattern recognition (ICPR), In *2012 21st international conference*, Tsubaka, 3660–3663. ISSN: 1051-4651.
3. BIG sportVU: A Multi-dimension Player at the Big Ten Tournament. n.d. Retrieved from http://www.stats.com/casestudies/BigTenSportVU_CaseStudy.pdf.
4. Erčulj, F., B. Dežman, G. Vučković, J. Perš, M. Perše, and M. Kristan. 2008. An analysis of basketball player's movements in the Slovenian basketball league play-offs using the sagit tracking system. Informally published manuscript, Facta Universitatis. Retrieved from http://www.facta.junis.ni.ac.rs/pe/pe200801/pe200801-08.pdf.
5. Bay, H., et al. 2008. Speeded-Up Robust Features (SURF) (Revised Edition) [Online]. Available: http://www.vision.ee.ethz.ch/∼surf/eccv06.pdf.
6. Yin, Fei, Dimitrios Makris, Sergio Velastin. 2007. Performance evaluation of object tracking algorithms. In *10th IEEE international workshop on performance evaluation of tracking and surveillance* (*PETS2007*), Rio de Janeiro, Brazil.

iXray: A Machine Learning-Based Digital Radiograph Pattern Recognition System for Lung Pathology Detection

Ria Rodette G. de la Cruz, Trizia Roby-Ann C. Roque, John Daryl G. Rosas, Charles Vincent M. Vera Cruz, Macario O. Cordel II and Joel P. Ilao

Abstract A radiograph is a visualization aid that physicians use in identifying lung abnormalities. Although digitized x-ray images are available, diagnosis by a medical expert through pattern recognition is done manually. Thus, this paper presents a system that utilizes machine learning for pattern recognition and classification of six lung conditions. Classified into two categories, namely histogram-based (normal, pleural effusion, and pneumothorax) and statistics-based (cardiomegaly, hyperaeration, and possible lung nodules). Using preprocessing and feature extraction techniques, the designed system achieves an accuracy rate of 92.59% for the histogram-based lung conditions using Sequential Minimal Optimization (SMO) and 67.22% for the statistics-based lung conditions using logic operations.

Keywords Pattern recognition · Sequential Minimal Optimization
Image histogram

1 Introduction

Over the years, medical records in hospitals have stated lung problems as some of the most common ailments and causes of morbidity being diagnosed.[1] Since the number of patients diagnosed with these lung diseases continues to rise,[2] early detection can make a big difference in a patient's life.

[1]Philippine Health Statistics. "Leading causes of morbidity". Philippine Department of Health. Retrieved June 3, 2017 from http://www.doh.gov.ph/node/1482.
[2]Global Health Observatory Data Repository. "Global Burden Disease 2015". World Health Organization. Retrieved June 3, 2017 from http://www.who.int/healthinfo/global_burden_disease/estimates/en/index1.html.

R. R. G. de la Cruz · T. R.-A. C. Roque · J. D. G. Rosas
C. V. M. Vera Cruz · M. O. Cordel II (✉) · J. P. Ilao
Center for Automation Research, De La Salle University-Manila,
2401 Taft Avenue, 1004 Manila, Philippines
e-mail: mac.cordel@delasalle.ph

J. Billingsley and P. Brett (eds.), *Mechatronics and Machine Vision in Practice 3*,
https://doi.org/10.1007/978-3-319-76947-9_7

It is a radiologist's job to examine chest X-ray images and this practice is one of the basic and practical procedures in diagnosing lung abnormalities. They apply their medical knowledge in manually recognizing discernible patterns associated with these lung diseases [7]. The concept of incorporating manual pattern recognition with medical technology is extensively practiced nowadays, as in the case of chest radiography and Computed Tomography (CT-Scan) [3]. Since pattern recognition is done manually, results become subjective and dependent on the interpretation of the reader. Furthermore, radiologists or medical experts in some rural areas are not readily available to interpret the plain radiographs for the early detection of common lung abnormalities. It is therefore desirable to develop a system that can objectively identify these anomalies in chest plain radiographs.

One of the possible solutions is to use image processing and machine learning for the automatic detection of lung abnormalities. Currently, there are only few works related to automated pattern recognition on digital radiographs. Tonpho et al. [11] investigated on the use of medical knowledge and balanced histograms to analyze chest X-rays by comparing the histogram distribution of the normal lungs to the one being analyzed. Their work used active contour to segment the two lung regions (left and right) in a chest X-ray. Center of gravity (CoG) was used for verification. Their method achieved an overall accuracy of 95% with 100 samples of lung images for testing. Another work is done by Chapelle et al. [1], which used Support Vector Machines (SVM) to classify images with the use of color histograms as its features. They concluded that heavy-tailed Radial Basis Function (RBF) kernels exceed the traditional polynomial kernel and Gaussian RBF kernel in performance.

Although automatic detection of abnormalities can be achieved, the gaps between the medical field and the auto recognition process must be considered. The medical field hasn't set any finite pattern for every abnormality because of the subjectivity of medical diagnosis, making it hard to model various patterns to describe the system.

This paper presents a system that utilizes image processing, feature extraction and machine learning for the classification of six X-ray lung conditions. Recognition rates, 10-fold cross validation, confusion matrix, ROC curve, were used to characterize the designed system.

In Sect. 2, the characteristics of the lung conditions and data set are described, and image enhancement and feature extraction techniques, and machine learning algorithms are discussed. In Sect. 3, the design of the system is described in detail. Section 4 provides the results and analysis on the output of the system based on the data set used. Finally, Sect. 5 concludes the possibility of automating the classification of lung conditions with the use of machine learning and logic operation.

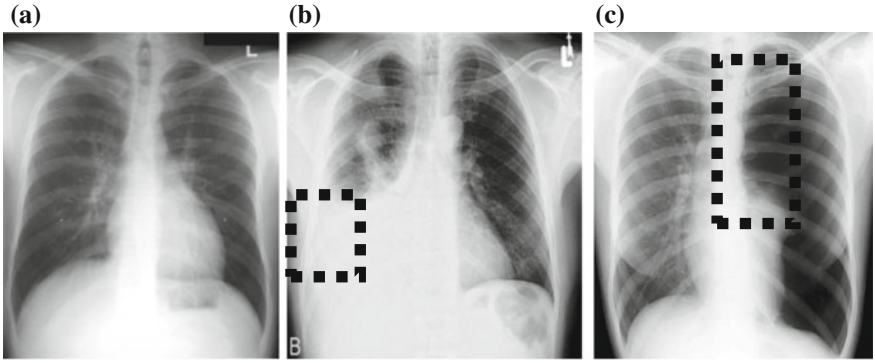


Fig. 1 Sample chest X-ray images of a **a** normal case, **b** pleural effusion case and **c** pneumothorax case. Regions enclosed in the boxes are the actual locations of the lung pathology

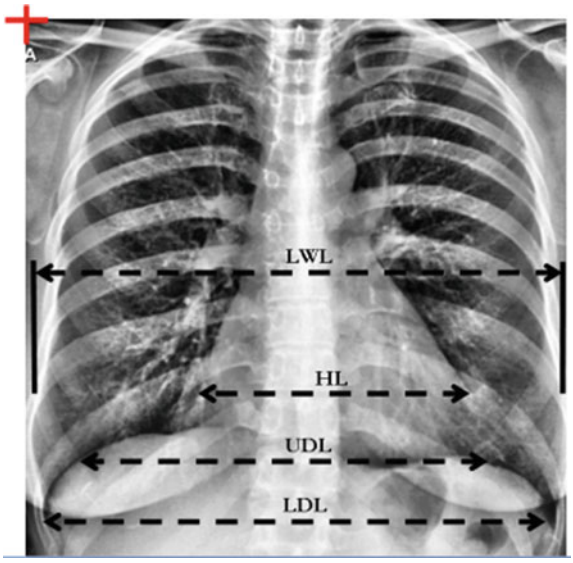


Fig. 2 Expected regions to measure for the cardiomegaly, hyperaeration and possible lung nodule cases

2 Data and Conceptual Framework

This paper focuses on six lung conditions, the of which are illustrated in Fig. 1a, b and c which correspond to chest X-ray images of normal, pleural effusion and pneumothorax cases, respectively. Figure 2 below shows the necessary measurements for detecting the other three cases, namely, cardiomegaly, hyperaeration and possible lung nodules.

In the normal case, the lungs are symmetrical and gray, and no collections of air or fluid are visible. The lung should not be smaller or blacker in appearance than the other, and lung markings should not be visible. The diaphragm is a dome-shaped

and denser (white) structure compared to the less dense (black) lungs. For the pleural effusion case, which is an abnormal collection of fluid in the outer linings (pleural spaces) of the lungs, the pattern is seen as an area of whiteness or opacity in the left or right basal part of the lungs. The pneumothorax case, which is the presence of air in the pleural spaces of the lungs, is seen as a pattern of blackness in the left or right basal part of the lungs [2].

In the cardiomegaly case, the heart is enlarged if the cardiothoracic ratio, or the ratio of the diameter of the heart (HL) from the lung wall lining (LWL) is greater than 0.5. The hyperaeration case statistics-wise, is diagnosed by the reduction in the lung markings and depression of the diaphragm. The ratio of the upper diaphragm length from the lower diaphragm length (LDL) should not exceed 0.7. For the possible lung nodules case, a detected circle or spot with a diameter larger than three centimeters is considered as unwanted 'mass' inside the lungs [2].

The digital radiographs used in this paper are retrospective. To ensure fully developed lungs, this study made use of digital chest radiographs of adults, i.e. ages between 18 and 50 years old. For the data used in the system, digital radiographs were gathered from the De La Salle Health Sciences Institute (DLS-HSI). The images were converted from a Digital Imaging and Communications in Medicine (DICOM) format to a Tagged Image File Format (TIFF) which is a lossless image format, considering the memory space that the system will consume in doing its routines. The number of images gathered for the histogram-based lung conditions are 58 each. Of the total number of images for each lung abnormality, 68.79% is used as the training data set while the remaining 31.03% is used as the testing data set. For the statistics-based lung conditions, 30 images for each condition were gathered and tested against 30 normal images.

Lung nodules are typically circular in forms such that circle detection was tested for the feasibility of automating the detection of possible lung nodules. Circular Hough Transform (CHT) is a mapping algorithm that is performed to detect the presence of circular shape or objects in an image. According to Pedersen [9], CHT is a transformation of a point in the image space to the parameter space defined according to the shape of the object of interest, in this case, a circle.

Since abnormality in a radiograph is manifested in the appearance of whiteness (mass or liquid) and blackness (air) on the image, grayscale information are mainly considered as primary features in this work. A histogram of a grayscale image represents the frequency distribution of gray levels. The features used in this study are the *mean* and the *skewness* of the histogram. It is said to be right skewed if there are extreme values towards the right side of the distribution and this has a mean greater than the mode. A left skewed distribution, conversely, has a mean which is less than the mode because the left side of the distribution contains extreme values.

Support Vector Machines (SVM) are related supervised learning methods used for classification and regression. Classification involves the separation of the data into training and testing sets. In the training set, each instance has one target value (class labels) and numerous attributes (features); these constitute the training vectors which are mapped into a higher dimensional space SVM then constructs a separating hyperplane with a maximum margin [4, 12]. A kernel function corresponds to a dot

product in a high dimensional feature space. Two out of four kernels utilized by SVM are commonly used and their parameters have a significant effect on decision boundary. The width parameter of the Gaussian RBF kernel and the degree of the Polynomial kernel control the flexibility of the resulting classifier in fitting the data.

The goal of SVM is to produce a model which predicts the target values of the test data given only the test data attributes. For this work, both Polynomial kernel and RBF kernel were used for comparison.

Polynomial kernel allows modeling of feature conjunction depending on the degree of the polynomial; higher degree polynomial kernels allow a more flexible decision boundary. The kernel is a directional function utilized for classification of normalized training data sets. The output of the system depends on the dot product inside the kernel. The function [5] can be represented by the formula below:

$$K(X, Y) = (\alpha \boldsymbol{x} \boldsymbol{T} \boldsymbol{y} + c)^d \tag{1}$$

From (1), $\boldsymbol{x}$ and $\boldsymbol{y}$ are the vectors in the input space, α represents its magnitude. The constant term c trades of the influence of higher-order versus the lower-order terms in the polynomial. Lastly, d is the polynomial degree.

RBF kernel permits a straightforward analysis of the interpretation produced by the hidden layer which is based on the dot product between the input vector and a weight vector. The hidden layer executes a non-linear transformation to the input space. The output layer does linear regression to predict the desired targets. The function [5] can be represented by the formula below:

$$K(X, Y) = e^{(-\gamma \times |\boldsymbol{x} - \boldsymbol{y}| \wedge 2)} \tag{2}$$

From (2), $\boldsymbol{x}$ and $\boldsymbol{y}$ are the vectors in the input space. Gamma (γ) is an adjustable parameter. Increasing the value of gamma will result into a smoother decision surface and a more regular decision boundary.

Sequential Minimal Optimization (SMO) is an algorithm that solves the SVM Quadratic Programming (QP) problems without using optimization steps. In relation to this, Platt [8] expands SMO definition by decomposing QP problems into sub-problems.

The advantage of SMO as explained in [8] is that sub-problems in the system can be analyzed to resolve the overall QP problem with the least time possible. Thus, there can be more optimization sub-problems to be understood using the said algorithm and additional storage is not needed for data gathering.

3 System Design and Methodology

Figure 3 shows the block diagrams for the designed system during the *training phase* and the *testing phase*. The *training phase* of the system uses the extracted features from the histogram-based lung conditions to create a model of the lung diseases.

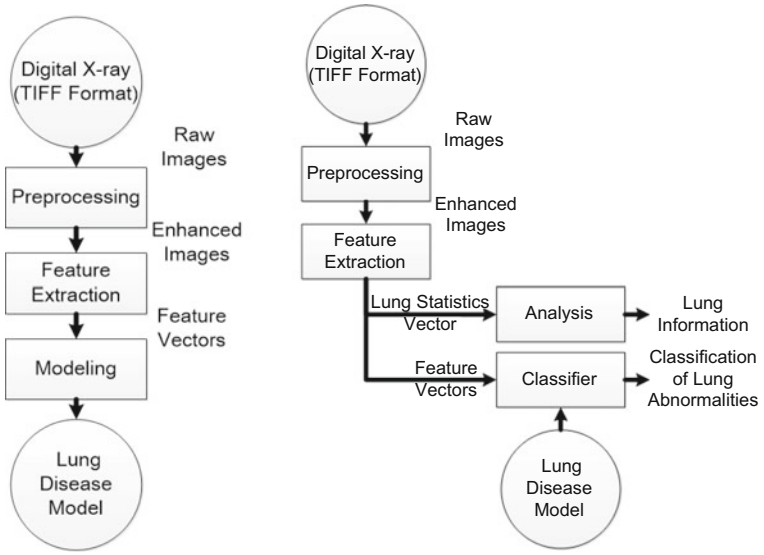


Fig. 3 Block diagram for the training phase (left) and testing phase (right)

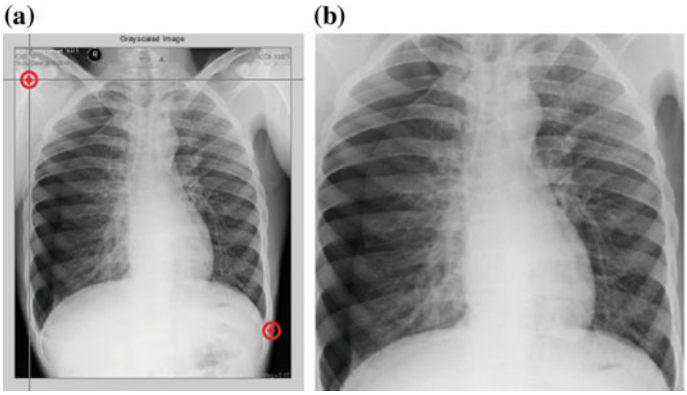


Fig. 4 Sample interface for the **a** selection of the Region of Interest (RoI), and **b** the cropped and resized portion of the selected region of interest

The *testing phase* uses the lung disease model from the previous phase to classify the histogram-based lung conditions given the extracted features. In addition, it implements logic operation to classify the statistics-based lung conditions.

This module provides an improved image such that features for extraction in the next block are emphasized. First, it applies *Image Normalization* where images are cropped and standardized to a certain dimension. The left side of Fig. 4 shows an example where the coordinates are defined and its right side shows a sample output normalized image. *Grayscaling* follows where images are converted to their 8-bit

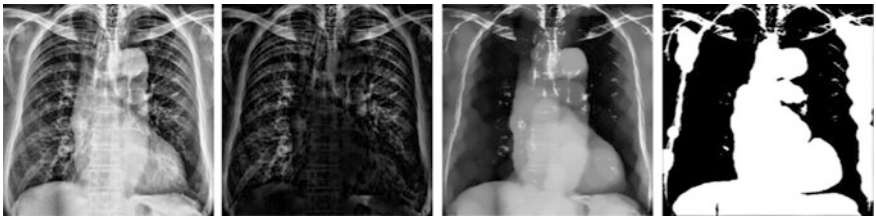


Fig. 5 A raw input image (left most), after top hat filtering (second from the left), and desired enhanced image (second from the right), and application of binarization (right most)

gray level equivalent to have uniformity in data. It is applied to remove the hue and saturation from the X-ray image. The grayscaled image then undergoes *Histogram Equalization* which uses Contrast-limited Adaptive Histogram Equalization to transform the values in the intensity image by improving the contrast per small region rather than the whole image.

Three image enhancement processes are applied to the images depending on the type of abnormality that is being analyzed. For the statistics-based lung conditions, image enhancement for the Feature Measurement is done for the hyperaeration case. It includes *Top Hat Filtering* (refer to Fig. 5 for illustration) which clips unwanted low-intensity pixels.

For the Lung nodule case, the same Top Hat filtering is also applied but with an addition of the *High Boost Filtering* (see Fig. 6 for illustration) that amplifies high-intensity elements found in ribs, lung walls and nodule-like entities inside the lung area. It also converts the histogram equalized image such that high spatial frequency components are heightened to make the nodules appear in high intensity for easier shape measurement.

For the histogram-based lung conditions, image division subdivide the cropped image into 100 smaller boxes of size 50 pixels by 50 pixels to locate further the region of abnormality. Furthermore, since patterns appear as the intensity of

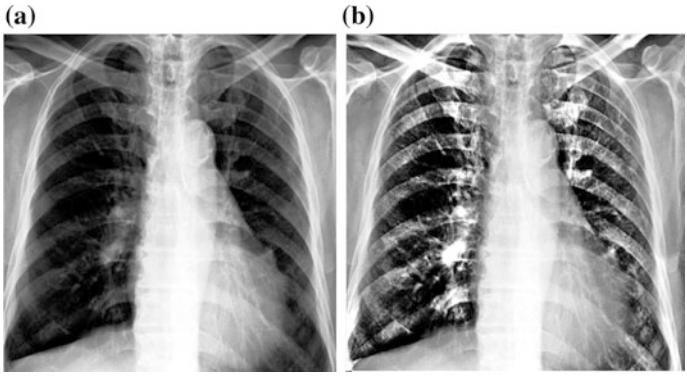


Fig. 6 Sample of a raw image **a** input image, and a **b** high-boosted image for lung nodule detection

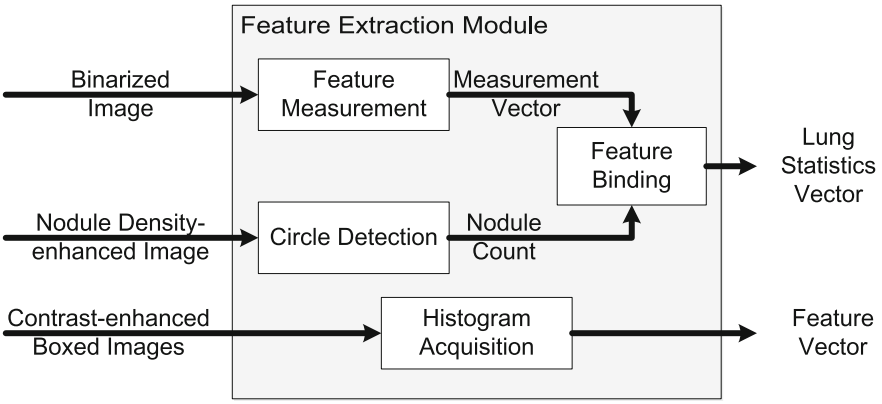


Fig. 7 Block diagram of the feature extraction module

whiteness and blackness in an X-ray image, the mean and skewness of the histogram for each smaller box are used as features for modeling and automatic classification of each condition.

The feature extraction module returns the merging Measurement Vector submodule that contains statistics from the measurement heart size, lung wall distance, measurement of the top and bottom of the diaphragm, and Circle Detection submodule that returns the number of detected possible lung nodules. Only the number of lung nodules is produced because most diagnosis only states the presence of a lung nodules rather than the size and its location. Lastly, Histogram Acquisition submodule returns Feature Vector and was retained (without vector reduction) in the design.

As shown in Fig. 7 feature binding is only used for the vectors that build up the Lung Statistics vector. It will not bind all features to minimize system confusion since only the histogram-based features will undergo machine learning.

Based from Fig. 2, on how to measure the essential vectors for the Cardiothoracic Ratio and the Diaphragm Ratio, the heart and lung walls are isolated from the smoothened image by binary padding to selected x and y values, removing other features and retaining the heart and the lung wall linings. Blobs are then formed to identify the heart and lung wall linings systematically. The system looks for the largest blob that corresponds to the heart, and its longest row, or the heart size (HL), obtaining the rightmost pixel equivalent to the apex of the heart. The lung wall linings are obtained from the row vector of the binary image equivalent to the longest row index in the heart blob. Through blobbing, the left and right walls are acquired. From there, the cardiothoracic ratio (CR) can be obtained using the heart size and the lung wall distance.

The diaphragm ratio (DR) can be computed through the same process of acquiring the cardiothoracic ratio. As shown in Fig. 2, the diaphragm area is isolated through binary padding, and then converted to a blob. The lower diaphragm length (LDL) is usually the longest row of the blob, adjacent to the ends of the left

and right lung spaces. The apex of the left and right slopes found on the upper area of the diaphragm is considered the endpoints of the upper diaphragm length.

The Circle Detection submodule also uses blobs to separate the regions of interest with the unnecessary features. Circular Hough Transform is utilized to detect circular entities in the image. The system, instead of finding the diameters and locations of the possible lung nodules as proposed, counts the number of possible lung nodules found in the image.

The Histogram Acquisition submodule returns the mean and skewness of the histogram, through the 100 sub-images from Image Division, which are used as representative information (blackness and whiteness) for each segment. Using mean and skewness values to represent the histogram reduces the vector size since the said technique is a linear transformation of the histogram data. It is also a way to avoid data overload the machine learning algorithm since a histogram vector of size equal to 25,600 (256 grayscale levels $\times$ 100 sub images).

The Analysis Module analyzes the Lung Statistics Vector that contains essential data gathered from the Feature Extraction module. It is done by using arithmetic and logic operations. In general, this module returns Lung Information specifying the detected number of possible lung nodules, the heart size, the cardiothoracic ratio, if the patient has cardiomegaly, and the diaphragm ratio, if the patient manifests hyperaeration.

The Modeling module and the Classification module are machine learning modules that make use of the feature vectors produced by the Feature Extraction module. The Modeling module is used for the training stage of the system. It outputs a lung disease model created by the system during its training phase. The Classification module used the lung disease model created by the training stage as its basis in matching or classifying the current image being tested to a specific lung abnormality. It then flags the image accordingly.

Table 1 True Positve (TP) and False Positive (FP) rates of RBF kernel with different *c*'s

c	*G*	TP rate	FP rate
1	0.01	0.907	0.046
10	0.01	0.926	0.037
100	0.01	0.852	0.074

Table 2 True Positive (TP) and False Positive (FP) rates of polynomial kernel with different *c*'s

c	*E*	TP rate	FP rate
0.01	1	0.889	0.056
0.1	1	0.926	0.037
1	1	0.87	0.065

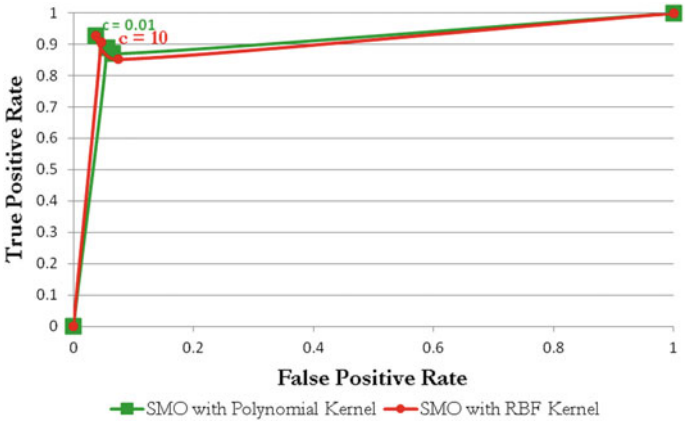


Fig. 8 ROC curve

4 Results and Analysis

This study used MATLAB for image preprocessing and feature extraction and WEKA for machine learning. With a sample size of 174 images for the histogram-based lung conditions, 40 images for each case were used for the training phase while 15 images for each case were used for the testing phase.

The test using RBF kernel and Polynomial kernel were tested with gamma value of 0.01 and exponential degree of 1, respectively, for different *c*'s.

Based on the results shown by Tables 1 and 2, the RBF kernel using c = 10 gave the highest True Positive (TP) rate of 92.59% and lowest False Positive (FP) rate of 3.7%, showing the most desirable curve in the receiver operating characteristic (ROC) plot in Fig. 8, while Polynomial kernel gave the same results using $c = 0.1$, with a TP rate of 92.59%.

The RBF kenel and Polynomial kernel obtained the same Sensivity rate of 97.2%, False Positive rate of 5.6%, False Negative rate of 2.8%, and Specificity rate of 94.4%. Given the results of the Histogram Acquisition tests, the system is good

Table 3 Sensitivity and specificity table for statistics-based lung conditions

Lung condition	Sensitivity (%)	FP (%)	FN (%)	Specificity (%)	Accuracy (%)
Cardiomegaly versus normal	66.7	42.9	33.3	73.3	70
Hyperaeration versus normal	93.3	46.7	6.7	53.3	73.3
Lung nodules versus normal	83.3	66.7	1.7	33.3	58.3
Abnormal versus normal	81.1	46.7	18.9	53.3	67.2

at detecting the true positives given its high sensitivity values and low false positive rates. Moreover, the system performs highly at avoiding false alarms with its high specificity rates. Having high sensitivity rate and specificity rate reflects a good system performance. Given the high specificity rate of the system, there is little chance for the system to overlook images present with an abnormality.

A test having a high specificity and sensitivity rate suggests that the system is good at avoiding false alarms. For the statistics-based lung condition tests, as listed in Table 3, the system is efficient in detecting both hyperaeration and Lung Nodule cases given its high sensitivity values of 93.3 and 83.3%, respectively. However their specificity values are relatively low. Conversely, the cardiomegaly cases obtained a sensitivity value of 66.7%, and a higher specificity value of 73.3%, which suggests that the system has yet to be modified to increase the detection rate of cardiomegaly cases, but it identifies normal cases against cardiomegaly cases well. The system misclassified 10 cardiomegaly cases as being normal cases. It is more desirable to classify normal images as having disease rather than failing to detect diseases. To have a good test, the system must have a low rate for False Negatives to avoid having undetected diseases. For the Lung Nodule cases, the tests obtained a specificity rate of 33.3%. The system fails to detect the majority of the normal cases as being normal, classifying them to be a Lung Nodule case.

The SMO function in WEKA is employed and two kernel functions are compared. Shown in Tables 4 and 5 are the number of correct and false matches for the histogram-based and statistic-based lung condition tests, respectively.

Given the correct matches and false matches of the tests using the 2 kernels, using Polynomial kernel gave a more desirable result as compared with using RBF fkernel. Although both kernels achieved the same accuracy rate of 92.59% for the normal versus pleural effusion versus pneumothorax case, the Polynomial kernel was able to correctly classify all images under the normal versus pleural effusion case and pleural effusion versus pneumothorax case while RBF kernel misclassified

Table 4 System accuracy and error rate in classifying histogram-based lung conditions

Histogram-based lung conditions	RBF kernel		Polynomial kernel	
	Correct match	False match	Correct match	False match
Normal versus pleural effusion versus pneumothorax	50 (92.59%)	4 (7.41%)	50 (92.59%)	4 (7.41%)

Table 5 System accuracy and error rate in classifying statistics-based lung conditions

Statistics-based lung conditions	Correct match	False match
Cardiomegaly versus normal	42 (70%)	18 (30%)
Hyperaeration versus normal	44 (73.33%)	16 (22.67%)
Possible lung nodules versus normal	35 (58.33%)	25 (41.67%)
Abnormal versus normal	121 (67.22%)	59 (32.78%)

1 image and 3 images for the normal versus pleural effusion case and pleural versus pneumothorax case, respectively.

The system's low performance in detecting cardiomegaly can be attributed to the preprocessing applied to the image because the edge for the lung wall is not correctly represented. In the process of the binarization, the system adds some details of the pleural spaces that are not part of the lung wall lining. In effect, the system cannot extract the approximate lung wall distance.

The error rate in detecting hyperaeration is primarily because of the diaphragm extraction techniques used. As discussed in the System Implementation, the point where the upper diaphragm lies is obtained by getting the average of all x values of the diaphragm outline. Averaging does not point to the top part of the diaphragm but rather it points at the middle part since averaging is a statistic technique for getting the central tendencies. Most of the time, the system ratio has a higher computed ratio since the upper diaphragm length that the system retrieved is not the shortest (i.e. the top diaphragm) since it is an average of the points along the x-axis.

The low accuracy in detecting possible lung nodules is due to cropping and blobbing issues. Stretching the image affects the shape of the possible nodules, making it hard to be recognized by the algorithm. In addition, other elements in the lungs such as blood vessels and air sacs may be perceived as circular entities in the image, causing the system to mistake them as lung nodules.

Table 6 Confusion matrix in classifying histogram-based lung conditions using RBF kernel

Lung conditions		SMO-based predicted class		
		Normal	Pleural effusion	Pneumothorax
Actual class	Normal	17	0	1
	Pleural effusion	0	18	0
	Pneumothorax	1	2	15

Table 7 Confusion matrix in classifying histogram-based lung conditions using polynomial kernel

Lung conditions		SMO-based predicted class		
		Normal	Pleural effusion	Pneumothorax
Actual class	Normal	17	0	1
	Pleural effusion	0	18	0
	Pneumothorax	2	1	15

Table 8 Confusion matrix for cardiomegaly and normal cases

Lung conditions	Cardiomegaly	Normal
Cardiomegaly	20	10
Normal	8	22

For the normal versus abnormal test, the system has a high error rate because normal images were immediately classified as hyperaeration. The reason for this is again attributed to the technique used in extracting the diaphragms.

As seen in Tables 6 and 7, it can be observed that both RBF kernel and Polynomial kernel performs well in classifying the images according to their correct labels; having correctly classified 50 out of 54 images. This is because the kernel parameters used in the testing was selected to be the best among other parameter sets. Given that the two kernels used the same set of images, it can be analyzed that the system gets confused when a disease is added to the matrix. It also means that it cannot differentiate the patterns that well when all three histogram-based lung conditions are being tested for.

From Table 8, it shows that it can classify images cases with cardiomegaly perfectly but is confused on labeling some normal images. This confusion can be attributed to the preprocessing applied to the image.

Table 9 shows that the system has difficulty with the latter test case, as it classifies 100% of the images incorrectly, categorizing normal lung images as having hyperaeration. This maximum error rate may be attributed to the state of the images before measurement, since in obtaining the upper region of the binary image of the diaphragm, the average row of intersection from the heart to the lowest point of the diaphragm on each side is acquired, and the images are unequal in terms of cropped area.

Table 9 Confusion matrix for hyperaeration and normal cases

Lung conditions	Hyperaeration	Normal
Hyperaeration	28	2
Normal	30	0

Table 10 Confusion matrix for lung nodules and normal cases

Lung conditions	Lung nodules	Normal
Lung nodules	25	5
Normal	20	10

Table 11 Confusion matrix in classifying statistics-based lung conditions

Lung conditions	Abnormal	Normal
Abnormal	73	17
Normal	42	48

Table 12 10-fold cross validation for the histogram-based lung conditions using SMO with RBF kernel

F	1	2	3	4	5	6	7	8	9	10	OA
AR	67	75	92	83	58	92	92	83	83	92	**81.7**

The value in bold is the overall accuracy of the validation.

Table 13 10-fold cross validation for the histogram-based lung conditions using SMO with polynomial kernel

F	1	2	3	4	5	6	7	8	9	10	OA
AR	83	92	83	100	75	100	92	75	83	100	**88.3**

The value in bold is the overall accuracy of the validation.

From Table 10, the system refers its output to the diagnosis of De La Salle Health Science Institute only. The table shows that the system detects circular entities in images that are diagnosed as normal. This is due to the pixelation of the images caused by image resizing. Circular Hough Transform detects circular entities formed by high-intensity pixels. For that matter, there can be blood vessels that are mistaken by the system to be lung nodules.

The system refers its output based on the results of the individual tests of the statistics-based lung conditions. The confusion matrix for abnormal cases as seen on Table 11 shows that the system is able to identify images positive of an abnormality or lung conditions with a high correct match of 73 but it performs not so satisfactorily in detecting those images not having diseases with 42 incorrectly classified images.

Ten (10)-fold Cross Validation is used to determine if the machine learning-based system used overfits the training data set. The overall accuracy results of the normal vs pleural effusion vs pneumothorax tests, as seen in Tables 12 and 13, decreased compared to the results obtained from the 10-fold cross-validation tests of each lung conditions case tested against a normal case which have overall accuracies ranging from 83.8 to 96.3%. The low accuracy rates means that the training data set for the system cannot distinguish patterns from normal and pneumothorax case.

From the tables above, an overall accuracy of 81.7 and 83.3% for both RBF kernel and Polynomial kernel were obtained despite having a 91–95% range when the lung abnormalities were classified separately.

J48 was used to make a Decision Tree from the classification of Histogram-based lung nodules to specify how the classifier uses the data's attributes such as mean and skewness of each of the image's box, in this case, in making a decision for the histogram-based lung conditions.

The J48 classifier works by determining which class falls under a certain range value of mean and skewness. It looks for a common ground among different instances in an attribute that effectively divides the set of sample, and from there, the classifier decides whether the value is categorized under a specific class. As an example, from the parent node 'mean_64' of the decision tree seen in Fig. 9, the system classifies instances with a mean value greater than 157.53 as pleural effusion, the value less than 157.53 is then further evaluated. This result can be related to the characteristic of a pleural effusion case, a whiteness of the pleural spaces of the lungs, which would usually have a higher value for the mean of its histogram levels compared to other cases because of images whiteness.

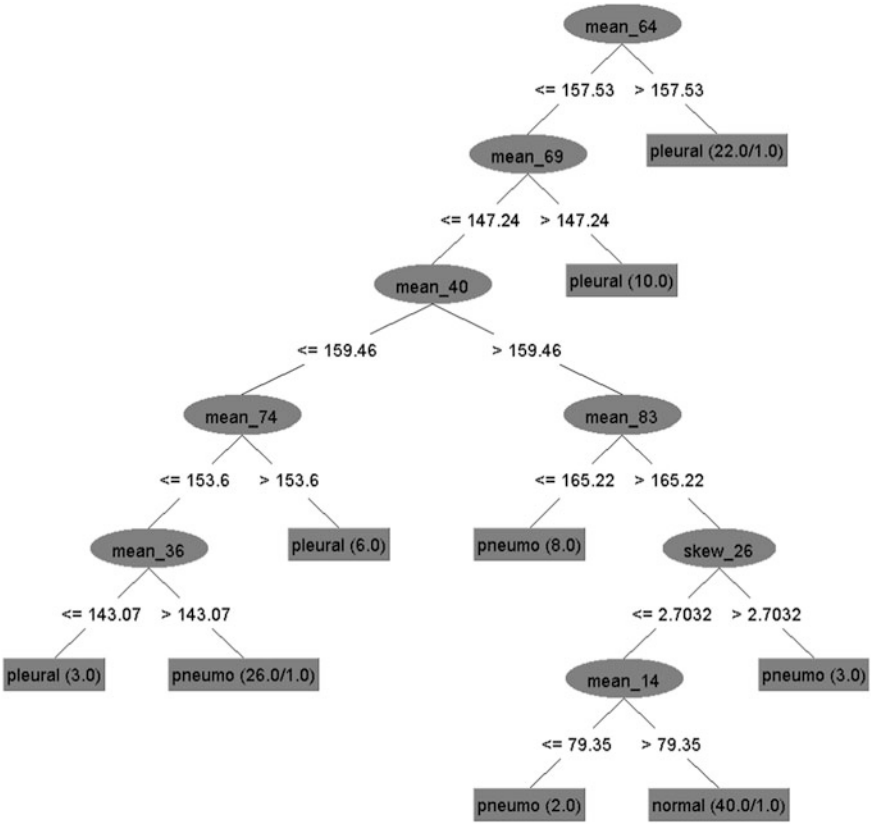


Fig. 9 J48 resulting pruned tree

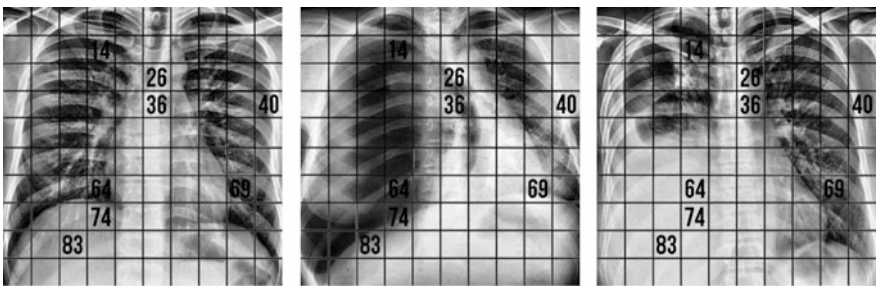


Fig. 10 Determinant boxes for normal image (left), pleural effusion image (middle), and pneumothorax image (right)

Figure 10 shows in what boxes in the image the J48 classifier decides on the mean and skewness attributes. As seen in the figures, the boxes analyzed consist of corners, edges and other distinctive locations in an X-ray image. As an example in comparing normal case versus pleural effusion case, it can be observed that in box 64, a normal case would have a lower mean value compared to the pleural effusion case because a pleural effusion case would have higher intensity values due to the white pixels covering the entire box. For the normal versus pneumothorax case, a pneumothorax case would have a lower mean value compared to the normal case because of the evident amount of black pixels in the same box.

The J48 classifier attained an accuracy rate of 74.08% with 40 correctly classified instances and an error rate of 25.9259% with 14 incorrectly classified instances in classifying histogram-based lung conditions. This accuracy is lower than the accuracy attained in using SMO which is a supervised learning method.

5 Conclusion and Future Work

This paper presented a machine learning-based system that classifies six lung conditions namely: normal, pleural effusion, pneumothorax, cardiomegaly, hyperaeration and possible lung nodules. For the preprocessing stage of Histogram-based Lung Conditions (normal, pleural effusion, and pneumothorax), we applied Contrast-limited Adaptive Histogram Equalization and box method. For the preprocessing stage of Statistics-based Lung Conditions (cardiomegaly, hyperaeration, and possible lung nodules), we applied Top Hat Filtering, High-Boost Filtering, and Circle Detection, respectively. For the classifier, SMO using Polynomial kernel and RBF kernel were both tested. The system obtained an overall accuracy rate of 92.59% for both kernels. With the accuracy rate of 92.59% for both kernels, the use of mean and skewness to represent an image's histogram is an effective way of linearly reducing the dimension of the histogram data and analyzing the variation among images.

For the statistics-based lung conditions, the use of blobs to represent an image's heart, lung wall and diaphragm is an effective means to measure the lengths of features in an image. The measurements obtained are needed to acquire the ratios that determine the presence of abnormalities. From here, the system's accuracy rate is 70% for the cardiomegaly case. Moreover, the test for the hyperaeration case has an accuracy rate of 73.3% which means that the system was able to classify majority of the images with an abnormality.

Circular Hough Transform is confused in detecting circular entities in an image as given by its 58.33% performance in classification. However, literature warns that CHT might not capable of distinguishing lung nodules from a blood vessel and an air sac. Comparing the system performance with other studies made on lung nodules classification, the system produced a lower accuracy rate against the 75 and 87.7% accuracy rates of past studies [6, 10].

Logic operation was used to classify the statistics-based lung conditions using the measurements, ratios and counts obtained. The performance of the system in classifying these lung conditions attained an overall accuracy of 67.22%. Given the results of ROC, the system needs improvement for the possible lung nodules test so that its specificity rate may be increased to perform better than Mousa and Khan [6] and Shimada et al. [10]. Thus, for the cardiomegaly test, its False Negative rate may decrease so it can detect abnormal images correctly.

For extensions of this research, an improved machine learning-based system can be achieved by using a larger database having solely high-quality images to have better and more reliable training data. Another recommendation is to have a further analysis on other ways of representing features and a more intensive research on how to effectively implement a dimension reduction scheme such as Principal Component Analysis and Linear Discriminant Analysis to prevent large memory consumption. As for the measurements, aside from presenting the number of detected possible lung nodules, the locations and sizes of the circular entities can also be considered.

It is also advised to consider other dimensions for Image Division other than 50 by 50 pixels per box, as larger divisions produce a smaller size for the histogram vectors, increasing the possibility of using the histogram itself as input for machine learning.

The system needs modification to address the difficulties in classifying hyper-aeration correctly because of issues regarding image quality and image cropping. To attain a better feature extraction and classification result, an optimized way of determining the correct feature points has to be implemented.

Moreover, we recommend using Histogram Equalization instead of Contrast-limited Adaptive Histogram Equalization for the Image Enhancement of the Feature Measurement submodule because Histogram Equalization intensifies the whiteness of the image as a whole while Contrast-Limited Adaptive Histogram Equalization is a selective enhancement tool. In addition, in order for measurements to be extracted accurately, the blobs formed by the heart and diaphragm should be solid in terms of its structure, forming smooth edges and filling holes.

The system also experiences difficulties in the classification of Lung Nodule-cased images primarily because of the inconsistencies of the data sets that were consulted from medical practitioners. Another reason of having a low accuracy rate in detecting possible lung nodules is because the system is not taught to identify lung nodules from a set of training images, rather, it detects circular entities in an image, which gives more room for error as blood vessels and air sacs may appear to be circular in shape. Thus, the correctness of the location and characteristics of the possible lung nodules can be questioned arguably. Hence, the group recommends the Interphysician Agreement among the doctors in order to decide what should be the final diagnosis for a given case. This mode of decision-making enables radiologists and other medical experts to discuss important findings, to be agreed upon by the majority based on the endorsement of the lead physician.

Acknowledgements We would like to give gratitude to Dr. Lourd Loreto, Dr. Adrian Rabe of the Philippine General Hospital and Dr. Jun Parungao of De La Salle Health Sciences Institute, for imparting with us the basic medical knowledge needed.

References

1. Chapelle, O., P. Haffner, and V. Vapnik. 1999. *SVMs for histogram-based image classification, IEEE Computational Intelligence Society*. Red Bank, New Jersey: Speech and Image Processing Services Research Laboratory, AT&T Labs-Research.
2. Dick, E. 2000. Chest X-rays made easy. *Student BMJ—The International Medical Journal for Students*. 8:316–317.
3. Duda, R., and P.E. Hart. 1973. *Pattern classification and scene analysis*. California: Medical eBook, Standford Research Institute.
4. Gu, L., X. Peng, Y. Sun, L. Qian, S. Wang, Q. Weng, and J. Xu. 2009. Computer-aided diagnosis: A support vector machine-based approach of automatic Pulmonary nodule detection in chest radiographs. In *IEEE Conference Paper – Proceedings of 2009 International Symposium on Bioelectronics & Bioinformatics*. Melbourne, Australia.
5. Howley, T., and M.G. Madden. 2005. The genetic kernel support vector machine: Description and evaluation. 24 (3–4):379–395. https://doi.org/10.1007/s10462-005-9009-3.
6. Mousa, W.A.H., and M.A.U Khan. 2002. Lung nodule classification utilizing support vector machines. In *IEEE Conference Paper—Proceedings of the International Conference on Image Processing*. Dhahran, Saudi Arabia: Department of Electrical Engineering, King Fahd University of Petroleum and Minerals.
7. Noriyasu, H. 2009. Pattern recognition in medical image diagnosis. In *Pattern recognition*, ed. Peng-Yeng Yin, ISBN- 978-953-307-014-8. Available from: http://www.intechopen.com/books/patternrecognition/pattern-recognition-in-medical-image-diagnosis.
8. Platt, J.C. 1999. Training of support vector machines using sequential minimal optimization. In *Advances in kernel methods*, 185–208. Cambridge, MA.
9. Pedersen, S. 2007. Circular hough transform. In *Aalborg University, vision, graphics, and interactive systems*.
10. Shimada, T., et al. 2002. Proposal of a nodule density-enhancing filter for plain chest radiographs on the basis of the thoracic wall outline detected by Hough Transform. *IEICE Transactions on Information and Systems—Special Issue on Measurements and Visualization Technology of Biological Information*, E85-D (1): 88–95.
11. Tonpho, T., A. Leelasantihan, and S. Kiattisin. 2010. *Investigation of chest X-ray images based on medical knowledge and balanced histograms*. Bangkok, Thailand: International Symposium on Intelligent Signal Processing and Communication Systems.
12. Üstün, B., W.J. Melssen, and L.M.C. Buydens. 2007. *Visualisation and interpretation of support vector regression models*, vol. 595, pp 299–309.

Real-Time 3D Mapping of Biopsy Fiducial Points Using Two Infrared Cameras

Bruce Shar and John Leis

Abstract A CT-guided biopsy is a specialised surgical procedure where a needle is used to withdraw a tissue or fluid specimen from a lesion of interest. The needle is guided while being viewed by the surgeon on a computed tomography (CT) scan. CT guided biopsies expose patients to a high dosage of radiation. They are lengthy procedures and the lack of spatial reference while guiding the needle down the predicted path are some of the difficulties currently encountered. To explore possible approaches to this problem, we investigate the use of two infrared cameras capable of imaging the biopsy needle area. These are then mapped into scaled 3D co-ordinate space using an extension of a previously reported method. The system is able to read, in real-time, infrared data from two cameras and import the data. The result is a scaled 3D estimate of the needle endpoints.

Keywords Biopsy · Guidance · Infrared · Computed tomography
DICOM

1 Introduction

Biopsies are a useful and commonly used tool in medicine, whereby a tissue sample is taken in order to be externally analysed to test for diseases such as cancer. Biopsies on tissue near the surface of the body can be easily performed, however when the tissue is deep inside the body the biopsy becomes much more difficult. In these cases a CT guided biopsy may be undertaken. A CT scan provides a cross

B. Shar (✉)
Clinical Informatics, Princess Alexandra Hospital, Brisbane, QLD, Australia
e-mail: bruce.shar@health.qld.gov.au

J. Leis
School of Mechanical and Electrical Engineering, University of Southern Queensland, Toowoomba, QLD, Australia
e-mail: leis@usq.edu.au

J. Billingsley and P. Brett (eds.), *Mechatronics and Machine Vision in Practice 3*,
https://doi.org/10.1007/978-3-319-76947-9_8

sectional view of the body, with a series of these images taken in what is known as a CT study, to visualise the location of both the suspect tissue and the needle being used to extract the biopsy sample.

A shortcoming of the current freehand biopsy method is the difficulty in accurately placing the needle at a desired angle. This is caused by the lack of spatial reference and also the weight of the needle hub, which can alter the angle during a scan. Another problem is that it takes multiple intermittent scans to accurately determine the needle trajectory, which causes unnecessary radiation exposure to the patient and the operator [1], while at the same time being quite time consuming and expensive. There is an excess dose for any CT study, depending on many variables including patient, equipment and operator. According to the radiation protection guideline As Low As Reasonably Achievable (ALARA), any method that can be implemented which reduces or eliminates this radiation dose must be made available for radiographic procedures. ALARA is not only a sound safety principle, it is a regulatory requirement for all radiation procedures [1].

