การจำแนกท่าทางของบุคคล 5 ท่าทางโดยไม่ขึ้นกับฉากหลัง ระยะห่างของกล้องและเครื่องแต่งกาย

นางสาวปิยะรัตน์ ศิลปศุภกรวงศ์

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิทยาศาสตรดุษฎีบัณฑิต
สาขาวิชาเทคโนโลยีทางภาพ ภาควิชาวิทยาศาสตร์ทางภาพถ่ายและเทคโนโลยีทางการพิมพ์
คณะวิทยาศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2555

FIVE-HUMAN-POSTURE CLASSIFICATION INDEPENDENT FROM BACKGROUND, CAMERA DISTANCE AND APPAREL

Miss Piyarat Silapasuphakornwong

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy Program in Imaging Technology
Department of Imaging and Printing Technology
Faculty of Science
Chulalongkorn University
Academic Year 2012

Thesis Title	FIVE-HUMAN-POSTURE CLASSIFICATION INDEPENDENT FROM BACKGROUND, CAMERA DISTANCE AND APPAREL
By	Miss Piyarat Silapasuphakornwong
Field of Study	Imaging Technology
Thesis Advisor	Professor Chidchanok Lursinsap, Ph.D.
Thesis Co-advisor	Associate Professor Aran Hansuebsai, Ph.D.
	Assistant Professor Suphakant Phimoltares, Ph.D.

Accepted by the Faculty of Science, Chulalongkorn University in Partial Fulfillment of the Requirements for the Doctoral Degree

_______________________Dean of the Faculty of Science

(Professor Supot Hannongbua, Dr. rer. nat.)

THESIS COMMITTEE

_______________________Chairman

(Associate Professor Pontawee Pungrassamee, M.S.)

_______________________Thesis Advisor

(Professor Chidchanok Lursinsap, Ph.D.)

_______________________Thesis Co-advisor

(Associate Professor Aran Hansuebsai, Ph.D.)

_______________________Thesis Co-advisor

(Assistant Professor Suphakant Phimoltares, Ph.D.)

_______________________Examiner

(Assistant Professor Chawan Koopipat, Ph.D.)

_______________________Examiner

(Assistant Professor Thanarat Chalidabhongse, Ph.D.)

_______________________External Examiner

(Kingkarn Sookhanaphibarn, Ph.D.)

ปิยะรัตน์ ศิลปศุภกรวงศ์ : การจำแนกท่าทางของบุคคล 5 ท่าทางโดยไม่ขึ้นกับฉากหลัง ระยะห่างของกล้องและเครื่องแต่งกาย. (FIVE-HUMAN-POSTURE CLASSIFICATION INDEPENDENT FROM BACKGROUND, CAMERA DISTANCE AND APPAREL.) อ.ที่ปรึกษาวิทยานิพนธ์หลัก : ศ. ดร. ชิดชนก เหลือสินทรัพย์, อ.ที่ปรึกษาวิทยานิพนธ์ร่วม : รศ. ดร. อรัญ หาญสืบสาย และ ผศ. ดร. ศุภกานต์ พิมลธเรศ, 120 หน้า.