There are some biopsy scenarios where the lesion being biopsied is difficult to access. In such scenarios, radiologists use double angle, i.e., the needle is angled in two planes, left/right and up/down. In order to hit the target, the angles need to be very precise. In an attempt to circumvent this problem, the CT gantry is tilted to the required angle. The radiologist then inserts the needle into the patient, and uses the gantry lasers as a guide for the up/down angle as required [2]. Laser guidance devices are also used to help guide probe placement in order to reduce procedure time and improve targeting accuracy [3]. Newer CT scanners, however, are often unable to tilt the gantry because of size and engineering challenges [4]. Consequently, there is a developing need for an alternative method of performing the double angled biopsies under CT-guidance.

2 Infrared Camera Interfacing

2.1 Camera Interfacing

The Nintendo Wii Remote infrared sensor was chosen due to availability, cost effectiveness, ease of interfacing, and high resolution [5]. The infrared sensor on the Wii Remote is capable of tracking up to four IR hotspots simultaneously, with positions output at 100 Hz in 1024×768 interpolated resolution [5]. Only the positions and size/intensity of the IR hotspots is output, with the raw video output from the sensor not available. A Bluetooth interface is included on the Wii Remote to enable communication.

An interface application was written to communicate with and configure the Wii Remote and access the IR hotspot position data, and a client program in MATLAB to provide additional abstract functionality.

The interface program spawns three threads—two responsible for interfacing with the Wii Remotes, and the last for communicating with the client program. Interfacing with each Wii Remote is performed using generic Bluetooth libraries, and input and output sockets.

The 'wiiuse—The Wiimote C Library' was used as the drivers for the system. These drivers are registered to/dev/wiimote, where the device is read from using MATLAB code by loading the wiimote drivers into the kernel, connecting from the client application to the driver, requesting data from wiimote, and finally processing and ploting the data. The followings steps are performed on launching the interface application:

1. A prompt is displayed for the user to make the Wii Remotes discoverable;
2. A socket connection is established with the Bluetooth HID (human interface device) on the host computer;
3. A scan for Bluetooth devices is performed, and input and output sockets created to connect to any discovered Wii Remotes;
4. The socket connection to the Bluetooth HID is closed;
5. Two polling threads are created and capture of IR data from the Wii Remotes started;
6. The user is prompted to assign left and right positions via button presses on each Wii Remote;
7. A client communications thread is created, and the user prompted to start the client program.

Sockets were chosen for communication with the client application, as they allow a simple means of interfacing with clients written in a variety of languages. The communication thread in the interface application first opens and binds to a socket, waits for the client application to connect, then enters a command/response loop. Command packets sent from client to interface application consist of a single byte, with "d" indicating a request for data and "x" indicating exit.

The response packet is 24-bytes for each Wii Remote, for a total of 48-bytes. Each 24-byte block of data from each Wii Remote consists of 6-bytes for each IR hotspot—two-bytes each for the u and v-positions, and two bytes for size/intensity.

The client application was written in MATLAB®. A socket connection to the interface application is established, and a display window opened containing a 3D plot of the IR hotspots. A loop is established that sends request packets, receives response packets, and calculates (x, y, z) coordinates for each of the two most intense hotspots. The loop exits, and the interface and client programs close when the display window is closed by the user. The (x, y, z) coordinates for each hotspot and the method for estimating the projection parameters is described in Sect. 2.2.

2.2 Camera Calibration

An aluminium jig as shown in Fig. 1 was machined to enable precise, repeatable location of two Wii Remotes in a binocular configuration. A Bridgeport milling

Fig. 1 Wii cameras were fixed using a jig

machine used to machine the jig was also employed to calibrate and estimate the projection (camera) parameters.

The focal length of the Wii Remotes and exact position of the IR sensors within them are unknown. As such an arbitrary reference origin was chosen, and the projection parameters calibrated with respect to this.

In order to estimate the projection parameters, the client application was modified to accept user input of (x, y, z) coordinates which represent the true value of the location of the infrared LEDs, then 100 data points were collected from each Wii Remote. The Wii Remote jig was positioned square to the mill at a distance of approximately a meter, and a single IR LED mounted to the mill table. The milling machine is capable of positioning an object to a precision of 0.01 mm, with accuracy better than 0.05 mm (positioning accuracy is closer to the 0.01 mm precision when not machining).

Forty-three positions spanning a range of ± 100 mm along each axis were collected, with 25 positions in the x–y plane corresponding to the designated zero z-coordinate, and 9 positions in each of the planes corresponding to the ± 100 mm z-coordinates. Outliers were removed from the collected (u, v) values, the mean of the remaining (u, v) values taken, and a derived equation used to solve for the projection parameters.

3 Mapping Algorithm

The problem of camera calibration for metric extraction has been addressed over many years by several authors, with extensive literature on the subject [6]. The area has historically attracted interest from two separate perspectives: computer vision [7, 8] and photogrammetry [9].

Given the requirements for mapping and calibration, and the need to have a simple and semi-automatic calibration system, we investigate the application of Zhang's methods for this problem [7, 10]. We first define the projection system

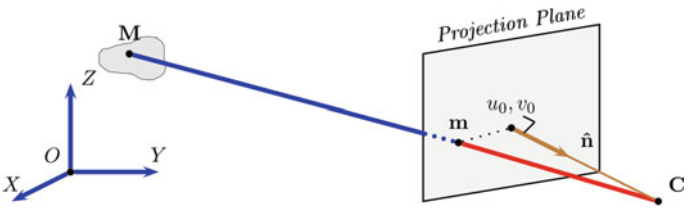


Fig. 2 Illustrating the mapping from observed camera point **m** to 3D point **M**

according to Zhang's result, and subsequently extend it to our case of two IR cameras. The basic problem may be summarised for one camera by referring to Fig. 2. The observed camera point **m** corresponds to 3D point **M**, but the mapping is not unique. Consider the situation where the camera (that is, the projection plane) is moved further away. The mapped point in the image plane may not alter. One approach is to utilize the precise knowledge of the camera position, comprising its center (u_0, v_0) and orientation with respect to the real-world *XYZ* axes. Furthermore, the scaling with respect to the real world is required. In general, this information is not available; in our situation, it is impossible to acquire. Thus, it is clear that at least two cameras are required to fully resolve the ambiguities.

4 Results

To check the correctness of the derived algorithm, a set of points in 3D space with known (x, y, z) coordinates were arranged and their locations also measured using the cameras. This provided (u, v) pixel locations from two cameras along with the corresponding 3D locations (Fig. 3).

In total 100 points were used, and the (u, v) coordinates of the points in the images recorded. The points formed a 15 × 20 × 15 grid in space, with each point 11 mm from the neighbouring points. The cameras were roughly 1 m from the grid, and each camera was roughly 2 m apart. An example training measurement is as follows:

(U1 V1)	(U2 V2)	(X Y Z)
(−17.00 −327.00)	(−181.00 −312.00)	(3.00 3.00 8.00)

To test the algorithm, leave-one-out cross-validation was used. This involved computing the position vector **p** for each camera on 99 of the points, then computing the position of the point that was left out. The error between the true location of this point and the estimated location was computed. This was repeated for every point, and the average error over the whole set of points was determined.

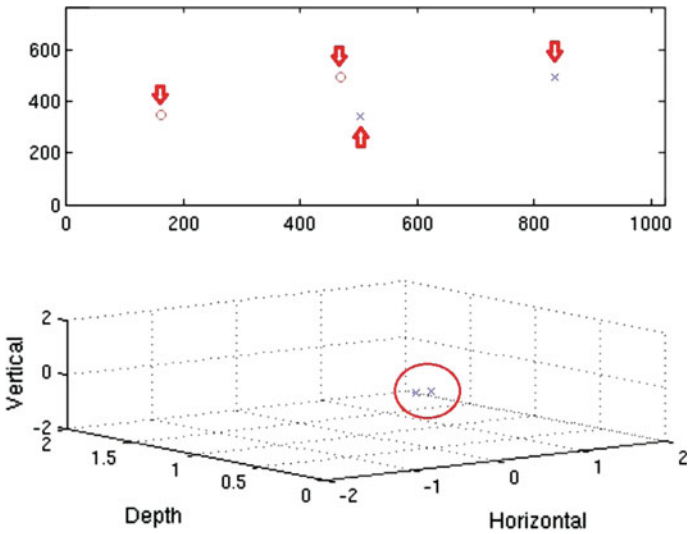


Fig. 3 Visualisation Tool: The upper plot shows the raw (u, v) coordinates returned by the Wii Remotes (crosses are from the left sensor and circles from the right). The lower plot is the (x, y, z) mapping of the (u, v) data for the two IR sources

This provides a good estimate of the error for all the points while ensuring that a point from the training set (used to build the model) is never used to test the model, as this would result in artificially low errors. The final mean error in position was 1.75 mm. This error represents the average Euclidean distance between the true points and their estimated locations.

This error has two main sources: modelling error and noise. Modelling error comes from non-linear effects which may be introduced by e.g. the camera lens which is not modelled by the linear algorithm. Noise comes from a mismatch between the measured and actual locations of the training points. Noise introduced in this way will increase the error when point locations are later estimated.

To check the correctness of the algorithm as derived, a test stub was developed to project a set of random points through a plane representing each camera. The ray-plane intersection point was used as the (u, v) camera point, and the projection algorithm as described was used to estimate the true (x, y, z) point.

The combined system was then tested with real data from the two IR sensors. This is shown in Fig. 4. The real-time data acquisition is able to translate to the estimated 3D position as the IR sources are moved.

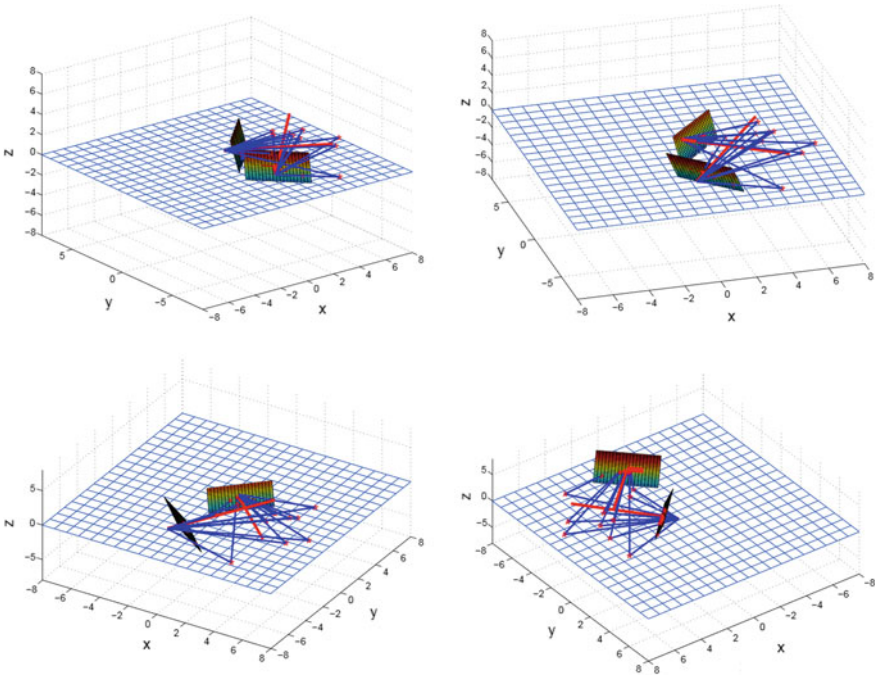


Fig. 4 Testing of the 3D mapping procedure with synthetic data. The camera planes are shown, with **u** and **v** axes as well as a perpendicular to each plane. The intersection of each ray with the camera plane is shown

5 Conclusion and Future Work

The problem of radiation dosage in CT-assisted scans was described, and a possible solution using an infrared camera pair to estimate the true biopsy needle position was investigated. A series of tests were conducted and the coordinates of the points in the images recorded. The error between the true location of this point and the estimated location was computed. The present system uses IR emitters on a test rig, and thus the final system will also require image processing of the needle reflection. To further increase the positioning accuracy, we propose the use of more than two cameras. The algorithm as derived should be amenable to this. It is anticipated that occlusion of the needle will occur in real-life situations, and this also suggests that the use of multiple cameras will be essential.

References

1. Teeuwisse, W.M., J. Geleijns, J.J. Broerse, W.R. Obermann, and E.L. van Persijn van Meerten. 2001. Patient and staff dose during CT guided biopsy, drainage and coagulation. *The British Journal of Radiology* 74.
2. Brant, W.E., and N.M. Major. 2006. *Fundamentals of body CT*, 3rd ed. Philadelphia: Saunders Elsevier.
3. Varro, Z., J.K. Locklin, and B.J. Wood. 2004. Laser navigation for radiofrequency ablation. *Cardiovascular and Interventional Radiology* 27 (5): 512–515.
4. Frederick, P.R., T.H. Brown, M.H. Miller, A.L. Bahr, and K.H. Taylor. 1985. A light-guidance system to be used for CT-guided biopsy. *Radiology* 154 (2).
5. Lee, J. 2008. Hacking the nintendo Wii remote. *IEEE Pervasive Computing* 7 (3): 39–45, July–Sept 2008.
6. Hartley, R., and A. Zisserman. 2003. *Multiple view geometry in computer vision*. Cambridge, UK: Cambridge University Press.
7. Medioni, G., and S.B. Kang. 2005. Camera Calibration. In *Emerging topics in computer vision*. Prentice Hall, New Jersey, ch. 2.
8. Trucco, E., and A. Verri. 1998. Stereopsis. In *Introductory techniques for 3-D computer vision*. Prentice Hall, New Jersey, ch. 7.
9. Fraser, C.S. 1997. Digital camera self-calibration. *ISPRS Journal of Photogrammetry & Remote Sensing* 52 (4): 149–159.
10. Zhang, Z. 2000. A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 22 (11): 1330–1334.

Towards an Automated, High-Throughput Identification of the Greenness and Biomass of Rice Crops

Rhett Jason C. Buzon, Louis Timothy D. Dumlao, Micaela Angela C. Mangubat, Jan Robert D. Villarosa and Briane Paul V. Samson

Abstract Plant phenotyping is a vital process that helps farmers and researchers assess the growth, health, and development of a plant. In the Philippines, phenotyping is done manually, with each plant specimen measured and assessed one by one. However, this process is laborious, time-consuming, and prone to human error. Automated phenotyping systems have attempted to address this problem through the use of cameras and image processing, but these systems are proprietary and designed for plants and crops which are not commonly found in the Philippines. In order to alleviate this problem, research was conducted to develop an automated, high-throughput phenotyping system that automates the identification of plant greenness and plant biomass of rice. The system was developed in order to provide an efficient way of phenotyping rice by automating the process. It implements various image processing techniques and was tested in a screen house setup containing numerous rice variants. The system's design was finalized in consultation with and tested by rice researchers. The respondents were pleased with the system's usability and remarked that it would be beneficial to their current process if used. To evaluate the system's accuracy, the generated greenness and biomass values were compared with the values obtained through the manual process. The greenness module registered a 21.9792% mean percent error in comparison to using the Leaf Color Chart. On the other hand, the biomass module yielded 206.0700% mean percent error using compressed girth measurements.

R. J. C. Buzon (✉) · L. T. D. Dumlao · M. A. C. Mangubat
J. R. D. Villarosa · B. P. V. Samson
College of Computer Studies, De La Salle University, Manila, Philippines
e-mail: rhettbuzon@gmail.com

L. T. D. Dumlao
e-mail: louis.dumlao@yahoo.com

M. A. C. Mangubat
e-mail: micaela.mangubat@gmail.com

J. R. D. Villarosa
e-mail: janrobertvillarosa@gmail.com

B. P. V. Samson
e-mail: briane.samson@dlsu.edu.ph

J. Billingsley and P. Brett (eds.), *Mechatronics and Machine Vision in Practice 3*,
https://doi.org/10.1007/978-3-319-76947-9_9

Keywords Automated phenotyping · Image processing · Greenness Biomass · Rice · Research optimization

1 Introduction

Plant phenotyping is the process of gathering the observable traits of a plant in order to assess its growth, health and development which is vital in the assessment of its more complex traits [1]. This process helps in finding out how to increase the yield and resilience of crops. In the middle of 1990s, plant breeders depend on intuition to select different traits that can increase crop yield or resiliency, but with the advent of modern genetics, scientists and farmers now have the capability to breed crops selectively with precision and accuracy [2]. However, they still need to grow, and analyze the genetic traits of these plants to aid in selective breeding making plant phenotyping vital in the selective breeding process [1].

2 Related Works

2.1 Related Works on Automated Phenotyping Systems

Existing technologies utilize image processing to automate plant phenotyping. The Scanalyzer 3D Phenotyping platform [3] developed by LemnaTec is an automated plant phenotyping system that can phenotype mature plants, such as corn, tomato and rice. It is capable of simultaneously phenotyping plants in large quantities by automatically moving them into a stereoscopic camera by placing all the plants in a conveyor belt. For rice, the system uses the HTS Bonit, an image processing software that analyzes the area, color, and height of the leaves [3]. While the Scanalyzer 3D has been properly tested and deployed, the platform is proprietary and the investment on infrastructure to deploy it is expensive [4].

PHENOPSIS [5] is an automated system by Optimalog that uses image processing for phenotyping Arabidopsis thaliana. It uses a mechanical arm to optimally position a displace sensor, and a camera on each plant to collect phenotypic data such leaf area, leaf thickness, and proportions of leaf tissues. PHENOPSIS is a proprietary solution designed for the French National Institute for Agricultural Research. The system only handles phenotyping of Arabidopsis thaliana and it is only designed to work with screen houses used by the institute [5].

The paper published by Tsaftaris and Noutsos [4] details a setup that is low cost and easy to deploy in nature, making it the most suitable model among all the other researched related systems to Luntian. In order to achieve such characteristics, the system discussed in the paper makes use of digital cameras which are inexpensive in terms of mass purchase and satisfactory in taking the required images. Additionally, each camera has been set up to utilize an open source firmware called

Canon Hack Development Kit. This firmware allows further manipulation of digital camera options which in turn, allow the researchers to tweak the settings to various factors that are present in the system environment. Examples of some manipulated settings are manipulation of the ultra intervalometer for taking time lapse images and utilizing the long exposure intervalometer for taking night time photos.

2.2 Related Works on Image Processing Pipelines

HTPheno is an image processing pipeline specifically made for plant phenotyping and was designed by Hartmann et al. [6]. HTPheno is not designed to any specific phenotyping setup and so it is made to be flexible and highly adaptable to different plant phenotyping setups and environments. HTPheno can analyze and collect 6 different plant phenotypic traits using only the top and side view images of a plant specimen.

The HTPheno pipeline makes use of different image analysis algorithms such as region definition, object segmentation, morphological operation and finally, compilation of the analysis results. For the software to properly analyze the phenotypic traits of a plant, calibration is first done using image segmentation, which partitions an image into different components or segments. This is done through color image segmentation with the multidimensional histogram thresholding approach in both Red Green Blue (RGB) and Hue Saturation Value (HSV) color spaces. Segmentation is done on these two color spaces instead of just one in order to accommodate varying light conditions. One drawback to this approach is the case wherein foreign objects close to the plant have a color similar to it, as that object will be segmented into the plant as well.

After all the segments in the image have been identified, the object of interest, the plant segment, is extracted in the image. To reduce the drawback of foreign objects being included in the plant segment, morphological opening is applied in the extracted image. Morphological opening performs erosion, or reducing the segment of interest by eroding its sides, and then dilation, which expands the segment of interest by enlarging or dilating its sides. Morphological opening removes small foreign objects in the plant segment because of erosion, and it smoothes the sides of the plant segment through dilation. This technique results in a plant segment with a lower noise level.

Finally, the plant segment is analyzed for phenotypic data. The plant segment is transferred back to the original image, which now forms an outline of the plant. From the segment, the software can calculate various plant phenotypic data and outputted for analysis.

Another system used an image processing pipeline for gathering phenotypic data for Arabidopsis thaliana [7].

To analyze and collect phenotypic data from an image, the first step is to convert it from colored to an 8-bit grayscale image, with the process assigning relatively greater pixel intensity for green pixels. Then, image segmentation is performed on the image using a binary mask. The binary mask is created by selecting pixels with

a grayscale intensity greater than 130, which means the brighter areas of the image, or areas of the images that were originally greener, are selected. Afterwards, holes inside this binary mask that are caused by either particles resting on the plant or spots in the leaves are filled. The binary mask is then converted into objects, to create the plant segment. To smooth the leaf edges and reconnect some objects of the plant segment that became disconnected during the segmentation process, the objects are also dilated and eroded. From this, phenotypic parameters are calculated for the plant segment and are saved into the database. The images generated by the pipeline are shown, and the user can choose to detect and remove incorrectly identified images, as well as make annotations of the image to include additional information.

Despite the ability of the current systems to automate plant phenotyping by way of image processing, these systems are not designed for crops that can be grown in the Philippines. Additionally, these systems are expensive to deploy. Thus, there is a research opportunity to create a phenotyping system that will automate the process through image processing but will be more adaptable to the Philippine setting.

3 Luntian

A system called Luntian, the Filipino word for "green", was conceptualized and developed which provides automated, high-throughput, phenotyping that can determine the greenness and biomass of rice crops. The manual process of phenotyping greenness and biomass takes a minimum of 24–48 h to complete. The time taken in manual phenotyping also increases significantly as more test beds are included in the phenotyping process. As the system is automated, it is going to have a relatively higher throughput than that of the manual process. In addition, the system also provides a way to phenotype different types plant specimens in batches, without losing considerable accuracy.

Luntian is designed to work together with a data gathering hardware setup that automatically captures images of plant specimens that is fit for image processing. The captured images are automatically sent to the system for preprocessing and phenotyping.

The system utilizes numerous image processing algorithms that are used to determine the greenness and biomass of plant specimens. Image processing algorithms are also used to normalize images and reduce the impact of changing environmental conditions to the phenotyping process. The system also reduces noise from the images that might affect the phenotyping process. Luntian is built on OpenCV, an image processing library that provides functions to perform the aforementioned algorithms.

Luntian implements a database that contains the phenotypic data from the greenness and biomass modules and the file path of the raw images gathered by the cameras. The actual raw images gathered are stored in a separate directory. This implementation allows the researchers to easily access these needed data in monitoring the progress of the rice crops.

Luntian is just one of two components in the Butil system, which aims to provide researchers an automated method of plant phenotyping. The other component is Seight, which, through the use of different image processing algorithms, obtains the plant height and tiller count from the images taken by the hardware setup. Luntian and Seight share the same image gathering setup, and both interface with the same data management module. Even with the separation of components, the intention is for Butil to be ran and used by crop scientists as one complete system.

3.1 System Functions

(1) *Data Capturing Module*: The Data Capturing Module is responsible for gathering the data for phenotyping. Researchers can schedule the phenotyping dates or capture appointments. These appointments will trigger the camera and a certain date and time, gather the raw images captured and store them into the database. Remotely triggering the cameras is done using the OpenCV API. The use of this library not only enables the cameras to be triggered through functions that are executed according to the scheduled camera appointment, but can also immediately send the captured images to the image processing modules since OpenCV is also used in these modules.

(2) *Preprocessing Module*: The Preprocessing Module uses the raw images gathered by Data Capturing Module and prepares and processes them for automated phenotyping. Color balancing is done through a technique called Normalization. This is done through shifting the values the Saturation and Value channels in the HSV color space to produce a softer curve in the histogram. Normalization in the system is done with the OpenCV function normalize() which attempts to normalize the value range of the two color spaces by shifting and scaling the values. After correcting the color balance, Segmentation then isolates the plant from the whole image, making phenotyping easier and more accurate. The segmentation algorithm used in this module is Otsu Thresholding which considers the darkness intensity of pixels in grayscale. Finally, noise filtering is done on the segmented image using morphological operations.

(3) *Greenness Module*: The manual method of phenotyping involves comparing the color of the plant with the Leaf Color Chart (LCC). The Four Panel LCC, shown on Fig. 1 determines the chlorophyll content or greenness of the plant in four values (2, 3, 4, 5). The Greenness Module analyzes the preprocessed images and determines the greenness of the plant by its LCC values. To determine how green the plant truly is, the greenness intensity of the plant will be looked at using the mean Hue value of the plant. Since the Hue of the plant is not reliant on the Saturation and Value spaces, it provides a more thorough conditions that will be less affected by lighting conditions. The mean Hue of the plant is compared to the Hue values of the four LCC panels. The LCC value of the plant is estimated by using the LCC Panel's Hue value closest to the plant's

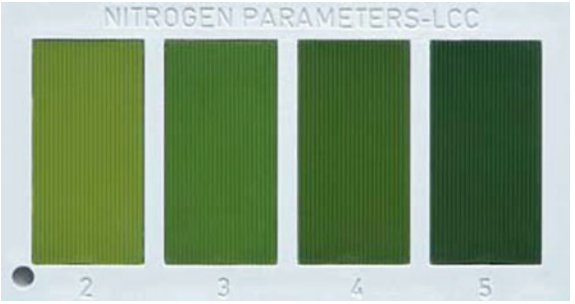


Fig. 1 The four panel leaf color chart

mean Hue value. After determining the LCC value of the plant specimen, the phenotypic data is saved to the database, allowing researchers to retrieve the estimated phenotypic data in the future.

During the process of development, the Hue value of the LCC has to be retrieved as comparison points to the mean Hue value of the plant. The Hue values of each LCC panel were not available in any of the manufacturer's documentation. To get the values, two methods were considered for the experimentation process. The first method is to get the Hue value by taking a picture of the LCC and defining that as a Hue value for all image samples. The second method to dynamically change the LCC value for each image sample.

In the first method, the Hue values of the LCC were retrieved by capturing a photo of the LCC and determining its Hue values in Adobe Photoshop. Since OpenCV stores Hue values in integers 0–180, and Adobe Photoshop determines Hue values in 0°–360°, the values retrieved in Photoshop is divided in half. The Hue values of each LCC panels are in Table 1.

In the second method, the Hue values were retrieved by dynamically getting the Hue values through sampling the LCC attached on the board. The LCC was attached to the segmentation board as a point of reference, and this region in the image samples was isolated. After isolating the LCC in an image each panel was sampled to dynamically retrieve the Hue values of each LCC panel.

The first and second algorithms of determining greenness were tested and compared with each other. The discussion of the results of both algorithms will be detailed in the Results section. Since the first algorithm obtained the lesser percentage error between the two, it was then used as the final algorithm in the system.

(4) *Biomass Module*: The Biomass Module analyzes the pre-processed images and approximates the biomass of the plant by estimate its plant volume. Plant radius

Table 1 Corresponding LCC and OpenCV hue values using the first method

LCC panel	OpenCV hue value
2	32
3	44
4	57
5	80

is first determined by counting the back pixels per row of the resulting binary image. The pixel counts are then averaged to get the mean pixel count. This mean count is converted from pixels to centimeters and used as the plant radius. Approximation of biomass is done by estimating the plant's volume when packed inside a cylinder. The formula for the volume of a cylinder is used to compute the approximate biomass of the plant.

$$volume = \Pi radius^2 height \quad (1)$$

The height will be retrieved from the *Seight* system which approximates the height of the plant.

(5) *Data Management Module*: The Data Management Module is the main interface of the system. The user can use this module to view the collected phenotypic data, and change system settings. This component is in the form of a web application in order to make the system accessible to the researchers in any location.

Subsection Physical Environment and Resources Luntian works alongside a hardware setup that will capture the images needed for phenotyping. As a proof of concept, the data capturing setup was tested on one test bed in a screen house. The test bed has two plants placed in front of the camera. These two plants serve as the representatives of the whole test bed and are used for phenotyping. A separation board was used to isolate the plant specimen from the background. Specifications of the spearation board is detail in Fig. 2.

IP cameras were used in order to make use of the remote capturing functionality available in the OpenCV library. The cameras must be positioned in the screen house properly for optimal data gathering. The measurement and specifications for the placement of the data gathering setup are shown in Fig. 3.

The IP camera that is used in the setup during the time of development and testing is D-Link DCS-932L. It has a resolution 640 × 480 Video Graphics Array (VGA) resolution and is compatible with the OpenCV framework.

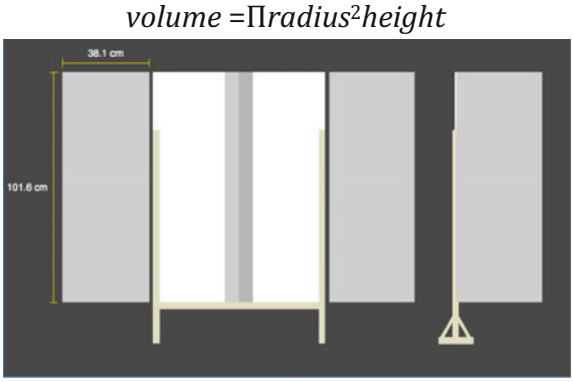


Fig. 2 Separation board specifications

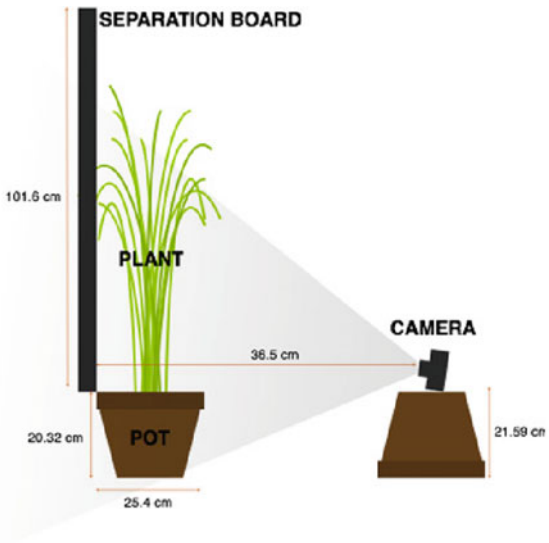


Fig. 3 Measurements and specifications for the data gathering setup

4 Results

4.1 Testing Methodology

For the purposes of testing the accuracy of the system's greenness and biomass algorithms, 28 plant specimen of C4 rice varying in size, greenness and plant stages were chosen for data capturing. These plants were chosen at random to give variety to different plants. At the time of capturing, 14 plants were on the mid-tillering growth stage and 14 other plants were on the tillering growth stage. The two growth stages give different plant structure and color to the plants and this gives a more thorough characteristic in the testing data. This can be seen in Fig. 4.

Following the planned data gathering setup, 28 images were captured by the system, one for each plant. After data capturing, crops were manually phenotyped for plant greenness (using the LCC), girth of the plant when tightly compressed, and girth of the plant when loosely compressed. In order to measure the girth when tightly compressed, the tillers are compacted as close as possible to each other to reduce the gaps in between then a tape measure is used for determining the circumference or girth. For the loosely compressed girth, the tape measure is placed around the edges of the plant without compacting the tillers. After assessing the measurements, the compressed girth was used in comparing the results since it is more appropriate to the formula used for computing the biomass which was defined in the previous chapters. The manual phenotyping process was done by the researchers to ensure the accuracy of the manual phenotypic data. The data taken from manual phenotyping were used to assess the accuracy of the system.

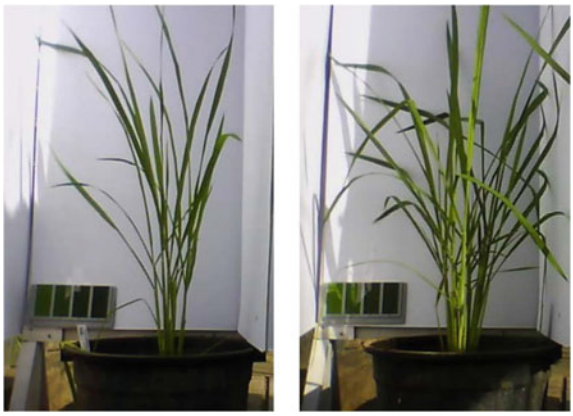


Fig. 4 Mid-tillering growth stage (left) and Tillering growth stage (right)

The mean error difference was computed by getting the mean of the difference between the manual phenotyping results and the automated phenotyping results. The percentage error for each specimen was computed using the formula below. The mean percentage error was computed by averaging the percentage error for each specimen.

$$percenterror = \frac{automatedval - manualval}{manualval} \times 100 \quad (2)$$

4.2 Greenness

In the development of greenness, two algorithms were tested in parallel so results can be compared with one another. The first algorithm is to rely on static Hue values of the LCC for all images. The other algorithm is to rely on dynamic Hue values of the LCC that changes for every image. Two results of the two algorithms will be discussed and compared in detail.

After processing the 28 images through the first algorithm, the system has registered a mean percentage error of 21.9792%. While the accuracy displayed by the system seems fairly high, the small range of values (2–5 LCC values) means the percentage error in the system can result in discrepancies between the actual value and approximated value significant enough for the greenness algorithm to become unreliable. The mean error difference of the LCC value generated by the system compared to the Manual Phenotyping method is 0.7083, so on average the LCC Value generated by the system can have an error up to more or less .71 the original value.

The second algorithm however, yielded a much higher mean percentage error of 33.2066%. This means that using the second algorithm with the dynamic Hue values has a less accurate estimation that using the first algorithm, although it has

the advantage of being more adaptable to different conditions. The mean error difference for the second algorithm is 1.0833, which is also higher than the first one. In the second algorithm, the error difference can mean it can over or underpredict by 1 LCC value.

Further analysis show that a number of factors that may affect the accuracy of the system. One major factor that affects results is the resolution of the camera. The 640 × 480 pixel resolution results in less pixels being used by the algorithm for determining greenness. The resolution also affects the segmentation of the plant (seen in Figs. 5 and 6) and this is because the delineation and edges of the plant are clearer and JPEG compression is less evident on larger images. Segmentation in lower resolution images can result in loss of data with the upper tillers not being segmented as part of the plant.

Another factor that can affect the accuracy of the greenness algorithm and its result is the effect of the daylight conditions inside the screen house. Shadows and uneven lighting can cause the color of the plant to vary when compared to actual inspection of the plant. Plant images were taken in daylight conditions at 2:45 PM.

Fig. 5 An image segmentation of the plant in an image with a resolution of 2000 × 3008

Fig. 6 An image segmentation of the plant using the 640 × 480 resolution image captured by the system

External lighting equipment was not used in the hardware setup, so daylight and shadows caused un-optimal lighting conditions.

4.3 Biomass

After processing the images with the algorithms, some images were removed due to excessive noise that will affect the results. A sample of removed images is shown in Fig. 7. It can be seen that even if the area is white, it was not successfully removed and was included as part of the plant. The error is the same for all removed images. The final dataset contains 23 samples.

Biomass results show that there is no exact trend in the differences of the measurements from the manual method with compressed radius, and the automated method values. There are also large differences between the values. This is caused by already having differences in the estimated radius values from the system, which greatly affected the estimation of biomass. Biomass values computed using compressed girth have a mean percentage error of 206.0700% while the mean percentage error of the radius for compressed girth is 66.7573%. Since there are already large percentage errors in estimating the radius, the values highly affected the results. Another factor that affected the results is the conversion of pixels to cm. Using the average number of pixels from all images is not sufficient to provide accurate results. Since the conversion was not exact, there was no exact trend in the differences of the results of the automation with the manually measured values.

The results show that there are outliers in the dataset. After further analyzing some of these outliers, it can be inferred that the presence of dark portions affect the measurements since these were not completely preprocessed in the binary conversion of the pictures. Despite the fact that the system was already improved by cropping the image first before converting to binary in order to reduce the lighter shadows, there were still dark portions which the system was not able to

Fig. 7 Sample of removed image because of improper thresholding

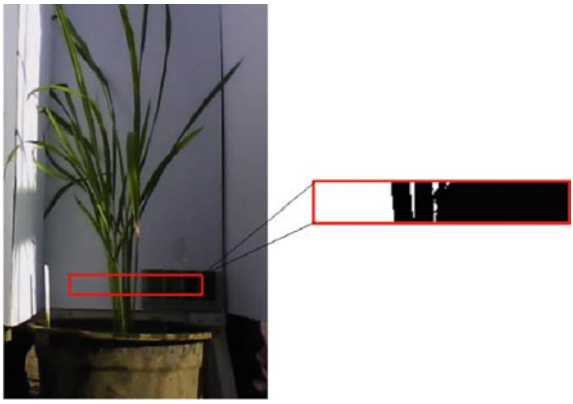


Fig. 8 Image sample with preprocessing issue

successfully differentiate from the plant. Figure 8 shows a sample image with its corresponding binary equivalent generated by the system. In order to address this issue, the images with excessive noise were removed as mentioned earlier. For further studies, the lighting setup should be improved for better results.

It can be seen that in the right portion of the cropped binary image, the dark portion behind the plant was not successfully preprocessed. Since this image is the one used for estimating the radius, the result was a greater radius value than the manually measured radius of the compressed and loose plant.

Another issue is that plants have different structures. There are some which have leaves in the region of interest which also became noise as shown in Fig. 9. In addition, there are also plants that appear wide in the front view but are actually linearly distributed. Automated results from such images may be larger than the actual since the tillers are not compressed. It could be possible to tie the tillers together before taking the picture in order to have more accurate results.

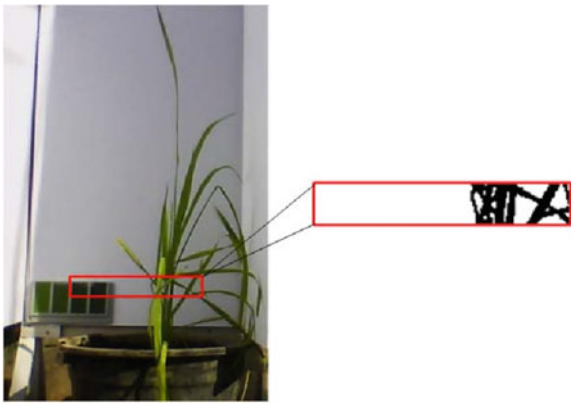


Fig. 9 Image sample with plant structure issue

5 Conclusion

The Luntian system was created for the benefit of experimental crop rice researchers to address their needs of an automated system that can speed up the phenotyping process. Throughout the study and the development of the system, consultations were made with the C4 rice researchers to match their manual phenotyping process and translate it to the automated system. Data from their manual phenotyping process were used as comparison data and metrics for determining the accuracy of the completed system.

Two greenness algorithms were tested in the research. The first algorithm, which relied on static LCC hue values for all images, yielded the most accurate results between the two in estimating the actual LCC values of the plant. The second algorithm, which relied on dynamic LCC values that is newly sampled for every image, yielded a less accurate result even though it has the advantage of being more robust than the first algorithm. While the determining greenness is shown to be accurate based on the results, both algorithms can prove to be unreliable to the researchers because of the relatively large average percent error especially when considering the level of accuracy needed for their research. The relative inaccuracy and unreliability of the algorithm is mainly rooted in how the image is captured, with factors such as uneven lighting and the low resolution of the camera used all playing a significant role in making both algorithms unreliable.

Based on the results, there are significant percentage errors in the biomass and radius estimations and there is no definite trend when compared with the compressed girth values. Further studies on the values may have to be done in order to discover a trend and use it to improve the algorithm. One recommendation is to use distribution fitting in order to find whether there are relationships between the data in order to improve the accuracy.

Overall, results yielded by the study can serve as a step towards automating plant phenotyping, which is a significant help for the researchers in developing their crops. In the future, effective automation would speed up the data gathering process, which would be more evident in larger set-ups. By curtailing the extent of human intervention in the phenotyping process, inconsistencies brought about by human error would consequently be reduced.

Acknowledgements The authors would like to thank Mr. Briane Samson, Dr. Florante Salvador and Dr. Joel Ilao from the College of Computer Studies of De La Salle University for their guidance throughout the duration of writing the research. They would also like to thank Mr. Alexis Pantola, for providing the Internet Protocol camera used in the development and testing of the system. Finally, they would also like extend their gratitude to the International Rice Research Institute especially C4 Rice Researcher Mr. Albert De Luna for working with us through consultations about plant phenotyping and the C4 Rice project, as well as giving feedback in the development of the Luntian system.

References

1. Helmert, M., and H. Lasinger. 2010. The scanalyzer domain: Greenhouse logistics as a planning problem. In *International Conference on Automated Planning and Scheduling*, May 2010.
2. Finkel, E. 2009. With 'Phenomics,' plant scientists hope to shift breeding into overdrive. *Science* 325 (5939): 380–381.
3. Eberius, M. 2014. Lemnatec HTS bonit: Image analysis for the quantification of rice in 2-D and 3-D assays. Lemnatec.
4. Tsaftaris, S., and C. Noustos. 2009. Plant phenotyping with low cost digital and image analytics. In *Information Technologies in Environmental Engineering: Proceedings of the 4th International ICSC Symposium*, vol. 4, p 239.
5. Granier, C., L. Aguirrezabal, K. Chenu, S.J. Cookson, M. Dauzat, P. Hamard, J.-J. Thioux, G. Rolland, S. Bouchier-Combaud, A. Lebaudy, B. Muller, T. Simonneau, and F. Tardieu. 2006. Phenopsis, an automated platform for reproducible phenotyping of plant responses to soil water deficit in arabidopsis thaliana permitted the identification of an accession with low sensitivity to soil water deficit. *New Phytologist* 169 (3): 623–635.
6. Hartmann, A., T. Czauderna, R. Hoffman, N. Stein, and F. Schreiber. 2011. Htpheno: An image analysis pipeline for high-throughput plant phenotyping. *BMC Bioinformatics* 12 (1): 148.
7. Arvidsson, S., P. Perez-Rodriguez, and B. Mueller-Roeber. 2011. A growth phenotyping pipeline for arabidopsis thaliana integrating image analysis and rosette area modeling for robust quantification of genotype effects. *New Phytologist* 191 (3): 895–907.

Maturity Analysis and Monitoring System for Sugarcane Crops

Jonathan Clark S. Camacho, Anthea Marina A. Co, Jan Percival S. J. Hao, Ana Carmela P. Salazar and Joel P. Ilao

Abstract The sugarcane stalk and foliage colors are strong indicators of maturity. While outdoor conditions affect the lighting of the captured sugarcane image, external factors like shadows and over exposure can easily be neglected in the Hue Saturation Value (HSV) color space. Through a series of experimentations in the HSV color space, the proponents found out that there is a significant shifting of Hue and Saturation values as sugarcane crops mature. The Hue and Saturation frequencies of both matured and not-matured sugarcanes were used as data for the maturity detection comparison module which utilizes RandomForest algorithm with 90.16% accuracy.

Keywords Sugarcane · Maturity analysis · HSV · Automation · Principal component analysis

1 Inroduction

Harvesting sugarcanes at the ripe age maximizes sugar recovery as the sucrose concentrations are almost the same percentage in all parts of the stalk. Physiologically, sugarcane ripening is reached when the stems reach their potential

J. C. S. Camacho (✉) · A. M. A. Co · J. P. S. J. Hao
A. C. P. Salazar · J. P. Ilao
College of Computer Studies de La Salle University,
2401 Taft Avenue, 1004 Manila, Philippines
e-mail: jonclark925@gmail.com

A. M. A. Co
e-mail: amc_anthea@yahoo.com

J. P. S. J. Hao
e-mail: janhao11_07@yahoo.com

A. C. P. Salazar
e-mail: acp.salazar@gmail.com

J. P. Ilao
e-mail: joel.ilao@delasalle.ph

J. Billingsley and P. Brett (eds.), *Mechatronics and Machine Vision in Practice 3*,
https://doi.org/10.1007/978-3-319-76947-9_10

sucrose storage, i.e., the point of maximum possible sucrose accumulation [5]. Yellowing and drying of leaves, metallic sound of mature canes when tapped, appearance of sugar crystal, glistening when a mature cane is cut in a slanting way and held against the sun are some of the visual indices of assessing the maturity of the cane [4]. The sugarcane stalk must suffer growth retardation, low temperatures and moderate drought among other factors to ripen [5].

Traditionally, specialized agriculturists perform ocular inspection and measurements on the sucrose using a Refractometer and Saccharometer during milling season. Refractometer is a laboratory tool that is used to measure an index of refraction [1]. Saccharometer is a device used to measure the concentration of sugar in a solution [4]. In Sugarcane mills, refractometry is a method used to test the brix, which is the solid concentration of a sucrose containing solution—measured in percentage by weight [3]. However, farmers and agriculturists estimate maturity using a less scientific approach through visual cues, such as stalk and foliage color, and internode distances.

This paper will discuss the implementation and performance analyses of the system. The Pre-processing Module, Feature Extraction Module, and Maturity Detection Comparison Module will be discussed in depth. The conclusions and recommendations for conducting such study will also be tackled in this paper.

2 Methods

2.1 System Set up

The system utilizes a camera that is mounted on a tripod to achieve camera stability. The camera is positioned two to three meters away from the target sugarcane subject to capture the upper half portion of the stalk. The images were captured in the JPEG RAW format with a minimum resolution of 1024 × 768 pixels and a maximum resolution of 3872 × 2592 pixels. A minimum of 1024 × 768 pixels picture resolution is needed to be able to process necessary features while keeping the resolution at the minimum.

The captured image is pre-processed and significant features were extracted using Matlab. The first principal component is saved in a CSV file format which will be fed to Java software application that will classify and label the sugarcane images as matured and not-matured through an implemented RandomForest machine learning algorithm. Figure 1 shows the system set up for this study.

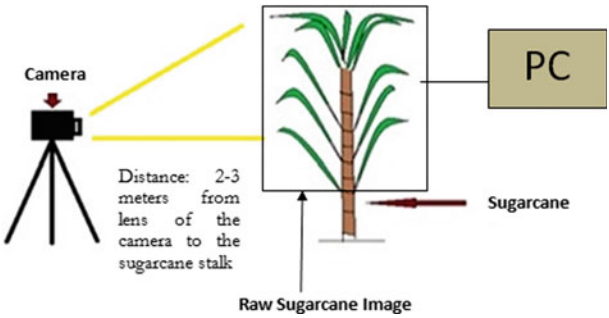


Fig. 1 Set-up of the maturity analysis and monitoring system for sugarcane crops

2.2 Database

The database is composed of 610 sugarcane image samples from Month 7 to Month 12. There are 73 matured sugarcane image samples and 537 not-matured sugarcane image samples. 70% of the sugarcane samples in the database are used as training data while the remaining 30% are used as test data.

2.3 Pre-processing

Pre-processing is important to properly isolate the sugarcane stalks and their foliage from the image background. The raw Red-Green-Blue (RGB) image of the sugarcane is converted to the Hue-Saturation-Value (HSV) color space because it is designed to approximate the way humans perceive color [6]. Figure 2 shows the flowchart of the implemented Pre-processing module.

There are three submodules under Pre-processing, namely: *RGB to HSV conversion*, *Hue Thresholding*, and *Masking of HSV Sugarcane Image with Binarized Sugarcane Image*. Sky and ground image regions provide irrelevant information to the sugarcane maturity classification process. The appropriate Hue Threshold range of values, from 0.05 to 0.5, which effectively removes the sky and the ground from the image, was determined through Trial and Error.

Binarized sugarcane image masks, shown in Fig. 3, are used to extract the relevant sugarcane regions in the HSV-converted sugarcane images. This process results in Masked Sugarcane image, an example of which is shown in Fig. 4.

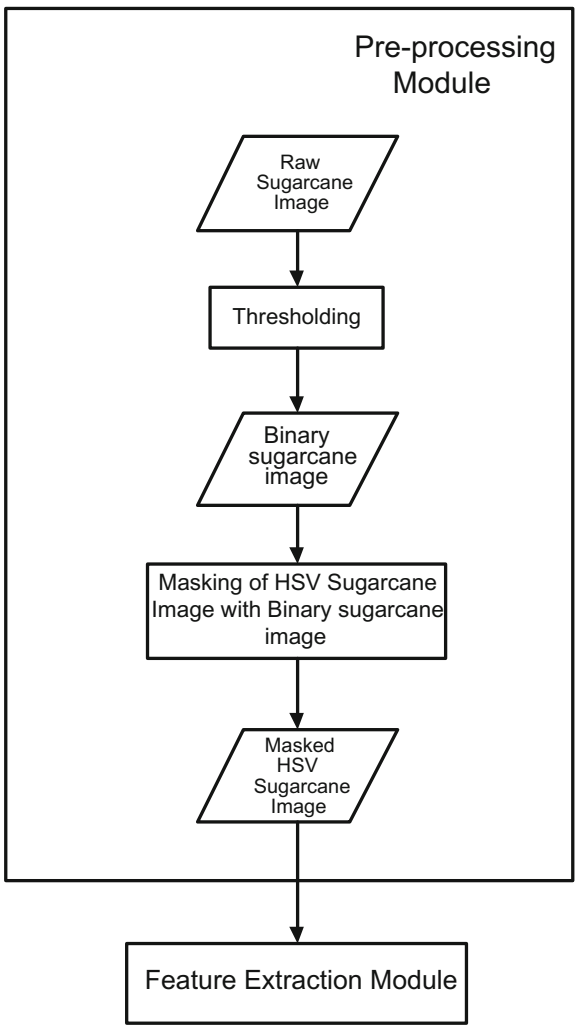


Fig. 2 Flowchart of implemented pre-processing module

2.4 Feature Extraction

Feature Extraction involves analyses and records the target features of the output image produced after Pre-processing. The target features were mainly the frequency of HSV values found in the sugarcane image, all of which are positive integers. There are four submodules under Feature Extraction, namely: Get HSV Values, Get Frequency Counts of Hue and Saturation Combinations, Normalize Frequency Counts of Hue and Saturation Combinations, and Principal Component Analysis (PCA).

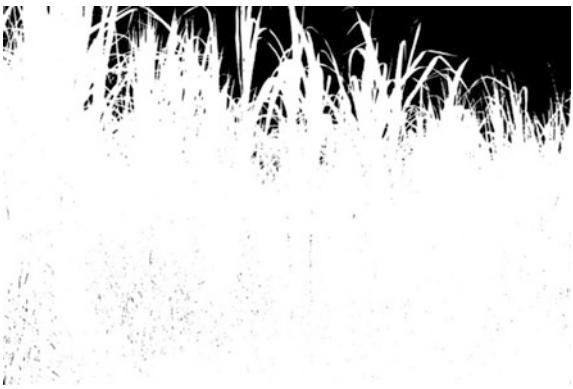


Fig. 3 Binarized sugarcane image

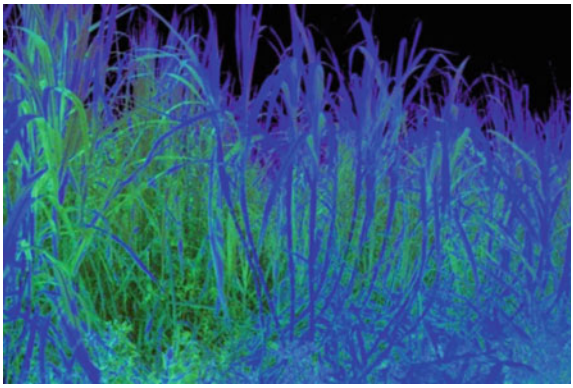


Fig. 4 Masked sugarcane image

The extraction was done per image sample and the features were recorded in CSV file format. The data was normalized to avoid imbalance throughout the data. As an effect, each sugarcane image has 4656 features. To save processing time, the data was reduced using Principal Component Analysis. Principal Component Analysis is a statistical technique used for handling high-dimensional data, and using the dependencies between the variables to reduce it in a lower-dimensional form without losing too much information [7]. The flowchart of the Feature Extraction Module is shown in Fig. 5.

The masked HSV sugarcane image is scanned pixel by pixel for the values of its Hue, Saturation, and Value components, to extract Hue-Saturation 2D histogram plots.

The pixels with specific combinations of Hue-Saturation components of an image sample are counted and recorded as frequency counts. This is done per individual image samples and the resulting vector of combined frequency counts is recorded in a general CSV file where individual image data are represented in rows.

The total number of pixels for every image may vary depending on the number of pixels removed after Pre-processing. In effect, there is an unbalanced level of

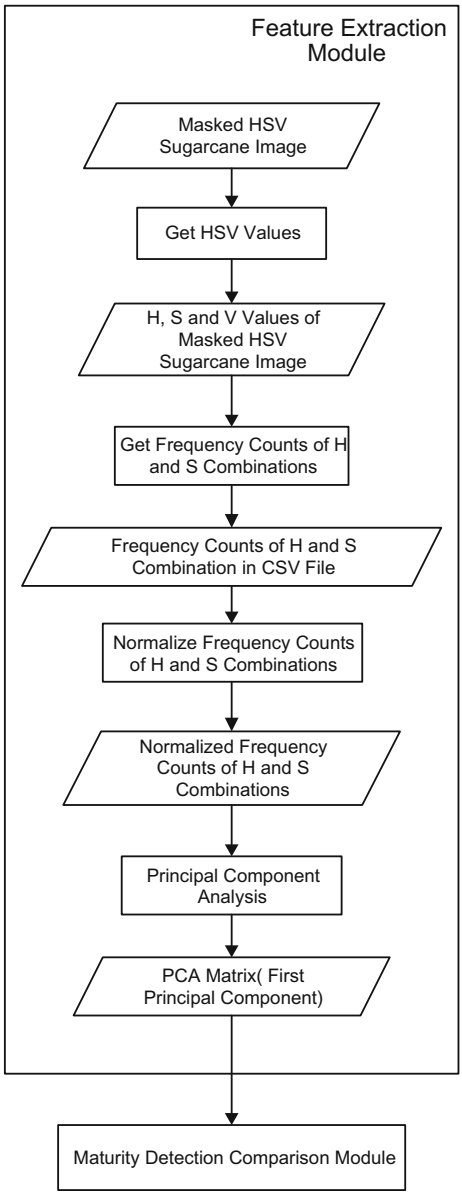


Fig. 5 Flowchart of implemented feature extraction module

HSV frequency throughout the data set. To address this problem, scaling through normalization is applied. Scaling of data reduces difficulties during computing in machine learning algorithms. In this study, we normalized all the feature vectors to each have a length of 1.

The main feature in the Feature Extraction module is the combined Hue-Saturation frequency counts. To obtain the combination of Hue and Saturation, consider a 3D matrix with x-axis and y-axis representing hue and saturation range respectively. A third axis represents the frequency count of the combination of the Hue and Saturation value. In the graph shown in Fig. 6, the feature vector is created by connecting the start point of the next column to the end point of the previous column.

HSV values were recorded with 2 decimal places. The range values of Hue are from 0.05 to 0.50 while Saturation values range from 0.00 to 1.00. With this combination, a feature vector of length 4646 for each sugarcane image samples is generated. The feature vector with this size will be full of redundancies therefore Principal Component Analysis (PCA) is applied.

Principal component analysis (PCA) is a quantitatively rigorous method in achieving simplification [6]. The principal components, as a whole, form an orthogonal basis for the space of the data. The goal of the PCA is to reduce the vectors of the data set. The approach is finding up to which principal component contains 95% of the data set's variance. The new set of principal components is obtained by the product of the normalized data set and the first to the thirty-ninth column of the coefficients matrix (see Fig. 7). Plotting the mean and error bars of the matured and not-matured samples of this new score matrix reveals that the first principal component has significant separation and distinction compared to the remaining components (see Fig. 8).

The first principal component is the product of the normalized data set and its most optimally weighted coefficient vector. The coefficients of the principal components are calculated so that the first principal component contains the maximum variance. Out of the 4646 combinations of Hue and Saturation, PCA found out that there is a range within these combinations that highly varies between matured and not-matured samples. It was revealed that the following range of Hue-Saturation values became important in determining the difference between matured and not-matured samples: 0.08–0.21 Hue with 0.30–0.88 Saturation, and 0.24–0.39 Hue with 0.13–0.61 Saturation. The first principal component, as the chosen feature

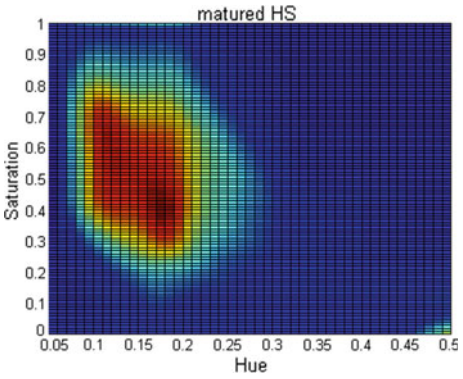


Fig. 6 Hue-saturation graph of a matured sugarcane image

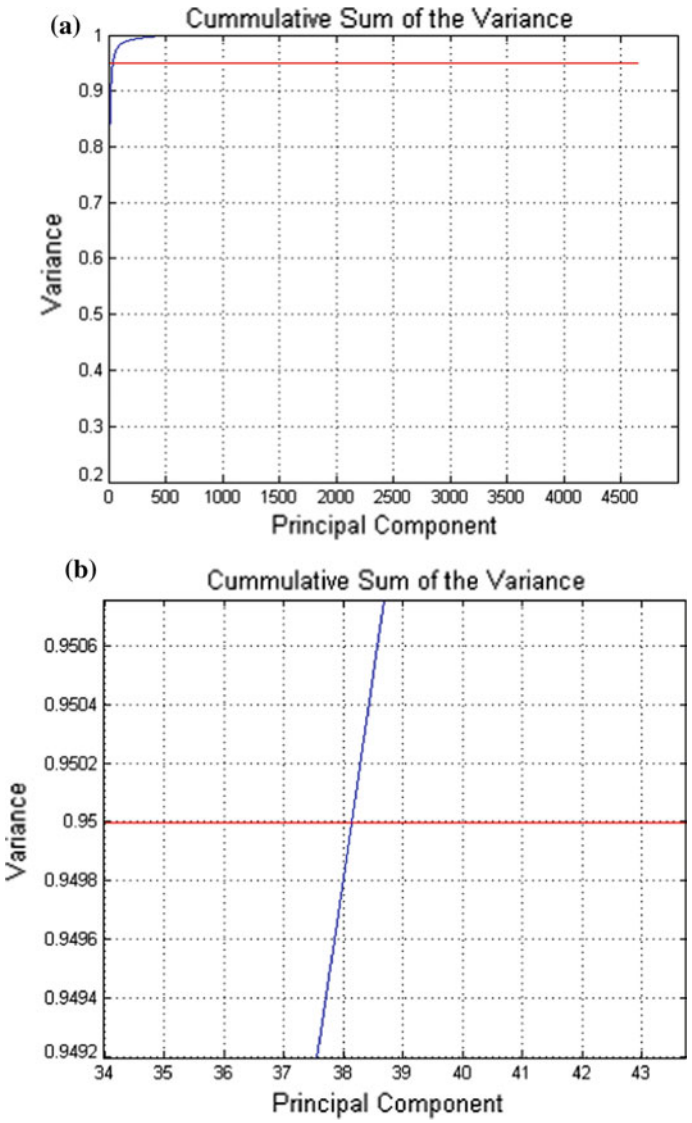


Fig. 7 **a** Shows the graph of the cumulative sum of variances/eigenvalues; **b** shows the intersection between the 95% threshold and the cumulative sum graph

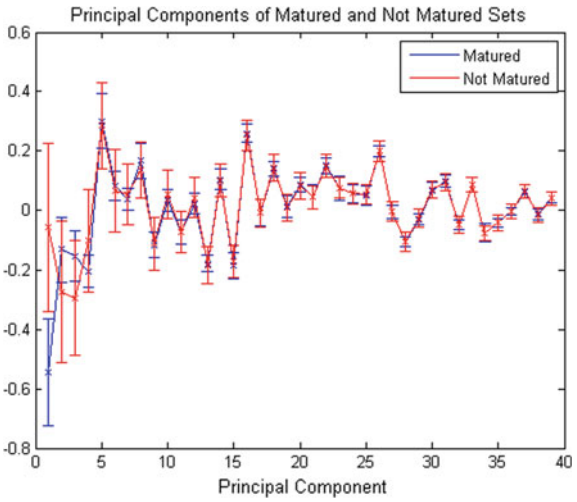


Fig. 8 Graph of the principal components

vector, now represents each sample in the whole data set. The vector is recorded in a CSV file as the final output of the Feature Extraction Module.

2.5 Maturity Detection Comparison

The accuracy of the system in classifying sugarcanes is entirely dependent on the machine learning algorithm used. Machine learning is important in the *Maturity Detection Comparison Module*. The flowchart of *Maturity Detection Comparison Module* is shown in Fig. 9. For this study, the machine learning algorithm implemented is RandomForest algorithm. RandomForest is a kind of a Decision Tree which utilizes bagging of un-pruned decision tree learners with a randomized selection of features at each split [4]. There are two phases to be considered to enable the system to classify sugarcane images as matured and not-matured: *Training phase* and *testing phase*.

The best algorithm to be used for the *Maturity Detection Comparison Module* is determined using WEKA software. The training data found in the CSV file, which is composed of 51 matured and 376 not-matured data, is fed to WEKA software and RandomForest machine learning algorithm is utilized to train and construct a classifier.

The best performing machine learning algorithm is used for the *testing phase*. The training data which consists of the first principal components of sugarcanes and its expert label was constructed to a trained classifier. This was used as the model for classification.

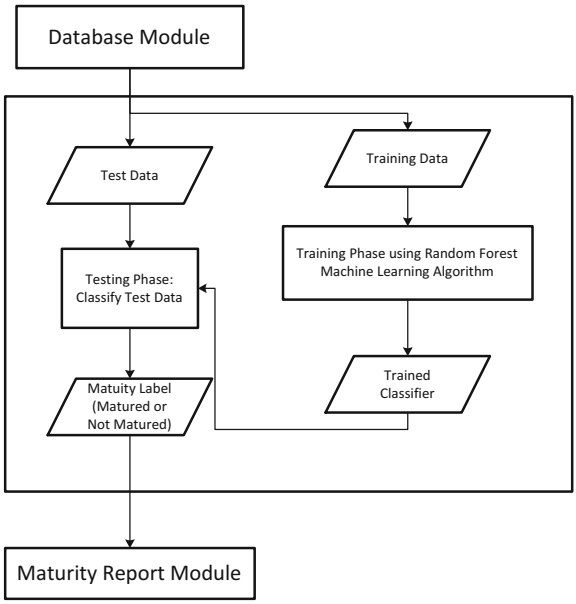


Fig. 9 Flowchart of maturity detection comparison module

3 Performance Analysis

3.1 Reference Image Set Verification

Reference Image set should consist of 100 matured sugarcane images. To consider a reference sugarcane image the image of the captured sugarcane is verified by an agriculturist. This was performed by visually inspecting the features of the sugarcane that indicate its maturity. Figure 10 shows a sample reference sugarcane image that is verified by agriculturists. The sugarcane is confirmed to be matured as it indicates evident signs of maturity: decreasing internode height towards the top, yellowing of the midrib of the leaves, and thrashing of matured leaves.

Fig. 10 Sample reference sugarcane image

3.2 Hue Thresholding

Though a series of experimentation, it was observed that there is a significant succession of change as the range of Hue Threshold values are adjusted. The raw sugarcane image is shown in Fig. 11 while the binary images produced by threshold value ranges from 0.05 to 0.5, 0.05 to 2, and 0.1 to 0.3 were shown in Figs. 12, 13 and 14 respectively. The threshold 0.05–0.5 has the least amount of pixel loss and able to properly isolate the significant portion of the sugarcanes and foliage. As a result, threshold value that ranges from 0.05 to 0.5 are selected to isolate the relevant features.

Fig. 11 Raw sugarcane image

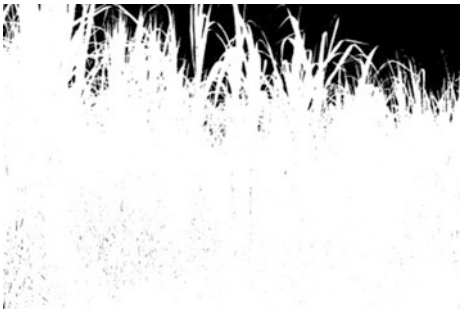


Fig. 12 Hue threshold value of 0.05–0.5

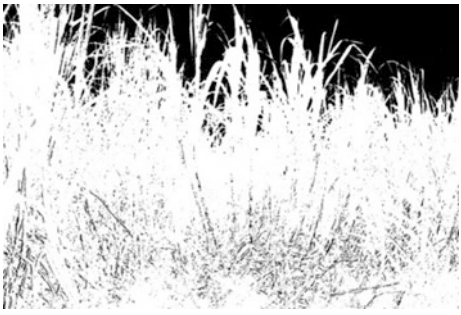


Fig. 13 Hue threshold value of 0.05–2

Fig. 14 Hue threshold value of 0.1–0.3

3.3 HSV Experimentation

The frequency of Hue-Saturation-Value of matured and not-matured sugarcane images were plotted with their corresponding hue, saturation, and value index. This experiment was done to determine features that will be used for the feature extraction module.

Figures 15, 16 and 17 show the plot of Not-matured and Matured Hue, Saturation and Value frequency with its corresponding H, S, and V index. The Hue and Saturation plot of the not-matured and matured sugarcane images shows significant difference than its Value counterpart values. The hue of the sugarcane sample shifted upward as it turns matured and its saturation shifted from left to right.

Figures 18 and 19 show the top view of the 3D histogram of not-matured and matured sugarcane images, respectively. Warmer colors indicate higher concentration of pixels with certain Hue-Saturation values. As observed in the figures, there is a transition of Hue-Saturation values in not-matured and matured sugarcane images.

Not-matured sugarcanes have hue values ranging from 0.1 to 0.27 and shifts to hue values 0.08–0.23 as it matures (see Fig. 18), while the saturation values of not-matured sugarcanes ranges from 0.18 to 0.6 shifts to ranges 0.25–0.78 as it matures (see Fig. 19).

3.4 Machine Learning Verification Using WEKA

Machine learning verification is performed to determine which machine learning algorithm in WEKA shows the highest accuracy in classifying the sugarcanes' maturity label. The test data contains 22 matured and 161 not-matured sugarcanes with their first principal component and corresponding expert's label, while the training data contains 51 matured and 376 not-matured sugarcane with their first principal component and corresponding expert's label.

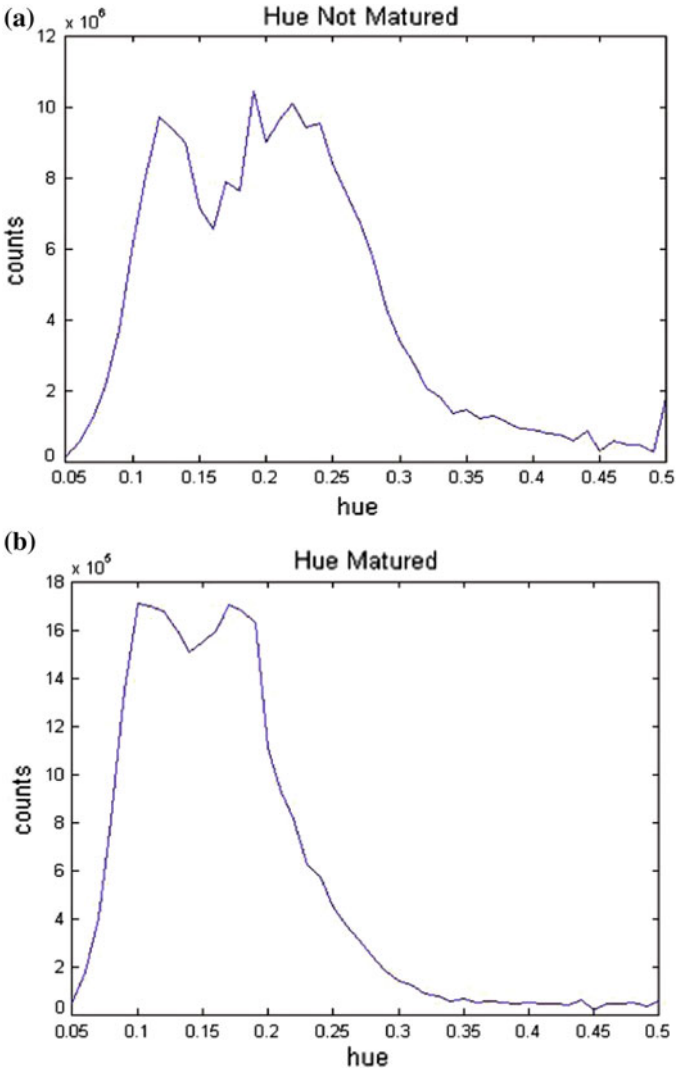


Fig. 15 Hue frequency plot of **a** not-matured and **b** matured sugarcane images

The training data is fed in WEKA to compare and verify the machine learning algorithms. The machine learning algorithms were verified using 10-fold cross validation. Table 1 shows the results of model building using DecisionStump, J48, RandomForest, RandomTree, and SVM algorithms in WEKA. In model building, SVM obtained the highest accuracy percentage among the other 4 machine learning algorithms.

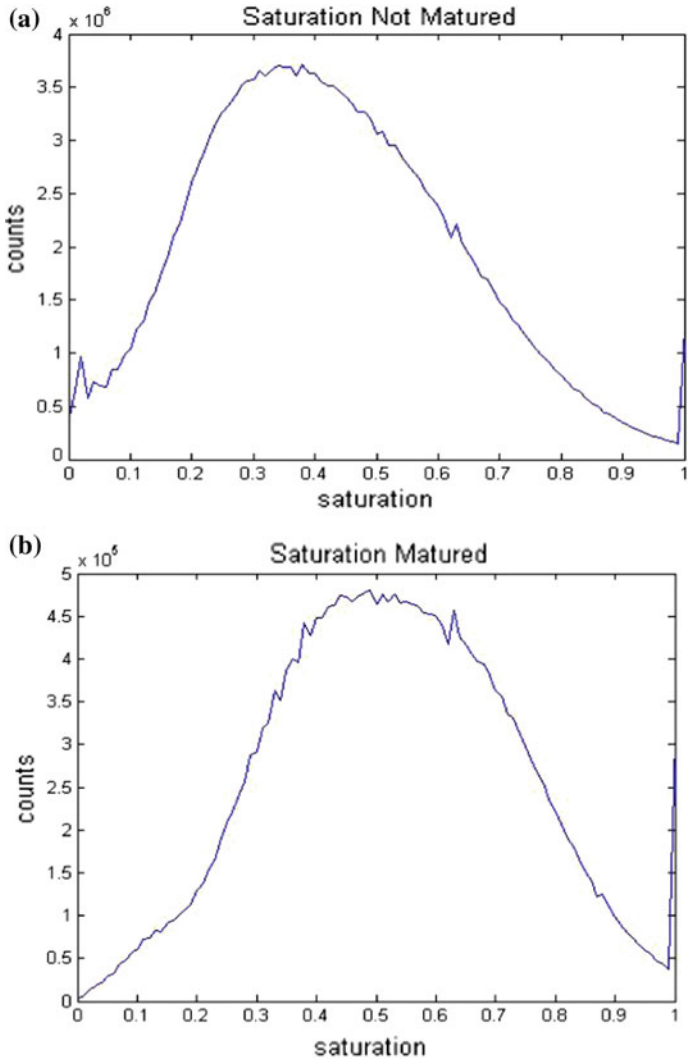


Fig. 16 Saturation frequency of **a** not-matured and **b** matured sugarcane images

Table 2 shows that SVM exhibits high accuracy compared to other machine learning algorithms in labelling maturity. It also has a high precision and recall percentages compared to the others which are ideal to lessen the False Positives ('not-matured' sugarcanes labelled as 'matured') and False Negatives ('matured' sugarcanes labelled as 'not-matured').

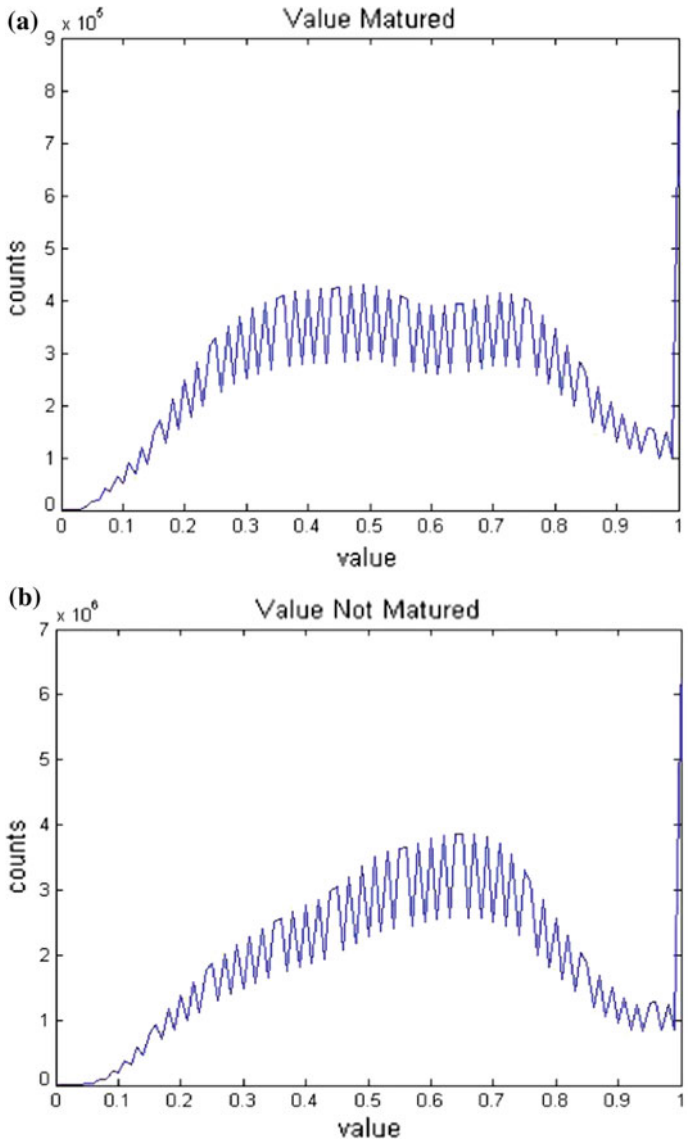


Fig. 17 Value frequency of **a** not-matured and **b** matured sugarcane images

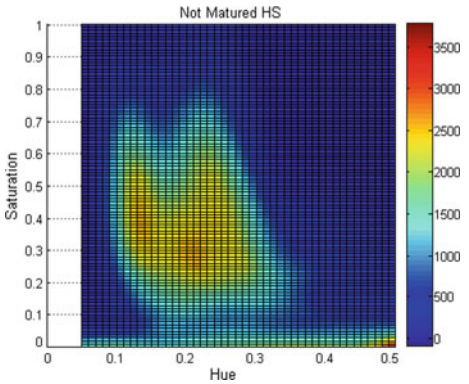


Fig. 18 Hue-saturation graph of not-matured sugarcane

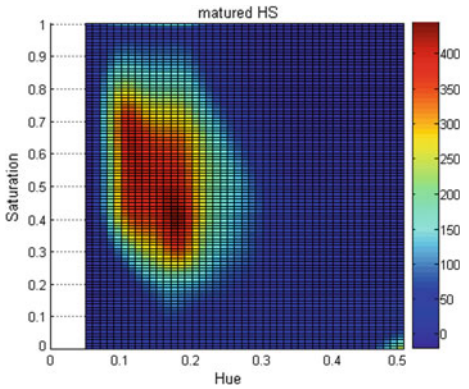


Fig. 19 Hue saturation graph of matured sugarcane

Table 1 Results of model building using machine learning algorithms in WEKA

Machine learning algorithm	Accuracy percentage (%)
Decision stump	90.16
J48	90.40
RandomForest	85.01
Random tree	84.78
SVM	91.33

Table 2 Comparison of machine learning algorithms in WEKA

Machine learning algorithm	% accuracy	% precision	% recall
Decision stump	89.62	54.10	90.90
J48	89.62	54.10	90.90
RandomForest	90.16	59.10	59.10
Random tree	90.16	59.10	59.10
SVM	92.34	62.50	90.90

3.5 System Accuracy Verification for Maturity Analysis

Since it is essential to know how accurate the system classifies and labels the sugarcane's maturity, verification of the system accuracy for maturity analysis is performed. The test data together with the expert label is fed to the machine learning algorithm to determine whether it can accurately classify the data or not.

As mentioned earlier, SVM is the best performing algorithm out of the five algorithms used in classifying sugarcanes. In order to confirm that SVM is the most fitting algorithm for the system, it is compared to the next best performing algorithm, RandomForest.

The results in the SVM machine learning algorithm are compared to the results in RandomForest algorithm. The test data contains 22 matured and 161 not-matured sugarcanes with their first principal component and expert label. While the training data contains 51 matured and 376 not-matured sugarcane with their first principal component and expert label.

The system accuracy percentage using unbalanced training data for RandomForest is 90.16%, while SVM is 92.34% (see Table 3). Sugarcanes that are labelled by the system as 'matured' are categorized as positive, while sugarcanes that are labelled by the system as 'not-matured' are categorized as negative. Table 4 shows the True Positive, False Positive, True Negative, False Negative, Precision and Recall using unbalanced training data. The implemented system makes use of unbalanced training data for both RandomForest and SVM algorithms.

Table 5 shows the system confusion matrix of the implemented system using RandomForest. 18 out of 183 sugarcanes were incorrectly classified by the system with 9 not-matured sugarcanes incorrectly labelled by the system as 'matured'.

Table 3 Results of WEKA system implementation using unbalanced training data

	Total test sugarcane	Correctly classified sugarcane label	Incorrectly classified sugarcane label	Percentage accuracy of correctly classified instances (%)
RandomForest	183	165	18	90.16
Support vector Machines	183	169	14	92.34

Table 4 Table for true positive, false positive, true negative, false negative, precision and recall based on the test data

	RandomForest	LibSVM
True positive (TP)	13/183	20/183
False positive (fp)	9	12
True negative (tn)	152	149
False negative (FN)	9	2
Precision percentage	59.10%	62.50%
Recall percentage	59.10%	90.90%

Table 5 System confusion matrix using RandomForest

		Predicted class	
		Matured	Not-matured
Actual class	Matured	13	9
	Not-matured	9	152

Table 6 System confusion matrix using SVM

		Predicted class	
		Matured	Not-matured
Actual class	Matured	20	2
	Not-matured	12	149

Table 6 shows the system confusion matrix of the implemented system using SVM. 12 out of 183 sugarcanes were incorrectly classified by the system with 12 not-matured sugarcanes were incorrectly labelled by the system as 'matured'. This is because the system treats its Hue-Saturation color space as matured. Sugarcanes in transition to maturity will not fully exhibit the color signature with that of the matured canes.

As mentioned earlier, Precision and Recall percentages are important factors for considering the best performing machine learning algorithm. In the system, RandomForest incorrectly classified 9 'not-matured' sugarcanes as 'matured' sugarcanes and 9 matured sugarcanes were incorrectly classified as 'not-matured' which is the reason it has a precision and recall of 59.10% (see Table 4). SVM's precision percentage is lower than its recall because 12 'not-matured' sugarcanes were incorrectly classified as matured and only 2 'matured' sugarcanes were incorrectly classified as not-matured. In classifying the test data using unbalanced training data, it can be observed that SVM outperforms RandomForest.

However, RandomForest gave more accurate results than SVM using monthly data sets as test data. Table 7 shows the accuracy of SVM in classifying data sets dropped during Month 11 and Month 12. Table 8 shows the number of matured and not-matured sugarcane images included in the monthly data sets.

Table 9 shows the results of RandomForest in classifying the monthly data sets. It can be observed that only 1 not-matured sugarcane was classified as matured sugarcane in Month 9. The precision and recall of RandomForest algorithm are both 83.33% with only 4 sugarcanes incorrectly classified in Month 11. In Month 12,

Table 7 Percent accuracy of RandomForest and SVM in monthly test data (unbalance training data)

	RandomForest % accuracy	SVM % accuracy
Month 07	100	100
Month 08	100	100
Month 09	98.92	100
Month 10	100	100
Month 11	95	86.87
Month 12	85.86	62.63

Table 8 Sugarcane images and actual labels per month (unbalanced training data)

	Total	Matured	Not-matured
Month 07	129	0	129
Month 08	82	0	82
Month 09	92	0	92
Month 10	109	0	109
Month 11	99	12	87
Month 12	99	61	38

Table 9 Precision, recall, true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) for RandomForest algorithm in monthly test data (unbalanced training data)

	Precision (%)	Recall (%)	TP	FP	TN	FN
Month 07	0	0	0	0	129	0
Month 08	0	0	0	0	82	0
Month 09	0	0	0	1	91	0
Month 10	0	0	0	0	109	0
Month 11	83.33	83.33	10	2	85	2
Month 12	88.52	88.52	54	7	31	7

Table 10 Precision, recall, true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) for SVM in monthly test data (unbalanced training data)

	Precision (%)	Recall (%)	TP	FP	TN	FN
Month 07	0	0	0	0	129	0
Month 08	0	0	0	0	82	0
Month 09	0	0	0	0	92	0
Month 10	0	0	0	0	109	0
Month 11	0	0	0	1	86	12
Month 12	62.24	100	61	37	1	0

RandomForest has a precision and recall of 88.52%. On the other hand, SVM, as shown in Table 10, has a precision and recall of 0% for Month 07 to Month 11. It misclassified 13 sugarcanes in Month 11. Although SVM achieved 100% recall, it has a 62.24% precision with 37 not-matured sugarcanes classified as matured sugarcanes. The system results of RandomForest and SVM algorithm were summarized and compared side by side with the actual maturity labels in Tables 15 and 16.

Table 11 Balanced training set

	Matured	Not-matured
Month 07	0	12
Month 08	0	12
Month 09	0	12
Month 10	0	12
Month 11	12	12
Month 12	61	13
Total	73	73

Table 12 Model building accuracy percentage of RandomForest and SVM for balanced training data

	Model building accuracy percentage
RandomForest	73.29
SVM	80.82

Table 13 Percent accuracy, true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) of RandomForest algorithm applied in a balanced training set

	% Accuracy	TP	FP	TN	FN
Month 07	86.92	0	17	112	0
Month 08	85.54	0	12	80	0
Month 09	91.40	0	8	84	0
Month 10	80.91	0	21	88	0
Month 11	86.87	12	13	74	0
Month 12	77.78	61	22	16	0

Table 14 Percent accuracy, true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN) of SVM applied in a balanced training set

	% Accuracy	TP	FP	TN	FN
Month 07	100	0	0	129	0
Month 08	100	0	0	82	0
Month 09	96.77	0	3	89	0
Month 10	100	0	0	109	0
Month 11	82.83	1	6	81	11
Month 12	62.63	61	37	1	0

To see how effective the algorithm is without the bias of the number of matured and not-matured sugarcanes, 73 matured and 73 not-matured sugarcanes were used for the balanced training set as shown in Table 11. Table 12 shows the accuracy percentage of RandomForest and SVM in model building. RandomForest has 73.29% accuracy while SVM has 80.82% accuracy. In the training sets presented previously, the number of not-matured sugarcanes is greater than the number of matured sugarcanes.

Tables 13 and 14 show the percent accuracy, true positives, false positives, true negatives, and false negatives of RandomForest algorithm and SVM using balanced training data. As observed in the tables, SVM is more accurate in classifying

Table 15 Actual maturity percentage

	Actual matured (number)	Actual maturity percentage
Month 07	0	0
Month 08	0	0
Month 09	0	0
Month 10	0	0
Month 11	12	12
Month 12	61	62

Table 16 System maturity percentage

	System matured (number)		System maturity percentage	
	RandomForest	SVM	RandomForest	SVM
Month 07	0	0	0	0
Month 08	0	0	0	0
Month 09	1	0	1	0
Month 10	0	0	0	0
Month 11	12	1	12	1
Month 12	61	98	62	99

Table 17 Human perception in classifying sugarcane maturity

Respondent	% Accuracy	Time in S
1	55.00	247
2	40.00	136
3	60.00	220
4	70.00	344
5	70.00	417
6	55.00	397
7	90.00	147
8	50.00	281
9	70.00	157
10	45.00	199
Average	60.50	254.5

sugarcanes during Month 7 to Month 10 while RandomForest is more accurate in classifying sugarcanes during Month 11 and Month 12.

Since one of the goals of the research is to historically monitor the sugarcanes, the results in the monthly data sets are more significant than the results gathered using the training set. Balanced training set reveals that RandomForest performs better than SVM during Month 11 and Month 12. RandomForest algorithm is the more fitting algorithm to be used in the system because it gave more accurate results

in Month 11 and Month 12 which are crucial months for the maturity of the VMC 87-599 variety. As shown in Tables 15 and 16, the sugarcane reach maturity on the 11th and 12th months.

3.6 System Benchmark

In addition to the testing done for system accuracy, the proponents of the study asked 10 sugarcane experts to classify 10 matured sugarcane images and 10 not-matured sugarcane images. This is done to see the efficiency and accuracy of the system against human perception.

In Table 17, The 10 sugarcane experts were able to classify the 20 sugarcane images in 254.5 s with an average accuracy of 60.50%. The system, which has 90.16% accuracy, was able to classify 183 sugarcane images in less than 2 s making it 29.66% more accurate and 252.5 s faster than human perception.

4 Conclusions

The database is composed of 610 sugarcane samples from Month 7 to Month 12. There are 73 matured sugarcane samples and 537 not-matured sugarcane samples. 70% of the sugarcane samples in the database are used as training data while the remaining 30% are used as test data.

The combination of the Hue and Saturation as illustrated from the top view of a 3D histogram, are said to distinguish matured and not-matured because of the shifting range values. The raw sugarcane images were converted to HSV and undergone masking during the Pre-processing. Not-matured sugarcanes have hue values ranging from 0.1 to 0.27 and shifts to hue values 0.08 to 0.23 as it matures (see Fig. 18). On the other hand, saturation values of not-matured sugarcanes ranges from 0.18 to 0.6 shifts to ranges 0.25–0.78 as it matures (see Fig. 19).

In Pre-processing, the sugarcane images were converted into the HSV color space. Background segmentation was done through Hue thresholding and masking HSV the image. The masked HSV sugarcane image was used in Feature Extraction where the target features were mainly the frequency count of Hue-Saturation values. The extraction was done per image sample and the features were recorded in CSV file format. The data was normalized to avoid imbalance throughout the data. Each sugarcane image has a feature vector length of 4656. To avoid redundant and irrelevant information, the data set's feature vector length was reduced to 1 using Principal Component Analysis.

Since machine learning determines the accuracy of the system in classifying sugarcane images, machine learning techniques such as Decision Stump, Random Tree, RandomForest, J48, and SVM were compared. Although SVM performs better than RandomForest in the unbalanced data sets, RandomForest outperforms

SVM in classifying sugarcane maturity using monthly data which is coherent to the goal of the research that crops be monitored monthly as it matures. The system has a 90.16% accuracy using the RandomForest algorithm.

As an addition to verifying the performance of the system, 10 sugarcane experts were asked to compare 20 sugarcane images (10 matured and 10 not-matured sugarcanes). The experts have an average accuracy of 60.50% and were able to finish the test at an average time of 254.5 s. This makes the system 29.66% more accurate and 252.5 s faster than human perception.

5 Recommendations

There are a number of things to be considered in study should it be pursued in the future, namely: the sugarcane variety, the camera and camera setting, lighting conditions, and weather. The best lighting condition for data gathering is between 6 AM and 8 AM.

Other features such as internodes, presence of flowers in some variety, leaf area index that indicate maturity may be considered by other future proponents. In some varieties, the presence of flowers suggests that the sugarcane is ripe for harvest. The structure of the stalk and foliage is also a good indicator of maturity.

It is important that parallel analysis be conducted to ensure the validity of the research. The parallel analysis may include refractometry and saccharometry testing of sugarcanes periodically during the course of the study.

Acknowledgements The authors would like to thank the following agriculturists and sugarcane plantation managers and owners for sharing their expertise on sugarcane farming: Mauro Merculio, Vivencio Sangil, Virgilio Obra, Dr. Norvie Manigbas, and Mr. Clarence Ortiz.

References

1. Henry, R., and C. Kole. 2010. *Genetics, genomics and breeding of sugarcane*. New York, USA: CRC Press.
2. HSV (Hue, Saturation, and Value) [Online]. Available http://www.tech-faq.com/hsv.html.
3. Lingle, S. 1988. Evidence for the uptake of sucrose intact into sugarcane internodes. *Plant Physiology* 90: 6–8 [Online]. Available http://www.plantphysiol.org/content/90/1/6.full.pdf.
4. Montillo, A. 2009. *Random forests* [Online]. Available http://www.dabi.temple.edu/~hbling/8590.002/Montillo_RandomForests_4-2-2009.pdf.
5. Silva, M., and M. Caputo. 2012. Ripening and use of ripeners for better sugarcane management. In *Crop Management and Tools for Higher Yield and Sustainability*, ed. F. Martin. https://doi.org/10.5772/1107.
6. Smith, L. 2002. *A tutorial on principal components analysis* [Online]. Available http://www.sccg.sk/~haladova/principal_components.pdf.
7. Shalizi, C., and B. Junker. (2010). *Principal component analysis* [Online]. Available http://www.stat.cmu.edu/~cshalizi/490/pca/pca-handout.pdf.

Towards an Automated Plant Height Measurement and Tiller Segmentation of Rice Crops using Image Processing

Karol Paulette Constantino, Elisha Jeremy Gonzales, Lordd Michael Lazaro, Ellen Chelsea Serrano and Briane Paul Samson

Abstract Plant phenotyping is the process of completely assessing the basic and complex characteristics of the plant, which includes height and tiller count. The International Rice Research Institute (IRRI) researchers does plant phenotyping to observe changes in the physical characteristics of the C4 rice crops after modifying its genetic makeup to increase yields without using too much water, land and fertilizer resources. As this advances, the traditional way of observing phenotypic data is still trailing behind. Automated plant phenotyping offers an effective substitute because it allows a regulated image analysis that can be reproduced due to the automation. This is to address the lack in accuracy, reproducibility and traceability in manual phenotyping. With this, an image processing system that automates the measuring of height and the counting of tillers of a rice crop, specifically the C4 rice, was developed. The system applies HSV and Thresholding for preprocessing, Canny Edge Detection (tiller) and Zhang-Suen Thinning Algorithm (height) for the plant structure and tracing and conversion for measuring the height. Tiller counting is done by counting the cluster of pixels in a given region of interest. Four experiments were conducted using different setups and different combinations of algorithms. The fourth experiment was able to get an average percentage error of 76.14% for the tiller count and 238.11% for the height measurement. Presence of shadows and hanging leaves heavily affected the results of this experiment.

K. P. Constantino (✉) · E. J. Gonzales · L. M. Lazaro · E. C. Serrano · B. P. Samson
College of Computer Studies, De La Salle University, Manila, Philippines
e-mail: karolconstantino@dlsu.edu.ph

E. J. Gonzales
e-mail: elishagonzales@dlsu.edu.ph

L. M. Lazaro
e-mail: lorddlazaro@dlsu.edu.ph

E. C. Serrano
e-mail: ellenchelseaserrano@dlsu.edu.ph

B. P. Samson
e-mail: briane.samson@dlsu.edu.ph

J. Billingsley and P. Brett (eds.), *Mechatronics and Machine Vision in Practice 3*,
https://doi.org/10.1007/978-3-319-76947-9_11

Keywords Image · Processing · Phenotyping · Plant · Height
Tiller · Segmentation

1 Introduction

Plant phenotyping is the process of completely assessing the basic and complex characteristics of the plant. The plants complex characteristics include its growth development and yield [1]. Some of the basic characteristics observed when phenotyping plants, specifically rice crops, are its height and tiller count. Plant height is related to the productivity and growth rate of a plant. Plants tend to grow to a certain height in each of its growth state [2]. However, plants drop in growth rate when plants have diseases or lack in water, which results to lower yield rates. In rice crops, more tillers would generally mean more yield [3]. Tillers are the grain-bearing part of rice crops and it possesses the leaves [4]. Leaf count indicates a plant's age and as it grows, its leaves will develop and grow in size [5]. Plants growing in good conditions develop leaves in a faster rate than those leaves growing in non-conducive environment for the plant.

1.1 Rice Crop Tillers & Panicle

This research focuses on phenotyping of rice crops. The two key part of a rice crop that is useful to this research are its tiller and panicle shown in Fig. 1.

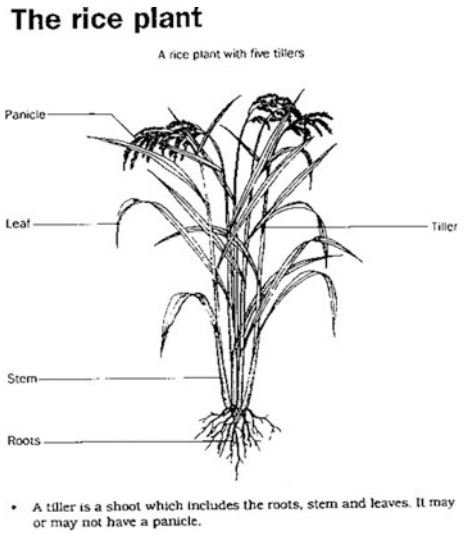


Fig. 1 Tiller of a rice crop [6]

Tillers in rice crops appear as soon the rice crop is selfsupporting [7]. The first tiller usually emerges when the seedling has five leaves. This signifies the start of the tillering stage. From the main stem, another stem would develop. This new stem would be the tiller. Once these tillers are fully developed, they would develop flowers. These flowers are more commonly called as panicles. Tillers are used to measure the height of the rice crop. The height of the tiller with the longest leaf would be the height of the plant.

Panicles are the flowers of the rice. They are placed at the end of the tillers. The seeds or grains produced in the fertilization of the rice plant grow in the panicles of the rice crops. Panicles are one of the identifying quality of a fully developed tiller.

According to Rousseau et al. [8], phenotyping of these traits requires properly tuned rating scales and well trained raters which adds to the cost and time. Furthermore, it is influenced by the subjectivity of raters making it lack in accuracy, reproducibility and traceability [8]. Fatigue of experienced staff can also be one cause of the degradation of accuracy and efficiency [9]. This is the current situation in the International Rice Research Institute (IRRI).

Automated image processing of the phenotyping of plants offers an effective substitute to manual visual assessment [8]. It also allows a regulated image analysis that can be reproduced because of the absence of subjectivity of manual visual assessment and it enables a high throughput because of the automation of the process. Moreover, calibrated protocols and data storage offer beneficial tools to trace or compare results [8].

1.2 Automated Plant Phenotyping

There are a number of researches and systems that has been done that uses image processing to automately measure height of plants and count its tillers. They used different tools to measure the height of an object and number of tillers of plants while applying image processing techniques.

Sritarapipat et al. [2] used marker bars and a formula where the values are taken from the processed image to get the height of the rice crops. They used band selection, filtering, and thresholding to remove noises affected by the rain, wind, and outdoor light, in the images. They took images of rice paddies in a very low plain called Tha Cheen River, a part of a rice field in Suphanburi, Thailand. These images were taken by two Single-Lens Reflex (SLR) cameras controlled by the field servers control unit. However, there are a lot of images being taken and storing these images in high resolution will take up too much storage space. Therefore, they took images in lower resolution at 10:30 a.m. daily to resolve the issues with regards to storage space and data transfer rate. Local Thai Meteorological Department staff recommended that images be taken in the morning as clouds and precipitation has less effect than other times of the day.

According to Sritarapipat et al. [2], the results were compared to the data of manually-measured rice crops. They were able to conclude that excess green had better results than red band. However, they also state that results are dependent on the marker used. Results might also change if there are plants present in the field image that are similar to rice crops, like wheat. The authors wrote that it will be identified as a rice crop. The system is limited by having only one marker bar and it should be in the same place always. Having multiple marker bars in one field image and bars varying in distances was not tested.