งานวิจัยนี้นำเสนอขั้นตอนวิธีเพื่อรู้จำท่าทางของบุคคล 5 กลุ่มท่าทางจากภาพวิดิทัศน์ เพื่อประยุกต์ขั้นตอนวิธีที่ใช้งานได้จริงในทางปฏิบัติของระบบเฝ้าระวังความปลอดภัย ในที่นี้องค์ประกอบหลายอย่าง เช่น เสื้อผ้า, ปัจจัยเกี่ยวกับกล้อง, ลำดับการทำท่าทาง, และขนาดร่างกายของบุคคลในภาพไม่ได้ถูกควบคุมหรือจำกัด เราได้กลุ่มของลักษณะสำคัญใหม่บนพื้นฐานของข้อมูลเชิงเรขาคณิตของโครงขอบของท่าทางทั้งแบบนิ่งและตามเวลาใหม่ ขั้นตอนวิธีของเราสามารถใช้ได้ทั้งท่าทางนิ่งและท่าทางที่เคลื่อนไหวซึ่งได้แก่ ท่าทางเดิน, ท่าทางยืน, ท่าทางนอน, ท่าทางนั่ง, และท่าทางงอตัว ขั้นตอนวิธีได้ถูกทดสอบสองลักษณะคือ ประสิทธภาพของลักษณะสำคัญที่นำเสนอ ซึ่งเปรียบเทียบกับงานวิจัยของ Juang, C.F. และChang, C.M. (ปี 2007) ซึ่งใช้วิธีระบบเครือข่ายประสาทที่มีความคลุมเครือ ในการแบ่งแยกท่าทางบุคคลเป็น 4 ท่าหลัก (ยกเว้นท่าเดิน) ได้แก่ ท่ายืน ท่านั่ง ท่างอตัว และท่านอน และ 5 ท่าหลักตามลำดับ และการประยุกต์ใช้กับข้อมูลหลายหลาย เพื่อทดสอบสภาพความทนทานขั้นตอนวิธีกลับกลุ่มข้อมูลที่หลากหลาย เราได้ทดลองวิธีการกับภาพวิดิทัศน์จาก 4 ฐานข้อมูล ได้แก่ ฐานข้อมูลท่าทางบุคคลของ Weizmann, ฐานข้อมูลพฤติกรรมที่ซับซ้อนของ UIUC, ฐานข้อมูลของห้องปฏิบัติการการเคลื่อนไหวกราฟิก CMU และฐานข้อมูลของเราเอง เมื่อทดสอบกับชุดข้อมูลของเราเอง ความถูกต้องเฉลี่ยคือ 93.27% แต่เมื่อทดลองวิธีของเรากับข้อมูลชุดอื่นๆ ความถูกต้องของเราสูงกว่า 95%

ภาควิชา วิทยาศาสตร์ทางภาพถ่าย	ลายมือชื่อนิสิต..
และเทคโนโลยีทางการพิมพ์	ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์หลัก........................
สาขาวิชา เทคโนโลยีทางภาพ	ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์ร่วม........................
ปีการศึกษา 2555	ลายมือชื่อ อ.ที่ปรึกษาวิทยานิพนธ์ร่วม........................

5173832323 : MAJOR IMAGEING TECHNOLOGY
KEYWORDS : HUMAN POSTURE RECOGNITION / COMPUTER VISION / IMAGE PROCESSING / SURVEILLANCE

PIYARAT SILAPASUPHAKORNWONG : FIVE-HUMAN-POSTURE CLASSIFICATION INDEPENDENT FROM BACKGROUND, CAMERA DISTANCE AND APPAREL. ADVISOR : PROF. CHIDCHANOK LURSINSAP, Ph.D., CO-ADVISOR : ASSOC. PROF. ARAN HANSUEBSAI, Ph.D. , ASSIST. PROF. SUPHAKANT PHIMOLTARES, Ph.D., 120 pp.

A five postures recognition algorithm from a human video was proposed. To make the algorithm practical to real surveillance applications, several contexts concerning clothing, camera factors, sequence of acting postures, and subject's body size are not neither constrained nor controlled. A new set of features based on only the geometrical information of stationary and temporal posture envelops was introduced. Our algorithm can handle both stationary and moving postures, which are walking, standing, lying, sitting, and bending. The algorithm was tested in the following two aspects such as the efficacy of proposed features, comparison with Juang, C.F. and Chang, C.M.'s research (2007) using Neural Fuzzy Network classifier in four (excluding walking) and five postures classification, respectively, and applicability to various data sets, testing the robustness of algorithm in various data sets. Four video clips from Weizmann human posture data set, UIUC complex active data set, CMU graphic lab motion database, and our own data set were experimented. With respect to our own data sets our average accuracy achieved 93.27%. But when our algorithm was tested will other data sets, our accuracy is higher than 95%.

Department : Imaging and Printing Technology	Student's Signature ....
Field of Study : Imaging Technology	Advisor's Signature ....
Academic Year : 2012	Co-advisor's Signature ....
	Co-advisor's Signature ....

ACKNOWLEDGEMENTS

This research is supported by Chulalongkorn University Graduate Scholarship to Commemorate The 72 nd Anniversary of His Majesty King Bhumibol Adulyadej, The 90TH Anniversary of Chulalongkorn University Fund. (Ratchadaphiseksomphot Endowment Fund.) and Center of Excellence in Mathematics.

The author would like to express her gratitude to her advisors, Professor Chidchanok Lursinsap, Ph.D., Associate Professor Aran Hansuebsai, Ph.D. and Assistant Professor Suphakant Phimoltares, Ph.D. for their tireless guidance, suggestions and assistance in this dissertation writing. Sincere appreciation are due to her thesis committee: to Associate Professor Pontawee Pungrassamee, M.S. for serving as the Defend Chair Lady, to Assistant Professor Chawan Koopipat, Ph.D., Assistant Professor Thanarat Chalidabhongse, Ph.D, and Kingkarn Sookhanaphibarn, Ph.D. for serving as thesis committee.

Many thanks due to Advanced Virtual and Intelligent Computing (AVIC) research center and Department of Mathematics and Computer Science, Chulalongkorn University for the allowance to use their research facilities and equipments for experiments. Appreciations are also due to Professor Chidchanok Lursinsap, Ph.D., Associate Professor Peraphon Sophatsathit, Ph.D., Asso.Prof. Suchada Siripant, Ph.D., and Assistant Professor Suphakant Phimoltares, Ph.D. for their help in training, information source, suggesting and supporting in experimental work.

Appreciation and gratitude are due to the Imaging Science and Printing Technology, Chulalongkorn University for research facilities, information source and equipments for experiments. Appreciations are also due to Assistant Professor Chawan Koopipat, Ph.D., Associate Professor Pontawee Pungrassamee, Associate Professor Aran Hansuebsai, Ph.D, Kingkarn Sookhanaphibarn, Ph.D., Prof. Mitsuo Ikeda, Ph.D. and Prof. Suda Kiatkamjornwong , Ph.D for their help in training, advising and moral supporting in doctorate degree study .

Finally, she would like to express her deep gratitude to her family for love moral support throughout the Doctoral degree study. She also would like to thank her mother and younger brother, Miss. Ratana and Mr. Chatturong Silapasuphakornwong, for their care, understanding and supporting in everything during the study and research period with deep love.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

Figure Page

CHAPTER I

INTRODUCTION

The human action analysis in video nowadays has been applied to the several useful applications, both in control and surveillance system [1, 2]. The main important idea of these applications is to recognize the human activities in the video scene which leads to identify and track each person. This includes action classiﬁcation and understanding his/her current behaviors for prediction in the near future process. The human action recognition is concealed in the several well-known applications ,i.e., teaching media, video game, smart-home and surveillance system, tele-medicine system, and home-care system, and so on. These applications helps making the everyday life of people convenient.