A system created by Ikiz [10] uses image processing techniques, like skeletonisation, to measure fiber length. The system would be depended on these factors: sample preparation, lighting technique, resolution, preprocessing algorithm, and processing algorithm. There were two levels of sample preparation that Ikiz [10] did, the one with fiber crossovers and one without. There were also two types of lighting, frontlighting and backlighting to create negative images since the only concern is to know the pixels of the fiber and the background. For preprocessing, both images that were assumed to contain single fiber and images that were assumed to have random crossovers, outline, thinning, and adding algorithms were applied.

The system that Ikiz [10] created achieved a 0.65 mm confidence level, higher than the 0.5 mm that is required. However, when running low resolution images through the adding algorithm, it sometimes results to negative bias. Added noises in the image, like dust or glare, causes points to disconnect resulting to shorter measurements.

Yang et al. [9] created the H-SMART system that performs x-ray computed tomography (CT) to automatically count the number of tillers in rice plant. Pots of plants pass through the x-ray system through a conveyor belt and stops at a rotation platform. There is a distance of 1122 mm between the focal spot and the center of rotation. All computations are done in 20 s and image processing in 5 s. Filtered Blackprojection (FBP) algorithm was used to create an image of the rice culms. Image processing methods, such as median filtering for image denoising, threshold operator for image segmentation, and some morphological operators are used to identify tillers in the FBP image. The researchers used three batches of rice, 50 pots each, from the tillering, heading, and flowering stages were taken and used to test the accuracy of the system. The system was able to have a 95% confidence in accuracy.

The systems, although well-advanced, are still noticeably young and can still be improved. Systems for measuring height are more common to researchers and many have tried different approaches and techniques but there are still a few things that can be improved. The use of marker bars and rulers, done by Sritarapipat et al. [2] and Lee et al. [11], with image processing techniques seems to be two of the most effective approaches. This is the same with the systems that count tillers, although there are fewer studies regarding this.

1.3 Importance of the Research

The problems presented by manual visual phenotyping, such as the additional cost and time due to the requirement of properly tuned rating scales and well trained raters which may be subjective and experiencing fatigue can cause the lack in accuracy, reproducibility and traceability [8] as well as the efficiency, led to the automation of phenotyping through image processing. It also speeds up the process of phenotyping which is a big factor for researchers. Rice crops, in particular, would benefit in the automation because of the growing population of the world, the demand for food increases [2].

Currently, in the Philippines, they are still using manual visual phenotyping and are experiencing the problems stated. The researchers proposes an image processing system, named Seight, that automates the measuring of height and the counting of tillers of a rice crop because although it's a fairly young field and many previous studies used image processing techniques for their systems, there are new and unused image processing techniques that can be used. Also, previous studies do not consider noises present in the image and its effect to the result. An appropriate infrastructure setup would be designed for the automated phenotyping. Seight would only focus in height measurement and tiller count.

2 Methodology

2.1 System Design

To develop the system, the researchers came up with a system design with 6 modules: Data Capturing Module, Data Management Module, Preprocessing Module, Height Module, and Tiller Module shown in Fig. 2.

(1) Data Modules: The Data Capturing Module is where images of C4 rice crops gathered for this research. These are currently cultivated in the International Rice Research Institute (IRRI) in screenhouse. In Fig. 3, it shows the setup used to capture the images. The distance of the camera from the board was also adjusted to 104.5 cm and the height to 53 cm. The angle of the camera was lowered to 5° and a colored card, which will be used as a marker, was placed at the top of the board. The camera used was a Canon 60D with an image size of

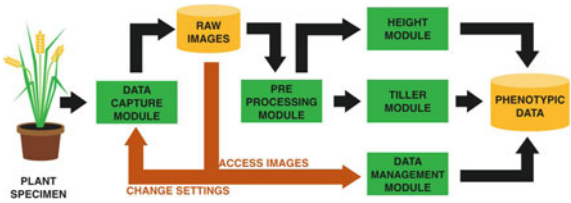


Fig. 2 System architecture

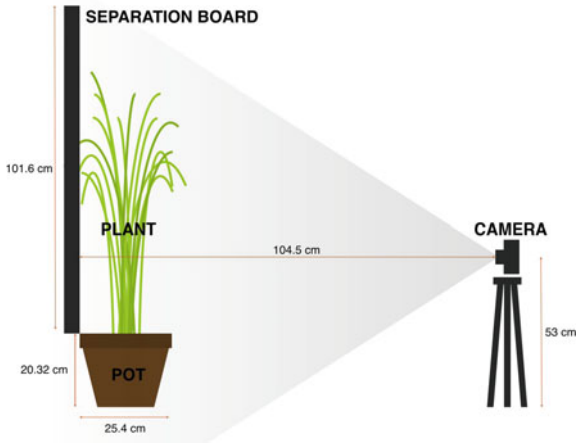


Fig. 3 Camera setup

1296 × 3110 pixels. The Data Management Module handles the plant images from the Data Capturing Module. The Data Management Module includes the database of the system and the settings to change the schedule of image capture of the plants. The user interface for this module is in a form of a website.

(2) Preprocessing Module: The Preprocessing Module prepares the raw images for analysis. Background removal, filtering and edge detection are the techniques that would be used for preprocessing. Background removal and filtering fall under noise reduction. These techniques would, respectively, lessen the noise and unnecessary details in the image. Edge detection is a subfunction of plant segmentation. Plant segmentation would segment the plants present in the image from each other. This allows the system to focus more on assessing a single plant which in turn increases in the efficiency of the assessment.

(3) Tiller Module: The Tiller Module determines the tiller of the plant. It is in this module that the tillers are counted. Edge detection and getting of the Region of Interest (ROI) were applied to the preprocessed image before counting the tillers.

(4) Height Module: The Height Module is responsible for determining the height of the plant in the image. Skeletonization and tracing were applied to the preprocessed image to get the height.

3 Seight

Four experiments were conducted using different setups and different combinations of algorithms.

3.1 Experiment 1

The first experiment was done as a part of the concept formulation of the methodology cycle. In this experiment, the design of the board and the specifications for the camera setup was formulated. This implementation of the designs formulated was not done inside the screenhouse of IRRI. The lemongrass and citronella plant were used as substitutes to the C4 rice plant.

Three boards, plain white background, white background with black stripes, and black background with white stripes, were used as part of the experiment for the board design. The board that is most appropriate for the study is the board with the plain white background. The striped board design was deemed unnecessary due to the new information gathered about height measurement. Initially, height was to be measured by taking the highest part of the plant and computing the height from that to a parallel point in the plant base row. However, since the method changed upon confirmation, the striped design became unnecessary.

The board's measurement were 76.2 cm by 101.6 cm. It is elevated from the ground at 20.32 cm, which is the height of the pot. The camera used for this experiment is Nikon Coolpix L21 digital camera that has an image resolution of 640 × 480 pixels. It was placed perpendicularly to the boards center with a distance of 200–300 cm. It was elevated 50–100 cm from the flat ground using a tripod.

3.2 Experiment 2

In the second experiment, the camera and board setup from the first experiment was followed in this experiment. A meter stick was placed on the left side of the board to aid in height measurement. The images for this experiment were taken using Nikon D40 with an image resolution of 2000 × 3008 pixels. This was done in the IRRI screenhouse.

The raw image of the plant was processed to isolate the rice plant from the background through thresholding in the HSV color space. Using the OpenCV function inRange() a minimum HSV color value and a maximum HSV color value was taken to determine which pixels to get in the image. The minimum HSV values were 0, 45 and 10 for hue, saturation and value, respectively. The maximum HSV values were 135, 255, 100.

To get the height measurement of the plant, its form is identified using the skeletonization technique, Zhang-Suen's algorithm. The system gets the highest and the lowest pixels in the image by iterating the image starting from the top left corner until the bottom right corner, moving to the right downwards. It takes the first white pixel it sees as the highest pixel. The same process was done to get the lowest pixel but the system starts at the bottom-right of the image to the topleft, moving left, upwards.

The Euclidean distance of the highest and lowest pixels would be identified. The Euclidean distance was converted from pixels to centimeters. The conversion from pixel to centimeters was based on the meter stick placed within the image. The pixels in a centimeter with reference to the meter stick are counted. This is then assigned as the conversion factor which is used to divide the height in pixels. The height of the plant was the result of the conversion.

The tiller count of the plant was computed by getting the edges of the plant in the image by applying Canny edge detection algorithm [12]. From the edge detected image and the preprocessed image, two regions of interest was created to focus on the lower part of the image. This region contains the section of the rice crop where the tillers are most visible. With the coordinates 0, 0 in the image is at the topmost-left, the starting coordinates of the region of interest is at 160, 515 with a width and height of 106 pixels.

The x variable was fixed at the middle row of pixels in the image and only the middle row of the regions of interest was tested. The tillers are identified by iterating through the middle row of pixels. Two regions of interest was used to count the tillers; the edge-detected image region of interest and the preprocessed image region of interest. When the system detected an edge on the edge-detected image, it looks at the neighboring pixel of the region of interest in the preprocessed image. If the neighboring pixel is not black, the tiller count was incremented.

3.3 Experiment 3

In the third experiment, the board and camera setup was modified to accommodate the actual setup of the plants inside IRRI's screenhouse. A separator was added in the board design to separate the two rice plants that was captured in the image. The new camera setups new height from the ground to the lens of the camera is 21.59 cm, the distance from the plant to the stand of the camera is 36.5 cm with an 8° vertical angle and a 0° degree horizontal angle. The camera used in this experiment was D-Link DCS-932L, an IP camera, with a 640 × 480 pixels resolution.

Due to the change in the quality of images, there was a need to add color balancing to the preprocessing module. The balancing of the images were done through equalization of the images histogram. The thresholding technique was also modified. In this experiment, there were two color spaces observed, HSV and YCrCb. Combinations of the equalization and thresholding in this color spaces were assessed to see what yielded the better result. The combination that yielded the better result was HSV for the equalization and YCrCb on thresholding. Table 1 show the values used for equalization and thresholding for this experiment.

The skeletonization algorithm used for this experiment was still the Zhang-Suen algorithm. The method of getting the height was getting the highest point in the image and the base as the lowest point. It uses converted Euclidean distance as the height. The reference used to get the conversion factor is the frame of the separation

Table 1 Table of values used for equalization and thresholding

Combination	Equalization	Thresholding algorithm values
HSV + HSV	Equalized in H channel using OpenCV equalizeHist() function	min: HSV(0, 45, 10) max: HSV(135, 255, 80)
HSV + YCrCb	Equalized in H channel using OpenCV equalizeHist() function	min: YCrCb(0, 0, 0) max: YCrCb(60, 255, 255)
YCrCb + HSV	Equalized in Y channel using OpenCV equalizeHist() function	min: HSV(0, 45, 10) max: HSV(135, 255, 100)
YCrCb + YCrCb	Equalized in Y channel using OpenCV equalizeHist() function	min: YCrCb(0, 0, 0) max: YCrCb(60, 255, 255)

board which has a width of 2.54 cm (1 inch). The pixels within that area is counted and used to divide the computed distance resulting to the height measurement of the plant in centimeters.

The algorithm for counting the tiller still remains the same. However, in this experiment, all rows of pixels were considered for the counting. The mode of the tiller count of the rows of pixels were considered by the system as the tiller count. The region of interest was changed to accommodate the change in resolution of the images. With the coordinates 0,0 in the image is at the topmost-left, the starting coordinates of the region of interest would be at 23, 380 with a width of 170 pixels and a height of 75 pixels.

3.4 Experiment 4

In the fourth experiment, the board and camera setup from the third experiment was followed in this experiment. However, additional side panels were added as walls to make the background of the plants similar on both sides. Figure 4 shows the schematics of the added design. As shown in the said figure, the side panels are 38.1 cm wide and 101.6 cm long.

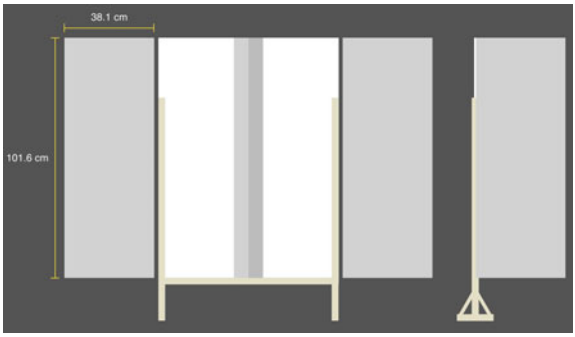


Fig. 4 Third board setup

The distance of the camera from the board was also adjusted to 104.5 cm and the height to 53 cm. The angle of the camera was lowered to 5° and a colored card, which will be used as a marker, was placed at the top of the board. Each image contains two pots of rice crops against the white board. These were taken using a Canon 60D with an image size of 1296 × 3110 pixels. A total of 28 images of rice plants in pots were taken.

Since the raw images taken include the background behind the marker board, the images were cropped. A quarter of the image was removed from each side. A tenth portion was removed from the bottom of the image to cut out the pot. Then, the resulting image was split in half to separate the two plants in the image. The images of individual plants were then normalized using the OpenCV function normalize(). Afterwards, the brightness were adjusted by +5 and the contrast was increased by 70% to make the plant more perceivable in the image. The plant was segmented from its background using HSV segmentation. The minimum HSV color values were 15, 35 and 40. The maximum color values were 70, 250, 240.

The height still used Zhang-Suen algorithm to skeletonize the image. After getting the skeleton, the system looked for the highest white-valued pixel. From there, it traced the tiller to the base. The neighboring pixels of the current pixel is checked to get the closest white pixel. The pixel which is directly below the current pixel is always checked first. The algorithm was also designed to address disconnecting lines in the image due to the skeletonization process done in the image. This is done by increasing the range the algorithm looks at. It can increase from checking 1 pixel up to checking 1% of the images height.

Euclidean distance was still used to get the height in pixels. It would be converted to centimeters using a conversion factor. A marker was used to identify the conversion factor to be used for converting the pixels to centimeters. A red marker with 4.7625 cm (1.875 inches) height and width was used. This was a quarter of a 3.75 inched note paper. This makes the marker have a height and width of 1.875 inches and was then converted into centimeters. It is placed on the top portion of the marker board. It would be detected in the image through HSV segmentation. Canny edge detection would be applied to the image from the segmentation. The largest contour was identified as the marker. The system gets the height of the contour. The height of the contour became the conversion factor.

The tiller module still used Canny edge detection algorithm and the algorithm used in counting the tillers in the previous experiment. In this experiment, the computation of the region of interest became dynamic. The starting y-value of the region of interest is the bottom one-eighth of the image, assuming that the tillers are in the bottom portion of it. The starting x-value of the region of interest is the middle column of the image subtracted with one-sixth of the width of the image. The width of the region of interest is one-third of the images width, while the height of the region of interest is one-eighth of the images height.

4 Results

For the fourth experiment, HSV was used for both preprocessing and segmentation of the raw images.

For the tiller count, it yielded a 76.14% average percentage error. It is because the bottom part of the images contains a significant amount of noise. A lot of unnecessary edges are detected, resulting in a higher tiller count than the actual count. Also, in the new set of images gathered, there were a lot of low hanging leaves, adding to the noise. That is why in Fig. 5, there are more instances where system counted more than the view count.view count. The view count is the number of tillers counted manually base from the visible tillers in the image.

For the height measurement, there are three major metrics that were used to analyze the results. The first one is the tracing length error which allows the evaluation of the tracing algorithm in the system separately from the conversion algorithm. It is computed by comparing the length of the traced tiller of the tracing algorithm in the system using raw images and the length of the manually traced tiller using an image which was generated by manually tracing the tiller with an image-editing software which resulted in a binary image with one traced tiller of one pixel width. The length of the manually traced tiller is considered as the measured value and the length of the system traced tiller is considered as the computed value when computing for the relative percentage error. The tracing relative percentage error is 22.5% with a standard deviation of 26 pixels that was rounded down from 26.15 pixels. Those with low trace length error have longer lengths than the measurements with the manually traced images while those with high error have short lengths.

In Fig. 6, the histogram shows the distribution of the relative percentage error, the standard deviation is a measure how spread out the numbers are. There is also a high frequency of relative percentage error that is from 0 to 25%. Some of the plants have high percentage errors because the separation board was detected as a

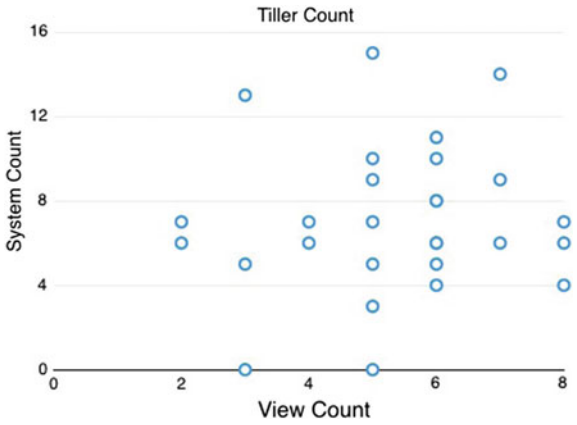


Fig. 5 Tiller scatter plot

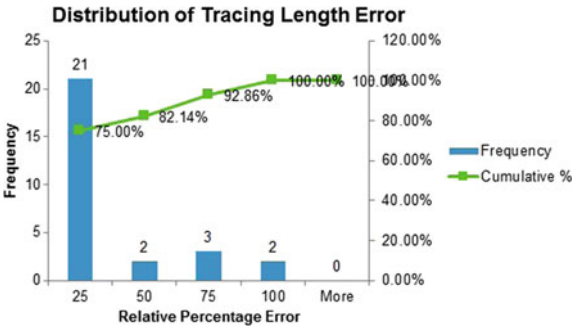


Fig. 6 Tracing length relative percentage error distribution

tiller. There are also big gaps in the skeleton that stopped the tracing algorithm. The small branches in the skeleton could be improved by adding a backtracking method in the tracing algorithm.

The next metric is the conversion error which is used to evaluate the conversion algorithm independently from the tracing algorithm. The length of the manually traced tiller is converted to centimeters using the conversion algorithm and it is compared to the actual length of the plant in centimeters. The actual length of the plant was taken by the IRRI researchers by using a meter stick to measure the length of the tallest tiller in the image. The manually traced tiller length was used because it is the ideal tracing of the tiller and with it the results will only be influenced by the conversion algorithm or an error in the actual physical measurement of the plant. The mean of the relative percentage error of the conversion is 230.73% with a standard deviation of 314.60 cm.

There are some are very high errors that reaches almost up to 1250% error. It is because it detects the noise around instead of the marker. Furthermore, as seen in Fig. 7, those that resulted in a low conversion error has a conversion factor of ranging from 30 to 35 of pixels per centimeter based from the marker and as the conversion factor decreases, the conversion error increases.

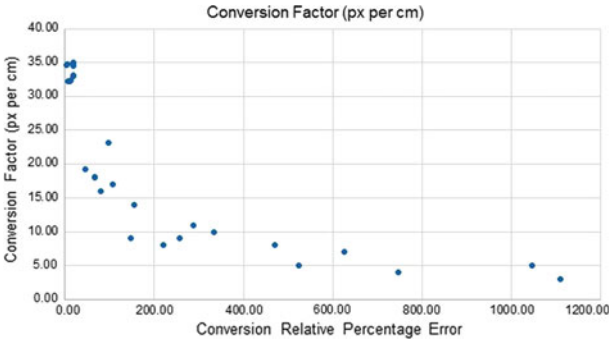


Fig. 7 Conversion factor and conversion error

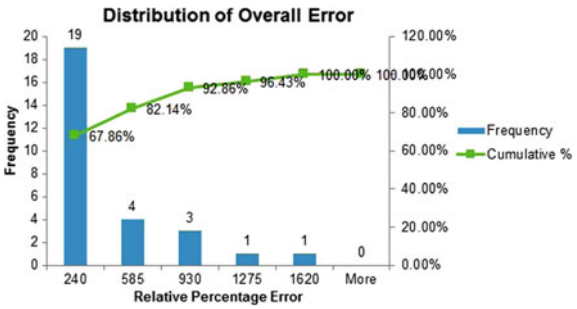


Fig. 8 Overall relative percentage error distribution

The final metric for the height measurement is the overall relative percentage error. It is computed by comparing the actual physically measured length of the plant and the system traced and converted length using the raw images. The mean of the overall relative percentage error is 238.11% with a standard deviation of 345.93 cm. The distribution of the overall error is positively skewed as seen in Fig. 8. Those that resulted in a very low conversion factor had very high conversion errors which in turn affected the overall percentage error.

5 Summary & Conclusion

Plant phenotyping is important in studying the growth and yield of the crops. The automation of plant phenotyping is essential in making it produce high throughputs and outputting results that are more accurate and reproducible. Seight is a system developed to automatically phenotype C4 rice crops. C4 rice is a rice that is being enhanced to increase yield. The system measures the height and counts the tiller of a rice crop through different image processing techniques. The plant is segmented from the image with HSV segmentation and thresholding. The height was measured by getting the highest pixel from the skeletonized image, identifying points in between until it reaches the base of the plant. It then gets the sum of the distances between points and converted from pixel to cm. The tillers were counted by setting an ROI in the image and all pixels within the ROI is checked. The tiller count is the mode of all the counted tillers for every row.

The second experiment has an average percentage error of the tiller count is 34.02 and 17.25% for the height measurement. For the third experiment, the combination with the lowest average percentage error of 33.47% was the HSV for preprocessing with YCrCb colorspace. On the other hand, the fourth experiment was able to get an average percentage error of 76.14% for the tiller count and 238.11% for the height measurement.

In conclusion, although the system is still lacking in accuracy, it was able to automatically count the number of tillers and measure the height of the plant, which is the main objective of the system. The fourth experiment was the best in terms of algorithms used for both the height measurement and tiller count.

References

1. [Online]. Available www.lemnatec.com/plant-phenotyping/.
2. Sritarapipat, T., P. Rakwatin, and T. Kasetkasem. 2014. Automatic rice crop height measurement using a field server and digital image processing. *Sensors* 14 (1): 900–926.
3. Li, X., Q. Qian, Z. Fu, Y. Wang, G. Xiong, D. Zeng, X. Wang, X. Liu, S. Teng, F. Hiroshi, et al. 2003. Control of tillering in rice. *Nature* 422 (6932): 618–621.
4. Tiller(Botany). infoSources.org, [Online; accessed 14 July [Online]. Available http://www.infosources.org/what\is/Tiller\(botany).html.
5. Wood, A.J., and J. Roper. 2000. A simple & nondestructive technique for measuring plant growth & development. *The American Biology Teacher* 62 (3): 215–217.
6. E.B. Radcliffe, and W. Hutchinson. 1999. Radcliffe's IPM world textbook. Minneapolis: University of Minnesota Press.
7. J. Maclean, Rice Almanac. 2013. Source book for one of the most important economic activities on earth. Philippines: IRRI.
8. Rousseau, C., E. Belin, E. Bove, D. Rousseau, F. Fabre, R. Berruyer, J. Guillaumes, C. Manceau, M.-A. Jacques, and T. Boureau. 2013. High throughput quantitative phenotyping of plant resistance using chlorophyll fluorescence image analysis. *Plant methods* 9 (1): 17.
9. Yang, W., X. Xu, L. Duan, Q. Luo, S. Chen, S. Zeng, and Q. Liu. 2011. Highthroughput measurement of rice tillers using a conveyor equipped with x-ray computed tomography. *Review of Scientific Instruments* 82 (2): 025102.
10. Ikiz, Y., J. Rust, W. Jasper, and H. Trussell. 2001. Fiber length measurement by image processing. *Textile Research Journal* 71 (10): 905–910.
11. Lee, J., E.-D. Lee, H.-O. Tark, J.-W. Hwang, and D.-Y. Yoon. 2008. Efficient height measurement method of surveillance camera image. *Forensic Science International* 177 (1): 17–23.
12. Canny, J. 1986. A computational approach to edge detection. *Pattern Analysis and Machine Intelligence, IEEE Transactions on* 6: 679–698.

Part III
Mechatronics in Human Interaction

The Use of Games Software to Enhance Educational Material

Samuel N. Cubero and John Billingsley

Abstract It is suggested that constructing course material using games frameworks might ameliorate problems of inattentive students. Four types of inattentive students are described and common reasons for their lack of attention in class are discussed. Each individual student possesses different levels of 'intelligence' or talents. Unfortunately, only a few of these are promoted, developed and assessed in most education. Fun and entertaining video games have clearly demonstrated the potential to hold the attention of students for long periods of time. This attribute can be of great value if imparted to teaching material. Some game design strategies are outlined, together with an overview of the tools that teachers will need to learn in order to develop such games.

Keywords Games · Education · Gamification · Edutainment · Intelligence

1 Introduction

As the years pass, more and more students are diagnosed with Attention Deficit Disorder (ADD) or Attention Deficit Hyperactivity Disorder (ADHD), finding it difficult to concentrate and pay attention in class. This paper proposes that teaching material should be imbued with the same excitement that gives computer games their addictive appeal, to present a feasible solution to this growing worldwide problem.

S. N. Cubero (✉)
Department of Mechanical Engineering, Petroleum Institute (PI),
Khalifa University of Science and Technology (KUST), P. O. Box 2533,
Abu Dhabi, United Arab Emirates
e-mail: sam@samcubero.com

J. Billingsley (✉)
Faculty of Health, Engineering & Sciences, School of Mechanical & Electrical Engineering, University of Southern Queensland (USQ), Toowoomba,
QLD, Australia
e-mail: john.billingsley@usq.edu.au

J. Billingsley and P. Brett (eds.), *Mechatronics and Machine Vision in Practice 3*,
https://doi.org/10.1007/978-3-319-76947-9_12

While the attention of many youngsters will be occupied in posting comments on social networking sites, the impact of the rapid development of computer games is unquestionable. Graphic effects are much faster paced and much more exciting than those found in movies and video games made a few decades ago. It is unfortunate that most school subjects and training courses today are taught much in the same way that they were taught a century ago. Although PowerPoint has assumed the role of chalk and a video recording might have replaced the human teacher, such a talking instructor can only deliver information by speaking words at a speed that is considered to be too slow and boring by today's players of video-games. Their senses are accustomed to constant stimulation and surprises.

Computer simulations have already been linked to many courses, but in the main, these still present the material in an academic manner that is often linear and predictable. For example, the Maersk oil drilling simulator [1] accurately mimics all the common operations on an offshore drilling rig platform and is extremely useful as a tool for training drill rig operators. However, just like most 'realistic' aircraft flight simulators, learning how to operate all the controls may be viewed by some as tedious and boring.

By exploiting the software tools available for creating online games, captivating tasks can be set that portray the true physics of the imaginary situation and demand that the student understand and demonstrate course principles to reach an exciting goal.

2 Why Are Students not Paying Attention?

Today, teachers and instructors face the worldwide problem of dwindling attention and concentration shown by students in a conventional classroom environment. This is exacerbated by the provision of lecture recordings on the web.

The 'inattentive students' fall into four typical classes.

(Type 1) 'Technology distracted students' are often seen sending text messages, viewing YouTube clips, or playing games on their mobile phones or laptops during class time.

(Type 2) 'Socializers who want attention' are other students who prefer chatting and socializing with their peers, rather than listening to the teacher.

(Type 3) 'The uninterested and the day-dreamers' are present in body and appear to be paying attention. However, their minds are elsewhere, focused on other things that are totally unrelated to what is being taught in class. These types of students do not find the taught subject matter, or the way it is being delivered, to be interesting enough or important enough to deserve their full attention.

(Type 4) 'Sleepers or sleep-deprived entertainment addicts'—those who appear to be tired, languishing, 'nodding off' with 'heavy eyelids' or sleeping with their heads on their desks. These kinds of students are physically and mentally too tired to concentrate and participate in class activities, due to insufficient sleep. Some students lack self-discipline and spend far too many hours each day playing addictive video games, or watching too many TV shows and movies.

Those others who are wide awake might find the rate of information delivery too slow and unchallenging.

These problems seem to be getting worse over the years, as young children and teenagers are exposed to new electronic devices with highly-addictive 3D video games that hyper-stimulate the human senses. Modern computer games are able to generate almost photorealistic characters at very high frame-rates. The speed at which stories are told has increased significantly over time. It could be deduced that today's students have become faster thinkers and faster learners, but this is seldom substantiated by assessment results. Why can students become experts at playing complex video games, while failing to become experts at completing their weekly assignments and school studies?

Why should games have the power to grip the attention of students in a way that seems to be lacking in teaching material? Could 'enjoyment' be the difference?

In both cases there is a clear objective, either to attain some goal by discovering and exploiting features within the course or game, or by collaborating with peers in a group activity. In principle there is an equivalence between finding that a sword's magic properties can only be invoked by knowing the words of a spell, or understanding that 'action and reaction are equal and opposite' when applying dynamics to an object.

Online games can be very addictive and can run for several days. A similar level of addiction to teaching material is usually only seen on the day before an examination or when an assignment is due.

3 Outdated and Inflexible Approaches to Teaching

Although most of the paperwork of textbooks and assignments has today been transformed into button clicking and computer keyboard work, the ways in which students are expected to learn new materials have changed little over the years. Most educational materials are still presented in a 'linear' fashion or in chronological order, just like the chapters of a printed book. The same is true of quiz questions.

The choice of topics, order of topics, style and method of learning for most students today is still largely dictated and controlled by a single teacher or instructor. These methods may not be compatible with every student's preferred or most effective style of learning. This 'one size fits all' approach to teaching is highly inflexible and is no doubt one of the major reasons that many students lose interest or find it difficult to learn in a classroom or lecture theatre environment.

4 Gardner's 'Theory of Multiple Intelligences'

Psychology research has identified different kinds of human intelligence, many of which are closely associated with a particular human sense, brain function or part of the human body. Howard Gardner [2], author of the book: *Frames of Mind: The Theory of Multiple Intelligences*, has identified nine different kinds of human intelligence which can be demonstrated and communicated with its own jargon:

1. Musical–rhythmic and harmonic (hearing, timing, ability to sing in tune to the correct pitch and timing, and/or play musical instruments. Ability to compose and play original music.) This is also related to 'Intelligence No. 5'.
2. Visual–spatial (vision, visualization, imagination; able to interpret and create complex drawings, charts, diagrams, maps, schematics, 3D models, etc.).
3. Verbal–linguistic (speaking, writing, learning and communicating, often in more than one language).
4. Logical–mathematical (problem solving, finding and testing relationships between objects and variables, connective thinking.) This is strongly related to 'Intelligence No. 2' (visual—spatial).
5. Bodily–kinesthetic (good hand-eye coordination and motor skills.) Somewhat related to 'Intelligence 1'.
6. Interpersonal (compassion, empathy, ability to make friends and maintain good rapport with others, 'people skills'—or 'emotional intelligence', teamwork skills, ability to influence and motivate others to change or take action [3]).
7. Intrapersonal (introspective or self-reflective abilities; having deep self-understanding and self-awareness; ability to control actions and emotions; ability to manage oneself, set goals and make decisions based on values and priorities [3]).
8. Naturalistic (caring about the environment, animals, plants, natural medicines, etc.).
9. Existential (mainly concerned with philosophy, religion, spiritual knowledge, etc.).

5 How Schools Have Failed Students

Most schools in the western world spend the majority of their effort teaching intelligences 3 and 4, 'verbal-linguistic' and 'logical-mathematical'. Skills relating to intelligences 1, 2 and 5 are regarded as arts, sidelined from the factors by which the school's performance is assessed.

Every person is unique and may be talented or highly skilled in a few of the nine intelligences, while being weak in many others. It is rare to find a person who is highly skilled in all areas. Young people who have spent a lot of time playing 3D video games usually have acquired very high 'visual—spatial' intelligence and are

able to absorb and interpret graphical information at a high rate. They prefer to learn with their eyes. Such students tend to become bored when listening to verbal lectures. Most average adults can read and understand between 400 and 600 words per minute [4], whereas most talk at a much slower rate, at only 100–200 words per minute.

With information delivered at a fraction of their maximum 'word comprehension rate', students may try to use their 'unused' brain capacity by socializing, checking SMS messages or playing games on their smart-phones.

Some students whose strength is 'Intelligence 5' learn and perform better when doing 'hands-on' laboratory work, in contrast to theoretical examinations and assignments. Brian Tracy [4] suggested that new information is easier to remember and recall if more than one type of sensory input is used when learning it. This is because more brain cells are used for storing such information. In addition, the more times that information is repeated or recalled, the easier it becomes to recall it.

Some students are naturally gifted at 'Intelligences 3, 6 and 7', and have developed into excellent communicators, speakers and team leaders, however, such intelligences are seldom evaluated in Engineering, mathematics and science subjects, nor do they help a student to earn very high grades in math and science subjects. Such students usually feel undervalued, unappreciated and unrewarded. They tend to drop out of High School or University to start a job or try to run their own businesses.

They perhaps bear out the message of Robert T. Kiyosaki's book, *Why A students work for C students and B students work for the government* [5]. The stories of Gates and Zuckerberg prompt us to question whether their abilities might have had no need of the educational system, or whether reforms to teaching are needed.

6 Education Tailored to the Individual Needs of Each Student

What can be done to improve interest in course materials? One solution is to remove the focus and attention from human teachers and to present all or most of the course materials and assessments in the form of a video game that is 'fun' yet challenging enough to keep students curious and interested. Using games to teach or train is also known as 'edutainment' or 'gamification', where complex or otherwise 'dull or dry' technical knowledge is presented within a very entertaining puzzle or challenge.

When designing such a game, players should be able to use several different kinds of 'human intelligences' in order to arrive at the same stage in the game, so that each student can learn easily using his or her preferred style of learning. A 'musical-rhythmic' learner would benefit from audio that could provide clues for completing a mission. A 'bodily-kinesthetic' learner might enjoy using a 'motion capture' (or 'mocap') suit or gloves to manipulate or assemble objects in a VR

(Virtual Reality) scene; while a student who is strong in interpersonal skills would prefer interacting with other characters in a game to learn important information for solving a puzzle or a challenge.

It is also important to give students the freedom to choose their goals, and to learn topics needed to attain them in any order that they prefer. If they encounter a mission that is too difficult to complete, the game should detect multiple failures or lack of progress and provide hints about which topic they need to master first. With this flexibility, advanced students can make faster progress than less knowledgeable students, enabling topics to be skipped at any time if desired. The same ultimate level of skill is to be reached by all players, although the effort required will vary with ability.

7 Popular Tools for Creating Computer Games

There are many levels at which interactive software can be developed and many different styles of gameplay. Some games still evoke nostalgia, such as the forty-year-old 'Leisure Suit Larry', which involved making a sequence of choices based on the experience gained from multiple failures. A framework for such a game can be a platform for iterative learning, only demanding planned pathways and some captivating still illustrations.

In contrast, those with abilities in JavaScript and HTML5 can create puzzle games and simulations 'from scratch'. It is deplorable that such films as 'Star Wars' portray dynamics that defy Newton's laws, where spaceships can change their course by banking in the style of an antique biplane. As an attempt to correct this misinformation, simple simulations can demonstrate dynamics as they really are, such as the 'Moon Lander' [6] modelled on a game dating back to the 1970s.

But such simulations might only hold the attention for a brief time. For a complete course, something is needed with a multitude of levels and the same ability to engross as 'World of Warcraft'. For example, players could explore a 3D game world, seeking and collecting (educational) information along the way so that they can successfully solve several different kinds of course-related problems in order to earn rewards, defeat 'boss monsters' or gain access to the 'next level' of the game (or course).

Presenting technical information in an entertaining manner can make the learning process much more enjoyable and pleasurable for students—perhaps even addictive. The main goal here is to make the learning (gaming) experience so enjoyable that students will want to know more and learn more, just to feel successful at learning new skills, solving difficult problems, or perhaps 'outscoring' their fellow students.

There is also great potential for creating 'multi-player' and team-based learning experiences. For example, students on a team could collect all the parts and then assemble, build and test-drive a steerable and rideable vehicle, similar to 'Car Mechanic Simulator' or 'Scrap Mechanic' [7, 8]. In a proper engineering course

that teaches machine design and usage of common engineering components, players should be able to select and connect realistic models of components based on calculations, performance limits and material properties. The combined behaviour of all selected and assembled engineering components will then determine the performance of the complete system or machine.

It is important that the simulated model of the 'entire machine' behaves just like a 'real-world' machine. If a component on a circuit board is drawing too much current and is over-heating, it should be seen and heard to catch fire and show smoke. Obvious indicators of success or failure are important to give players real-time feedback.

While the top box-office games creators are jealous of their software, others are keen to see their framework exploited to create new games, in exchange for a proportion of any revenue gained. This could however open the question of their entitlement to revenue from educational software that is not marketed as such, but which forms part of a purchased course. Nevertheless, of the hundred or so game engines that are listed in Wikipedia, many have free public use licenses.

At the time of this writing, the two most popular 3D game engines being used to produce commercial quality games are 'Unreal Engine' [9] and 'Unity 3D' [10]. The 'Unreal Engine' is one of the most popular frameworks for creating commercial-quality 3D games. Knowledge of the C++ programming language or 'Blueprints' flowcharting is needed to develop games on Unreal Engine 4. Although its development tools can all be downloaded and used at no cost, a 5% royalty must be paid to its developer (Epic Games) if the game is of commercial quality. In contrast, all games made with 'Unity' engine are royalty-free, but game developers must purchase a licence with a one-time fee. While Unreal Engine can be used to create many different kinds of 2D games, Unity is far more popular among developers for creating 2D games for mobile phones and tablets, perhaps because it has a larger online community of developers and contributors. Games made with the Unity 5 Engine can be developed using the C# or Java language (to make script code), which are easier to learn and use than C++.

Perhaps the easiest way to make highly customized 3D games is with 'GameGuru' and AGK 2 (BASIC or C++) programming [11], however, animated characters and 3D assets must still be created and loaded into AGK 2. Modelling and animation software can be used to create animation sequences for rigged models and textures that can be imported into a game engine. Most modern game engines also come with add-on commands or libraries to handle Virtual Reality headsets, providing a highly immersive viewing and playing experience.

Creating a popular 3D computer game is a very complex task, requiring the breadth of talent involved in producing a 'blockbuster' Hollywood movie. Games frameworks can be used to minimise the effort needed for game programming. Such frameworks allow game developers to focus mainly on creating the game and the story. Copyright-free clipart and 3D models are also freely available, or can be purchased from asset stores online. What is essential is the 'story' that will carry the student along. An outline example is given in the appendix.

For those who want to create their own illustrations, there are countless 'apps' and software suites. But if 'flat' illustrations are not enough, 3D packages can render and animate characters as they move and rotate. Many are free, such as 'Blender', but they may have a steep learning curve.

8 Current 'State-of-the-Art' in Edutainment

Simulation software has already been mentioned, where applications such as 'Flight Simulator' can illustrate authentic dynamics to train the reactions of the student. Other software mimics traditional learning games, such as 'Pelmanism' for training memory. All the classic educational games for children, encouraging reading, arithmetic and logical deduction, can be delivered on the computer.

These have the advantage that the student can 'get it wrong' with little or no penalty, restarting the game for a better attempt. But have these the allure of the 'shoot-em-up' 'kill the dragon' games that are so addictive?

9 'Best Practice' Principles for 'Gamification'

Amy Jo Kim [12] listed seven main concepts necessary to create a successful and engaging 'gamified' experience that will keep players interested and keen to come back for more. These include:

1. Design for the 3 key stages of your player lifecycle: Beginner, Expert, and Leader/Master.
2. Build a system that is easy to learn but hard to master.
3. Use game mechanics to guide the way towards mastery.
4. As players progress, increase the complexity and challenges.
5. Implement intrinsic motivators like power, autonomy and belonging.

Players should feel compelled to complete challenges and difficult challenges, not because they will gain artificial 'external' rewards or bonus points, nor out of fear of punishment, but to enjoy 'intrinsic' or 'internal' rewards, such as a sense of accomplishment, self-improvement or satisfaction after successfully solving a difficult problem.

Several different methods for effectively 'gamifying' course materials are described by various researchers [13–18], and many of these didactic techniques can be applied in the development of educational software games. This is still a fairly new field of research (probably less than 10 years old). Not much detailed research has yet been published describing the effectiveness of using entertaining 3D video games (not just 3D simulators) to teach engineering and technical courses.

10 Conclusions

Learning happens naturally via 3 E's, in this exact order:

1. Exposure (sensing and absorbing new information—either self-taught or in school)
2. Experimentation (trying one's own variations on an example—hands-on, even if it involves many failures); and
3. Experience (repetition and practice).

If students are not paying attention, they will not absorb any new information that is presented. If video games can hold the attention of students, then they can overcome this barrier and students can begin to experience the 3 E's of learning.

Games can be designed to exploit all nine different types of human 'intelligence', as described by Gardner. That is the challenge awaiting all teachers.

Acknowledgements The authors wish to acknowledge the financial support of the General Studies Department, Arts and Science Program, Petroleum Institute, Abu Dhabi, United Arab Emirates (UAE).

Appendix: A Game that Teaches First Year Dynamics Topics

You are at the controls of a capsule orbiting a black hole at a radius of 800 km. The mass of the black hole is similar to that of the Earth and your time per orbit is around four minutes. Your capsule has a 'universal propulsion drive' that can accelerate it in any direction, but to save you from injury it is limited to 8 g.

In level one, your mission is to pursue and destroy another capsule in the same orbit. Rotational inertia is included in the model, so that directing your thrust is not simple. When you apply that thrust, the acceleration will take you in unexpected directions. When you fire a projectile, its path will be an orbital curve. Success will take you to level two.

Suddenly a missile appears 400 km away in the same orbit. Its not-so-simple guidance system accelerates it towards you at 10 g. If it can get within 1 km it will detonate and destroy you. How can you escape? It can accelerate faster than you can.

Perhaps you can 'dodge aside', as you might in a bullfight, letting the missile hurtle past you. Maybe you can coax the missile into greater and greater sweeps as it pursues you, but can you keep that up for long? How can you put it out of action?

Close to a black hole, 'gravity gradient' can become significant. Here it is only around a six-thousandth part of a *g* per metre, but the gradient is proportional to the inverse cube of the radius. If you can coax the long, thin missile to within 100 km the gradient will be close to one tenth of a *g* per metre, not too uncomfortable for you but enough to pull the missile into a vertical attitude despite its attempts to rotate.

But be careful! At 50 km the gradient is two thirds of a *g* per metre, rather uncomfortable. At 25 km the gradient of five *g* per metre will pull your capsule apart. Not far below that it would pull your head off.

To immerse the player will call for the help of games creation software, to convey the three-dimensional ambiance and give intuitive access to the controls. Help for the player can be on hand from 'virtual manuals' that can be grasped, opened and studied while the action is paused.

As the game progresses, the story-line can be 'embroidered' to match the escalating engineering challenges. Development will commence at www.essdyn.com/sim/blackhole.htm.

References

1. Maersk Drilling Simulator website: URL http://www.maerskdrilling.com/en/about-us/drilling-simulator.
2. H. Gardner. 1983. Frames of mind: The theory of multiple intelligences, Basic Books. ISBN 0133306143.
3. B. Tracy. 1986. 10 keys to personal power, YouTube URL https://www.youtube.com/watch?v=IFoUOafvFNY.
4. B. Tracy, Accelerated learning techniques (Audio tape), Nightingale-Conant; USA, 1996 (ISBN-10: 0671536842, ISBN-13: 978-0671536848)
5. Kiyosaki, R.T. 2013. *Why A students work for C students and B students work for the government*. USA: Plata. ISBN 978-1-61268-076-7.
6. Moon lander and other simple simulations. URL http://www.jollies.com.
7. Car Mechanic Simulator: URL https://www.youtube.com/watch?v=Lr0mcRpYhSc.
8. Scrap Mechanic website, URL http://www.scrapmechanic.com.
9. Unreal Engine 4, URL www.unrealengine.com.
10. Unity 3D website, URL https://unity3d.com.
11. GameGuru, AGK (AppGameKit) & Dark Basic website, URL www.thegamecreators.com.
12. Amy Jo Kim, Smart gamification: Seven core concepts for creating compelling experiences. URL https://www.youtube.com/watch?v=F4YP-hGZTuA.
13. Robert Torres—Transforming Education with Gamification, URL https://www.youtube.com/watch?v=h_okPIMyv9c.
14. Sebastian Deterding, Meaningful play: Getting gamification right, Google Tech Talk, January 24, 2011: URL accessed June 30, 2015. URL https://www.youtube.com/watch?v=7ZGCPap7GkY.
15. K.M. Kapp. 2012. The gamification of learning and instruction: Game-based methods and strategies for training and education, USA: Pfeiffer, ISBN-13: 978-1118096345, ISBN-10: 1118096347.
16. K.M. Kapp, L. Blair., and R. Mesch. 2013. The gamification of learning and instruction fieldbook: Ideas into practice, USA: Pfeiffer. ISBN-13: 978-1118674437, ISBN-10: 111867443X.
17. B. Burke. 2014. Gamify: How gamification motivates people to do extraordinary things, Bibliomotion, ISBN-10: 1937134857, ISBN-13: 978-1937134853.
18. L. Sheldon. 2011. The multiplayer classroom: Designing coursework as a game, Cengage Learning PTR, 1st ed. ISBN-10: 1435458443, ISBN-13: 978-1435458444.

Multi-channel Electro-tactile Feedback System for a Prosthetic Hand

Koren Ward and Daniel Pamungkas

Abstract Amputees with prosthetic hands often have difficulty holding and manipulating objects due to the absence of the sense of touch. This is especially true with tasks that require some degree of skill and tactile feedback to perform. To provide prosthetic hands with touch sensing and tactile feedback, researchers have been experimenting with various types of force and/or tactile sensors together with various methods for delivering the tactile information to the brain. Although some success has been achieved recently with force sensors and implanted electrodes, these systems are expensive, surgically invasive and can represent an infection risk where cables are attached to nerves through the skin. Also, non-invasive tactile feedback methods involving temperature, vibrations or electro-mechanical force feedbacks, can be somewhat awkward and ineffective due to being cumbersome or unable to deliver appropriate sensations. To address some of these issues we have developed an electro-tactile feedback system for prosthetic hands. Our system is comprised of force sensors that can be placed almost anywhere on a prosthetic hand, and TENS electrodes that can be placed on the wearer's arm. Our system is inexpensive, multi-channel and easily fitted to existing prosthetic hands. Experimental results are provided that show how this form of tactile feedback can enable a user to feel various objects touched or gripped with a robotic humanoid hand.

Keywords Prosthetic hand · Electro-tactile feedback

K. Ward (✉)
University of Wollongong, Wollongong, Australia
e-mail: koren@uow.edu.au

D. Pamungkas
Politeknik Negeri Batam, Batam, Indonesia
e-mail: daniel@polibatam.ac.id

J. Billingsley and P. Brett (eds.), *Mechatronics and Machine Vision in Practice 3*,
https://doi.org/10.1007/978-3-319-76947-9_13

1 Introduction

The human hand and associated sensory nerves have evolved over time to provide humans with considerable dexterity for performing object manipulation, as detailed in [1]. This level of dexterity requires precise control of hand and finger muscles with feedback from a complex array of sensory nerves within the hands [2]. By grasping objects and receiving tactile feedback, humans are able to perceive various properties of an object (e.g. shape, weight, texture) that can facilitate both the manipulation and classification of objects, as explained in [3]. This complexity poses challenging problems toward the development of prosthetic hands and the rehabilitation of amputees who have lost one or both hands.

It is estimated worldwide over 3 million amputees are living with the loss one or both hands. Most prosthetic hands available today provide limited control of artificial fingers and no somatic sensory feedback. Consequently, amputees have to rely mainly on visual feedback and careful control when using a prosthetic hand to pick up or manipulate objects. This can make a prosthetic hand feel unnatural, awkward and distracting which can sometimes result in the amputee refusing to use the prosthetic hand, as explained in [4].

To address this problem, we have been experimenting with the development of an electro-tactile feedback system for prosthetic hands. Our proposed system is comprised of resistive film force sensors, that can be placed almost anywhere on the prosthetic hand, and TENS electrodes that can be placed on nearby skin (e.g. upper or lower arm).