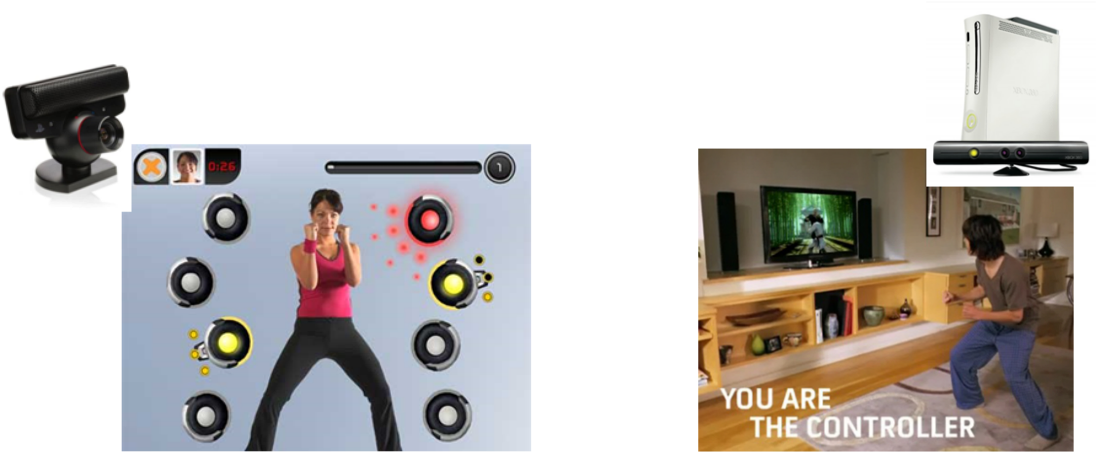
In many systems available nowadays, the users can control them by only performing the required actions. For example, the teaching media [3, 4] helps the student learn the dancing and yoga by themselves. Firstly, the important postures of dancing of student are recognized. After that, the system will compare the student's postures with the original recorded postures of professional. The scores of dance-performing are depended on the similarity of student and teacher dancing. In game industry ﬁeld, the players can play games by their actions, captured from the recording camera of video game equipment. The human postures are recognized for controlling the interaction of the games and players directly. The game's equipments and software in the trend "playing with action" have been developed continuously. The examples of developed video game equipments of two big game's companies for capturing the players' behaviors are SONY Playstation Move© and Kinect Xbox 360©, as shown in Figure 1.1a[1] and Figure 1.1b[2], respectively.

In the smart-home system, the human behavior in video surveillance within the house is recognized to control the other systems, especially the electric system in the house. For example, the light in the pathway will be turned on when people walking pass through that area. When the host is sitting on the sofa in the living room, the television will be turn on. Moreover, when everybody is sleeping in the bedroom, every light and other electric appliances in the house will be turned off automatically [5].

Furthermore, the video surveillance within the house is also applied to be the home-care system which aims to take care and look after the children and elderly to do something at home more safety such as, cooking, dining, taking a bath, sleeping and so on [6, 7]. An advantage of

[1] www.xspblog.com/2008/06/09/playstati...n-3-eye/, available on 3/7/2010

[2] www. mmomfg.com/2010/06/13/kinect-off...me-0613/ and www.epicbattleaxe.com/tag/kinect/, available on 3/7/2010



(a) (b)

Figure 1.1: The sample applications of human action recognition in video game. (a) SONY Playstation Move©, (b) Kinect Xbox 360©.

this system is able to control from other faraway place, such as hospital or their relative houses and offices,through the internet and mobile network [8]. So, when the accident happening at home, such as falling or something burning, the warning signal will be sent to the emergency section, close to hospital, and their relatives immediately. The patients will be rescued in time [9, 10, 11, 12, 13, 14, 15]. In addition, the idea of the long distants controlling through the internet network are adapted to the "tele-medical" system that linked between the patient living in theirs house and the doctor working in the hospital [16]. The doctor can follow the symptoms of patients from the surveillance data that is sent though the internet to the hospital. This remote medical diagnosis and treatment helps improving the quality of life of some patients, who were ill with some diseases, such as sleep apnea, insomnia, parkinson, including the convalescent patient. As these diseases will be identified the symptom from the recognized risk postures, thus, the patients can live in their home and also in the care of the doctor intimately at the same time [17, 18, 19].

Moreover, the extreme target, the requirement is to predict and to analyse the actions that will be happen in the near future for planing and protecting the criminal . For instance, the suspected person detection application aims to detect the person who performs the suspected behavior, caused by the robbery and terrorism. However, at present, the human behavior recognition in the surveillance system is only able to simulate the 3D-animation of the robbery occurrence. The accused person identification in the crime is still the challenge research problem [20].

All examples are implied that the applications adapted from the human behavior recognition are very useful and important for improving the quality of life in many things such as living in everyday life, improving the medical technology, rescuing the sufferer in time, and reducing the criminal. These are the reasons why the research topic of human behavior recognition is still required to study and developed continuously. The effective algorithm that is fast processing and more accuracy is necessary. Many researches have been developed several techniques and methods for action estimation.

The collection of human information is an important task for setting the techniques and methodology for analysis. Several information source such as, sensor suit [21], direction detector sensor [22, 23], sensor matrix bed [24, 25, 26, 27], pressure sensor [28, 29], accelerometer [30, 31, 32], brain-wave and heart rate sensor [18, 19, 33, 34], and frames of video [35, 36], are chosen as the human behavior information collection. In each information source, it has the different approach and method for analysis. Nevertheless, the action recognition from the video directly has been challenged as problem. It is because of the difficulty to know the exact position and direction of the human in the frame without any marks or signals from the sensors sending to the system. To effectively recognize a human action, the combination of several postures, from the direct scene, many techniques of segmentation, feature extraction, and classification are used for identifying each person in the frame by tracking them, and predicting their postures, respectively.

This dissertation proposes the human posture recognition which is considered in five basic postures such as walking, standing, lying, sitting and bending. This research aims to develop the proposed method in condition, robust to the personal apparel, shapes and size, and invariant to the camera factors as viewing angles and distances. It is possible to detect the human posture directly from the video. There is neither any sensors and the special suit nor any signal sending equipments. People can wear the normal clothes, do not hold the sensor equipments, and perform the normal behavior in front of the camera in every distance and viewing angle. In the experiment, the new human behavior video dataset was constructed for testing the capability of the proposed method, which was larger and more various than the previous research. Moreover, the other three datasets were also tested on the proposed method for applicability testing to confirm that the proposed method is suitable for adapting in the real human action video. Our method was designed in order to be convenient to develop and apply to the real life application in the near future. In addition, the general algorithm was enhanced to detect five sleep-positions to apply in sleep-monitoring application, an example of real-life applications.

1.1 Objectives

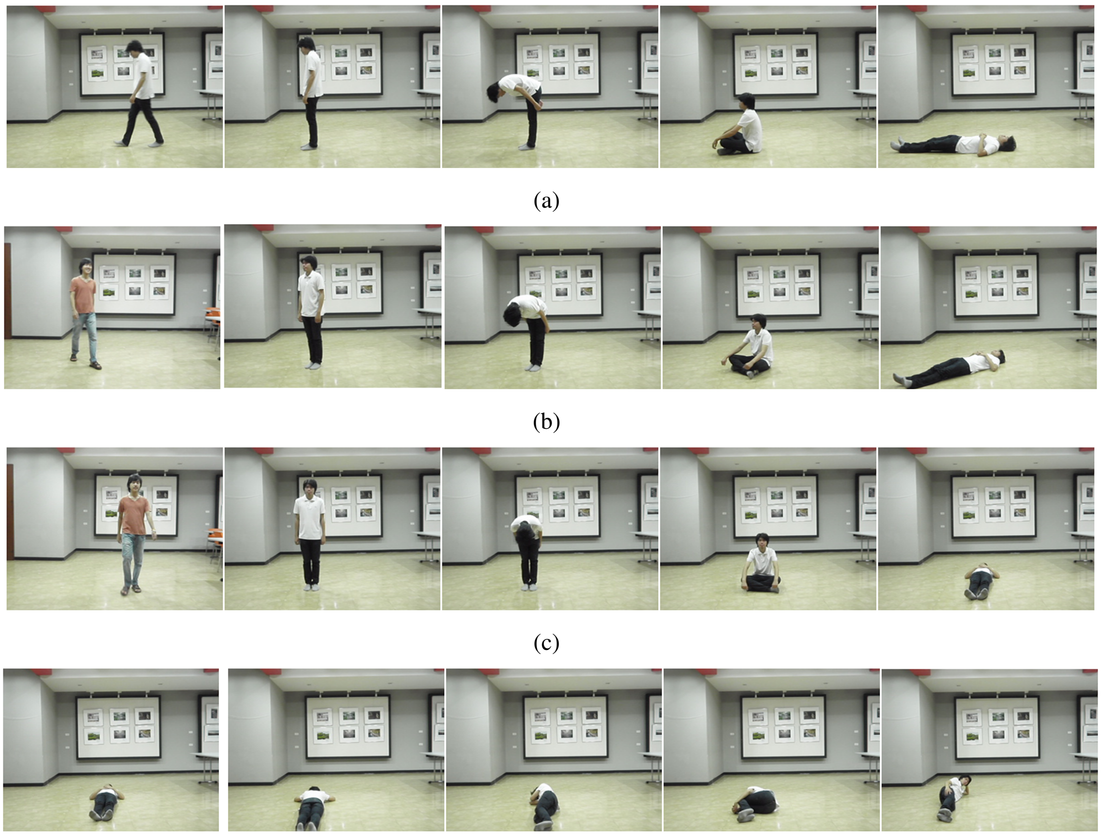
The main objective of the proposed method is to recognize common daily of acting human postures and categorized an image sequence into five main postures directly. The major experimented postures are *standing, sitting, bending, lying, and walking*. However, the *lying* postures may be perceived as *sleeping* postures. It is because they have the same action unless the condition of the brain-waves is measured. Hence, in this study, the words *lying* and *sleeping* implied the same action. For *lying* posture, the following five detailed postures were studied in this dissertation: (1)supine position (face up position); (2)prone position (face down position); (3)log position (lying on either left or right side with the arm down the side); (4) fetal position (lying curled, with limbs bend close to the torso); and (5) reclining position (the posture of Buddha sleeping). Some examples of the studied postures are illustrated in Figure 1.2. The postures in Figures 1.2a, 1.2b, and 1.2c are walking, standing, bending, sitting, and lying in $0°$, $45°$, and $90°$, respectively. The postures in Figure 1.2d are five detailed lying postures, i.e. supine, prone, log, fetal, and reclining positions.