Our system is inexpensive, easily fitted to existing robotic or prosthetic hands and capable of delivering multiple channels of stimulus from the fingers and palm of the prosthetic hand. In addition, each channel is capable of producing a variety of sensations by modulating both the frequency and intensity of the signal. We show how this information can assist when gripping and manipulating objects with a robotic or prosthetic hand.

This paper is organized as follows: Sect. 2 provides a brief background review of related work. Section 3 presents the implementation details of our proposed electro-tactile feedback system. Section 4 presents some preliminary experimental results that demonstrate how our tactile feedback system can provide useful tactile feedback when gripping and manipulating objects. Section 5 provides some concluding remarks and a brief overview of future work.

2 Background

Researchers have been investigating various methods for providing force feedback from prosthetic hands. Most feedback systems involve the use of various types of force sensors embedded in a prosthetic hand combined with various methods for delivering the tactile information to the brain, see [5] for a comprehensive review.

Force sensing is generally achieved by using pressure sensitive resistive films, back EMF from finger actuators, or hydraulic fluid within rubber membrane fingertips combined with pressure transducers, see [6].

Interfacing force sensors to the amputee is achieved either by surgically implanting electrodes that stimulate sensory nerves (see [7] and [8]), or through non-invasive feedback methods involving the use of vibrators [9], air pressure [10] or spatially-mapped tactile displays involving pressure, vibration, shear force or temperature, see [11, 12].

Although some success has been achieved recently with force sensors and implanted electrodes, these systems are expensive, surgically invasive and can pose an infection risk where the cables emerge from under the skin. Non-invasive tactile feedback methods involving temperature, vibrations or electro-mechanical force feedbacks have less bandwidth, but have been shown to improve both the use and the sense of ownership of the prosthetic hand by making it feel less like a tool and more like a natural part of the amputee's body. See [13–15].

Saunders and Vijayakumar investigated the utilization of vibro-tactile feedback for informing a user of the forces applied by a robotic hand when gripping an object [13]. This involved fitting eight motor driven vibrators to the user's arm between the wrist and elbow. A light gripper force activates the vibrators nearest to the wrist, whereas, a stronger gripper force activates the vibrators nearer to the elbow. They reported that subjects could grip, lift up and put down objects more effectively with this feedback system.

Similar results were achieved in [14] with the development of a higher bandwidth vibrating tactor. This was constructed from three DC vibration motors and was able to generate different sensations by using a combination of different frequencies and amplitudes from the vibration motors.

Kim and Colgate also developed a compound 2-DoF tactor which was able to deliver more information from a robotic hand like low and high touch pressure [15]. Due to the size, and to make it more effective, they chose to mount their tactor on the skin of the user's chest.

The main criticism of vibro-tactile feedback systems is their low bandwidth and limitations in reproducing natural touch sensations. To address this issue some researchers have devised prosthetic hand feedback systems that apply forces to the skin rather than vibration. For example, Antfolk et al. [16] developed a mechanical force feedback device for delivering force sensations from a prosthetic hand to the wearer. Their proposed system used five servo motors to deliver force information from five pressure sensors mounted on the fingers of the prosthetic hand. A button is fitted to each servo motor to deliver applied pressure to the user's skin on the user's forearm. Similarly, Ajoudani et al. [17] used a combination of DC motors and pulleys to deliver grip-force information from a prosthetic hand to the user by applying pressure to the upper arm.

Even though these force feedback systems can enable the user to distinguish finger pressure or the grip force of a prosthetic hand they are limited in bandwidth

and somewhat cumbersome and therefore can restrict movement and cause the prosthetic hand to feel unnatural to the user. To address these issues we have been developing an electro-tactile feedback system for prosthetic hands.

Previously, electro-tactile stimulation systems have been devised for providing substitute visual perception to the blind, e.g. [18, 19]. Furthermore, Kim et al. [20] and Perović et al. [21] experimented with multi-electrode electro-tactile feedback to determine its suitability for haptic perception. Their results show that electro-tactile feedback has potential for delivering haptic sensations from devices such as prosthetic hands but the information can be difficult to resolve when too many closely spaced electrodes are used.

To address the low bandwidth of vibro-tactile feedback systems and to improve on previous work with electro-tactile feedback we have developed a versatile configurable multi-channel electro-tactile feedback system. Our proposed system is comprised of adhesive force sensors that can be placed anywhere on a prosthetic hand, and Transcutaneous Electrical Nerve Stimulation (TENS) electrodes that can be placed almost anywhere on the user's skin. Our system is inexpensive and can be easily fitted to existing prosthetic hands or built into new prosthetic hands. Experimental results are provided that show how this form of tactile feedback can enable a user to feel various objects touched or gripped with a robotic artificial hand.

In the following section we provide the implementation details of our proposed electro-tactile feedback system followed with our preliminary experimental results.

3 Prosthetic Hand and Electro-tactile Feedback

3.1 Overview

Our electro-tactile feedback system is comprised of force sensors, that are placed on the fingers and palm of a prosthetic hand, interface circuits for processing the sensor data and TENS electrodes that are placed on nearby skin. To test our electro-tactile feedback system we fitted the force sensors to a humanoid robotic hand that was interfaced to a data glove. We also implemented a control panel in software on a PC to monitor the sensor data and deliver appropriate pulses to the TENS electrodes fitted to the user's right arm. The robotic hand was manually positioned with the user's right hand and controlled with the user's left hand via the data glove. This arrangement enabled the user to both control the hand and experience feedback from the electro-tactile feedback system.

3.2 Robotic Hand and Tactile Sensors

The robotic hand was comprised of an EH1 Milano series anthropomorphic hand from Prensilia, as shown in Fig. 1. The EH1 robotic hand is approximately the same size and configuration as an adult male forearm and hand and has six motors and tendons for manipulating the fingers and thumb. Five motors are utilized for bending the five fingers and one is used for abduction/adduction of the thumb making the EH1 robotic hand capable of manipulating and gripping a wide variety of objects.

To provide the EH1 hand with tactile force sensing, we fitted 16 polymer film force sensors to the fingers and palm of the EH1 hand, as shown in Figs. 2 and 3. Each force sensors was custom cut from a FlexiForce FSR408 sensor strip supplied by Interlink Electronics, as shown in Fig. 2a. To enable wires to be attached to the force sensors, thin copper conductors were inserted and bonded to each sensor, as shown in Fig. 2b. Figure 2c shows a finger tip with force sensors fitted. Each force sensor has approximately infinite resistance when no force is applied, 50 K ohms when light pressure is applied and less than 5 K when pressed firmly.

Each finger on the EH1 hand was fitted with three force sensors, as shown in Fig. 3. The sensors were positioned on the distal, middle and proximal phalanges of the fingers, as shown in Fig. 3a. An additional larger force sensor was fitted to the palm of the hand, see Fig. 3a. The force sensors were also covered with a thin layer of neoprene to improve the hand's grip and to even out pressure on the sensor surfaces when objects are held, as shown in Fig. 3b.

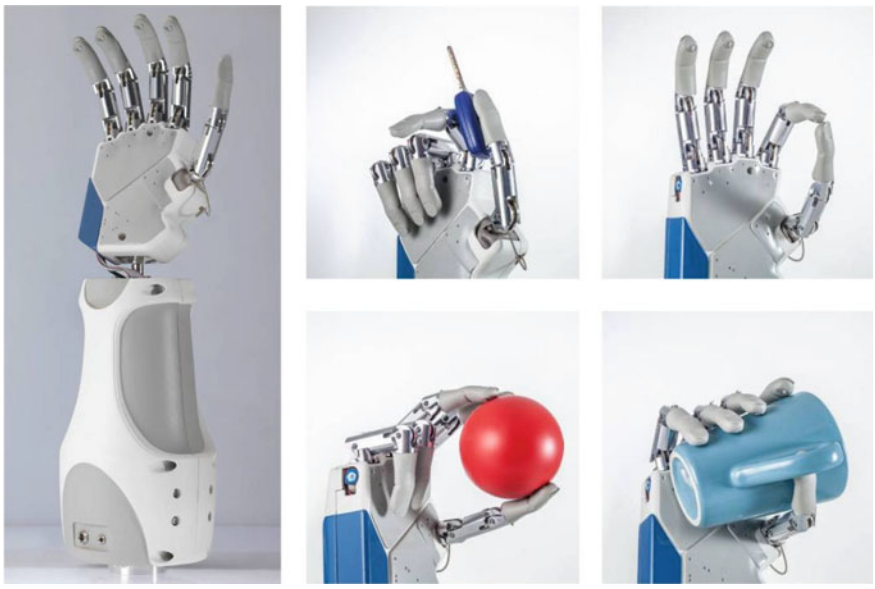


Fig. 1 EH1 Milano robotic hand

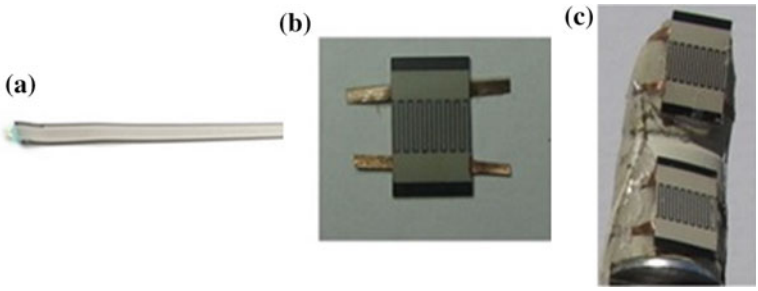


Fig. 2 **a** FSR480 force sensor strip. **b** Force sensor cut from FSR480 strip. **c** Cut force sensors mounted on a robotic finger

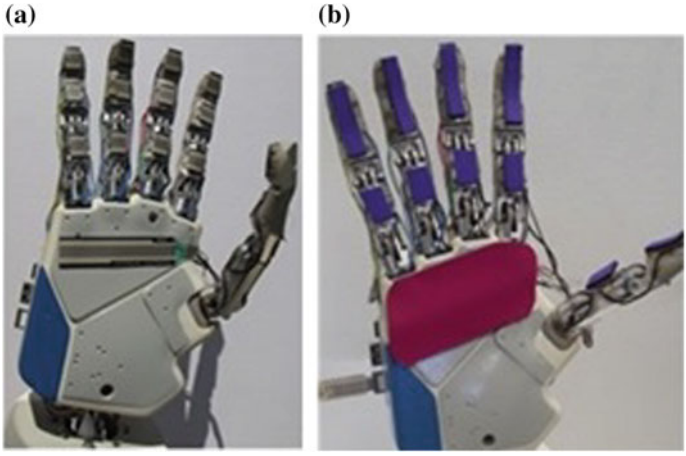


Fig. 3 EH1 robotic hand showing: **a** fitted force sensors and **b** neoprene covering

The force sensors are connected to the analog inputs of a microprocessor control board via voltage divider circuits. The control board samples the analog sensor data 20 times per second and sends it to a PC for further processing and then onto the user as electro-tactile feedback, as explained in the following section.

3.3 Electro-tactile Feedback System

To deliver the tactile information from the computer to the user, a custom-built wireless TENS electro-tactile feedback system was devised, as shown in Fig. 4. This feedback system is capable of providing six channels of electrical stimulus to the user's skin with controllable frequency and intensity.

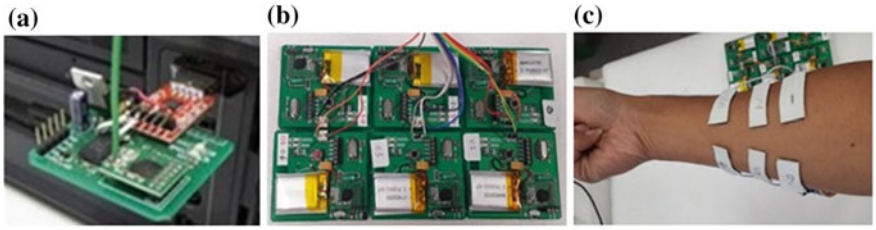


Fig. 4 TENS feedback system: **a** transmitter **b** receivers **c** arm with electrodes fitted

The electro-tactile feedback system consists of a USB transmitter, shown in Fig. 4a and six TENS receiver units, shown in Fig. 4b. The transmitter unit transmits data wirelessly from the computer to the receiver units which convert it into electrical pulses that are delivered to electrodes adhered to the user's skin, as shown in Fig. 4c. A common ground electrode is also adhered to the back of the user's arm. Although the electrodes could be placed almost anywhere on the user's skin, we chose this arrangement to try to approximate sensory hand tactile stimulation.

The mapping between the EH1 robotic hand sensors and the electrodes adhered to the user's lower arm is shown in Fig. 5. This arrangement allows the user to receive six separate channels of stimulus via sensory nerves in the skin (five for each finger and one for the palm). As the stimulus is relatively mild, painless and adjustable for user comfort, it did not result in any significant muscle contractions during our experiments.

Since each EH1 robotic finger has three force sensors delivering tactile information to one TENS electrode, we mapped three separate stimulation frequencies to each sensor. Namely: distal phalange 100 Hz, middle phalange 60 Hz, proximal phalange 30 Hz and the palm 20 Hz. To avoid confusion, rather than mix the

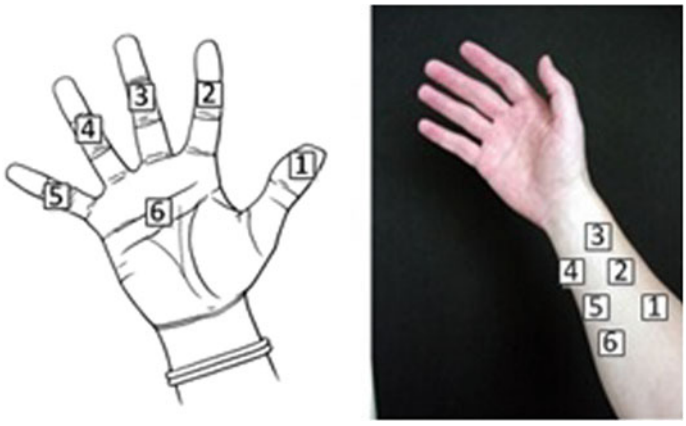


Fig. 5 Mapping between sensor regions and electrodes

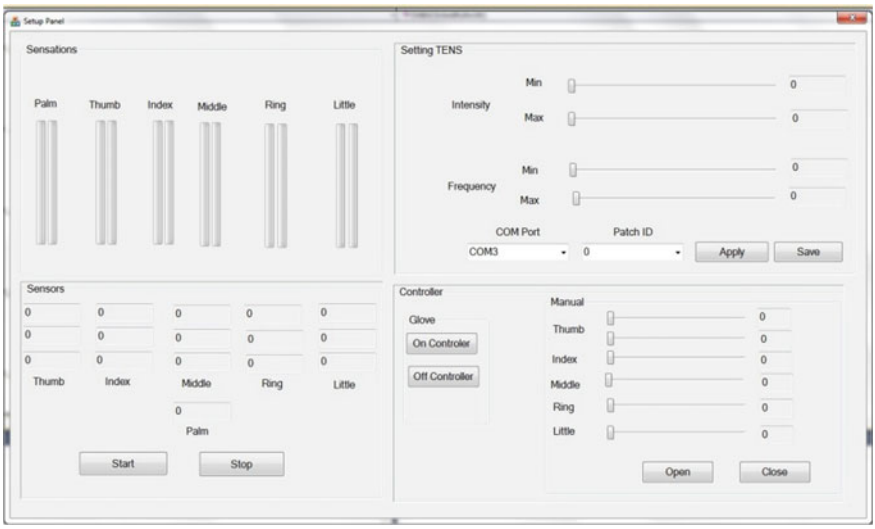


Fig. 6 User Interface

frequencies coming from separate activated sensors on each finger, we chose to deliver only the frequency from the sensor with the most applied force.

The intensity of the pulses delivered to each TENS electrode depends on the amount of force applied to the associated sensors. Again, only the finger sensor with maximum applied force is gated through to the TENS electrode. For simplicity, we divided the intensity into four levels to represent zero, light, medium and high forces.

To adjust the TENS settings and monitor the sensor and feedback data, a graphical user interface was implemented, as shown in Fig. 6. The user interface has controls for setting the maximum stimulus delivered to each finger and palm as well as indicators for monitoring the raw sensor data and the pulse intensity and frequency sent to the TENS electrodes.

4 Experimental Results

To demonstrate the potential of our electro-tactile feedback system, we fitted a handle to our EH1 robotic hand so that it could be easily held and positioned with the user's right hand. On the left hand of the user, we fitted a P5 data glove linked to the EH1 hand, as shown in Fig. 7. The electro-tactile feedback electrodes were fitted to the user's right hand, as described in Sect. 3.3. With this arrangement the user could position the EH1 hand with his/her right hand, control the fingers of the EH1 hand with his/her left hand, and experience tactile feedback from the EH1 hand via the TENS electrodes adhered to his/her right arm. Both the grip forces

Fig. 7 P5 data glove for controlling the robotic hand

from the EH1 hand's sensors and the electro-tactile feedback delivered to the user via the TENS electrodes could also be observed on the control panel, as shown in Fig. 8.

We conducted two experiments with five users to see if the electro-tactile feedback made it easier for a user to handle objects with the robotic hand. Prior to the commencement of each trial each user was asked to adjust the maximum level of intensity of the electro-tactile feedback signals to suit their preference. The first experiment involved picking up and putting down various objects with different grips. The second experiment involved gripping, holding and manipulating objects that have the similar size and shape but different weights.

For the first experiment, different objects, including a mobile phone, tennis ball and jam jar were placed on the table. The supervisor then demonstrated, with his own hand, how he wanted each object to be picked up and put down with the robotic hand.

Figure 8 shows examples of a user grasping a tennis ball with (a) five fingers, (b) three fingers and (c) two fingers. The green vertical bars on the user interface show the intensity and frequency of the electro-tactile feedback from the palm, thumb, pointer, middle, ring and little fingers respectively.

After 20 min picking up and putting down objects with the robotic hand, both with electro-tactile feedback turned on and off, each user was asked to comment on any effect the electro-tactile feedback had on performing these tasks. All users

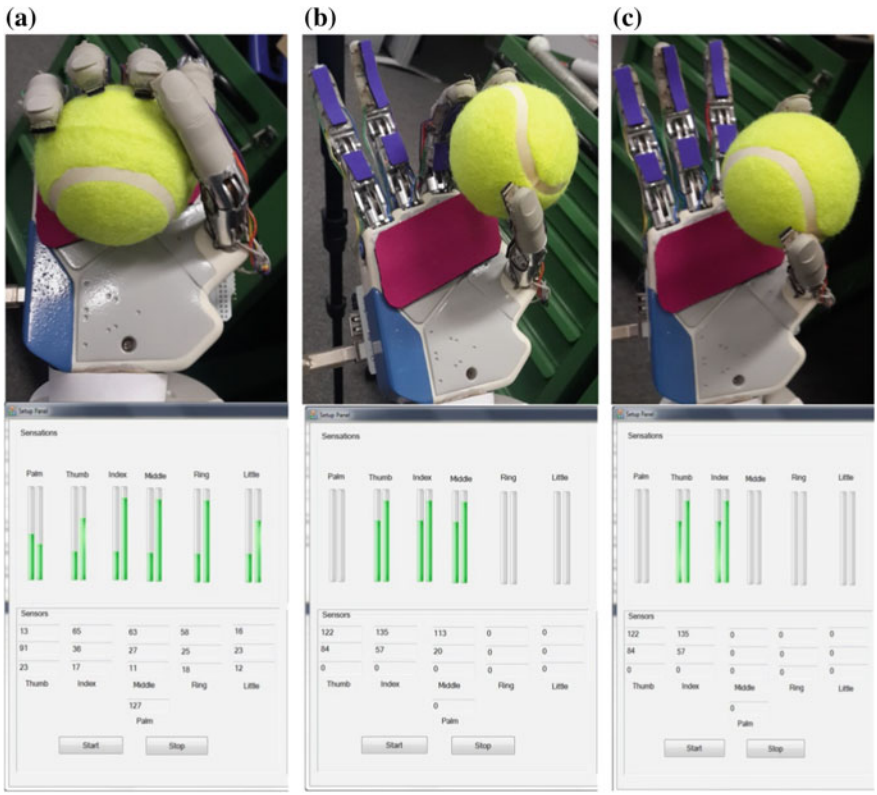


Fig. 8 Robotic hand gripping tennis ball with: **a** five fingers **b** three fingers **c** two fingers

reported that the electro-tactile feedback improved their ability to pick up, hold and put down objects. The general opinion was that the electro-tactile feedback made them more aware of the object being held by the robotic hand with less need to use their eyes to see what the robotic hand was doing.

For the second experiment, objects with similar size and different weights were placed on the table (e.g. metal and plastic pipes, full and empty bottles) and the users were asked to alternate between picking up lighter and heavier objects. The users were also asked to apply only sufficient force to prevent the objects from slipping from the robotic hand's fingers and to manipulate the objects around within the robotic hand by moving the robotic hand's fingers.

We found, without electro-tactile feedback, most users frequently dropped the heaver objects and often applied more force than necessary to pick and manipulate the lighter objects. When the electro-tactile feedback was turned on, all users reported that they were able to quickly learn how much force to apply to pick up and manipulate objects without slippage. For example, Fig. 9a, b shows a golf ball and a lighter plastic ball being manipulated between the thumb and pointer fingers

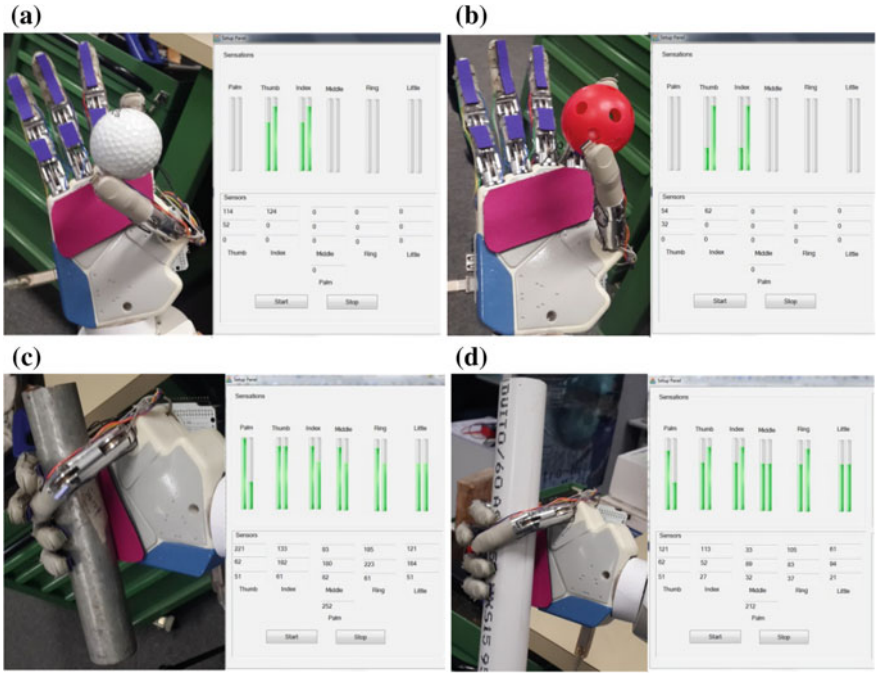


Fig. 9 Robotic hand gripping: **a** A golf ball. **b** Aplastic ball. **c** Asteel pipe. **d** A PVC pipe

and the different forces applied by the fingers. Similarly, Fig. 9b, c shows a heavy metal pipe and a lighter plastic pipe being held by the robotic hand and the appropriate finger forces applied to maintain grip of these objects.

5 Conclusion

The development of a prosthetic hand capable of the same tactile sensations as a natural hand remains a major challenge facing prosthetic technologies. In this paper we present some preliminary results of our prototype electro-tactile feedback system for robotic and prosthetic hands. Our proposed electro-tactile feedback system is comprised of force sensors that can be placed almost anywhere on a prosthetic hand and TENS electrodes that can be placed on the user's arm or elsewhere. Our system has benefits in that it is inexpensive, multi-channel and can be fitted to existing robotic or prosthetic hands with relative ease. Although more extensive experimentation is needed to fully evaluate our system, our preliminary experimental results show that this form of tactile feedback can assist a user of an anthropomorphic robotic hand to become more aware of objects held and manipulated with the robotic hand. For future work we intent to conduct further

experiments to see if this form of sensory feedback can enable the user to become more spatially aware of the robotic hand and its interactions with objects. We also intend testing our system on amputees with prosthetic limbs.

References

1. MacKenzie, C., and T. Iberall. 2010. *The grasping hand.* Amsterdam, The Netherlands: Elsevier.
2. Johansson, R.S., and J.R. Flanagan. 2009. Coding and use of tactile signals from the fingertips in object manipulation tasks. *Nature Reviews Neuroscience* 10: 345–359.
3. Klatzky, R.L., and S.J. Lederman. 1995. Identifying objects from a haptic glance. *Percept Psycophys* 57: 1111–1123.
4. Biddiss, E., D. Beaton, and T. Chau. 2007. Consumer design priorities for upper limb prosthetics. *Disability Rehabilitation: Assistive Technology* 2: 346–357.
5. Jiminez, M.C., and J.A. Fishel. 2014. Evaluation of force, vibration and thermal tactile feedback in prosthetic limbs. In *IEEE international conference on haptic interfaces for virtual environment and teleoperator systems (Haptics)*, 437–441.
6. Fishel, J.A., and G.E. Loeb. (2012). Sensing tactile microvibrations with the BioTac—Comparison with human sensitivity. In *Proceedings IEEE/RAS-EMBS international conference on biomedical robotics and biomechatronics*, 1122–1127.
7. Tan, D.W., M.A. Schiefer, M.W. Keith, J.R. Anderson, J. Tyler, and D.J. Tyler. 2014. A neural interface provides long-term stable natural touch perception. *Science Translational Medicine* 6 (257): 257–268.
8. Raspopovic, S., et al. 2014. Restoring natural sensory feedback in real-time bidirectional hand prostheses. *Science Translational Medicine* 6, no. 222: 222ra19.
9. Pylatiuk, C., and A. Kargov, Schulz, S. 2006. Design and evaluation of a low-cost force feedback system for myoelectric prosthetic hands. *JPO: Journal of Prosthetics and Orthotics* 18, no. 2: 57–61.
10. Antfolk, C., et al. 2012. Sensory feedback from a prosthetic hand based on air mediated pressure from the hand to the forearm skin. *Journal of Rehabilitation Medicine* 44 (8): 702–707.
11. Kim, K., and J.E. Colgate, Peshkin, M.A. (2007). A pilot study of a thermal display using a miniature tactor for upper extremity prosthesis. In *Proceedings frontiers in the convergence of bioscience and information technologies*, 531–536.
12. Kim, K., et al. 2010. On the design of miniature haptic devices for upper extremity prosthetics. *IEEE Transactions on Mechatronics* 15 (1): 27–39.
13. Saunders, I., and S. Vijayakumar. 2011. The role of feed-forward and feedback processes for closed-loop prosthesis control. *Journal of NeuroEngineering and Rehabilitation* 8: 60.
14. M. D'Alonzo, C. Cipriani, Carozza M.C. (2011). Vibrotactile sensory substitution in multifingered hand prostheses: Evaluation studies. In *Proceeding ieee international conference on rehabilitation robotics rehab week*, Zurich, Switzerland.
15. Kim, K., and J.E. Colgate. (2012). Haptic feedback enhances grip force control of sEMG-controlled prosthetic hands in targeted reinnervation amputees. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 20 (6): 798–805.
16. Antfolk, C., et al. 2013. Transfer of tactile input from an artificial hand to the forearm: experiments in amputees and able-bodied volunteers. *Disability and Rehabilitation: Assistive Technology* 8 (3): 249–254.
17. Ajoudani, A. et.al. (2014). Exploring Teleimpedance and Tactile Feedback for Intuitive Control of the Pisa/IIT SoftHand. *IEEE Transaction On Haptics* 7 (2).

18. Kaczmarek, K.A., J.G. Webster, P. Bach-y-Rita, and W.J. Tompkins. 1991. Electrotactile and vibrotactile displays for sensory substitution systems. *IEEE Transactions on Biomedical Engineering* 38 (1): 1–16.
19. Meers, S., and K. Ward. 2004. A vision system for providing 3D perception of the environment via transcutaneous electro-neural stimulation. In *Proceedings of the 8th IEEE international conference on information visualisation*, London, 546–552.
20. Kim, G., Asakura, Y., and R. Okuno, Akazawa, K. 2005. Tactile substitution system for transmitting a few words to a prosthetic hand user. In *Proceeding of the 2005 IEEE engineering in medicine and biology 27th annual conference*, 6908–6911, Shanghai.
21. Perović, M., M. Stevanovic, T. Jevtic, M. Strbac, G. Bijelic, and C. Vucetic. 2013. Electrical stimulation of the forearm: a method for transmitting sensory signals from the artificial hand to the brain. *Journal of Automatic Control, University of Belgrade* 21: 13–18.

Temporal Difference (TD) Based Critic-Actor Adaptive Control for a Fine Hand Motion Rehabilitation Robot

Xianwei Huang, Fazel Naghdy, Haiping Du, Golshah Naghdy and Catherine Todd

Abstract Robot assisted post-stroke rehabilitation training is an effective approach in delivering the highly intensive repetitive training, aiming to retrain the neural pathways in the brain thus to restore and improve the affected mobility skills. The adaptive control of robotic devices, especially assist-as-needed control providing exact assistive force intensity along the intended motion trajectory for fine motion, can be a complex but effective method. A temporal difference based critic-actor reinforcement learning control method is explored in this study. The effectiveness of the method is verified through Matlab simulation and implemented on a hand rehabilitation robotic device. Results suggest that the control system can fulfil the control task with high performance and reliability, thus holding the promise of improving the fine hand motion rehabilitation training efficiency.

Keywords Stroke recovery · Rehabilitation robot · Adaptive control
Reinforcement learning · Neural network

1 Introduction

The impairment of fine hand motor skills is a common phenomenon among stroke and spinal cord injury (SCI) patients, which may severely lower their quality of life by affecting their capability of performing delicate activities of daily life (ADL) such as eating, drinking and handwriting. Moreover, post-stroke restoration and enhancement of impaired fine hand motor skills is a complex process due to fine motion requires exact intensity of finger strength and the coordinated movements between them. Recent neural science research suggests that robot assisted therapy can be an effective approach in providing highly repetitive training to achieve better rehabilitation results by triggering neuroplasticity in the brain.

X. Huang (✉) · F. Naghdy · H. Du · G. Naghdy · C. Todd
University of Wollongong, Wollongong, NSW 2522, Australia
e-mail: xh962@uowmail.edu.au

J. Billingsley and P. Brett (eds.), *Mechatronics and Machine Vision in Practice 3*,
https://doi.org/10.1007/978-3-319-76947-9_14

In this study, a 5 degree of freedom (DOF) hand rehabilitation robotic device named Amadeo (Tyromotion GmbH, Graz, Austria) [1, 2] is used as the experimental platform of this study. Aamdeo can provide position-based both passive and active assistive training modes that emphasize the flexion and extension of each finger. During training, the device is attached to the finger tips using a small magnetic disc and cohesive tape for connection with the robot. The moving finger slides can transfer bend or stretch movements to all the fingers at once or independently as shown in Fig. 1.

The standard control algorithm of Amadeo is based on position control with constant assistive force intensity, ignoring the actual need of the patient. This is the major limitation of the Amadeo's standard training protocol because the constant assistive force intensity can be excessive at some point during the movement of the finger. Thus in clinical practice, it makes the training task less challenging and may compromise the training effectiveness. Dynamic assistive force intensity, or adaptive assist-as-needed (AAN) control seems a desirable solution in order to achieve better training results with less training sessions required and to reduce the and associated costs of rehabilitation training.

Regarding these limitations, this study proposes a Temporal Difference (TD) based critic-actor structured reinforcement learning control method that provides AAN extension/flexion assistive force during training. The remaining part of the paper is structured as follows. The control strategy, TD critic-actor neural network (NN) theoretical base is explained in Sect. 2. Details of reinforcement learning neural network (RLNN) controller online learning are given in Sect. 3. Section 4 covers validation method of experimental work. Finally, conclusions are made in Sect. 5.

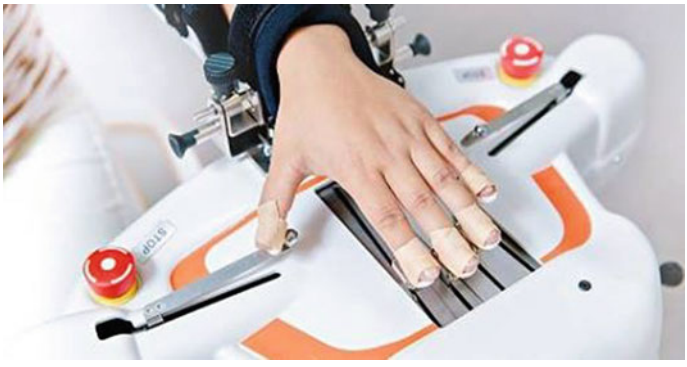


Fig. 1 Hand training with Amadeo

2 Control Strategy

The control strategy consists of the following steps: (1) Amadeo device stays in free mode and continuously checks if the position of any of the sliding actuators are changed as the result of the force applied by the subject. As soon as a position change is detected, the algorithm tracks the movement of the corresponding finger and predicts the next position of the finger and the intensity of the force applied by the finger. (2) If there is no change of position, the detection of force on any of the actuators means that the subject is trying to move the slides but the force is not adequate. The algorithm generates the amount of force required to move the slides. If no force detected, this implies that the subject is not able to apply any force on the slide. Hence, the algorithm generates just enough force to compensate the friction force to actuate the slides. (3) If the force intensity applied by the subject is adequate to move the slides, the subject operates independent from the control algorithm.

The proposed critic-actor based RL control algorithm can be seen in Fig. 2. The critic NN estimates the one-step ahead position of the actuator and force applied by the subject, the actor NN updates and implements the control policy. According to the agent-environment based RL structure [3], in this case, input r(t) is the desired position and velocity for an actuator that connected to the finger. Control signal u(t) or action (a) is the position and velocity to the actuator in the form that the machine can execute. Output y(t) or state (s) is the position of an actuator and the force applied to it. And the reward (r) is a smaller error between the predicted and actual output than that of the previous time step.

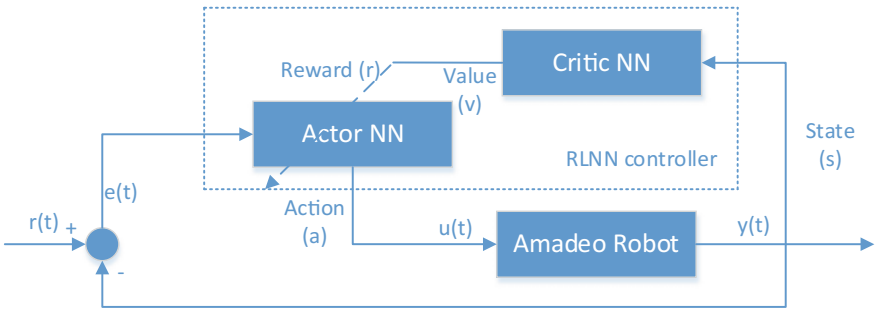


Fig. 2 Critic-actor based RL control scheme

2.1 Temporal-Difference (TD) Learning

The prediction of the output is estimated through the model-free TD learning method, which is a prediction method used for solving the reinforcement learning problem. The TD-error at time t is defined as:

$$\delta_t = r_{t+1} + \gamma V(s_{t+1}) - V(s_t) \tag{1}$$

where γ is the discounting factor, δ_t is measured between the value at state s_t, and the value at the subsequent state s_{t+1}, plus any reward r_{t+1} accumulated along the way. A TD-error δ_t between the predicted value and the subsequent value is generated at each time step, the value function for every state V(s) is updated in proportion to both the TD-error and the eligibility of the state [4].

TD (λ) combines basic TD learning with eligibility traces to further the speed of reinforcement learning. TD (λ) algorithm learns by reducing the error between estimates made by the agent at different times. The common approach in TD (λ) algorithm is the n-step reward estimation:

$$R_t^{(n)} = r_{t+1} + \gamma r_{t+2} + \gamma^2 r_{t+3} + \cdots + \gamma^{n-1} r_{t+n} + \gamma^t V_t(s_{t+n}) \tag{2}$$

The estimate averages all the n-step estimates, each of which is weighted proportional to λ^{n-1}, where $0 \leq \lambda \leq 1$ is the trace decay parameter.

The binary reinforcement signal can be either "0" or "−1" corresponding to success or failure respectively. In the learning process, there are two situations where the learning agent-Amadeo should be given a negative reward. One is the predicted position changing direction is contradictory with the detected moving trajectory of the subject, the other situation is the error between predicted and detected force value is larger than 1.0 N. These two factors imply the force tracking and prediction ability of the RL controller. When any of these happens, Amadeo will receive a (−1) reward, representing failure.

2.2 Actor-Critic Algorithms

The actor-critic learning algorithm is a part of the TD learning method. During the learning process, the critic generates an effective reinforcement signal and sent it to the actor, which learns and executes the behavioural output. The policy evaluation step is performed by observing the results of applying current actions and the state value function of each state is estimated from the reinforcement feedback [4]. The TD based actor-critic learning method works by selecting an action from the current policy. Critic-actor algorithm makes a TD-update to the critic's value function V that is defined by:

$$V(s_t) \leftarrow V(s_t) + \alpha_c \delta_t, \tag{3}$$

where α_c is the learning rate for the critic NN, and δ_t is the TD-error. Similarly, α_a stands for the learning rate of the actor NN, the policy values of the actor NN are updated by:

$$P(s_t, a_t) \leftarrow P(s_t, a_t) + \alpha_a \delta_t \tag{4}$$

2.3 Actor–Critic NN Design

In actor and critic NN, the inputs of the critic NN are position and force of the control feedback loop and the outputs of the RLNN controller are position and velocity signals that drive the connected slides of each finger. There is only one hidden layer with 10 neurons in the hidden layer thus to reduce the computing complexity and ensure the real-time property of the control system. In the RLNN controller, NN weights are used as function approximation to the value function. The back propagation procedure is used to make the TD update according to the following equation for every output unit:

$$\Delta w_t = a(Y_{t+1} - Y_t) \sum_{k=1}^{t} \lambda^{t-k} \nabla_w Y_k \tag{5}$$

where w_t is the vector of neural network weights being tuned, Y_t is the prediction for the output at time step t, $\nabla_w Y_k$ is a set of partial derivatives for each component of the weights w, α is the standard learning rate, and λ is the factor controlling how much future estimates affect the current update [5].

2.3.1 The Critic NN

The critic NN generates a scalar evaluation signal to tune the actor NN. The critic itself consists of a NN which approximates an evaluation function based on performance measurement. The tracking error between the estimated and actual output value $e(t)$ can be considered as an instantaneous utility function of the plant performance. The reinforcement signal R is defined as:

$$R = W_C^T \sigma(V_C^T e_n) + \psi \tag{6}$$

where σ is the nonlinear activation function, W_C is the number of neurons in hidden layer of the critic NN, and e_n is the input to the critic NN, ψ is the auxiliary term. The critic NN weight update rule is a gradient-based adaption given by:

$$w_c(t+1) = w_c(t) + \Delta w_c(t) \tag{7}$$

$$\Delta w_c(t) = a_c(t)\left[-\frac{\partial E_c(t)}{\partial w_c(t)}\right] \tag{8}$$

where a_c is the learning rate of the critic NN at time t, which usually decreases with time to a small value, and w_c is the weight vector of the critic NN. The output of the critic NN, the value function approximates the discounted total reward. Specifically, it approximates at time given by:

$$R(t) = r(t+1) + a\sum_{t+1}^{k} r \tag{9}$$

where $r(t+1)$ is the reward at time $t + 1$, $a(0 < a < 1)$ is the discount factor.

2.3.2 The Actor NN

The principle in adapting the actor NN is to indirectly back propagate the error between the desired ultimate objective, denoted by E_A and the approximate function from the critic NN. In the actor network, the state measurements are used as inputs to create a control as the output of the network. The weight estimation error for the actor NN is defined as:

$$\hat{E}_A(t) = W_C - \hat{W}_A(t) \tag{10}$$

The critic NN and the actor NN are tuned sequentially, with the weights of the other NN being kept constant. The weights in the actor NN are updated similarly to the ones in the critic NN.

$$w_a(t+1) = w_a(t) + \Delta w_a(t) \tag{11}$$

$$\Delta w_a(t) = a_a(t)\left[-\frac{\partial E_a(t)}{\partial w_a(t)}\right] \tag{12}$$

where $a_a > 0$ is the learning rate of the actor NN at time t, which usually decreases with time to a small value, w_a is the weight vector of the actor NN.

3 Online Learning

The RLNN controller needs to be trained to learn the dynamics of the control system, thus to provide the AAN adaptive control through interaction with the patient. The TD (λ) algorithm is implemented with the critic NN for policy evaluation and the actor NN for policy implementation. The critic NN and the actor NN are tuned online using the output of Amadeo, the weights of one NN are held constant while the weights of the other are tuned until convergence. This procedure is repeated until both NN have converged.

3.1 Variables Identification

The online learning and simulation are carried out using MATLAB/SIMULINK. The neural network is trained using back propagation, we set time step t = 0.01 s, after several simulation experiments for each variant to determine the optimal value of the RL parameters.

The learning rete α of the critic and actor NN are set to be the same value. We train for 2000 time steps to smooth out the random element. The ε-greedy decision mechanism allows us to see the underlying relationship that α and λ have. The step size parameter is held constant, for each value of λ, we use α with increments of 0.01. Each (λ, α) pair is regarded as an independent learning algorithm. The average error from 10 episodes between predictions is shown in Fig. 3.

Figure 3 shows the relationship of average error and the learning rete α value, suggesting that the larger the λ value, the smaller the α value, yields a better performance. Indeed, for $\lambda \geq 0.9$, The TD learning process turns to be unstable when $\alpha > 0.1$. The best performance with the lowest average error is identified when $\alpha = 0.45$, $\lambda = 0.6$. According to the simulation result, the rest of the variables and their initial values are shown in Table 1.

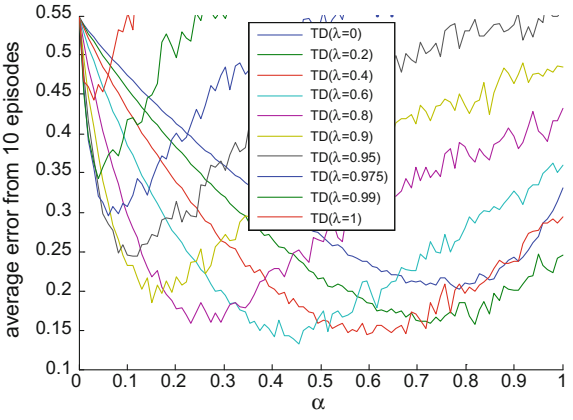


Fig. 3 Online TD (λ) learning to decide the optimal value of λ and α

Table 1 Variables and their initial value in RLNN controller

Parameter	Symbol	Value	RL component
Critic network learning rate	a_c	0.45	NN
Actor network learning rate	a_a	0.45	NN
Number of hidden neurons	n	10	NN
Initial weight	w_i	0.1	NN
Decay rate	λ	0.6	TD (λ)
Discount factor	γ	0.95	TD (λ)
ε-greedy	ε	0.05	TD (λ)

3.2 Tuning and Convergence

The learning procedure was executed as follows: we start by creating a sequence of finger slides positions beginning with the starting position and ending in the final position when the training task is over. For each of these positions we use the back propagation procedure of the neural network to compute the TD (λ) update, as described in Sect. 2.2. The learning process is repeated until we can no longer improve the RLNN controller. Then we evaluate the error function on the validation set and choose the network with the best performance. The network is then further trained using early stopping. The initial cross validation ensures that the network is likely to yield good convergence.

The initial weights of both NN are set to be 0.1 in order to allow spontaneous, unlearned actions when policy arises. Weights of the critic and actor NN are increased for neurons if stimulus and action traces coincide with an increased effective reinforcement signal, whereas weights are decreased if traces coincide with an effective reinforcement signal decreased below baseline. The changes of weights for both NN are shown in Figs. 4 and 5 respectively.

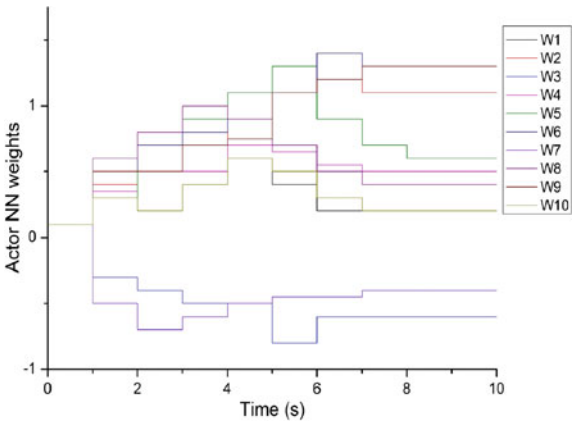


Fig. 4 Actor NN weights

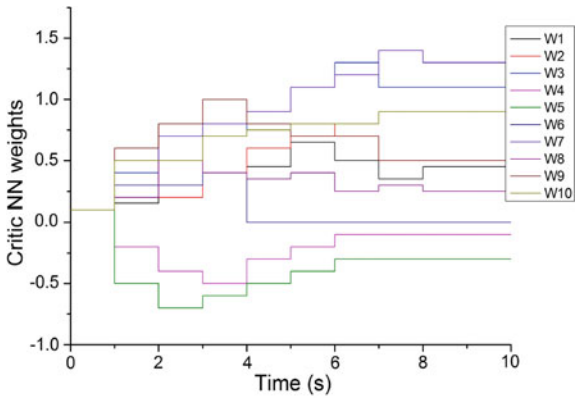


Fig. 5 Critic NN weights

In the learning process, if the performance of the RL algorithm is improving, then Amadeo will get less negative reward, thus the cumulative reward value increases. The reward against learning time is shown in Fig. 6.

It can be seen that with the process of online learning, the RLNN controller gradually adapts the control system, thus less negative reward value from the critic NN is received. Once the learning process is finished, the resultant policies with its corresponding parameters are updated. Figure 7 shows the critic NN control policy update difference and reward difference along with the iteration numbers.

As it can be seen from Fig. 7, the x-axis shows the training epochs/iteration number and y-axis shows the reward difference and policy update difference. At the end of 20,000 iterations, the reward difference becomes less than one and the policy update difference becomes small enough after 20,000 iterations. This indicates that the updated control policy is gradually fitting the critic and actor NN. The learning process is repeated until it no longer improves the RLNN controller.