1.2 Scope and Research Problems

An interesting literature on home care surveillance system [5, 6, 8, 10, 37, 38] is the inspiration of this dissertation. The human posture recognition from the video clips which consider both static posture as well as motion posture for human behavior predication is the goal of this research. In the experiment, the five main postures, i.e. standing, lying, sitting, bending and simply walking, and five sub-postures of lying were concerned. The proposed method was independent from background, cloth texture, body size and camera distance by using the morphological geometry derived from the length and width of human body observation. The more various and greater number of samples than Juang and Chang's study [10] were used in the training and testing data set. Finally, the efficiency of recognition was evaluated by the percentage of accuracy. The changing posture and simple walking recognition parts are expected to increase the efficiency of the system.
To achieve the objectives, three main research problems are queried.

1.2.1 How is the person represented in a frame?

The human representation in the video frame, as in the image, is the first important thing that we have to do in the posture recognition. Before the human is represented, we will find the actual region of the human in the frame. A difficult thing is "how to separate the region of interesting (ROI) from the background". This is still a classical question so far [39, 40, 41,



(a)

(b)

(c)

(d)

Figure 1.2: Varied styles of main postures in different viewing angles of the camera, (a) to (c) as five-main posture such as walking, standing, bending, sitting, and lying, respectively. (a) At 0 degree. (b) At 45 degrees. (c) At 90 degrees. (d) Five types of concerned lying postures (*left to right*) supine, prone, log, fetal, and reclining at 90 degrees.

42, 43, 44, 45, 46, 47]. The method that distinguish the ROI from the background is called *Segmentation*. Hence, to answer the first research question, the segmentation method will be applied to this research work.

1.2.2 How are the suitable features for identifying the main posture?

As being the camera-based system, none of sensor and marker indicate the human information and send it to our system. So, we will find something that can represent and identify the characteristic of each main posture. An attribute, called *feature*, is extracted from the region of interesting (ROI) will be represented the human information of our system. The feature can be extracted from many approaches by for example: extracting from the histogram, using geometrical morphology, computing the coefficients from discrete cosine transform (DCT) or discrete Fourier transform (DFT), using color values from different color system, representing direction by vector, decreasing the dimension of data by using PCA, etc.

A good feature should have the obviously different value in the different class. In the recent years, there are many proposed features for posture classification such as, centroid, length-width ratio, posture angle, and DFT-coefficients, etc. [10, 11, 13, 48, 49]. The separate class will be distinguished by using these features.

However, the effective features that can be distinguished the five main postures correctly are so important. The high performance of classification also leads to the high accuracy in applications. Meanwhile, the efficient features should use in less amount to decrease the cost for computation and trend to work in real-time. So, to fining the effective features for posture classification is an goal of this dissertation.

1.2.3 How are the main postures distinguished?

Generally, the method for categorizing the data as the classes is the classification model. There are many methods both required training set, called *Supervise Learning*, and *Unsupervise learning*, that can adjust the class by itself. The effective classification model should help classifying the data in the suitable classes. Hence, in this research, we develop the posture classification model to categorize the five main postures and five sleep positions in order to achieve the high accuracy of recognition.

1.3 Constraints and Overview Setting

The proposed posture recognition system is designed to recognize the main five human postures, including the sleep posture which contains five sub-postures. These five postures are the most common postures daily acting by a human. However, some postures cannot be distinguished due to the body's skeleton. For example, walking and standing postures are rather difficult to recognize if each posture is taken from a single image. A series of continuous movements must be recognized in case of walking posture. But standing posture needs only a single image. Our proposed technique considers both stationary and moving postures in time domain.

The following constraints were involved in our recognition process to enable the efficiency to cope with realistic situations beyond the other previous works.

1. A subject performing the postures did not wear any special suit or sensor instruments.
2. The postures could be varied by each individual's styles and viewing angle of the camera.
3. Viewing angles were ranged from $0°$ to $90°$.
4. The camera was settled in a tripod with 90-110 centimetres height from the floor.
5. For testing the effect of perspective, a volunteer performed the individual postures with distances of 3 to 9 meters apart from the camera and captured on the 0, 45 and 90 degrees of viewing angles, depending on different purposes of testing set.
6. Each volunteer had different body sizes and could wear in different cloth textures in varied apparel.
7. The video file was captured at resolution 320x240 pixels at the rate 15 frames per second. Each training video was about 1-2 minutes length and each image file was captured in true colors of RGB color space.
8. The most contexts in the background image were not be altered by any means throughout the analysis and experiments.

A challenging problem that other researches did not concern is the problem of various postures in $90°$ viewing angle. Previously, the general posture recognition methods did not support in this point. For the example in Blank's research [50], the robustness of the posture recognition method was tested from $0°$ to $81°$ viewing angles. Jung and Chang [10] concerned the posture at 0 degree. The postures at 90 degrees were not considered. The task of this problem is related to the following three main issues.

Firstly, regardless of the floor line and texture of the background, it was noticeable that some lying postures in $90°$ resemble the other postures in some other viewing angles. For example, in Figure 1.2, the sitting posture resembled the fetal lying posture. The bending posture confused with the log lying posture. The standing posture was similar to the prone and supine lying postures.

Secondly, some parts of body were partially occluded due to the perspective of the camera. Considering the bending posture in Figure 1.2c, we could see that the torso was occluded by the head and the waist was partially occluded by the head and some part of the torso. This problem usually occurs when the photograph of the bending person was shot from the back.

Lastly, in Figures 1.2a and 1.2c, we found that the widths of a body part shot in between 0° and 90° are obviously different. The minimum width occurred when the shooting angle was 0 degree but the maximum width occurs at 90 degrees. Hence, using only the length and width of the body as features might not be enough in the step of classification. In addition, the envelop of human body, when lying in supine and prone postures and being shot in 90°, had a perspective view. This maked it difficult to extract the feasible features of these two postures. The other systems did not consider this issue due to this reason. But this complication concerning 90° camera shooting was solved in our system.

1.4 Contents of The Thesis

This thesis consists of 5 Chapters, introduction, theory and literatures review, proposed method, experimental and results, and discussion and conclusion. The theoretical background, scientific rationale is described in the Chapter II. In addition, the several literatures and the basic famous traditional techniques in ROI segmentation, feature extraction, and classification are also reviewed in this chapter. For the developed processes of proposed method, the details are explained in the Chapter III. An example of enhancement of the general algorithm for adapting to application is also presented in this Chapter. For Chapter IV, two main experiments are examined and their results are discussed : 1)The capability experiment, comparing the efficiency of proposed method and the traditional method, and 2)applicability experiment, testing the proposed method with the other human actions datasets. Finally, the conclusions and the suggestions of the future development are presented in Chapter V.