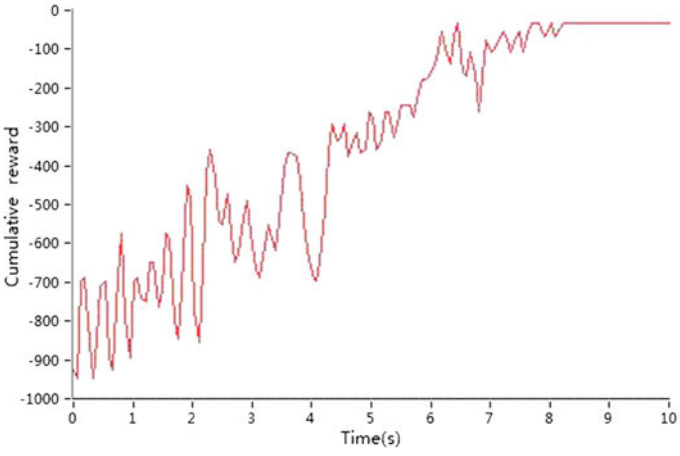


Fig. 6 Cumulative reward of the critic NN

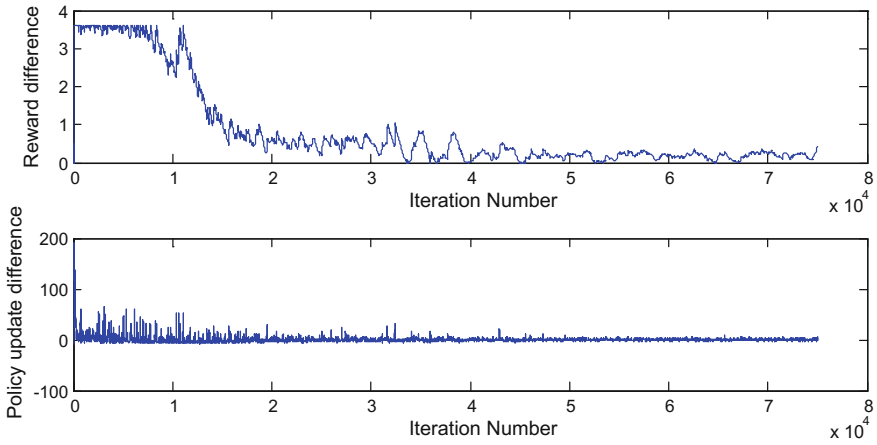


Fig. 7 Policy update difference and reward difference

The TD reinforcement learning algorithm is based on policy iteration and implemented using an actor/critic structure. Both actor and critic approximation NN are adapted simultaneously. The algorithm starts by evaluating the cost of a given initial admissible control policy and then uses this information to obtain a new and improved control policy, that is, the new policy that will have a lower associated cost compared with the previous control law. These two steps of policy evaluation and policy improvement are repeated until the policy improvement step no longer changes the actual policy, thus converging to the optimal controller. In order to validate the force tracking ability of the proposed controller after online learning, a random step signal, resembling the sudden change of force applied by the patient, is used as the input signal as shown in Fig. 8.

It can be noted that the controller is able to track the sudden change of force input, which represents the force exerted by the patient in real rehabilitation application. However, we can also note that there is some delay between them, around 0.5 s when started, and is reduced to around 0.1 s at 10 s after starting. The force tracking error keeps lowering at the same time, suggesting that the RLNN controller is adapting to the dynamics of the control system during online learning process.

4 Implementation Results

Tyromotion GmbH, the maker of Amadeo provides a development environment based on LabVIEW that receives and executes external control signals in the form of position and velocity from a host machine and provides the current position and force feedback produced by the actuators through a UDP communication protocol.

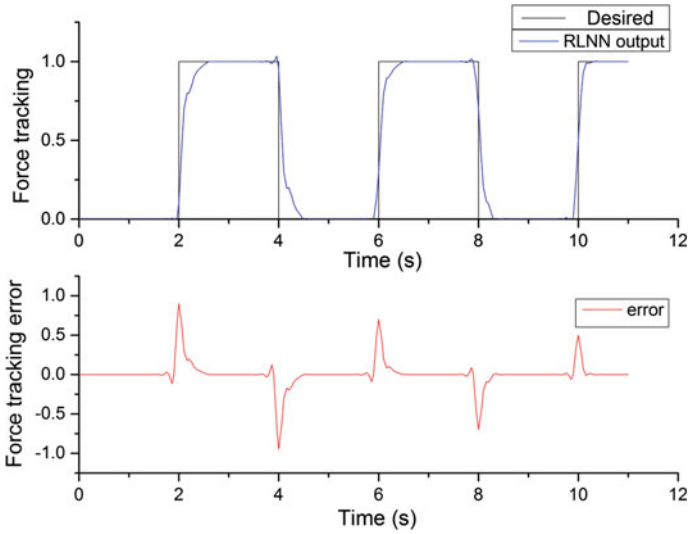


Fig. 8 RLNN controller force tracking and tracking error

The TD (λ) based RLNN control algorithm is coded in Matlab using 'MATLAB Script' and embedded in the LabVIEW environment, which provides accessibility to call custom-defined .m functions in LabVIEW files. This structure enables easy update and revision of the control algorithm.

To validate the effectiveness of the proposed control method in full grasping/extending ROM and force intensity, the following tests are performed on a post-stroke patient who suffered from a stroke 62 months before the tests and with partial control ability of his hands and fingers. An initial set of position and velocity and the desired endpoint location were given as input to the network: position = 0 (ranges from 0 to 100, where 0 is the furthest end and 100 represent the nearest end of the finger slides ROM), velocity = 1(ranges from 0 to 10, where 0 represent static and 10 is the full speed). Then, the current state force and position feedback at the end of each time step is used to update the inputs of the critic NN. Real-time assessment of force strength during full extension and grasping is shown in Fig. 9.

As shown, the force intensity varies between −20 and 20 N, where negative denotes the movement of extending and positive represents grasping. It can be seen that the actual output of force curve (F_Output) complies with the pattern of the estimated force (F_Estimate) in both directions, which means the RLNN controller is able to track and estimate moving trajectory of the finger. Also, the error between them turns to be smaller after 7 s, resulted from the online learning achieved by the RLNN controller. The provided assistive force is roughly the same intensity of that is needed after the online learning process.

In implementation, according to the ROM of each finger, this position value ranges from 0 to 100, the output position of the control system is denoted as 'P_sent', while the actual output of Amadeo robot pattern is depicted as 'P_receive'

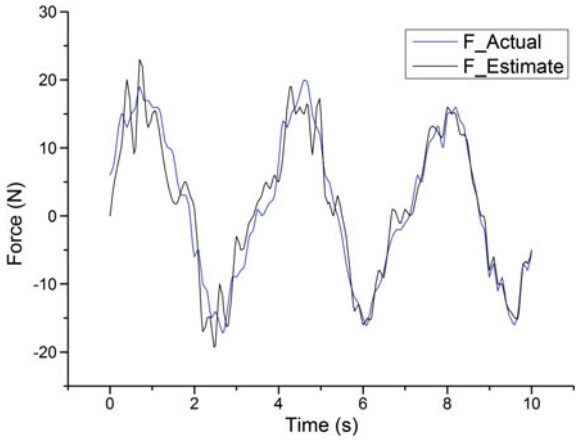


Fig. 9 Estimated and output force curve

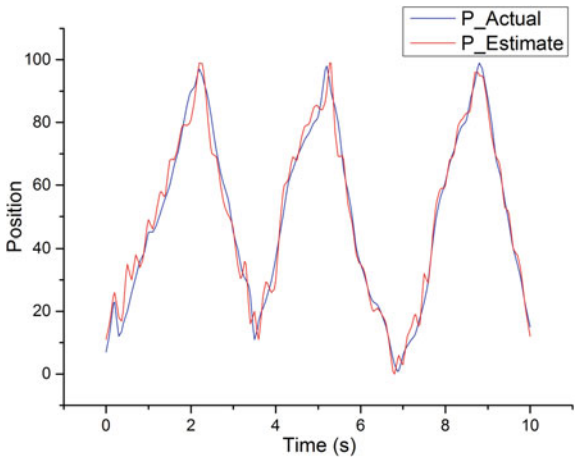


Fig. 10 Send and receive position and Amadeo output position

as shown in Fig. 10. The largest difference between these two values occur around 3 s after the start of the experiment. It can also be noticed that, compared with the total ROM of 80, the error suggests the proposed control method is able to assist the subject's movement in the intended trajectory within a timely manner.

5 Conclusion and Future Work

This study focuses on the development of an adaptive TD (λ) based actor-critic RLNN controller for the purpose of post-stroke fine hand motion rehabilitation training. Results suggest that this controller can fulfil the control requirements of Amadeo robot. However, the RLNN controller may need a large amount of online

learning to show significant improvement. Furthermore, extensive clinical experiments on post-stroke patients and evaluation analysis should be conducted before a firm conclusion can be drawn.

References

1. Sale, P., V. Lombardi, and M. Franceschini. 2012. Hand robotics rehabilitation: Feasibility and preliminary results of a robotic treatment in patients with Hemiparesis. *Stroke Research and Treatment* 2012: 1–5.
2. Hwang, C.H., J.W. Seong, and D.S. Son. 2012. Individual finger synchronized robot-assisted hand rehabilitation in subacute to chronic stroke: A prospective randomized clinical trial of efficacy. *Clinical Rehabilitation* 26: 696–704.
3. Kaelbling, L.P., M.L. Littman, and A.W. Moore. 1996. Reinforcement learning: A survey. *Journal of Artificial Intelligence Research* 4: 237–285.
4. Lewis, F.L., D. Vrabie, and K.G. Vamvoudakis. 2012. Reinforcement learning and feedback control: Using natural decision methods to design optimal adaptive controllers. *Control Systems, IEEE* 32: 76–105.
5. Papahristou, N., and I. Refanidis. 2011. Training neural networks to play backgammon variants using reinforcement learning. In *Applications of evolutionary computation*, vol. 6624, 113–122, eds. Di Chio C., S. Cagnoni, C. Cotta, M. Ebner, A. Ekárt, and A. Esparcia-Alcázar, et al. Heidelberg: Springer.

A Mechatronic Solution for Stroke Rehabilitation in a Reaching Task

P. N. Brett, X. Du, N. Mikov, A. Mohagheghi and T. Korff

Abstract This paper describes a mechanically simple, tactile sensing solution as a mechatronic system for discriminating human motion in the process of stroke recovery in reaching task. Three seated healthy volunteers and five stroke survivors were included in trials and demonstrated that the system can discriminate motion of arm and trunk of the body. The results show robustness with respect to variations in volunteers with a accuracy above 90% for both types of participants.

Keywords Tactile sensing · Discrimination · Stroke recovery

1 Introduction

A common therapy aiding stroke survivors is the reinforcement of types of motion that utilise the plasticity of the brain in the recovery process. As an example, this paper describes the distributive approach to tactile sensing applied to monitor arm movement in a reaching task. The evaluation is on the extent, trajectory and type of arm/body motion utilised. This information can be used as feedback to the stroke survivor and therapist to encourage the activity, to feedback advice and provide a measure of improvement.

There is growing incidence of stroke in populations of the world. In the UK alone there are 152,000 reported annually. Currently there are 1.2 million stroke survivors [11], many of which are undergoing the repetitive therapy of reinforcing motion. The therapy is challenging for the survivor as it requires many hours each day. A sensing and guiding system as described in this paper is an ideal tool to

P. N. Brett (✉)
University of Southern Queensland, Queensland 4350, Australia
e-mail: Peter.Brett@usq.edu.au

X. Du · N. Mikov · A. Mohagheghi · T. Korff
Brunel University, London, UK

J. Billingsley and P. Brett (eds.), *Mechatronics and Machine Vision in Practice 3*,
https://doi.org/10.1007/978-3-319-76947-9_15

accomplish this approach tirelessly. The stroke survivor needs to maintain concentration and therefore part of this work is to devise schemes to entertain and encourage interest that will engage the person fully in the recovery task. In contrast with other research [8, 9, 17], on assisting in repetitive motion therapies, the approach introduced in this paper is aiming toward a low-cost system that is able to achieve the same as alternative, more elaborate equipment.

This paper describes a tolerant system using the distributive approach to tactile sensing that can discriminate types of motion within the context of the therapeutic process. For the reaching task sensing output is used to describe the type and extent of motion and utilises a chair, as the subject is seated. The distributive approach to tactile sensing is applied to a deforming structure responding to the varying applied loads. In this application, the applied load is distributed under the seated person, reacting to weight and inertial forces from motion of the body. The sensing approach utilises the coupled output transients of a limited number of sensory elements reflecting strain or deflection applied to the deforming continuum structure of the chair: In this example, discrimination of motion is achieved in real-time by monitoring the coupled time-series of the sensory data. Coupled features in the data are anticipated and recognised through an algorithm to discriminate the type, extent and changes in movement taking place. The real-time information is displayed to the user/therapist. The experimental device has been tested on three healthy volunteers and the results are discussed here. The work has subsequently led to pilot patient trials.

2 Methodology and Experimental Setup

Tactile sensing refers to the sense of touch in the biological somatosensory system [14, 15]. In engineering, tactile sensing refers to extracting information from the interaction between a surface and contacting objects [7]. This information, such as force and pressure, has been used for precise control of machines and work pieces [5], toys and consumer devices, and more recently in application bioengineered systems such as prosthetics and exoskeleton devices [14], and surgical tools [6, 10]. Much tactile sensing is applied as a point measurement [5, 15]. The distributive approach applied to tactile sensing has been explored extensively by the research team. It assumes that loading occurs somewhere in the range of a deforming structure. The coupling of sensory transients through the deforming structure enables recognition of changing loading without specific reference to location of application. This enables a mechanically simple setup to be achieved in contrast with sensor arrays, for example Hammond et al. [4]. The complex part in developing an operational system is the correlation between events and coupled set of sensory transients.

Within a particular motion task, the types of movement can be recognised automatically, and their extent too. The types of complex motion in gait have also been recognised automatically using this approach [2, 3]. As in the case with all sensing elements, sensitivity is increased with greater transmission between measurands and the sensing element.

This sensing approach has been found powerful in applications where discrimination or recognition can form the basis of outputting a description of an object event or process. This can be used well where the goal is to identify or contrast as opposed to measure. Examples are surgical robotic tools where the approach is used to recognise tissues and tissue structures ahead on the tool path by using the tissue as the coupling medium [13]. Other examples are endoscopes and catheters [12], and earlier investigations on controlling robotic grippers for handling deforming mediums such as food and dough [1], discriminating cells [18], and in the discrimination of types of gait mentioned earlier [2]. In the latter three applications, the sensing medium has been an artefact, and often a flat plate, to which a low number of sensing elements have been deployed to enable robust sensing over the full extent of the deforming medium. For applications in discriminating human motion, the approach is tolerant to the build of the person, or indeed their placement on the sensing medium [2, 3, 16]. The approach is ideal for application to imprecision associated with multi-dimensional, variable human activity and the wide range of potential motion that can take place around a specified task.

For the reaching task, the sensing medium is an office chair (Fig. 1), consisting a rigid metal frame with cushioned bottom seat and back support. To investigate sensitivity of performance to sensor placement, sixteen infrared proximity sensors integrated into the chair were used to detect deformation of the seat. The aim is to discriminate the nature of reaching performed by the seated person. A typical non-linear characteristic of sensor output is shown in Fig. 2. In this application, as expected, non-linearity of sensor characteristics augments automatic discrimination over a greater range of motion types.

Fig. 1 The converted sensing chair

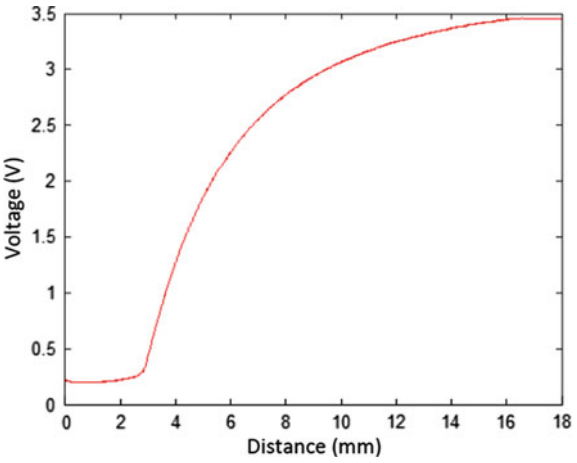


Fig. 2 Voltage-distance characteristic of IR displacement sensor

The reaching task (reaching to a target while seated) is used in therapy and tests the fundamental ability of the central nervous system to integrate multiple degrees of freedom (joints, body segments) in motion. In healthy individuals, arm and trunk movements are coordinated when reaching to a target placed within 90–100% of maximum arm reach. The trunk does not contribute to the movement. Contribution of the trunk to arm movements starts beyond the singularity of maximum reach, where movement of the trunk starts before, or simultaneously, with the reaching motion, and finishes after the desired reach is achieved. This pattern of coordination between the arm and trunk is disrupted after stroke leading to trunk recruitment in executing the reach to a target placed within the reach of the arm. Muscle weakness, spasticity, and abnormalities in the regulation of spinal reflexes are amongst suggested mechanisms that likely contribute to the impaired control of movement after stroke.

To discriminate trunk angles and the contribution to the reaching task, healthy subjects were tasked with reaching a target placed at three distinct locations beyond the arm length. Motion was performed using maximum contribution from the arm and different contributions from the trunk (minor, moderate and maximum). In order to unify performance within the reaching task of these healthy volunteers, targets were placed at distinct locations defined based on their individual maximum arm length. The length was measured from the acromion to third metatarsal anatomical landmarks. Using this metric, the full length of the arm corresponds to 120% of this distance. Figure 3 illustrates the three discriminative conditions corresponding to reaching to a target placed at distances equal to: 120% of the arm length involving no or minimal trunk contribution (Fig. 3a); 150% of the arm length involving moderate trunk contribution (Fig. 3b); the maximum reach of the participant involving maximum contribution from the trunk (Fig. 3c). The anatomical statistics of the three volunteers are shown in Table 1.

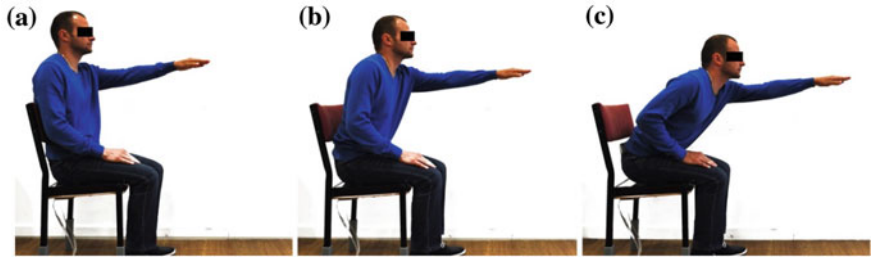


Fig. 3 Reaching task conditions involving arm contribution and different target locations

Table 1 Anatomical statistics of the three volunteers

Volunteer	Gender	Age	Weight (kg)	Height (cm)	Arm length (cm)	Max. reach (cm)
A	Male	26	68	175	67	124
B	Female	27	52	165	59	115
C	Male	33	81	182	71	130

3 Results

Typical sensing transients from one of 4 sensing elements used in combination to discriminate the motion are shown in Fig. 4 for each of the three reaching conditions respectively. The transients represent the deflection of the surface. For condition A (120% arm length), it can be observed that the sensing transient has low amplitude and, as would be expected, there is increased amplitude with the increased reach, and consequently greater trunk contribution to the motion. In all three conditions, sensing transients show a down-drift characteristic (indicated in Fig. 4b) that is associated with initiation of trunk bending. In condition B and C,

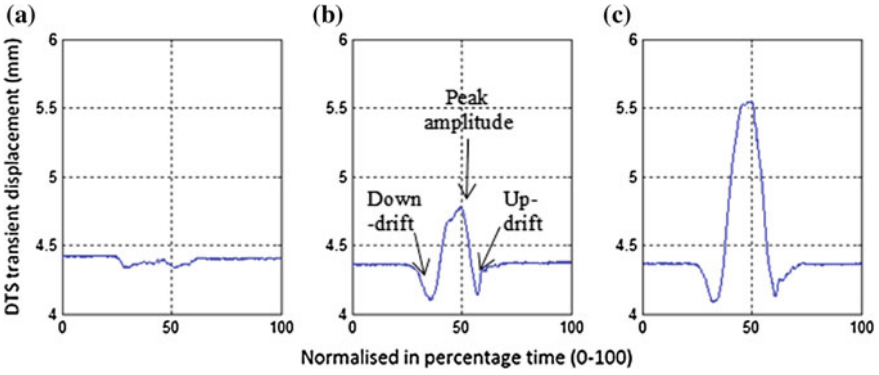


Fig. 4 Sensing transients during the reaching tasks under condition **a** 120% of the arm length, **b** 150% of the arm length, **c** the maximum reach

the peal amplitudes of the deflection correspond to the reaction when the target is reached and there is acceleration back to the initial position. It also can be seen that there is an up-drift stage (indicated in Fig. 4b) in the time series training sets, indicating that the trunk is accelerating back to the initial position.

The peak amplitude correlates with the level of trunk bending in the task. This feature combined can be used to generate patterns for discriminating bending conditions in reaching tasks. It is important to note that the same algorithm used to interpret the multi-data sensory time series offers a robust, automatic prediction independent of the weight and range of seated position of the subject on the chair, and is also tolerant to variations in motion as the reaching task is performed. The algorithm used in this study is an Artificial Neural Network (ANN) derived from MATLAB Neural Network Toolbox. A three-layer feed-forward ANN was defined with 10 hidden neurons. The ANN was implemented, trained using independent experimental data and then verified on three healthy volunteers with different height, and body build. Each volunteer performed the reaching task under three conditions, 15 times respectively. Table 2 shows the results classified by the neural network are largely successful for identifying trunk contribution in the reaching task conditions A, B and C.

A pilot clinical study was performed with five chronic stroke survivors. The physical relevant statistics of the volunteer participants are presented in Table 3. The stroke survivor participants are at differing levels of impairment for the motion of the upper limb based on Fugl-Meyer (FM) upper extremity test score (summarized in Table 3). The target to be reached was placed at different locations according the maximum reach of each participant respectively. For the stroke

Table 2 Results showing reaching condition classifications by the ANN

	Number of correct classifications (out of 15)		
Volunteer	Condition A (120% reach)	Condition B (150% reach)	Condition C (maximum reach)
1	15	14	15
2	15	15	15
3	15	15	15
Average classification accuracy–99.3% (134 correct classifications out of 135)			

Table 3 Physical records of the healthy and stroke volunteers

Volunteer	Gender	Age	Weight (kg)	Height (cm)	Side tested	Arm length (cm)	Max. reach (cm)	Fugl-Meyer score
S1	Female	52	50	160	Right	68	43	20/60
S2	Male	28	80	178	Left	71	127	58/60
S3	Female	56	54	168	Left	65	120	59/60
S4	Female	48	57	155	Left	62	106	59/60
S5	Male	49	83	175	Right	69	98	38/60

Table 4 Results of classifications of reaching to the critical boundary with the ANN

Stroke participants	Fugl-Meyer score	Average trunk angle (degree) to the critical boundary	Relevant discrimination condition	Number of correct classifications (out of 15)
S1	20/60	21	Condition B	10
S2	58/60	4.3	Condition A	15
S3	59/60	4.3	Condition A	15
S4	59/60	6	Condition A	15
S5	38/60	30.9	Condition B	15
Average accuracy 93.3% (70 correct classifications out of 75)				

participants who can reach beyond their arm length (S2–S5), the target was placed at a critical boundary which is equal to their arm lengths. For the participant who could not reach an arm length (S1), the target was placed at the maximum reach position, measured as 63% of arm length (Table 4).

The identical ANN introduced in previous the paragraph was tested for classification of reaching. For participants S2, S3, and S4 of mild motor impairment (FM scores are above 50), the results show similarity with healthy participants (all classified as Condition A). For participants S1 and S5 of moderate impairment (FM scores are between 20 and 40), the critical boundary is classified as Condition B. The average accuracy of classifications in this trial is 93.3%. It is identified that the accuracy for stroke participant S1 is 66.7% (10 out of 15), which is far below the accuracy for other participants at 100%. The FM score of stroke participant S1 is 20 out of 60, which is the lowest among all stroke participants. The second lowest score is 38 (participant S5). This suggests that to improve the performance of the ANN, further data will be needed for lower scored patients.

4 Conclusion

A sensing system for discriminating the contribution of different segments of the body in a reaching task has been devised. The application is intended for use in assessing and monitoring progress in stroke recovery. The development of the system has followed requirements of design principles for low cost taking into account robustness to variation in motion, build and initial posture of a group of volunteers. Presented results show that sensing transients from the distributive tactile sensing chair have different characteristics when performing reaching tasks with different trunk bending contribution. A neural network was applied to discriminate the contribution of trunk bending in three reaching conditions with 99.3% accuracy for healthy participants and 93.3% for stroke survivor participants. This can be used to evaluate the level of motor impairment for people with stroke hemiparesis.

References

1. Brett, P.N., R.S.W. Stone, and B.S. Evans. 1998. A system for the automatic handling of compact shaped dough products. *International Journal of Mechatronics* 8 (2): 85–102.
2. Elliott, M.T., X. Ma, and P.N. Brett. 2009a. A smart sensing platform for the classification of ambulatory patterns. *Proceedings of the Institution of Mechanical Engineers, Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, 223(H5): 567–575.
3. Elliott, M.T., I. Petra, X. Ma, P.N. Brett, and D.J. Holding. 2009b. Quantifying sway through surface deflection patterns: A novel approach using distributive tactile sensing. *Part H: Journal of Engineering in Medicine*.
4. Hammond, I.I.I., L. Frank, Rebecca K. Kramer, Qian Wan, Robert D. Howe, and Robert J. Wood. 2014. Soft tactile sensor arrays for force feedback in micromanipulation. *IEEE Sensors Journal* 14 (5): 1443.
5. Hanna, Y., M. Boukallel, and K. Althoefer. 2011. Tactile sensing for dexterous in-hand manipulation in robotics—A review. *Sensors and Actuators A: physical* 167 (2): 171–187.
6. Hui, X., A. Jiang, L. Seneviratne, and K. Althoefer. 2012. Pixel-based optical fiber tactile force sensor for robot manipulation. In Sensors, IEEE, 1–4.
7. Lee, M., and H. Nicholls. 1999. Review article tactile sensing for mechatronics—A state of the art survey. In *Mechatronics* 9 (1): 1–31.
8. Murphy, M. Alt, C. Willén, and K. Sunnerhagen. 2011. Kinematic variables quantifying upper-extremity performance after stroke during reaching and drinking from a glass. *Neurorehabilitation and Neural Repair* 25 (1): 71–80.
9. Nasr, N., S. Torsi, P. Wright, S. Mawson, and G. Mountain. 2010. Supporting the self-management of stroke by applying a user-centred design approach. *Clinical Rehabilitation* 24: 276–287.
10. Shimachil, S., S. Hirunyanitiwatna, Y. Yasunori Fujiwara, A. Hashimoto, and Y. Hakozaki. 2008, June. Adapter for contact force sensing of the da Vinci® robot. *The International Journal of Medical Robotics and Computer Assisted Surgery* 4 (2): 121–130.
11. Stroke Association. 2017. Stroke association, United Kingdom key statistics. https://www.stroke.org.uk/sites/default/files/state_of_the_nation_2017_one_pager.pdf.
12. Tam, B., X. Ma, and P.N. Brett. 2009. An experimental investigation of a fibre optic sensing method to discriminate contact and motion of a flexible digit typical of invasive clinical environments, accepted. *Part H: Journal of Engineering in Medicine*.
13. Taylor, R., X. Du, D. Proops, A. Reid, C. Coulson, P.N. Brett. 2010. A sensory guided surgical micro-drill. In *Proceedings of the institution of mechanical engineers. Part C: Journal of Mechanical Engineering Science* 224 (7): 1531–1537.
14. Tiwana, M.I., S.J. Redmond, and N.H. Lovell. 2012a. A review of tactile sensing technologies with applications in biomedical engineering. *Sensors and Actuators A: Physical* 179: 17–31.
15. Tiwana, Mohsin I., Stephen J. Redmond, and Nigel H. Lovell. 2012b. A review of tactile sensing technologies with applications in biomedical engineering. *Sensors and Actuators A: physical* 179: 17–31.
16. Tongpadungrod, P., D. Rhys, and P.N. Brett. 2003. The optimisation of a distributive contact tactile sensing technique for measuring the position of a static point load. *International J of Sensors and Actuators, Part A* 105: 47–54.
17. Volpe, B.T., H.I. Krebs, N. Hogan, L. Edelstein, C. Diels, and M. Aisen. 2000. A novel approach to stroke rehabilitation robot-aided sensorimotor stimulation. *Neurology* 54 (10): 1938–1944.
18. Wu, Z., X. Ma, P.N. Brett, and J. Xu. 2009. Vibration analysis of submerged rectangular microplates with distributed mass loading. In Proceedings of royal society, A. 465 (2104): 1323–1336 (April 8 2009).

Fuzzy-CPG Hybrid Control of a Hip Joint Walking Assist Device Based on Soft Sensing of Muscle Forces

Xingsong Wang, Fengpo Du and Jigang XU

Abstract To meet the vast request of the elder people with moving function impairment in the coming aging society, a prototype of hip joint walking assist device was develop. Based on the detailed analyzing of the variation characteristics of the muscle force during one gait cycle, a hybrid control algorithm was developed, which melted fuzzy control and CPG-based learning control. Fuzzy control algorithm is developed respecting the angular velocity, angular acceleration of the hip-joint of the device and the muscle force during the gait period. To releasing of the dragging force between the device and the wearer, the CPG learning algorithm is designed to study the output of the fuzzy control algorithm in stable walking state and thereafter taking over the stable state control of the device. Experiments showed the effectiveness of the proposed method.

Keywords Hip joint assist device · Fuzzy-CPG hybrid control
Soft sensing · Muscle force

1 Introduction

The aging population is one of the most critical challenges for our current industrialized societies. Caused by a low birth policy and long life expectation, the aging population will threaten the sustainability of our social welfare. A possible scenario for the next years is that aging-related gait syndromes will lead to an increase of the number of people needing assistance in their daily living activities, e.g. basic mobility, personal hygiene and safety awareness. In this scenario, it is plausible that people will become progressively more reliant on technology to meet their desire to live independently, actively and satisfactorily. Among all the assistive devices springing up, wearable robotic orthoses (namely "exoskeletons") were proposed as

X. Wang (✉) · F. Du · J. XU
School of Mechanical Engineering, Southeast University, Nanjing, China
e-mail: xswang@seu.edu.cn

J. Billingsley and P. Brett (eds.), *Mechatronics and Machine Vision in Practice 3*,
https://doi.org/10.1007/978-3-319-76947-9_16

a solution by many research teams active in the field of medical robotics to assist people affected by gait disorders.

Robotic exoskeleton combines the intention of the wearer and the robotic dynamics together to assist human to achieve normal movement. Thus an exoskeleton for gait assistance is generally anthropomorphic in nature, "worn" by the user, and fits closely to his or her body [1, 2]. Given the close interaction with the user, the robot should be light-weight and take into account the user's joints range of motion, anthropometry, and kinematics to provide a comfortable physical human–robot interface. Furthermore, the actuation and control of the robot should allow the user to implement his or her own movement without hindrance while receiving a certain assistance safely.

Robotic exoskeletons have been developed for a large number of applications, with the goal of assisting or enhancing human activities. For the lower extremity, devices have been developed to apply assistive torques to the biological joints to augment healthy individuals or assist those with disabilities to walk [1–7]; assist with load carriage by providing a parallel path to transfer load to the ground [8, 9], thus off-loading the wearer and bypassing their musculature; or finally, provide gait retraining or rehabilitation for those with disabilities [10].

It is worth to cite also single-joint active orthoses, such as the SERKA, an active knee orthosis addressing stiff knee gait in stroke patients [11] actuated by a cable-driven rotational SEA, the Dynamically Controlled Ankle–Foot Orthosis [12] and the Adaptive Ankle–Foot Orthosis by Blaya et al. [13], which are examples of simpler active orthoses making use of SEAs to assist push-off or to correct dropped foot gait. Examples of active orthoses for the hip flexion–extension assistance are the devices introduced by do Nascimento et al. [14] and the hip exoskeleton designed by Ferris et al., both powered by artificial pneumatic muscles.

Here, we introduce the design of a light-weight active pelvis orthosis (APO), for assisting hip flexion–extension with the rest two degrees of mobility passive. The device was typically designed with rigid links, and DC-motors-harmonic-reducer driven joints. Due to the simplicity of the structure, the device hence ensures good kinematics compatibility, enhancing comfort of human-robot physical interaction, avoiding limitations and constrains to user's gait pattern, and addressing anthropometric variability.

To achieve unified movement and relieve mutual interference between exoskeleton and the wearer, the control strategy should be integrated with the intention of the wearer along with smooth continuity. There are lots of control methods being developed in the limb powered exoskeleton. Each has its own advantages and never being satisfied. The control of exoskeletons still needs good efforts for bettering their performances.

In this paper, due to the innovative muscle force soft sensing method, a fuzzy-CPG-based hybrid control algorithm of the hip powered exoskeleton is developed based on the angular velocity, angular acceleration of the hip-joint and the thigh muscle force.

2 Soft Sensing of Thigh Muscle Forces

The walking movement of human is the result of the muscle contraction or extension which drives the bone to move cooperatively. It is the most accurate and complex movement that derived from the body and the nervous system. In order to develop the coordination control strategy to actuate the hip exoskeleton, we need to collect interactive information between the wearer and the exoskeleton which typically represented by the variations of muscle forces of thigh in one gait cycle.

2.1 Airbag Based Soft Sensing of Muscle Force

As shown in Fig. 1, an airbag-based muscle force sensing system was developed, which was an indirect measurement system used to get the variation of the muscle forces. The measuring system mainly consists of two the airbag force sensors, a data acquisition board and a wifi-based data transmission modular. The muscle force data gotten in the experiments is transferred to a PC with data analysis software embedded in it. The change of the gas pressure in the airbags can be used to qualitatively calculate the variation of thigh muscle force.

2.2 Muscle Force Data Acquisition Experiment

The diagram of the muscle force data acquisition experimental is shown in Fig. 2. Two airbag force sensors are bound on the thigh flexor and extensor. And a

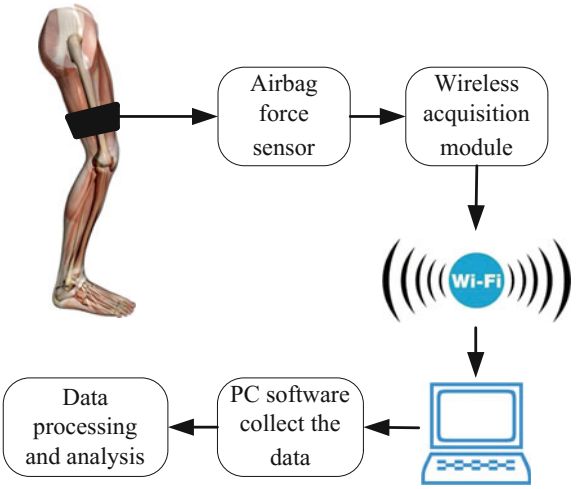


Fig. 1 Muscle force measuring system

PC-based data acquisition and analysis software displays the outputs of the two sensors in real-time. The difference between the output voltages of the two sensors actually reflects the force changes of thigh muscle pair.

The experiment with normal gait walking was carried out on a treadmill. The voltage difference between the two airbag sensors was shown in Fig. 3. In which, the positive value represented that the flexor played a dominate role. Otherwise, the extensor played a dominate role.

2.3 *Experiment Analysis*

In order to facilitate the subsequent analysis and the developing of control algorithm, we intercepted a typical gait cycle in Fig. 4 and transformed the abscissa to the gait cycle. In order to make the curve more intuitive, the variation curve of the muscle forces was shown simultaneously with walking gaits of two feet in Fig. 4.

Comparing the curve of measured muscle force with reference to the gait status, the following phenomena could be found:

(1) In the support phase of left leg, flexor force increases gradually in from stride1 to stride 2 point, and reaches the first just after stride 2. Then the flexor force decreases from stride 2 to stride 3, and when the flexor decrease to zero at stride 3, the extensor takes over the supporting work. Analogously, the extensor strength decreases from stride 4 to stride 5, and reaches zero at stride 5.
(2) In the swing phase of the left leg, the flexor force increases gradually from stride 5 to stride 6, and reaches maximum at stride 6. From stride 6 to stride 1, the extensor strength decreases to zero.
(3) Comparing the gait status of two legs shown in Fig. 4a, one can find that at stride 3 of left leg and stride 6 of right leg, the extensor force of left leg is zero but the flexor force of the right leg is the largest. The muscle forces of two legs are complementary.

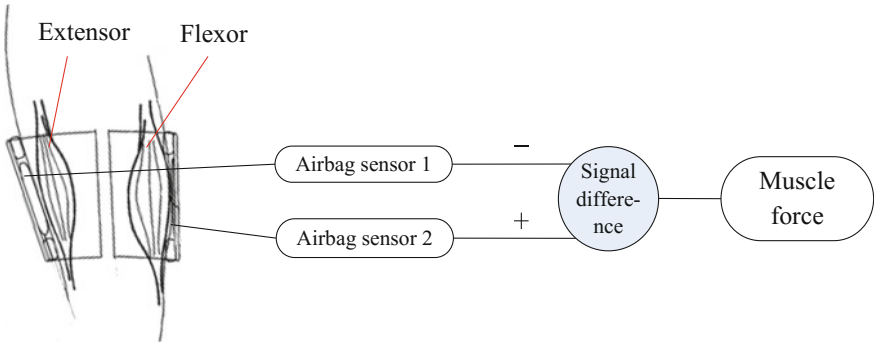


Fig. 2 Collecting experiment

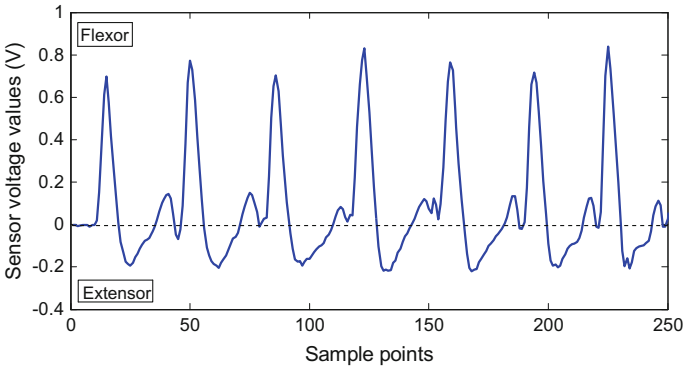


Fig. 3 The curves of muscle force

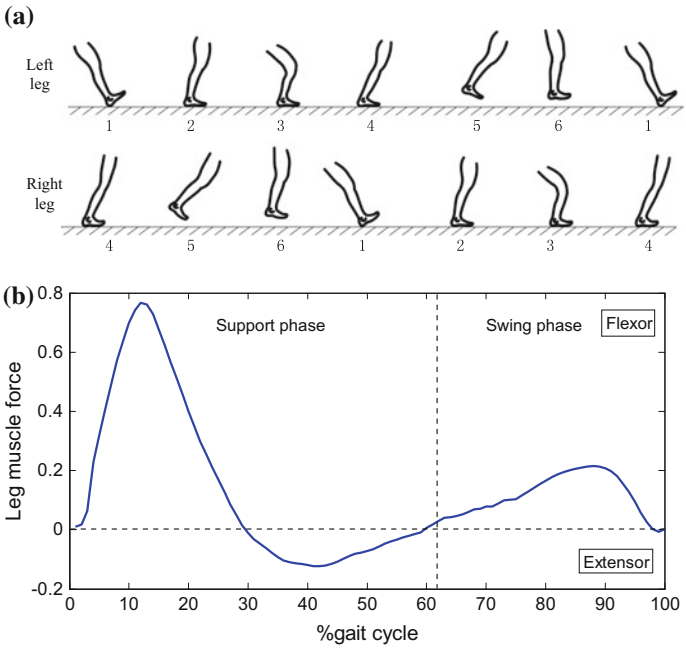


Fig. 4 The muscle force of the left leg

(4) Also from Fig. 4a, one can infer that for one leg, the direction of angular velocity of the hip joint and the direction of muscle force are opposite. Therefore the flexor does negative work to store a large amount of energy and then release at stride 3.

3 Fuzzy Control Algorithm

Assisting the human movement is the basic function of the powered hip joint exoskeleton. Therefore, it is important for the controller to correctly judge the switching points such as start, stop or direction change of leg movement during the man-exoskeleton walking. Fuzzy control algorithm, which only based on intuitive concepts, were employer here for this purpose with only considering kinematic features of hip movement and the changing curve of the thigh muscle force with no need to build an accurate control model of the whole man-machine system.

3.1 Fuzzy Controller

The input of Exoskeleton fuzzy control system is the gait intention of the wearer, which is the interaction between the wearer and the exoskeleton and can be regarded as the feedback error of the entire man-machine system. Let variable e be the angular velocity of the bar of the device binding on human thigh bone which can be gotten from the encoder installed in the hip joint of the device. e and its rate of change, the angular acceleration $\dot{e}$ are the input variables of the fuzzy controller. The angular velocity reflects the movement frequency of the thigh very well. While the angular acceleration reflects the motion state of the device. The language inputs of variables e and $\dot{e}$ are E and EC, respectively. And the output variable of the fuzzy controller is the control input u of the motor driver, whose language variable is U.

Based on experimental results of ordinary walking, the basic domain of the hip joint angular velocity e for normal case is [−200°/s, 200°/s], and the basic domain of the angular acceleration $\dot{e}$ is [−2000°/s^2, 2000°/s^2], the basic domain of the input u is [−5 V, 5 V]. The domain of discourse of e is X = {−3, −2, −1, 0, 1, 2, 3}, domain of $\dot{e}$ is Y = {−3, −2, −1, 0, 1, 2, 3}, domain of u is Z = {−3, −2, −1, 0, 1, 2, 3}. Thus the corresponding language variable of E is A = {PB, PM, PS, ZO, NS, NM, NB}, language variable of EC is B = {PB, PM, PS, ZO, NS, NM, NB}, language variable of U is C = {PB, PM, PS, ZO, NS, NM, NB}. The configuration of each variable is the same, with maximum value of membership function is 1. Here we select variable A which represent e:

$$\begin{aligned}
\mu_{PB}(x) = 1 \leftrightarrow x = +3,\\
\mu_{PM}(x) = 1 \leftrightarrow x = +2,\\
\mu_{PS}(x) = 1 \leftrightarrow x = +1,\\
\mu_{NB}(x) = 1 \leftrightarrow x = -3,\\
\mu_{NM}(x) = 1 \leftrightarrow x = -2,\\
\mu_{NS}(x) = 1 \leftrightarrow x = -1,\\
\mu_{ZO}(x) = 1 \leftrightarrow x = 0.
\end{aligned}$$

From the domains of the input/output variables and their corresponding language variables, the quantitative factors of each variable can be found:

Quantitative factor of angular velocity e is:

$$k_e = 3/200 = 0.015$$

Quantitative factor of angular acceleration $\dot{e}$ is:

$$k_{ec} = 3/2000 = 0.0015$$

Quantitative factor of output u is:

$$k_u = 5/3$$

As shown in Fig. 5, we use the triangular membership function.

3.2 Fuzzy Rules

In designing of fuzzy controller, the main factor to achieve good control performance is the determination of fuzzy rules. The change of the angular velocity and angular acceleration obtained by processing of human hip joint in one gait cycle, and the change of the thigh muscle force collected by experiments in a gait cycle are summarized and shown in Fig. 6.

By comparison of the relevant values in Fig. 6, the following important inference and the fuzzy rules of Table 1 of can be obtained:

(1) It is a high relevant between the thigh muscle force and the change of hip angular velocity, the angular velocity of the process will be used as the main reference of fuzzy control rules in the swing phase.
(2) In the two main conditions of heels on the ground and toes off the ground, the direction of motion is in the switching state where the angular velocity of hip joint is zero. As a result, the angular acceleration is needed to determine the direction and the size of muscle force.

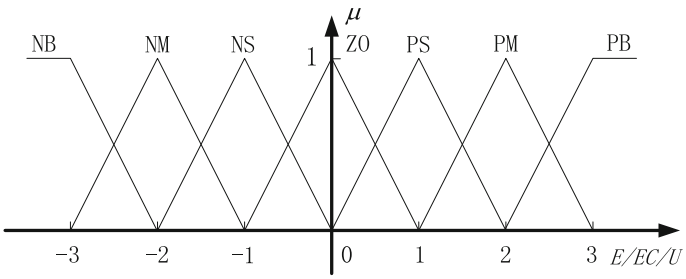


Fig. 5 Membership functions of E/EC/U

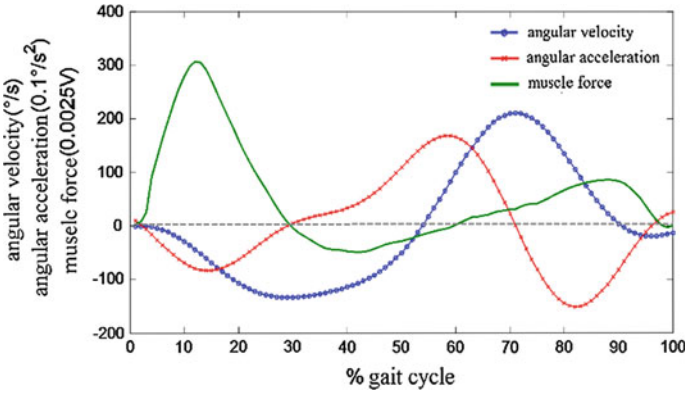


Fig. 6 Hip characteristic in a gait cycle

Table 1 Fuzzy rules

ω α	NB	NM	NS	ZO	PS	PM	PB
NB	ZO	NM	NS	NS	PS	PM	PB
NM	ZO	NM	NS	NS	PS	PM	PB
NS	ZO	NM	NS	NS	PS	PM	PB
ZO	ZO	NM	NS	ZO	PS	PM	PB
PS	ZO	NM	NS	PS	PS	PM	PB
PM	ZO	NM	NS	PS	PS	PM	PB
PB	ZO	NM	NS	PS	PS	PM	PB

(3) From the heel on the ground to the soles of feet on the ground, the muscles do negative work which will affect the stability of walking. Through several experimental studies, the bones do positive work to human besides the above, in which the amount of force is equal to the size of flexor muscle force.

The fuzzy control rule table is composed of 49 fuzzy control rules, and each rule determines a fixed fuzzy relation, which can be represented by R_1 to R_{49} respectively, the fuzzy relationship of fuzzy control system is shown:

$$R = R_1 \cup R_2 \dot{\cup} \ldots \cup R_{48} \cup R_{49}$$

According to the relationship between input and output of the fuzzy inputs and rules in Table 1, the values of fuzzy output can be obtained by the Mandani method:

$$U^* = (E^* \times EC^*) \circ R$$

4 CPG Control Algorithm

Although the fuzzy control algorithm can judge the intention of human walking and control the exoskeleton following the wearer's movement very well. But if the exoskeleton working in this mode for a long time, it may cause great fatigue to the wearer. In order to walk for a long time, considering the features of human walking, a CPG-based learning algorithm was introduced in the proposed hip-joint walking assist device.

4.1 Mathematic Model of CPG

CPG shows strong nonlinear dynamic properties. In recent years, the researchers used single or multiple nonlinear oscillators to simulate CPG. The most significant feature of the nonlinear oscillator is the existence and stability of its limit cycle. The system trajectories are close to the limit cycles. There are many kinds of mathematical models of CPGs. Hopf oscillator and Matsuoka oscillator are currently most commonly used models. We use the Hopf oscillator because it is relatively simple and its amplitude as well as frequency is easy to control. The differential equations of the Hopf oscillator are shown below:

$$\dot{x} = \left(\mu - r^2\right) - \omega y + \varepsilon F$$
$$\dot{y} = \left(\mu - r^2\right) + \omega x$$

Where $r = \sqrt{(x^2 + y^2)}$, $\mu > 0$, μ controls its amplitude. And ω is the natural frequency of the oscillator shown in Fig. 7.

4.2 Learning Algorithm of CPG

Because of the entrainment effect of the oscillator, the oscillation frequency is synchronous with the input signal. When the disturbance is over, the system will return to its own inherent limit cycle and forget the disturbance caused by entrainment effect. However in this application, we hope the selected CPG system remember the frequency characteristics of the input signal. From the existing oscillators, we use the following changing rules to adjust the frequency ω of the output signal. Then the differential equations for the modified oscillator model are shown below.

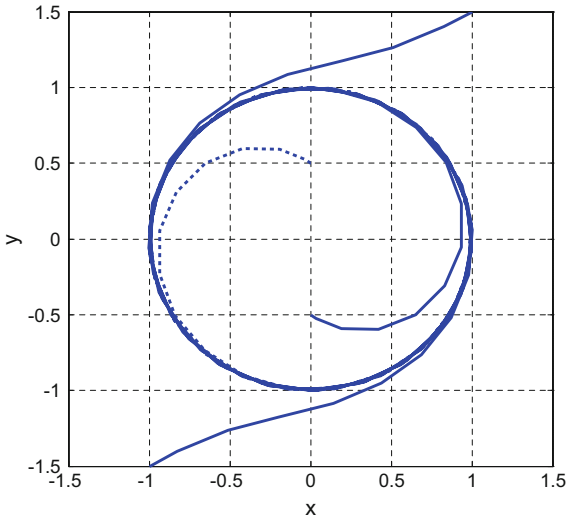


Fig. 7 The limit cycle of Hopf oscillator model

$$\dot{x} = \left(\mu - r^2\right) - \omega y + \varepsilon F$$
$$\dot{y} = \left(\mu - r^2\right) + \omega x$$
$$\dot{\omega} = -\varepsilon F \frac{y}{r}$$

From simulation results, we find that the modified Hopf oscillator model can learn any harmonic frequency. But if the input signal is a non-harmonic periodic signal, this modified CPG oscillator can only adapt to the component with nearest frequency.

As the oscillations of a Hopf oscillator are harmonic, an appropriate linear combination of several Hopf oscillators can be reproduced by periodic signals.

$$x(t) = a_0 + \sum_{n=1}^{\infty} A_n \cos(n\omega_0 t + \varphi_n)$$

In this way, each oscillator encodes one frequency component of the teaching signal. We also associate to each oscillator a variable encoding for the phase difference between the oscillator and the first oscillator, thus enabling us to reproduce any phase relationship between the oscillators. Figure 8 shows the structure of the network. The equations describing this CPG are as follow:

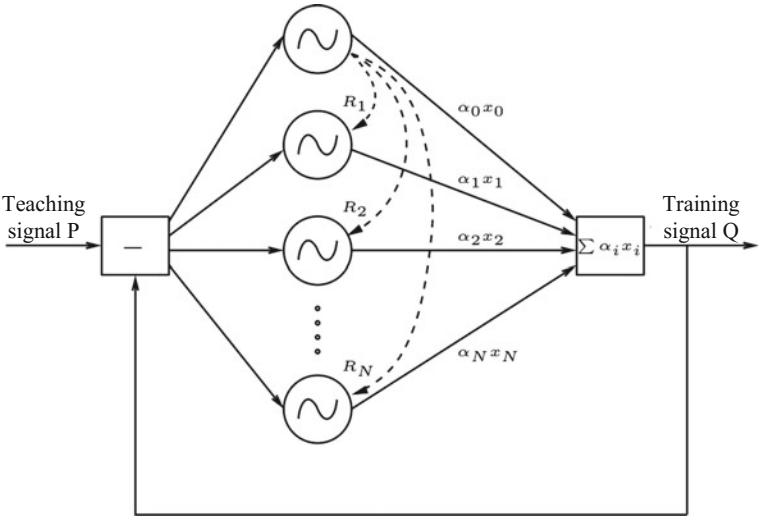


Fig. 8 Enhanced Hopf oscillator model

$$\dot{x}_i = \gamma\left(\mu - r_i^2\right) - \omega_i y_i + \varepsilon F(t) + \tau \sin(\theta_i - \psi_i)$$
$$\dot{y}_i = \gamma\left(\mu - r_i^2\right) + \omega_i x_i$$
$$\dot{\omega}_i = -\varepsilon F \frac{y_i}{r_i}$$
$$\dot{\alpha} = -\eta x_i F(t)$$

where,

$$\dot{\psi}_i = \sin\left(\frac{\omega_i}{\omega_0}\theta_0 - \theta_i - \psi_i\right)$$
$$\dot{\theta}_i = \mathrm{sgn}(x_i)\cos^{-1}\left(\frac{y_i}{r_i}\right)$$
$$F(t) = P(t) - Q(t)$$
$$Q(t) = \sum_{i=0}^{N} \alpha_i x_i$$

Where $F(t)$ represents the negative feedback in which average is the remaining of the teaching signal $P(t)$. The CPG continuously learns unless $F(t) = 0$. $Q(t)$ is the signal that the system learns from the teach signal. η is the learning constant, and ψ_i is the phase difference between oscillator i and i + 1.

5 Hybrid Control Algorithm

CPG learning algorithm and the fuzzy control algorithm both have their own advantages and disadvantages. For instance, fuzzy control algorithm is not suitable for the wearer to walk a long time with the assist device. Similarly, if only controlled with a CPG algorithm derived from a given trajectory, the assist device could not adapt to the variance during walking of the wearer, and thus the ride-comfort of the device will be bad. Based on these reasons, we integrate the two algorithms to improve the comfort with long time wearing. The hybrid control algorithm is shown in Fig. 9.

During the first 5 s of a walking state change, i.e. start, walking up/down stairs, etc., which is estimated by the velocity of acceleration of the hip joint rotation, fuzzy algorithm dominate the control of the device. And after 5 s of relatively calm walking, CPG algorithm starts learning the stable gait cycle and thereafter takes over the control efforts. This hybrid control method promises to reduce fatigue of the wearer for long time walking.

6 Experiment Studies

As is shown in Fig. 10, an experiment platform was developed to verify the validity of the algorithm performed on a prototype assist device. The experiment platform is composed of a walking gait data acquisition and analysis instrument, a treadmill and the prototype hip joint assist device. The speed of and the slope of the treadmill are all adjustable for generating different gaits and strides.

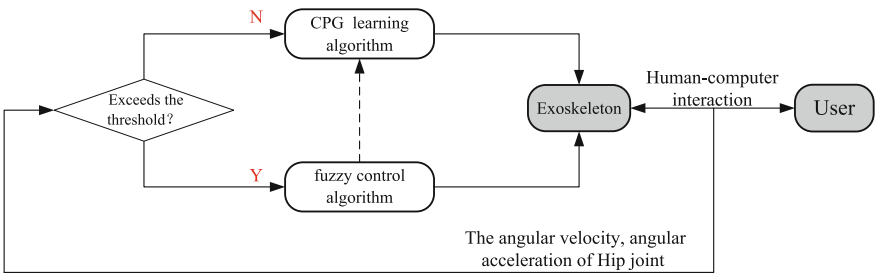


Fig. 9 Coordination control strategy

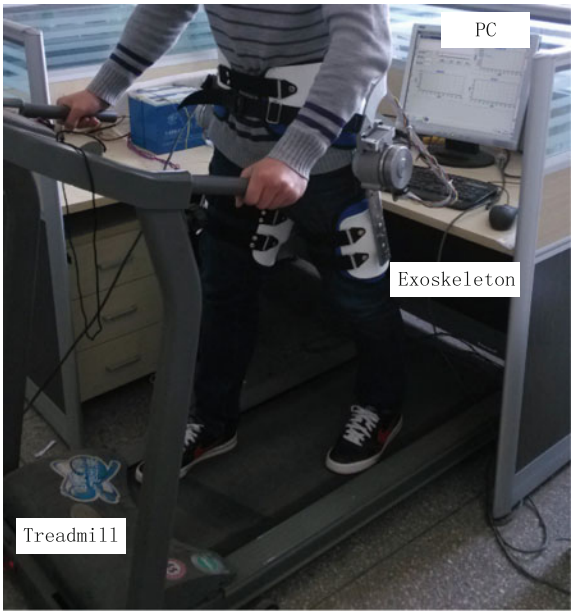


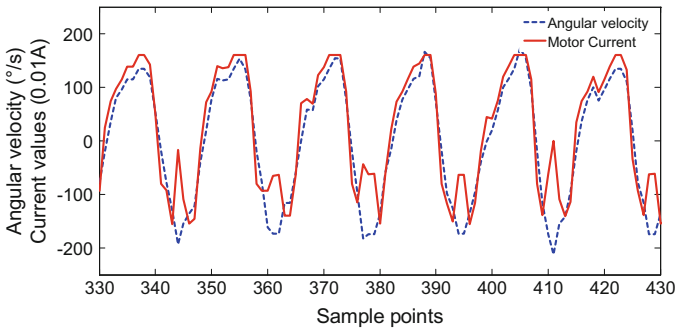
Fig. 10 Experiment platform

6.1 Fuzzy Control Experiment

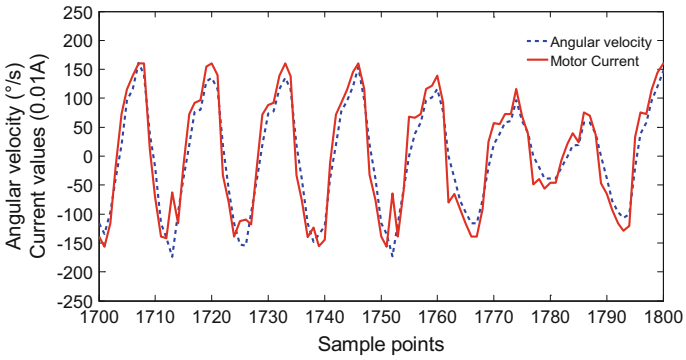
Experiments of different gait periods were done to verify the validity of fuzzy control algorithm. Figure 11 showed the tracking performances of angular velocity of the device for gait cycle periods of 2 and 1.3 s, respectively. It is obvious that fuzzy algorithm can effectively track and assist motions of any gaits, which provide a reliable base for CPG learning algorithm. On the other hand, as depicted in Fig. 11b, the decrease in stride leads to the decrease of the output torque of the motor instantly (Fig. 12).

6.2 CPG Learning Experiment

To verify the validity of the CPG algorithm and the control performance with different threshold values, the wearer was asked to walk at similar gait period with treadmill speed been changed accordingly. Several experiments of different stride frequency were done and the results are shown in Fig. 13. The CPG algorithm is capable of learning the characteristic of human walking and then automatically helps the wearer in the following walking.



(a) the gait cycle is 2s



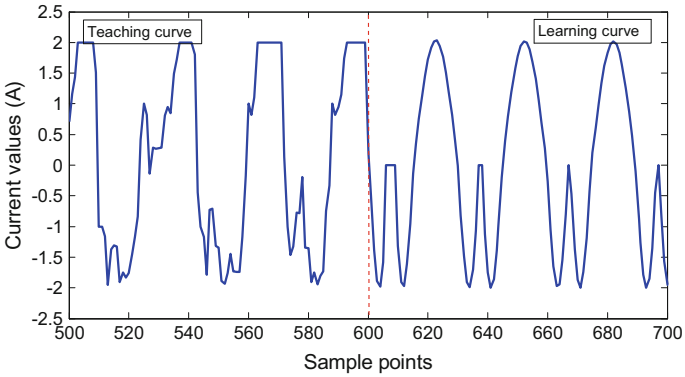
(b) the gait cycle is 1.3s

Fig. 11 Fuzzy control experiment

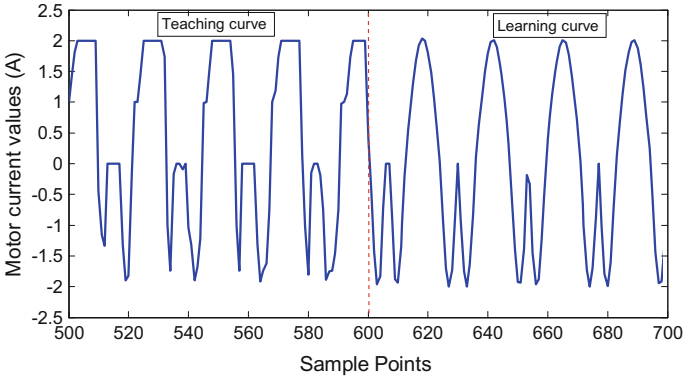
6.3 Hybrid Control Experiment

In order to verify the benefits of the hybrid strategy based on dedicated selected thresholds value, several experiments were conducted. After learned the human gait cycle, the wearer is assisted by the device controlled by CPG algorithm. At this circumstance, if there occurs an emergency (such as speed up), the threshold of hip angular acceleration will overshot, and the fuzzy control algorithm will be triggered. The experimental results are shown in Fig. 13.

When the wearer encounter a sudden event, which induced an overshoot of angular acceleration, the hip joint assist device will jump out of from the CPG oscillation, resume to work in the fuzzy control mode again. If the walking cycle is stable again, the CPG will relearn the gait cycle characteristics, and then a new control loop start again, which completes the man-machine hybrid control.



(a) The treadmill speed us 1km/h



(b) The treadmill speed us 2km/h

Fig. 12 CPG learning experiment

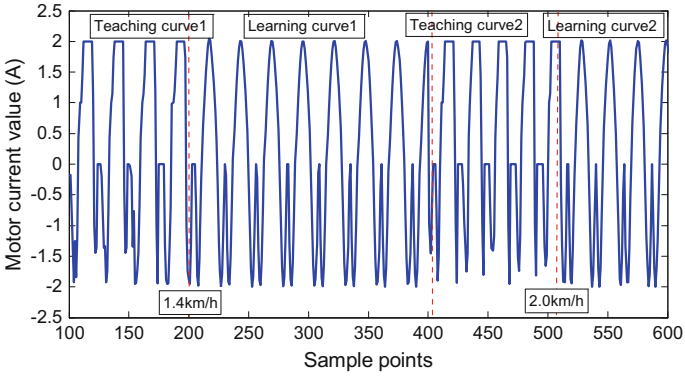


Fig. 13 Coordination control experiment

7 Conclusions

A coordination control system for an exoskeleton suit which takes full advantage of the steer ability of human and the load ability of the robot to achieve the goal of assistance is developed. A fuzzy control algorithm is developed based on the angular velocity, angular acceleration of the hip-joint and the muscle force during the gait period. The CPG learning algorithm is designed to study the output of the fuzzy control algorithm. According to the two algorithms the coordination control concept based on threshold is proposed. Besides these, the software-hardware platform of the system is set up for experimental research.

References

1. Dollar, A.M., and H. Herr. 2008. Lower extremity exoskeletons and active orthoses: Challenges and state-of-the-art. *IEEE Transactions on Robotics* 24: 144–158.
2. Herr, H. 2009. Exoskeletons and orthoses: Classification, design challenges and future directions. *Journal of NeuroEngineering and Rehabilitation* 6: 21.
3. Kawamoto, H., S. Lee, S. Kanbe, and Y. Sankai. 2003. Power assist method for HAL-3 using EMG-based feedback controller. *In IEEE international conference on, systems, man and cybernetics*, 2003, 1648–1653.
4. Pratt, J.E., B.T. Krupp, C.J. Morse, and S.H. Collins. 2004. The RoboKnee: An exoskeleton for enhancing strength and endurance during walking. In *Proceedings of 2004 IEEE international conference on robotics and automation ICRA'04,* 2430–2435.
5. Quintero, H.A., R.J. Farris, C. Hartigan, I. Clesson, and M. Goldfarb. 2011. A powered lower limb orthosis for providing legged Mobility in paraplegic individuals. *Topics in Spinal Cord Injury Rehabilitation* 17: 25–33.
6. Sawicki, G.S., and D.P. Ferris. 2009. A pneumatically powered knee-ankle-foot orthosis (KAFO) with myoelectric activation and inhibition. *Journal of Neuroengineering and Rehabilitation* 6: 23.
7. Sawicki, G.S., and D.P. Ferris. 2009. Powered ankle exoskeletons reveal the metabolic cost of plantar flexor mechanical work during walking with longer steps at constant step frequency. *Journal of Experimental Biology* 212: 21–31.
8. Kazerooni, H., and R. Steger. 2006. The berkeley lower extremity exoskeleton. *Journal of Dynamic Systems, Measurement, and Control* 128: 14.
9. Walsh, C.J., K. Endo, and H. Herr. 2007. A quasi-passive leg exoskeleton for load-carrying augmentation. *International Journal of Humanoid Robotics* 4: 487–506.
10. Shorter, K.A., G.F. Kogler, E. Loth, W.K. Durfee, and E.T. Hsiao-Wecksler. 2010. A portable powered ankle-foot orthosis for rehabilitation. *Journal of Rehabilitation Research & Development* (Accepted).
11. Sulzer, J.S., R.A. Roiz, M.A. Peshkin, and J.L. Patton. 2009. A highly backdrivable, lightweight knee actuator for investigating gait in stroke. *IEEE Transactions on Robotics* 25 (3): 539–548.
12. Hitt, J., A.M. Oymagil, T. Sugar, K. Hollander, A. Boehler, and J. Fleeger. 2007, April. Dynamically controlled ankle–foot orthosis (DCO) with regenerative kinetics: Incrementally attaining user portability. In *Proceedings of IEEE international conference on robotics and automation*, 1541–1546.

13. Blaya, J.A., and H. Herr. 2004. Adaptive control of a variable-impedance ankle–foot orthosis to assist drop-foot gait. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 12 (1): 24–31.
14. do Nascimento, B.G., C.B.S. Vimieiro, D.A.P. Nagem, M. Pinotti. 2008. Hip orthosis powered by pneumatic artificial muscle: Voluntary activation in absence of myoelectrical signal. *Artificial Organs* 32 (4): 317–322.
15. Righetti, L., J. Buchli, A.J. Ijspeert. 2006. Dynamic hebbian learning in adaptive frequency oscillators.*Physica D:Nonlinear Phenomena,* 216 (2): 269–281.
16. Righetti, L., A.J. Ijspeert. 2006. Programmable central pattern generators: An application to biped locomotion control. In *Proceedings of 2006 IEEE international conference on Robotics and Automation IEEE ICRA*, 1585–1590.

Sign-Language Recognition Through Gesture & Movement Analysis (SIGMA)

Clement Ong, Ian Lim, Joshua Lu, Claudine Ng and Thomas Ong

1 Introduction

The 2000 Census of Population and Housing, published by the Philippine Statistics Authority, revealed that 1.23%, or 942,098 people had a disability. Of this number, the statistics office declared that 3.4% were totally deaf, and 5.4% had an oral (speech) defect [8]. Combining these percentages results in about 82,904 potentially speech-impaired persons in the Philippines. Prominently, sign language comes to the aid of the deaf and speech impaired. Filipino Sign Language is currently being used by 54% sign language users in the Philippines. FSL is the ordered and rule-governed visual communication which has risen naturally and embodies the cultural identity of the Filipino community of signers [1].

Sign language translation is needed everywhere. In education, social services, and most importantly health care services. In the Philippines however, most healthcare workers do not understand sign language. They provide daily treatment and care for many people, some of who are speech impaired. It should be noted that most common errors in the field of medicine is because of miscommunication [3]. When the clients and healthcare providers do not share a common language, a qualified sign language interpreter can facilitate communication. They are necessary in situations where important information need to be exchanged such as taking a patient's medical history, giving diagnosis, performing medical procedures, explaining treatment planning, and during emergency situations [9]. However, there is a lack of sign language interpreters in the field of medicine.

Using computers to recognize sign language and convert this to text is not new; many successful systems have been developed, although only a handful for FSL.

C. Ong (✉) · I. Lim · J. Lu · C. Ng · T. Ong
Center for Automation Research, De La Salle University,
2401 Taft Avenue, 1004 Manila, Philippines
e-mail: clem.ong@delasalle.ph

J. Billingsley and P. Brett (eds.), *Mechatronics and Machine Vision in Practice 3*,
https://doi.org/10.1007/978-3-319-76947-9_17

All systems use either an instrumented glove to read hand pose and motion (gestures), or use computer vision to do the same [2, 5, 12].

Computer vision approaches tend to require significant amounts of computing power, while glove-based approaches provide direct reading but are inherently unable to provide hand versus body position information due to the relative measurement nature of the glove sensors.

Sign-language Recognition through Gesture and Movement Analysis (SIGMA) combines a data glove with image processing. The guiding principle of SIGMA is to use the inherent direct reading capabilities of an instrumented glove with "just enough" video processing to provide sparse hand position information, relative to the signer's body, over time. The end-goal is to produce a system that will be within the computational capabilities of a smartphone, without the need for additional computing or data resources.

2 The SIGMA System

SIGMA aims to facilitate limited communication between the speech-impaired and those who are not, particularly in the medical field where proper communication is important. To achieve this goal, a prototype data glove was developed to capture hand posture and attitude information. A complementary vision system is used, limited to detecting the hand position with respect to the body. The system is designed to recognize a vocabulary comprising the letters A to Z, numbers zero to nine and 30 words used for assessing health, as expressed in Filipino Sign Language.

The prototype data glove is instrumented with two types of sensors—a set of nine resistive flex sensors (one on the thumb and two for the other fingers) and an Inertial Measurement Unit (IMU). These sensors are interfaced to a small, single-board computer, whose analog input channels were increased by adding an analog multiplexer. The SBC times the data acquisition, sets the acquisition sequence, controls the analog multiplexer, digitizes the signals and does signal conditioning in the digital domain. After conversion, the data is processed and sent to the PC for training or recognition.

The flex sensors utilized are passive resistive units and are held to each digit of the glove using string (Fig. 1) The sensor changes resistance proportionately to the amount of bend. As the sensor is flexed, the resistance varies from 25–125 KΩ. The sensor is 5.5 cm, which ensures that it reads the bending of only one joint. The flex sensors were used as part of a voltage divider so that it generates voltage values rather than resistance values.

An Inertial Measurement Unit (IMU) is a sensor package that measures linear and angular motion usually with a triad of gyroscopes and triad of accelerometers [3]. The accelerometer generates three signals describing the accelerations along each of its axes produced by, and acting on the device. The gyroscope also outputs three signals that describe the angular rate about each of the sensor axes. IMUs can

output either analog or digital (encoded) signals, and may also include a geomagnetic set of sensors, as well as having a microprocessor to fuse the all sensors signals and directly output motion information. The prototype glove uses an Analog Devices ADXL345 3-axis accelerometer and an ITG3205 3-axis angular sensor (gyro) combination on a single board. The output of these sensors are encoded and sent via I2C communications.

Data acquisition is accomplished through the SBC, which is an Arduino-compatible device. The Arduino is responsible for the collection of data from the different sensors connected to it. It is used to manage the sensor data accurately by using proper timing to decide which of the sensor data will be converted into digital value. After data acquisition, the microcontroller filters all the data before sending it to the computer by serial communication. A static Kalman filter is used to reduce the noise from the accelerometers and compensates for the drift of the gyroscopes over time. A running average algorithm is applied to the ADC readings from the flex sensors.

The vision system utilizes a webcam with a resolution of 640 × 480. It detects the hand position with respect to the signer's body. To reduce image processing requirements, red and blue wrist bands are worn on each hand. The wrist bands serve as high-visibility markers for the vision system; the colors were chosen for best webcam contrast and visibility [4]. The process begins by manually tagging the coordinates of the body of the signer, then determining the Euclidian distance to each hand location.

The system uses the Georgia Tech Gesture Toolkit (GT^2k), which is based on the Hidden Markov Toolkit (HTK), which was developed by the Cambridge

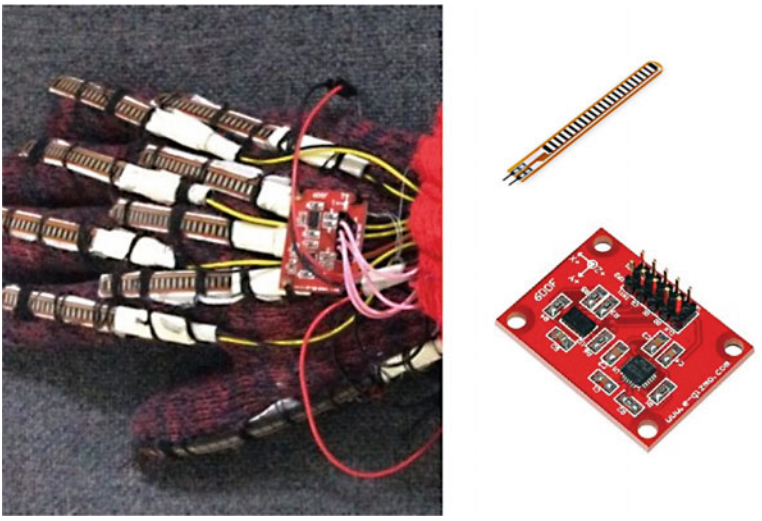


Fig. 1 Left: Prototype data glove, showing sensor placement. A single flex sensor is used for the thumb, while a pair instruments each of the other fingers. Top right—a resistive flex sensor; Bottom right, an Inertial Measurement Unit (IMU) module

University Engineering Department. HTK was originally designed for speech recognition but has found application in many other areas, including speech synthesis and gesture recognition [11].

The feature vectors derived from the glove and video information is used to generate a Hidden Markov Model (HMM). The HMM models are generated using the Baum-Welch algorithm. Each hand gesture has a unique gesture model which consists of the following information: number of states, number of features, state transition probability matrix, observation probability matrix, and initial probability vector. A total of 30 training samples were used for each hand gesture. Each training sample consists of 17 features: Positions of right and left hand, six IMU values, and nine flex sensor values. In addition, the number of states used to represent the gestures in the vocabulary vary from four to seven (4–7), depending on the gesture complexity.

Four states are used for gestures that are static, i.e. the letter "C", as shown in Fig. 2, right; five states gestures are used to represent dynamic gestures (hand attitude/arm movement) which have the same hand form (i.e. fingers stay in the same position) throughout the signing, six states are used for gestures that consists of dynamics which have changing hand form, and seven states for dynamic gestures with multiple movements.

Gesture recognition is accomplished using the Viterbi algorithm—a dynamic programming algorithm that it uses maximization instead of summation at the recursion and termination steps. It also keeps track of the arguments that maximize $\delta_t(i)$ for each iteration of t and i, then stores them in an $N \times T$ matrix Ψ. This matrix

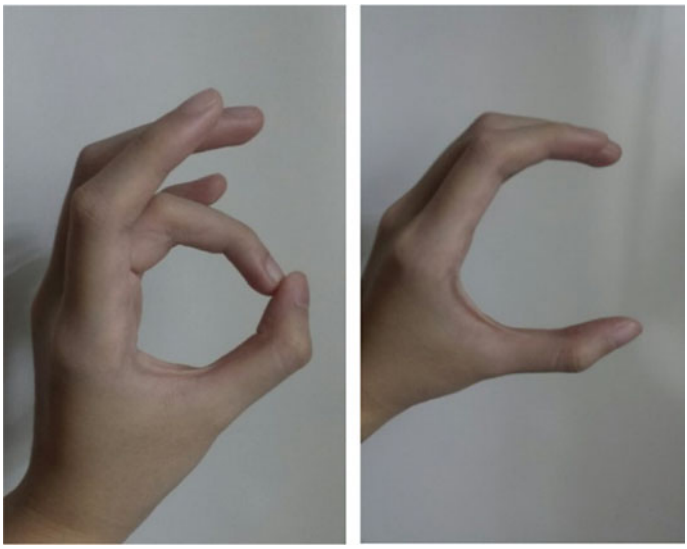


Fig. 2 Hand pose for "7", left, versus "C", on the right. The finger curl and hand attitude are similar for the two symbols

is used to retrieve the optimal state sequence at the backtracking step. The best state sequence S* or the Viterbi Path maximizes the likelihood of the state sequence for a given observation sequence [10].

3 Performance

The system recognition rates were determined by first testing system performance with a limited vocabulary, then on the full vocabulary. The limited vocabulary test was split into two—one where letters A–Z and numbers 0–9 comprised the vocabulary, and the other composed of only the words [6, 7].

For the A–Z and 0–9 test, six untrained signers were invited to perform the gestures five times each under supervision. The data recorded was used to create a set of models using split validation which randomly selects 2/3 of the data for training and the rest for testing. A confusion matrix was generated, showing an overall recognition rate of 71.77%.

The gesture with the lowest recognition rate is the number 7 with a percentage of 10.00% and is often recognized as the letter C. This can be attributed to the similar hand pose and finger positions for the two symbols, as shown in Fig. 2. The flex per finger as measured by the glove for the two symbols is shown in Fig. 3; note the similarity of the shape of each line. It is also visible in the graph for the letter "C" that very mild or nil bending is shown around the knuckles of the pinky, ring and middle fingers. The insensitivity of the system to the bending of the knuckle joints is attributed to how well the glove conforms to the hand of the signer. If the glove is loosely fit, there is a tendency for the glove material to comply more than the sensor itself, resulting in strongly reduced flex values.

Another gesture in which the system gets confused is with the number "2" and the letter "U", with an accuracy of 40%. These two gestures have the exact same

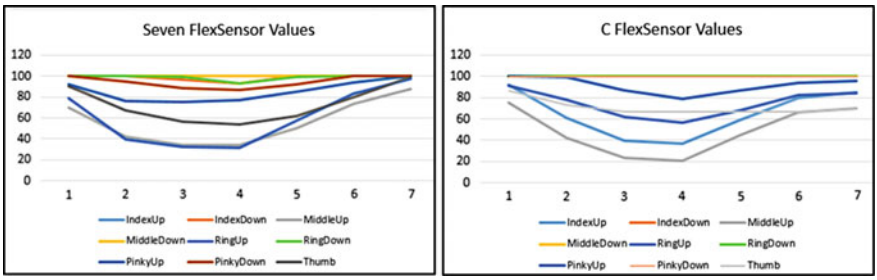


Fig. 3 Flex sensor readings for "7", shown on the left, versus "C", on the right, showing strong similarity in form. There are 9 traces, one for each flex sensor. Vertical axis is the raw ADC value, where higher readings mean less flex

Fig. 4 Hand pose for "2" and "U", left and right images, respectively

hand form, which means that the flex sensor values from the data glove are identical. As seen in Fig. 4, the only difference between them is the index and middle finger positions—number "2" has the fingers apart while letter "U" has the fingers together. The data glove does not have sensors capable of detecting if these two fingers are together or apart, which results in a possibly wrong recognition.

On the other hand, there are multiple gestures which achieve high accuracy such as the letters "E" (100%), "P" (91.7%), "Q" (93.8%), "Y" (85.7%), and "Z" (100%), and the number "4" (92%). The hand forms of these characters and number are unique enough to differentiate from the others.

To characterize the system recognition on a limited vocabulary of 30 words, 10 untrained signers were invited to perform the gestures for the words, three times for each word. Split validation was also used with the same 66–33% ratio. Recognition results have an overall rate of 80.59%. The lowest recognition rate is the gesture for "always" which has a percentage of 28.6%, attributed to its hand form being similar to "tomorrow". The additional downward movement of the hand for "tomorrow" is the same movement one would do at the conclusion of any word, including "always". The IMU values for these two words are shown in Fig. 5.

The words that achieved a high recognition rate are "Cough", "Doctor", "Physical Exam", "Tuesday", "Wednesday", "Saturday", and "Sunday". This is because the length of feature vectors, hand forms, motions, and positions are unique between these words.

A test consisting of the whole vocabulary of 30 words, 26 letters and 10 numbers was conducted which involved the participation of four signers who do not have expertise in the field of sign language but were proficient enough to be acknowledged to have correct signing. The system was able to accurately recognize the vocabulary 79.44% of the time. It can be noted that there are gestures, letters and numbers which were successfully recognized at a rate of 90% and above. Similar to tests with neophyte signers, the gesture for "always" garnered the lowest percentage of accuracy.

Despite the larger vocabulary, the resulting accuracy of recognition for this test is higher compared to the limited vocabulary of only letters and numbers signed by neophytes. The accuracy of recognition for the larger vocabulary compares

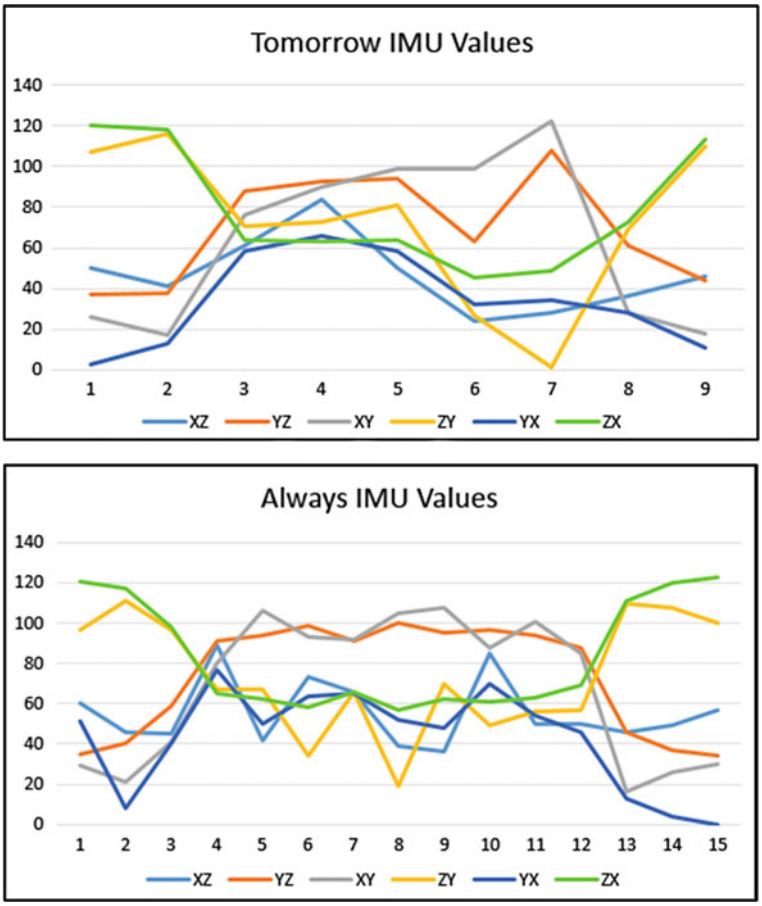


Fig. 5 "Tomorrow" versus "Always" IMU values, as digitized by the Arduino

favorably with the words-only tests with neophyte signers. This reinforces the notion that consistency of signing can allow bigger vocabularies without penalty in recognition rate.

The specificity of the answers expected from the questions for which this system was developed can be used to get correct recognition, if candidate words are presented by the system. Contextual information can often be used to determine which of two possible recognized words or symbols is the actual signed one (i.e. the answer of "2" or "U" to a question like "how many times did your stomach ache today"). Recognizing this, the system can be set to output the two most likely candidates for the gesture performed—this results in a slightly better recognition rate of 80.55%.

4 Feature Reduction

The importance of each sensor to the performance of the system was determined by removing that input from the input vector, retraining the models and rerunning the recognition tests.

Table 1 shows how well the system performs when a video or IMU information is removed. Compared to the system's performance with complete features, accuracy unsurprisingly decreases when a feature is removed.

Removing video from the system the system results in a higher accuracy compared to the system without the IMU. This shows that having the IMU is more descriptive than the imaging because the IMU not only shows the orientation of the hand but also the acceleration, while the video processing is limited to only determining the position of the hand versus the signer's face.

Table 2 shows that gestures that are signed with both hands have higher recognition rate when the vision system is present (see No IMU column) except Allergy, Blood Pressure, Pain and Temperature. These gestures can be misrecognized because they have a more complex movement therefore require the IMU to determine hand orientation.

Similar tests were run with each of the flex sensors removed (Table 3); the flex sensor number and their locations are shown in Fig. 6. From the results, removing a certain flex sensor can cause individual gesture's recognition rate to unexpectedly increase, however this is balanced out by a decrease in some other gesture—removal of any flex sensor results in an overall decrease when the full vocabulary is tested, as shown in Table 4.

Table 1 Removal of vision and IMU effect on performance

	No vision	No IMU	No vision and IMU	Complete (Reference)
Accuracy	64.61	58.84	48.56	80.55

Table 2 Two-handed gesture performance without vision or IMU information

	No vision	No IMU	No vision and IMU	Complete
Allergy	90.0	80.0	66.7	85.7
Blood pressure	85.7	72.7	75.0	100
Doctor	100	100	50.0	100
Drug test	25.0	57.1	37.5	20.0
Pain	90.9	58.3	14.3	84.6
Physical exam	75.0	77.8	50.0	100
Sometime	26.7	58.3	20.0	81.8
Temperature	62.5	27.3	20.0	100
X-ray	28.6	75.0	25.0	85.7

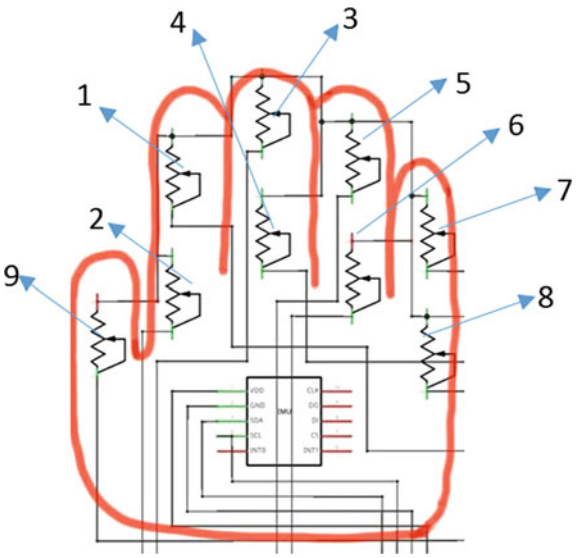


Fig. 6 Flex sensor numbering and location on glove

Table 3 Effect of flex sensor removal on recognition rates for select letters and words

	T	A	H	L	Nurse	Allergy	Doctor
No Flex 1	44.4	53.3	86.7	40.0	83.3	77.8	100
No Flex 2	53.8	71.4	93.8	80.0	73.3	88.9	100
No Flex 3	80.0	50.0	66.7	36.4	50.0	77.8	100
No Flex 4	33.3	71.4	72.7	55.6	75.0	100	100
No Flex 5	80.0	69.2	100	77.8	83.3	75.0	100
No Flex 6	100	85.7	100	62.5	53.8	90.0	100
No Flex 7	50.0	57.1	100	70.0	80.0	90.0	100
No Flex 8	81.8	55.6	85.7	44.4	87.5	100	100
No Flex 9	75.0	28.6	100	40.0	100	91.7	100

Gestures like the letter “H” and “L” still had a high recognition rate which is at 93.08 and 80% respectively because these two letters possess unique hand forms. Other gestures such as “Doctor” and “Allergy” are not affected by the removal of a

Table 4 Flex sensor removal impact on overall recognition

Flex sensor# removed	1	2	3	4	5	6	7	8	9
Accuracy %	71.25	70.23	70.79	70.53	74.96	73.78	73.25	72.93	69.27

one flex sensor. This is due to the gestures being relatively unique that taking a single flex sensor would barely change the hand form as seen by the glove. From the tests, the thumb flex sensor appears to be the most significant—its removal results in the recognition accuracy dropping to 69.27%.

5 Summary and Future Work

SIGMA combines a data glove housing a 6-DOF IMU and nine flex sensors with image processing. Information is combined to derive more accurate hand pose over time, favoring direct reading to reduce computational cost. With these sensors, the system is able to recognize the gestures 79.44% of the time for a limited vocabulary of 26 letters, 10 numbers and 30 healthcare-related words signed in Filipino Sign Language.

The overall system is expected to be enhanced in several areas if it is to cope with a larger vocabulary. The data glove can be improved further by adding sensors to measure sideward movement of fingers, or in a simpler form, a means by which the system can tell if two fingers are touching, i.e. a simple switch. The sensor can be placed in between the index and the middle finger or at the tips of the finger. More advanced IMUs are now on the market, providing true motion representation by combining the sensors with a motion computation engine, i.e. the MPU6050. Improvements can also be made in the recognition of alphabets.

One of the limitations of the current system is the need to sign one symbol at a time, i.e. having to stop to "punctuate" the end of a gesture. Continuous gesture recognition for alphabets can be achieved by comparing incoming data flow with the set of pre-recorded gestures by using two types of outputs: likelihood = how close the performed gesture to the trained model, and time index = the time progression of the performed gesture. Through continuously comparing data to the pre-recorded gesture using a moving window, HMM will be able to classify an alphabet per window. Looking at the likelihood, it could help identify if it is an alphabet or not. Since the likelihood parameters will be continuously updated, it would be possible to recognize a gesture immediately, without waiting for its completion. In addition, by continuously capturing data, the signer would not be restricted to go back to the starting position each time a letter is signed.

More robust modeling, coupled with other Machine Learning (ML) techniques are might probably reduce data representation requirements while still increasing vocabulary size and recognition accuracy. These software improvements however

will require significantly more computing power, thus a tradeoff between this and host computer requirements will always have to be made.

Acknowledgements We would like to thank Maria Elena Lozada and John Xandre C. Baliza of De La Salle—College of Saint Benilde School of Deaf Education and Applied Studies for their tutelage of Filipino Sign Language.

References

1. Abuan, M. (2009). Calls made for a national language for the deaf. The Carillon [Online]. Available http://archive.is/mrGi4.
2. Aguilos, V., C. Mariano, E. Mendoza, and J. Orense. 2007. *A portable letter sign language translator*. Manila: De La Salle University Library.
3. Aoki, N., K. Uda, T. Kiuchi, and T. Fukui. 2008. Impact of miscommunication in medical dispute cases in Japan. *International Journal for Quality in Health Care* 20 (5): 358–362.
4. Bose, A. 2013. How to detect and track red objects in live video in MATLAB Arindam's Blog, November, 10 2013. [Online]. Available http://arindambose.com/blog/?p=72.
5. Krishna, R., S. Lee, S.P. Wang, and J. Lang. 2012. Sign language translation. [Online]. Available: http://people.ece.cornell.edu/land/courses/ece4760/FinalProjects/s2012/sl787_rak248_sw525_fl229/sl787_rak248_sw525_fl229/index.htm.
6. Martinez, L. 2004. *An introduction to Filipino sign language*. Manila: Philippine Deaf Resource Center.
7. Martinez, L. 2012. Primer on Filipino sign language. [Online]. Available http://opinion.inquirer.net/41909/primer-on-filipino-sign-language. (Accessed October 22, 2014).
8. Philippine Statistics Authority. 2000. Census of population and housing, special report on persons with disabilities. [Online] Available https://psa.gov.ph/sites/default/files/2000CPHSpecialReportonPersonWithDisability.pdf. ISSN 0117-1453.
9. Registry of Interpreters for the Deaf. 2007. Interpreting in health care settings. [Online]. Available https://drive.google.com/file/d/0B3DKvZMflFLdYVVBd0RIWDlOMW8/view.
10. Rodriguez-Fuentes, L.J., and M.I. Torres 2003. Comparative study of the Baum-Welch and Viterbi training algorithms applied to read and spontaneous speech recognition. In *Pattern Recognition and Image Analysis, First Iberian Conference*, IbPRIA.
11. Westeyn, T., H. Brashear, A. Atrash, and T. Starner. 2003. Georgia tech gesture toolkit: Supporting experiments in gesture recognition—International conference on perceptive and multimodal user interfaces.
12. Zimmerman, T., J. Lanier, C. Blanchard, S. Bryson, and Y. Harvil. 1987. A hand gesture interface device. In *Conference on Human Factors in Computing Systems and Graphic Interface*, Redwood City.

Design and Experimental Demonstration of a Mechatronic Solution for Endovascular Catheters

P. N. Brett, X. Du, M. Z. Assadi, F. Rodriguez y Baena, F. Liu, R. Hinchliffe and M. Thompson

Abstract This paper describes a mechatronics approach that provides vascular surgeons with the perception of movement and tissue interaction in the vicinity of the tip of a catheter in endovascular procedures. The current system described is experimental and used in phantom units. It integrates 3D visualization generated from scan with real-time tactile sensing in the vicinity of the tip of the catheter to update on the nature of tissue interaction, the curvature and relative orientation of the catheter sleeve and guide wire. This approach offers superior perception by the clinician, in contrast with current application of catheters used in this application. By being well informed of conditions at the working environment of the catheter tip the clinician will be able to administer therapies with greater precision in the surgical task and within a reduced operating time. The approach will reduce risk for patients and significantly reduce risks for the clinician, who is currently exposed to high doses of ionizing radiation during the process of catheter guidance.

Keywords Smart robotic device · Surgical · Catheter

P. N. Brett (✉)
University of Southern Queensland, Queensland, QLD 4350, Australia
e-mail: peter.brett@usq.edu.au

X. Du · M. Z. Assadi
Brunel University, London, UK

F. Rodriguez y Baena
Imperial College, London, UK

F. Liu · R. Hinchliffe
University of Bristol, Bristol, UK

M. Thompson
St. George's Hospital, London, UK

J. Billingsley and P. Brett (eds.), *Mechatronics and Machine Vision in Practice 3*,
https://doi.org/10.1007/978-3-319-76947-9_18

1 Introduction

Endovascular techniques are frequently used to treat cardiovascular, neurovascular peripheral vascular diseases and offer far less invasive therapy than in open surgery. In 2012 there were a total of 92,445 percutaneous coronary interventions (PCI) within the UK alone [1]. For the clinician, the current approach is hampered by complexity of techniques that enable perception by the clinician using 2D fluoroscopic X-ray imaging to assist navigation. The use of ionising scanning is not acceptable for the operator who is exposed frequently, and the visual presentation is neither ideal nor clear to interpret. In the visual presentation, the guide wire is visible and, in some instances the catheter is also visible. A good deal of skill is exercised in the interpretation. Greater clarity will be expected to lead to more precise, safer and shorter procedures.

Remedies avoiding these aspects have been sought. The robotic systems from Corindus Vascular Robotics [2], and the Magellan Robot system of Hansen Medical [3], are currently experimental systems for endovascular surgery that track the placement of the catheter in the vasculature and attempt to reduce frequent use of X-ray scanning. Schemes to assist tracking that use ultrasound [4], electromagnetic elements [5] and force control [6] are advancing with further opportunity, and more recently sensing and actuation approaches to minimise tissue interaction are under investigation. The attributes of an ideal system are single use within an acceptable cost frame, applicable to the wide range of catheter sizes, and readily usable without excessive training and set-up overhead. The range of cross-section size of catheters is important as currently actuation at a small scale within 1–3 mm diameter is proving difficult to achieve in practice.

In the work described here the experimental demonstration system uses three strain sensing elements to discriminate catheter and guidewire behaviour, and types and positions of tissue interaction. Combined with a 3D visualisation of the vasculature, it offers potential to give clear perception of the working environment and interaction with tissues in the vicinity of the tip of the catheter. This demonstrates its use in planning motion at critical stages. The approach used offers potential to deal with perturbations in the 3D projection of the catheter at the tip to account for relative motion between the tip of the catheter and tissue boundaries. It is able to define motion of the cable relative to the sleeve which may have changed relative orientation during the process of placement in the vasculature. These attributes can reduce the number of stages where x-ray scanning is necessary and provide the surgeon with a clear view in which to operate with the catheter device.

The scheme for sensing, that is used in the work described can be applied as a passive system where the surgeon imparts motion carrying out the normal decisions based on clear information on tissue interaction and positioning of the tip relative to target tissues. The system can be configured to reduce velocities imparted by the operator, or to freeze motion at critical stages. By devising a system that looks and feels the same as a current catheter the expectation is the unit will be readily adopted for use by vascular clinicians without the requirement of extensive resource

to support the activity. The protocol in the operating room will be similar. Single use is a real possibility as the catheter embodies few sensing elements and is expected to retail in a realistic range for health provision budgets. Taking this approach, the system can be realistic for widespread use and benefit for procedures across the normal range of catheter sizes.

2 Experimental System

The system and process is illustrated in the schematic of Fig. 1. The clinician enters the vasculature in the normal approach and is in control through a handset enabling the advance and rotation of the catheter and guidewire. The handset can automatically restrain or prevent further motion according to the state of the working environment at the tip. This is to maintain safe operation and, for example, avoid circumstances leading to penetration. A 3D rendered internal image of the vasculature is generated from initial scan data and is used to place the catheter such that its position and behaviour in the vicinity of the tip is illustrated in real-time [7], showing interaction with tissues. Given this information it is possible to compute various visual aspects of the working environment to aid planning of motion of the catheter tip, and to advise the operator how to move the guide wire in order to access the required precise trajectory. Important when inserting into a branch vessel. 3D views such as lateral, target, cockpit each have relative advantages to control in different circumstances. Three strain sensing elements placed in the sleeve of the catheter near the tip are used to discriminate motion and curvature of the catheter and guide wire, and interaction types with the tissues in 3 dimensions in real-time. The distributive tactile sensing scheme used [8] works well in these circumstances in real-time. Feedback to the operator is both visual and audible. Colour coding illustrates no-tissue contact, rubbing, and critical or unsafe configurations and circumstances. The operator is offered guidance to control the catheter sleeve and guide wire to follow the required trajectory and to reduce interaction that if occurring.

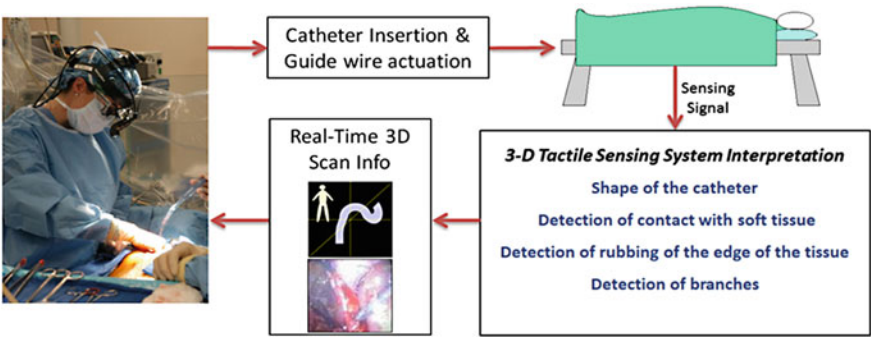


Fig. 1 System schematic

In the experimental set-up for demonstration, rapid-prototyped static phantom units were produced to represent the environmental circumstances of bends, vessel branches and changes in vessel dimension or shape, and the goals encountered in placing stents in vessel branches. The perception and control by surgeon operators is considered the essential factor at this stage of advancing the system design. Experiments investigated operator response to conditions presented.

A typical set-up is shown in Fig. 2, where the operator is guided by the 3D visualization screen image. Here the challenge was to identify the vessel branch using 3D visualisation, and to follow guidance indicators on screen to then move the catheter sleeve, and synchronously both orientating and feeding the guidewire. The combination of real-time derived tactile and positional sensory information with pre-operative scan data enables automatic update on perturbations in scan data. Contact between catheter tip and tissues is determined using the tactile approach described earlier and served to discriminate sliding as opposed to penetrative pose at the tip, the presence of edges, at the position of a branch, and the orientation of the guidewire within the sleeve. Planning of wire trajectory onto the new vessel axis of a branch was also demonstrated with ease.

The visualization component is built on the VTK open source library (Kitware Inc). As shown in Fig. 3, most of the screen space is allocated to CT/MRI imaging. New functionality of the smart catheter is imported to the system through a plugin mechanism. A user interface widget allows user to control the catheter, which docks at on left side. Currently, the system can visualize the CT and catheter, interactive 3D camera movement, and real-time catheter tracking animation. It also includes 3D visualization of the catheter, user commands to define a target position, distance to the target and a first person view of the catheter. The graphical front end provides a colour coded catheter display that changes to convey information on collisions and is accompanied by prompting sounds, bending behaviour, rubbing and general risk in an intuitive and minimalistic way. Planning information, as defined by the surgeon prior to surgery or at any point during the procedure, is displayed as overlay on screen to help guide the manual insertion process.

Fig. 2 The mechanical experimental set-up

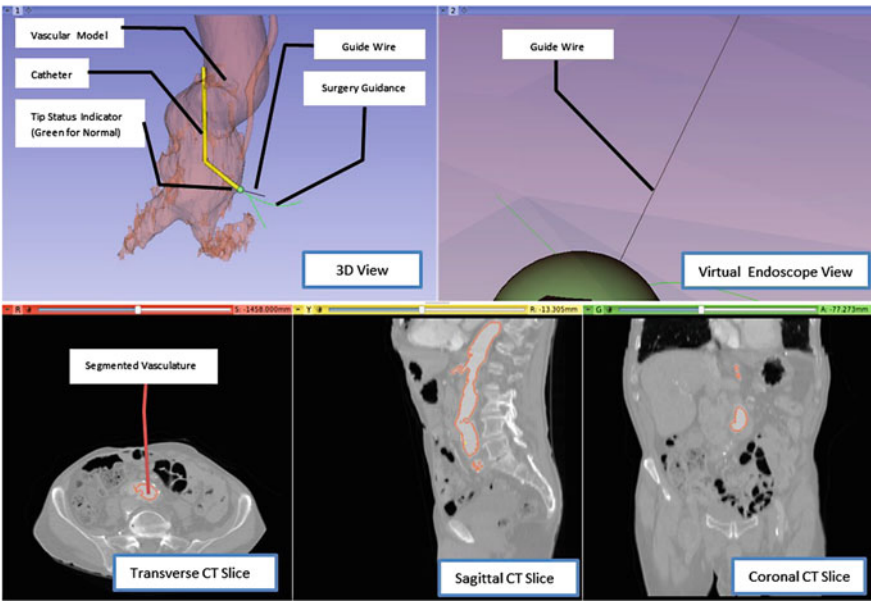


Fig. 3 The real time user graphic interface

Using this approach in laboratory phantoms has demonstrated that the aims of the system can be achieved. Mechanically it is a relatively simple construction that aligns with existing manufacturing processes. Currently the catheter has been constructed as a composite fabrication at 3 mm in diameter. With robust manufacturing processes, and through the principles used, it envisaged that smart catheters can be constructed over a scale working from 1 mm in diameter upwards.

3 Conclusion

The work presented demonstrates how mechatronics as an approach can be used to design an enhanced Catheter system for endovascular processes. The mechanical design is relatively simple, and relies on embedded software and visual feedback to enhance the perception of the surgeon of the process.

Experimental evaluation of the approach has been by phantom units, testing surgeon perception and capacity to interpret navigation cues. In contrast with complex, and expensive, systems capable of assisting in complex endovascular processes, this system is aimed at the majority of processes, such as the placement of stents. The ability to navigate into branches is relatively straightforward and is achieved with little training as there is familiarity with the process. In practice it is expected that the eventual system will be in a similar cost range to existing

catheters. This is an important factor as the system is to provide the benefits of robotics technology to the majority of endovascular processes enabling precision and consistency of the result without inhibitions of high training overhead, set up time and prohibitive costs. Technology is being used already to enhance delivery while providing a system that will both enhance and reinforce surgeon skill.

References

1. National Audit of Percutaneous Coronary Interventional Procedures. 2012. *National Audit of Percutaneous Coronary Interventional Procedures Public Report Annual Public Report*, British Cardiovascular Intervention Society.
2. Kapur, V., N.R. Smilowitz, and G. Weisz. 2014. Complex robotic-enhanced percutaneous coronary intervention. *Catheterization and Cardiovascular Interventions* 83(6): 915–921.
3. Riga, C.V., C.D. Bicknell, A. Rolls, N.J. Cheshire., and M.S. Hamady. 2013. Robot-assisted fenestrated endovascular aneurysm repair (FEVAR) using the Magellan system. *Journal of Vascular and Interventional Radiology*, 24(2):191–196.
4. Kesner, S.B., and R.D. Howe. 2014. Robotic catheter cardiac ablation combining ultrasound guidance and force control. *The International Journal of Robotics Research* 33 (4): 631–644.
5. Adeline, S., B. Kramer, P. Chinnadurai, S. Walker, M. O'malley, A. Lumsden, and Bismuth, J. 2016. Flexible robotics with electromagnetic tracking improves safety and efficiency during in vitro endovascular navigation. *Journal of vascular surgery*.
6. Hedyeh, R., C.J. Payne, and G.-Z. Yang. 2014. Current and emerging robot-assisted endovascular catheterization technologies: a review. *Annals of biomedical engineering* 42 (4): 697–715.
7. Fangde, L., C. Burrows, and Rodriguez, y Baena F. 2013. Deformation-as-control for a biologically inspired steerable needle. In *Robotics and Biomimetics* (ROBIO), 2013 *IEEE International Conference*, 848–853. Piscataway: IEEE.
8. Tam, B., X. Ma, D.J. Webb, D.J. Holding, and Brett, P.N. 2010. Discriminating contact in lumen with a moving flexible digit using fibre Bragg grating sensing elements. In *Proceedings of the Institution of Mechanical Engineers, part H: journal of engineering in medicine* 224 (6): 765–774.

Unsupervised Habitual Activity Detection in Accelerometer Data

Carolyn Domingo, Solomon See and Roberto Legaspi

Abstract The activity of the user is one example of context information which can help computer applications respond better to the needs of the user in a seamless manner based on the situation without needing explicit instruction. With potential applications in many fields such as health-care, assisted living and sports, there has been considerable interest and work done in the area of activity recognition. Currently, these works have resulted in various successful approaches capable of recognizing common basic activities such as walking, sitting, standing and lying, mostly through supervised learning. However, supervised learning approach would be limited in that it requires labeled data for prior learning. It would be difficult to provide sufficient amounts of labeled data that is representative of free-living activities. To address these limitations, this research proposes motif discovery as an unsupervised activity recognition approach. Habitual activities would be detected by finding motifs, similar repeating subsequences within the collected accelerometer data. A 3D accelerometer sensor worn on the dominant arm is used to record, model and recognize different activities of daily living. The raw accelerometer data is then processed and discretized in order to perform motif discovery. Results have shown motif discovery to increase the performance in varying degrees (5–19%) depending on the discretization technique used.

C. Domingo (✉) · S. See
De La Salle University, Manila, Philippines
e-mail: carolyndomingo@gmail.com

S. See
e-mail: solomon.see@gmail.com

R. Legaspi
Transdisciplinary Research Integration Center, Research Organization of Information and Systems, Tokyo, Japan
e-mail: roberto.legaspi@gmail.com

J. Billingsley and P. Brett (eds.), *Mechatronics and Machine Vision in Practice 3*,
https://doi.org/10.1007/978-3-319-76947-9_19

1 Introduction

Activity Recognition is an active area of research that involves recognizing the actions and goals of one or more agents from a series of observations [2]. An action is defined as a single person's activities that may be composed of multiple gestures or atomic movement organized temporally, such as 'walking', 'waving', and 'punching'. As computers become more prevalent in that they appear in many facets of our lives, being embedded into our phones, cars, clothing and homes, user's activity provides important context information that can help enhance the ability of these pervasive systems to properly react and adapt to the current situation.

This research concentrates on providing a model for detecting habitual patterns in a user's activity, which can be served as an activity diary. The diary would show activities that the user performed during the day, along with details such as the time it was performed and the duration of the activity. This would increase the user's awareness of his daily activities and promote a better lifestyle. There are several applications that can be built upon this activity diary. Finding the habitual activity of a person can be utilized in assisted living systems, providing supervision by monitoring a person's activity and detecting any anomalies that might occur for emergency help services. In healthcare, patients are often required periodic evaluation to monitor their condition, but these visits only offers doctors only a glimpse of the user's state. Many research designs and clinical studies can benefit from knowledge about the daily activities of subjects. Compare to self-reporting, a habitual pattern recognition system would ideally recognize the user's activity and its duration in a manner that is less intrusive. Not only can remote healthcare monitoring be applied for early diagnosis, but also for rehabilitation. There are many applications wherein habitual pattern recognition can be used to increase the lifestyle quality of people.

While activity recognition has for many years been treated as a computer vision problem. With the advancement in the semiconductor industry, sensors have become more low-cost, low-power, and much more prevalent. Sensor-based activity recognition exploits the emerging sensor network technologies to monitor an actor's behavior along with their environment. Currently, a wide range of sensors, including contact sensors, RFID, accelerometers, audio and motion detectors, to name a few, is available for activity monitoring.

Wearable sensors mounted on the person would be the most direct method for recognizing human activity as it allows for continuous recording of activities across different locations and is independent from external infrastructure. Several researches [5, 15, 19] conducted have all shown that motion or inertial sensors to consistently perform better compare to other sensors. Accelerometers, in particular, proved to be especially popular among many activity classification systems as it was shown to be the most information-rich and most accurate [19]. They react fast to activity changes and reflect well the type of activity being performed. In fact, most current generation smartphones have come equipped with precise and

reliable accelerometers capable measuring movements and tilt relative to gravity. The availability of such devices brings about an exciting opportunity for accessible data mining tool for a wide audience.

2 Review of Related Literature

2.1 Supervised Learning

If these situations to recognize are known a priori, the problem can be tackled as a classification task. Supervised learning requires the use of labeled data upon which an algorithm is trained. Following training, the algorithm is then able to classify unknown data.

Physical activities such as sitting, standing, walking and running are some of the most common activities that are performed by people and were used for several researches. Classification models, when trained and evaluated using data collected with a single accelerometer, have shown promising performances with high success rates: 94–100% [16], 93% [4], 90.8% [11].

For more complex activities, research by Bao and Intille [3] seeks to recognize common everyday activities. Aside from simple activities such as standing, walking and sitting, it also included complex activities such as folding laundry, brushing teeth, watching TV and vacuuming. The data was collected from 20 subjects performing 20 everyday activities under a semi-controlled setting. 5 sensors were used, and are attached to the right hip, dominant wrist, non-dominant upper arm, dominant ankle, and non-dominant thigh. Result shows an accuracy of 80%.

In another research [21] that also tries to recognize daily life activities with wearable sensors. Two dataset were used in the study. The first had 8 labels (computer use, watching tv, grooming, phone use, hygiene, meal preparation, eating, dishwashing) with data recorded from accelerometers worn on the wrist, hip, and thigh. The second dataset has 20 activity labels (e.g. having lunch, having dinner, picking up cafeteria food, sitting/having a coffee, washing dishes, washing hands, personal hygiene, using the toilet) with data recorded from accelerometer on the wrist and hip pocket. Using SVM, an overall accuracy of 71.3% was reached in the first dataset while 76.3% was reached in the second dataset.

In [6] meanwhile concentrated on recognizing sports exercises, which are walking, running, Nordic walking, football & basketball & floor ball, spinning & cycling, gym training, roller skating, tennis & badminton, and aerobics. 12 subjects wore accelerometers attached to the hip and wrist along with a GPS device. The research collected data under laboratory and out-of-laboratory setting. Under laboratory setting, a 90% accuracy was achieved, while a 72% was achieved outside laboratory setting.

In all these works, the systems are only capable of detecting activities the models were trained with. For the purpose of an activity diary, it would be difficult to know

beforehand all the activities that will be needed for training. When a new activity is encountered, additional labeled activity sets would have to be collected to adapt or retrain the models. It is impractical to train a model for all possible free-living activities. Annotations can also be very subjective, and it's difficult to distinguish activity due to the complexity of human behavior. Activity labels would have to be carefully chosen since the way people view the activity can differ from the patterns the machine would find. This is why domain experts are sometimes needed the help design the system. Inaccurate annotations could end up hurting the performance of the system. Supervised activity recognition systems are too constrained and difficult to scale up.

2.2 *Unsupervised Learning*

In an unconstrained environment, it is quite natural that the input signal can consist of as of yet unseen observations and therefore not accounted for in the training phase. It is unsuitable to put a constraint that the subject will always performs movements that were already seen in the training phase. This is one of the severe limitations in supervised learning.

Unsupervised Learning is much less predominant in Activity Recognition as it needs to identify new unknown actions as accurately as possible. But when implemented properly, it can offer a more flexible support.

Some works [13, 20] made used of Subsequence Time Series clustering (STS) clustering, which is a popular pattern discovery techniques for time series. STS clustering employs a clustering technique like k-means clustering to the subsequences of time series which are generated using a sliding window technique. In [20], subject were asked to perform nine sport activities for ten minutes each, following a predetermined route. An EM clustering algorithm was used to form homogeneous groups of the different sports activity. 85% accuracy was achieved through this method. In [13], a Gaussian Mixture Model (GMM) was used to detect abnormal activities. Data was collected by volunteers who were asked to perform 20 instances of falling over the collection period. Results showed a low false positive rate of 6–7 s per day. However, clustering algorithm does not take into account the temporal pattern of the data. Although these previous researches were able to report high performances, it might be due to the dataset they used. The data collected in [20] was from a specific set of activities collected under a controlled setting, while in [13] only focused on detecting falls.

In other works, [18] uses a Hidden Markov Model (HMM) to cluster free-living activities and detect unusual events. The dataset was a selected detailed hour from a single user performing free-living activities which included sleeping, reading, exercising, driving and others. A best overall purity score of 0.81 was achieved. Another work [22] using HMM reached an accuracy of 95.22% for detecting a six activities performed in sequence (standing, jumping running, walking, stair up, stairs down). The same author conducted another research [23] with similar dataset

and approach with the only difference being that a few activities (sitting and laying on the ground) was added. The HMM was able to achieve a 91% accuracy. The method used in these researches seems much more suitable for activity recognition since the HMM algorithm is a temporal statistical model. However, all these works had the number of clusters and duration of activities known beforehand when building the model, and the activities that were recognized were limited. Their method was not done completely in an unsupervised manner, but instead semi-supervised.

The few works done in unsupervised activity recognition mostly made use of simplified scenarios, defining only a small set of activities to be recognized. The STS clustering would group together similar patterns together, but these patterns would not necessarily belong to a single type of activity. It also does not take into consideration that an activity would be composed of several different motions.

2.3 Time Series Motif Discovery

Motifs is developed primarily in the field of bioinformatics, and refers to sequence pattern that is widespread and has been proven or assumed to have a biological significance. Motif discovery refers to the method of finding these matching motifs. Although the literature on finding motifs in DNA and protein sequences is extensive [9], most of the algorithms look for pre-specified patterns within long sequences. These methods are very fast and efficient, but because they require a pattern to be specified a priori, they are not useful as general tools for detecting all possible motifs in a sequence.

The notion of motifs has carried over to time series data, and has been first defined by Line et al. [14]. Motif discovery is basically the process of finding recurring subsequences in a time series. Time series motif discovery has been successfully applied in fields such as finance, speech recognition, and medicine. Since activities are series of actions or motions arranges temporally, these activities would exhibit a certain pattern that can be found through motif discovery. There was even a research [10] done in activity recognition using the notion of motif to characterize event-subsequences though video was used as data.

3 Experiment

3.1 Data Collection

Subject acceleration is collected using accelerometer from a smartphone attached to the volunteer's dominant upper arm. The sensor data collected will include time and acceleration along the 3-axis (x, y, z) with a sampling rate of 100 Hz.

According to research [12], the upper arm has been identified as the optimal location for an accurate and comfortable wearable body monitor. It has been observed that the upper arm was gender neutral, had a relatively large surface to fit the sensor, low mobility and unobtrusive.

The volunteer was instructed to perform routine activities to perform. Though the details such as how long, how or what order to perform the activities were up to the volunteer. A video recording of the volunteer's session was recorded, and is to be used as reference when evaluating the system. Approximately 5 h of data was collected from the volunteer.

Ideally, the system would be used to detect activities of daily living (ADL) would be composed of activities or tasks that people undertake routinely in their everyday life. These are often used in healthcare as a measurement of the functional status of a person. The ability to perform these specific tasks is indicative of a person's independence.

- Self-care (Grooming, Eating)
- Communication (Writing, Keyboard use)
- Physical Activity (Standing, Sitting, Walking, Carrying, Lifting, Climbing)

An annotator will view the recorded session, and build a time diary of the volunteer's activity. From the video, the following activity labels were used: *walk, offscreen, walk and prepare coffee, prepare coffee, move item, move box and board, walk and move camera, walk and move item, drink, sit, touch face, paperwork, talk, reach down, desktop, adjust phone, walk and move paper, move paper, walk and prepare food, headset, eat, and wipe table*.

Unfortunately, a limitation of video recording was that the user is limited to certain space for data collection. The label *offscreen* typically involves the user leaving the room to the bathroom or retrieving items outside the room.

3.2 Evaluation

The average cluster purity (ACP) and average event purity (AEP) metrics will be used to measure the model's performance in recognizing and clustering the same activities together [1]. The ACP refers to the degree to which a cluster is limited to only one event, while the AEP indicates how an event is limited to only one cluster. These can be computed as follows:

Cluster purity

$$p_i = \sum_{j=1}^{N_e} \frac{n_{ij}^2}{n_i^2} \tag{1}$$

Average cluster purity

$$ACP = \frac{1}{N}\sum_{i=1}^{N_C} p_i * n_i \tag{2}$$

Event purity

$$p_j = \sum_{i=1}^{N_C} \frac{n_{ij}^2}{n_j^2} \tag{3}$$

Average event purity

$$AEP = \frac{1}{N}\sum_{j=1}^{N_e} p_j * n_j \tag{4}$$

where

N Total number of instances
N_e Total number of events
N_C Total number of clusters
n_i Total number of instances in cluster i
n_j Total number of instances of event j
n_{ij} Total number of events j classified into cluster i.

The Average Clustering Event Purity (ACEP) is a measure that combines both ACP and AEP to evaluate the quality of a clustering solution and can be computed as:

$$ACEP = \sqrt{ACP * AEP} \tag{5}$$

This same concept will be used to evaluate the performance motif discovery with each subsequences falling under a motif be treated as a cluster.

4 Algorithm

4.1 Segmentation

Two different window sizes were applied to the data. Based on the 100 Hz sampling rate utilized in data collection, a smaller window size of 200 (2 s) and a larger window size 1000 (10 s) was used with a 50% overlap. Feature can then be extracted from each window frame.

4.2 Feature Extraction

Since activities are composed of several gestures arranged temporally, the raw data (acceleration across the X, Y, Z axes) would have been insufficient to describe the action being performed. Instead, there needs to be a way to characterize or summarize the data within a single frame while preserve useful information as the amount of data is being reduced, because statistical calculations require all points to be available for analysis. Extracting key signal features will enable advanced processing algorithms to discover useful context information. The features that were considered are as follows:

- Mean
- Standard Deviation
- Pairwise Correlation
- Dominant Frequency
- Spectral Energy
- Spectral Entropy

The features were selected based on previous research [3, 21] that worked on detecting activities of daily living.

4.3 Discretization

As this task of unsupervised activity recognition will be approached as a motif discovery problem. The collected data will need to be processed and transformed into a discrete representation that would allow for pattern matching needed in motif discovery. Different method were considered to translate the vector into a symbol. These are namely: EM clustering, semi-supervised classification, and a modified Symbolic Aggregate Approximation (SAX).

4.3.1 Subsequence Time Series (STS) Clustering

Clustering is the process of organizing objects into groups whose members are similar in some way. It is a common method for determining the intrinsic grouping of unlabeled data. Subsequence Time Series clustering (STS) clustering is a popular pattern discovery techniques for time series. STS clustering employs a clustering technique like k-means clustering to the subsequences of a time series which are generated using a sliding window technique.

Clustering was performed using the EM clustering implemented by WEKA, which is k-means clustering with expectation maximization to determine the size of the clusters.

4.3.2 Semi-supervised Classification

In supervised learning, the activities that are to be detected are used to build a model. Instead of collecting and training large quantities of different activities of daily living. A data model of some common basic arm movement is built to discretize the acceleration data into a discrete string, with each symbol of the string representing a basic arm movement. Motif discovery will then later try to find patterns of arm movement that forms an activity.

Originally, 11 classes were formulated based on common arm gesture observed during data collection, but after a short sanity test as explained further in the results section, two less distinctive class (reach down and turn) were removed. This leaves the following 9 classes:

1. Arm down still
2. Arm raised still
3. Arm down move
4. Arm raised move
5. Reach up
6. Swing back and forth
7. Move arm side to side
8. Walk
9. Walk while carrying item

To generate the model of basic arm movements, a decision tree (DT) and support vector machine (SVM) classifier was build using the features mean and standard deviation extracted from each segmented window. The classifier with the best performance, was then used to classify each window, thus discretizing the acceleration data (Fig. 1).

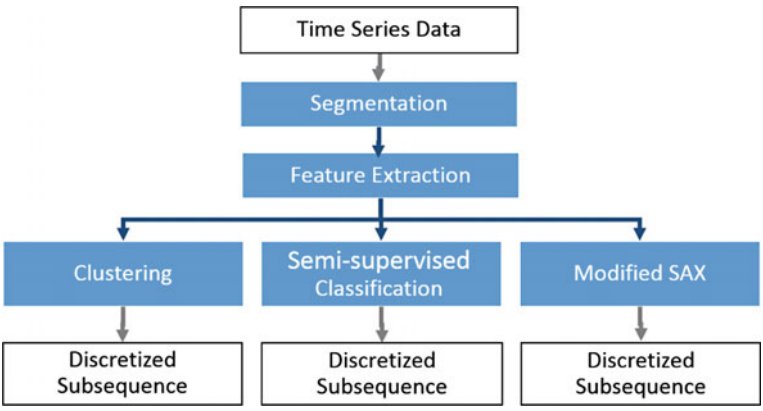


Fig. 1 System flow, from segmentation to discretization

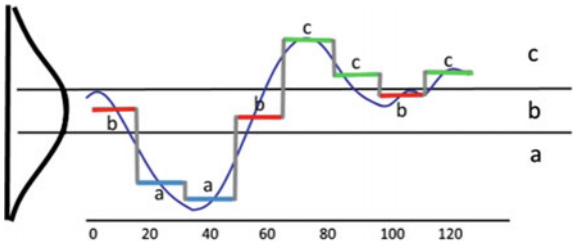


Fig. 2 SAX

4.3.3 Symbolic Aggregate Approximation (SAX)

SAX is a simple dimensionality reduction for time series mining. First, Piecewise Aggregate Approximation (PAA) is performed, where the times series is divided into equal sized segments and the mean value of the points that lie within the frame is computed. SAX then converts the PAA coefficients into symbol string, by determining the number and the location of the breakpoints. This number of breakpoints is related to the desired alphabet size, and locations of these breakpoints are typically placed produce equal size areas under the Gaussian curve [7]. This discrete representation allows researchers to make use of data structures and algorithms in other fields such as bioinformatics or text mining (Fig. 2).

To apply SAX on the collected multidimensional acceleration data, some modifications were made. First the data is normalize, since it was observed that the entire acceleration in the X-axis is shifted lower or higher on different sessions. In addition to the mean, PAA is also performed on the total standard deviation as it has been shown in the results section that the using only the mean as features are not enough to distinguish arm movements. Adding the single feature total standard deviation was shown to improve performance significantly. To keep the alphabet size from getting too large, only one breakpoint is used for each axis and two breakpoints for the standard deviation. So there will be two areas for each axis and three for the standard deviation; assigning a symbol to each permutation of where the PAA coefficient may lie, this leads to an alphabet size of 24 (Fig. 3).

4.4 Motif Discovery

Time series motifs are approximately repeated subsequences of a longer time series stream. When a repeated structure is discovered, it suggests that there may be an underlying reason for why the pattern might be conserved [17] (Fig. 4).

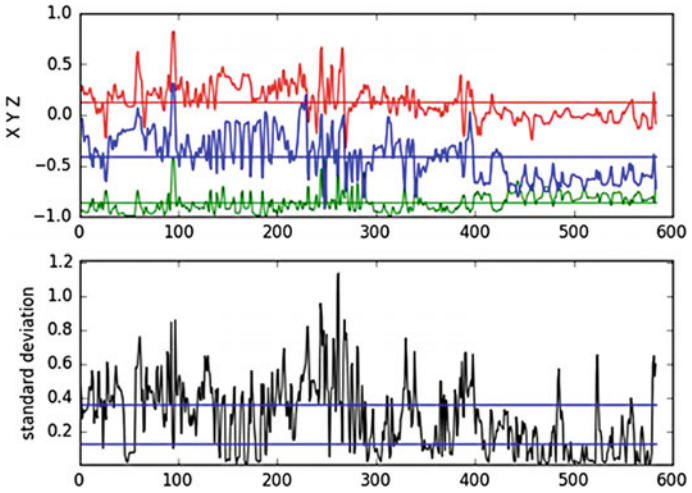


Fig. 3 Mean acceleration along the X, Y and Z axis (top) and the total standard deviation (bottom) representing a 1 h record session with straight lines as the breakpoints

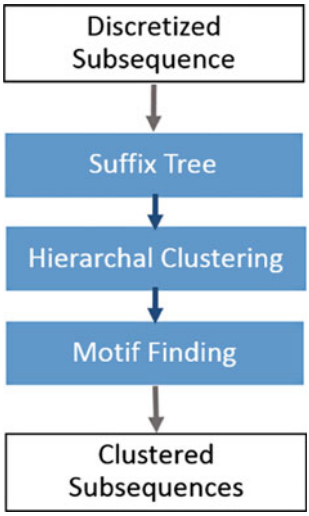


Fig. 4 System flow of motif discovery

4.4.1 Suffix Tree

To perform initial pattern discovery, suffix tree is used to find all recurring common subsequences of a minimum length to serve as a candidate motif.

Suffix tree is a compressed ordered tree with all the paths from the root to the leaves corresponding to all the suffixes of a given text. Suffix referring to substring

containing the last letter of a string. Each edge contains a substring and each leaf contains the position of the suffix. The suffix tree exposes string properties like longest or most frequent substring, allowing for fast implementations of many important string operations [8].

4.4.2 Hierarchical Clustering

The exact matches found in the suffix tree are then put through hierarchal clustering in order to cluster together similar subsequences.

Hierarchical clustering is performed by first computing the distance between each pair of object. Dynamic Time Warping (DTW) is used since activities are temporal sequences which may vary over time or speed.

Once pairwise distance is computed, objects with the smallest distance are considered to be very similar and are merged together. This is done iteratively until a stop criterion of a certain distance or cluster size is reached.

To determine the distance criterion, the data is first clustered to the point that no singleton cluster is left. If one plots the distribution of the distances in the remaining distance matrix, the plot would typically form a concave down pattern as seen in Fig. 5. Since the objective in clustering is to find the most homogeneous clusters that are as distinct as possible from other clusters, the mode in the distribution is used as the distance criterion on where the hierarchical tree should be cut.

Each cluster formed from hierarchical clustering would then represent a motif.

4.4.3 Motif Finding

With the motif now identified, the time series data can now be divided into different activity. Since there could be some overlap when matching subsequences to the ones in the cluster, a simple greedy heuristic is employed to label the temporal data.

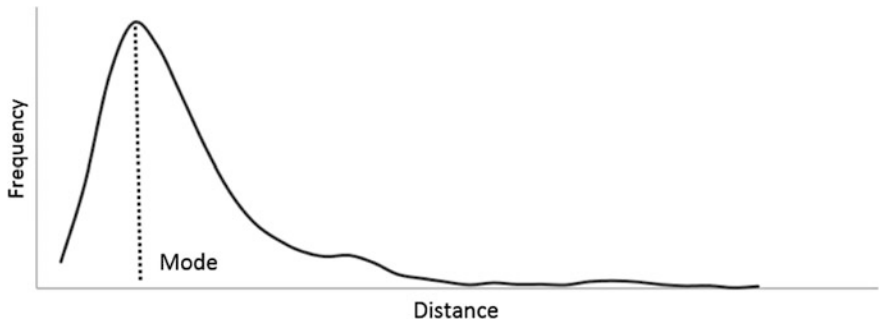


Fig. 5 Distribution plot of distance matrix

With the location and frequency of all the recurring subsequence pattern derived from the suffix tree build earlier, a heuristic is built where the subsequence pattern by sorted by significance are then mapped onto the timeline.

The pattern are sorted by frequency, followed by length, then by position as it has been stated that the most significant motif in a given time series is the pattern with the highest count of non-trivial matches [14].

The cluster number of the pattern is then mapped onto the timeline. Once a pattern is mapped, the algorithm check if the neighboring sequences also fall under the same cluster number, expanding the motif sequence. Afterwards, the algorithms moves onto the next most significant pattern until the entire timeline is mapped or until all the motif patterns are exhausted.

5 Results

5.1 Semi-supervised Classification

Running the data through DT and SVM with tenfold cross validation, the resulting performance was quite high. Showing that the data is collected can be used to detect a person's action or activities.

Though the performed did drop when testing the model using data collected from a different user (Tables 1 and 2).

Closer inspection of the confusion matrix shows that reach down and turn were not very distinct classes. Reach down and walk are always classified together, and

Table 1 Performance of arm data model in tenfold cross validation

	TP rate	FP rate	Precision	Recall	F-measure
DT (2 s)	0.917	0.008	0.917	0.917	0.916
DT (10 s)	0.905	0.010	0.906	0.905	0.905
SVM (2 s)	0.860	0.015	0.861	0.86	0.857
SVM (10 s)	0.829	0.020	0.797	0.829	0.805

Table 2 Performance of arm data model against test set

	TP rate	FP rate	Precision	Recall	F-measure
DT (2 s)	0.502	0.05	0.548	0.502	0.487
DT (10 s)	0.491	0.051	0.473	0.491	0.465
SVM (2 s)	0.527	0.049	0.605	0.527	0.518
SVM (10 s)	0.431	0.060	0.456	0.431	0.392

Table 3 Performance of 9 classes arm data model in tenfold cross validation

	TP rate	FP rate	Precision	Recall	F-measure
DT (2 s)	0.945	0.007	0.946	0.945	0.945
DT (10 s)	0.928	0.009	0.929	0.928	0.928
SVM (2 s)	0.889	0.015	0.891	0.889	0.887
SVM (10 s)	0.866	0.018	0.866	0.866	0.861

Table 4 Performance of 9 classes arm data model against test set

	TP rate	FP rate	Precision	Recall	F-measure
DT (2 s)	0.575	0.051	0.630	0.575	0.565
DT (10 s)	0.551	0.053	0.570	0.551	0.515
SVM (2 s)	0.641	0.046	0.703	0.641	0.610
SVM (10 s)	0.670	0.043	0.731	0.670	0.644

turn is classified to arm down still and walk. The new results of DT and SVM are shown in the Tables 3 and 4.

Removing the non-distinct classes have led to a noticeable increase in the performance, more so by SVM, which increased by 14.4 and 24.2% for 2 and 10 s respectively. Despite having a higher tenfold cross validation performance, DT has very low accuracy when tested against the test set. Therefore, SVM will be used to discretize the acceleration data since it performed best against the test set.

5.2 SAX

SAX originally only makes use of the mean of the window, but a short test run on using only the mean as features to classify the arm movement data from semi-supervised shows a significant drop in performance, especially on the test set where the accuracy all fell below 50% (Tables 5 and 6).

Since using only the mean of each window is not enough to distinguish different arm movement, additional features, specifically standard deviation, were considered

Table 5 Performance of arm data model in tenfold cross validation using mean as features

	TP rate	FP rate	Precision	Recall	F-measure
DT (2 s)	0.834	0.021	0.833	0.834	0.832
DT (10 s)	0.861	0.018	0.862	0.861	0.861
SVM (2 s)	0.661	0.044	0.683	0.661	0.652
SVM (10 s)	0.611	0.055	0.661	0.611	0.579

Table 6 Performance of arm data model against test set using mean as features

	TP rate	FP rate	Precision	Recall	F-measure
DT (2 s)	0.408	0.075	0.428	0.408	0.398
DT (10 s)	0.381	0.076	0.449	0.381	0.340
SVM (2 s)	0.435	0.073	0.429	0.435	0.399
SVM (10 s)	0.495	0.067	0.420	0.495	0.418

Table 7 Performance of arm data model in tenfold cross validation using mean and total standard deviation as features

	TP rate	FP rate	Precision	Recall	F-measure
DT (2 s)	0.927	0.009	0.927	0.927	0.927
DT (10 s)	0.895	0.014	0.895	0.895	0.895
SVM (2 s)	0.871	0.017	0.873	0.871	0.866
SVM (10 s)	0.858	0.019	0.860	0.858	0.854

Table 8 Performance of arm data model against test set using mean and total standard deviation as features

	TP rate	FP rate	Precision	Recall	F-measure
DT (2 s)	0.575	0.051	0.630	0.575	0.565
DT (10 s)	0.551	0.053	0.570	0.551	0.515
SVM (2 s)	0.641	0.046	0.703	0.641	0.610
SVM (10 s)	0.670	0.043	0.731	0.670	0.644

to be included in a modified SAX discretization. This would however increase the alphabet size of SAX too much since if even just dividing each dimension (x mean, y mean, z mean, x stdev, y stdev, z stdev) by two would lead to an alphabet size of 26 or 64.

To minimize the features, the total of the three standard deviation are added together and tested to see if it offers any significant help in distinguishing activities.

Table 9 Purity score summary of different discretization technique and motif discovery

	Discretization step	After discretization (%)	After motif discovery (%)	Difference in purity (%)
2 s	Clustering	30.10	35.09	+4.99
	Semi-supervised	28.27	34.51	+6.23
	Modified SAX	22.78	34.61	+11.82
10 s	Clustering	31.47	38.96	+7.48
	Semi-supervised	29.12	38.37	+9.25
	Modified SAX	23.89	43.00	+19.11

Just by including one feature (total standard deviation), the performance increases significantly with close to the result when using all three standard deviation as seen in Tables 7 and 8. Therefore, total standard deviation was included in the SAX.

5.3 Comparison of Different Discretization Techniques

Table 9 shows the comparison of the different discretization before and after performing motif discovery. The 10 s window always outperforms their 2 s counterpart. This indicates that the 2 s window size might be too sensitive to the movements made by the user, some of which are not very useful to describe the activity. From empirical video analysis, these subsequences formed with 2 s window tend to be smaller only a few minutes at most, whereas activities would usually last much longer. There also wasn't much difference in purity score in the 2 s window.

From among the 10 s, the best performance was from SAX (10 s) with a purity score of 43%. This is despite having a very low score from before motif discovery. The modified SAX in both window size both receive a large increase in performance of 11 and 20% for the 2 and 10 s windows respectively. This suggests that motif discovery is finding useful recurring patterns in the data rather than increasing score from forming larger segments, which would have only given a modest increase in purity.

5.4 Video Analysis

Since annotating the user's activity can be subjective, a change in labels might lead to different scores. Making the evaluated purity scores misleading.

To confirm the performance of the system and purity score evaluation are reliable, video analysis is performed. The resulting clustered subsequences from motif discovery are reviewed and re-annotated. This would of course lead to a higher

Table 10 Purity scores when compare against annotated time diary and against re-annotations

	Discretization step	Against time diary (%)	Against re-annotations (%)
2 s	Clustering	35.09	35.90
	Semi-supervised	34.51	38.03
	Modified SAX	34.61	38.49
10 s	Clustering	38.96	40.65
	Semi-supervised	38.37	45.04
	Modified SAX	43.00	46.51

Table 11 Sample of video annotations

Video #	Time	Activity label	Arm movement
Video 1	08:36–08:56	Move item	Arm rest back
Video 1	22:58–23:08	Move item	Arm rest back
Video 3	29:59–30:05	Desktop	Arm rest back
Video 3	30:10–30:17	Desktop	Arm rest back
Video 3	33:40–34:07	Desktop	Arm rest back

purity score as the annotator is likely to give a better fitting labels. However, there should not be much difference in the scores.

Table 10 indeed supports that while a change in annotation can change the performance, the difference is rather insignificant, the new scores remain close to the original scores with only 1–7% difference.

5.4.1 Observations

Some observations that were made from video analysis, are that the occurrence of the somewhat pure clusters, wherein the activities in a single cluster are pure per video. This is likely due to the sensor not being strapped on the exact same way during each recording. Even after offsetting the data to align the data better, it was not enough to counteract completely the sensor displacement.

The routines it detected tend to be short actions like pouring a pitcher or taking a drink. The routines being detected also rely heavily on the arm movement. There were a few clusters that would have been considered very pure when looking solely at the arm movement. An example would be in Table 11.

The accelerometer does not capture enough information to distinguish activities with similar arm movements. For instance, the activity ‘walk and prepare coffee’ was often mixed with other similar activities such as ‘walk and move paper’ or ‘walk and move camera’.

5.5 Benchmark

To determine how well the system compare to using a supervised learning algorithm, a classifier is built using Decision Tree (DT) and Support Vector Machine (SVM) with mean and standard deviation as features.

Table 12 Purity scores when compared against annotated time diary and against re-annotations

	Algorithm	F-measure (%)	ACEP (%)
2 s	DT	63.00	49.85
	SVM	44.30	52.72
10 s	DT	58.50	50.77
	SVM	42.30	53.13

The data is divided into an 80 and 20 split to serve as training and test set. The resulting confusion matrix was then used to the cluster compute cluster and event purity (Table 12).

From F-measure, the best result was from DT (2 s) with 63%; however, its purity score was actually only 50%. The best performance in terms of purity is SVM (10 s) with 53%.

The difference in the two measurement (F-measure and Purity) is because F-measure only concerns itself with which how many of the instances are classified correctly and incorrectly into a class. Meanwhile, in clustering, the clusters are not assigned to a class, therefore there are technically no correctly or incorrectly classified instances. Instead, cluster purity determines how well the cluster was able to separate the data with each clusters mostly containing instances from a single class. The SVM algorithm which tends to create a more generalized classifier was able to get a higher purity score.

Using 53% as a benchmark for the performance, then the 43% from motif discovery does not seem as bad in that regard, considering it was an unsupervised method. Given only the accelerometer data from the upper arm, the activities that can be detected are too limited. Adding more sensors and information would be the best way to increase performance.

6 Conclusion

This research present an approach for unsupervised activity recognition using motif discovery that allow support for variable-length motifs. The dataset consists of the acceleration recorded on the dominant upper arm as the volunteer perform activities of daily living. Video recording of the session was also taken as reference for evaluation.

Before motif discovery, the data was first pre-processed by segmenting and extracting features from each sliding window. Different discretization step were then employed, these are namely: clustering, semi-supervised classification, and a modified SAX. Clustering was the initial discretization method performed.

Motif discovery is performed in three main steps. The first is suffix tree to help find candidate motifs. This is followed by hierarchical clustering using Dynamic Time Warping as the distance measure. This step groups together similar strings. Lastly, motif finding is done to map the motifs onto the timeline.

When the clustering and motif discovery were applied to data, it was shown that clustering was rather low, and that motif discovery improves the data by a few percent. Since motif discovery is shown to improve the purity score, other discretization techniques were implemented and evaluated. Motif discovery is after all highly dependent on the discretization step as it is its input.

Semi-supervised classification was to address the problem of clustering which allowed the clustering to determine the how the data was to be grouped. In semi-supervised classification, a model of basic arm movement is built and is used

guide how data is discretized. Each discrete symbol would then be representative of some basic arm movement, and an activity would then be composed of certain pattern of arm movement. Though the results show a similar performance to clustering, the model can still be further improved with by collecting more data to create a better classifier. More classes would also be needed to build a more general model of basic arm movements, since the ones listed here are just the ones observed being performed from the video. It does not represent all possible arm movement.

Lastly, the modified SAX which discretizes the data based on where the mean and standard deviation of a window lies on the data distribution. It is the simplest method and had the lowest score before motif discovery, but had yielded the best results at 43%.

The output from the first two discretization techniques were ineffective representation of the data. Likely, during discretization, as the data dimension is being reduced, some useful information is being lost. Meanwhile, SAX, which is the most direct translation of the data, provided the best results after motif discovery. Though the modified SAX performed best, further improvements can be made by adjusting the size of alphabet, number and location of breakpoints, or making use of other features.

When supervised learning was performed as benchmark, the purity score from SVM was only 53%, indicating that detecting of activities of daily living using only the acceleration of the upper arm is too difficult. The detected routines are only based on how the dominant arm is moving. It does not distinguish very well actions with similar actions by the upper arm. In order to narrow down more specifically what type of action the user is performing, more information would be needed. Future research may consider adding more sensors as well as including other contextual information such as time and location. When using accelerometer sensor, it is important that the sensor placement be as consistent as possible, even small displacement in the sensor can affect the accuracy of the data.

References

1. Ajmera, J., H. Bourlard, I. Lapidot, and I.A. McCowan. 2002. *Unknown-multiple speaker clustering using hmm*. IDIAP: Technical Representative.
2. Altun, K., B. Barshan, and O. Tuncel. 2010. Comparative study on classifying human activities with miniature inertial and magnetic sensors. *Pattern Recognition* 43 (10): 3605–3620.
3. Bao, L., and S.S. Intille. 2004. Activity recognition from user-annotated acceleration data. In: *Pervasive computing*, 1–17. Springer.
4. Bonomi, A.G., A. Goris, B. Yin, K.R. Westerterp, et al. 2009. Detection of type, duration, and intensity of physical activity using an accelerometer. *Medicine and Science in Sports and Exercise* 41 (9): 1770–1777.
5. Chen, L., J. Hoey, C. Nugent, D. Cook, and Z. Yu. 2012. Sensor-based activity recognition. *Systems, Man, and Cybernetics, Part C: Applications and Reviews, Transactions on IEEE* 42 (6): 790–808.

6. Ermes, M., J. Parkka, J. Mantyjarvi, and I. Korhonen. 2008. Detection of daily activities and sports with wearable sensors in controlled and uncontrolled conditions. *Information Technology in Biomedicine, Transactions on IEEE* 12 (1): 20–26.
7. Fuad, M.M.M., and P.F. Marteau. 2013. *Towards a faster symbolic aggregate approximation method*. arXiv preprint arXiv:1301.5871.
8. Giegerich, R., and S. Kurtz. 1997. From ukkonen to mccreight and weiner: A unifying view of linear-time suffix tree construction. *Algorithmica* 19 (3): 331–353.
9. Gusfield, D. 1997. *Algorithms on strings, trees and sequences: Computer science and computational biology*. Cambridge University Press.
10. Hamid, R., and S. Maddi, A. Bobick, I. Essa. 2006. Unsupervised analysis of activity sequences using event-motifs. In: *Proceedings of the 4th ACM international workshop on Video surveillance and sensor networks*, 71–78. ACM.
11. Karantonis, D.M., M.R. Narayanan, M. Mathie, N.H. Lovell, and B.G. Celler. 2006. Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring. *Information Technology in Biomedicine, Trans-actions on IEEE* 10 (1): 156–167.
12. Kasabach, C., C. Pacione, M. Des, and A. Teller, D. Andre. 2002. Why the upper arm? Factors contributing to the design of an accurate and comfortable, wearable body monitor. In *Whitepaper, Bodymedia, Inc. Citeseer*.
13. Lee, M.S., J.G. Lim, K.R. Park, and D.S. Kwon. 2009. Unsupervised clustering for abnormality detection based on the tri-axial accelerometer. *ICCAS-SICE* 2009: 134–137.
14. Lin, J., E. Keogh, and S. Lonardi, P. Patel. 2002. Finding motifs in time series. In *Proceedings of the 2nd workshop on temporal data mining*, 53–68.
15. Logan, B., J. Healey, M. Philipose, and E.M. Tapia, S. Intille. 2007. A long-term evaluation of sensing modalities for activity recognition. In *UbiComp 2007: Ubiquitous computing*, 483–500. Springer.
16. Mathie, M., B.G. Celler, N.H. Lovell, and A. Coster. 2004. Classification of basic daily movements using a triaxial accelerometer. *Medical & Biological Engineering & Computing* 42 (5): 679–687.
17. Mueen, A., and E. Keogh. 2010. Online discovery and maintenance of time series motifs. In: *Proceedings of the 16th ACM SIGKDD international conference on knowledge discovery and data mining*, 1089–1098. ACM.
18. Nguyen, A., and D. Moore, I. McCowan. 2007. Unsupervised clustering of free-living human activities using ambulatory accelerometry. In *Engineering in medicine and biology society, 2007. EMBS 2007. 29th annual international conference of the IEEE*, 4895–4898, IEEE.
19. Parkka, J., M. Ermes, P. Korpipaa, J. Mantyjarvi, J. Peltola, and I. Korhonen. 2006. Activity classification using realistic data from wearable sensors. *Information Technology in Biomedicine, Transactions on IEEE* 10 (1): 119–128.
20. Siirtola, P., P. Laurinen, E. Haapalainen, and J. Roning, H. Kinnunen. 2009. Clustering-based activity classification with a wrist-worn accelerometer using basic features. In *Computational intelligence and data mining, 2009. CIDM'09. IEEE symposium on IEEE*, 95–100.
21. Stikic, M., D. Larlus, S. Ebert, and B. Schiele. 2011. Weakly supervised recognition of daily life activities with wearable sensors. *Pattern Analysis and Machine Intelligence, Transactions on IEEE* 33 (12): 2521–2537.
22. Trabelsi, D., S. Mohammed, and Y. Amirat, L. Oukhellou. 2012. Activity recognition us-ing body mounted sensors: An unsupervised learning based approach. In *Neural networks (IJCNN), The 2012 international joint conference on IEEE*, 1–7.
23. Trabelsi, D., S. Mohammed, F. Chamroukhi, L. Oukhellou, and Y. Amirat. 2013. An un-supervised approach for automatic activity recognition based on hidden markov model regression. *Automation Science and Engineering, Transactions on IEEE* 10 (3): 829–835.