CHAPTER II

THEORY AND LITERATURES REVIEW

2.1 Theoretical background and scientific rationale

Human behavior analysis in computer vision has been adapted to several applications such as surveillance system, control system, medical image analysis and home-care system [2], as concluded as Figure 2.1. Surveillance system aims to record and detect some information of the human behaviour in a period of time. These involve in human detection, human identification, human tracking, 2D-3D reconstruction, and behaviour prediction [51]. In computer games, the interaction between players and actions in the game is a good example of the control system application [2]. The software in the game might recognize the human performance, to represent and reconstruct their actions, in order to control the other part of the game to run the story all over the game [52]. For medical image analysis, the system aims to find the informal or abnormal behaviour of the patient for disease diagnosis. The suspicious behaviour detection at the beginning helps planing the treatment in time. Lastly, in the home-care system, it is designed to watch out the accident in the house, such as the falling or injury from the accident in the kitchen, and to monitor the risky behaviour of the people who has to stay at home alone. Moreover, in the near future, the system can be developed to be remote monitoring [9, 53] and controlling by Smartphone [54].

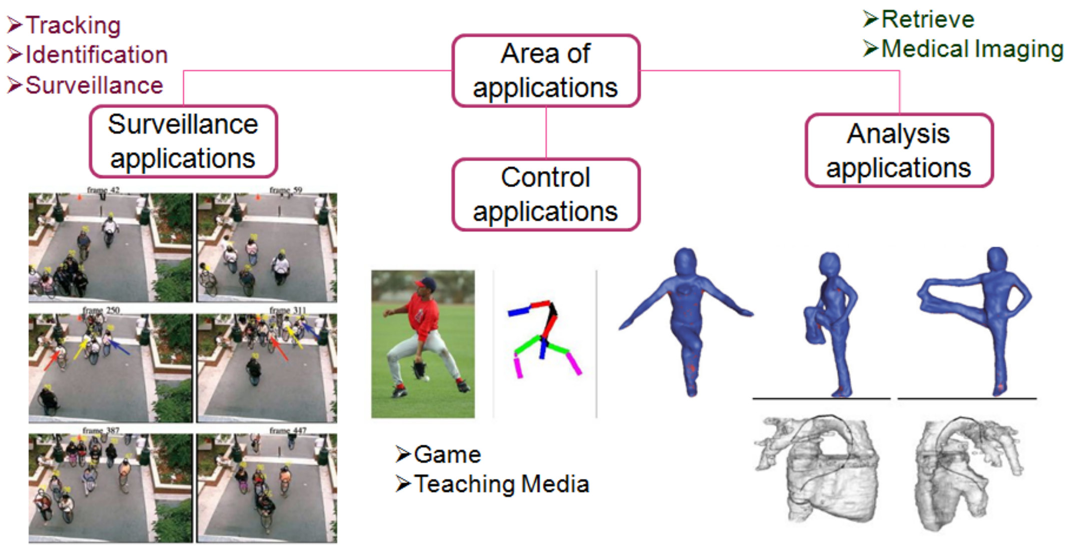


Figure 2.1: Applications of human posture recognition in video.

These applications that are compose of several techniques which lead to understand the human routines for improving the quality of the human lifes [1, 2, 51, 55].

The main process of the human action recognition composes of two main parts which are image representation and action classification [1] as illustrated as framework in Figure 2.2. The image representations are divided into two categories: global and local representation. The difference between global and local representation is the process workflow. The global representation uses a top-down process. Firstly, the background is known and registered by analysing for a period. Then, the human region is localized and tracked in each frame, using the background subtraction method. Lastly, the Region of Interest (ROI) is encoded as the whole. However, the global representation is sensitive to the viewpoint, noise and occlusion. On the other hand, the local representation uses a bottom-up process. The local patches are calculated from the spatio-temporal interest points and these patches are merged in the final process using the approaches which focus on the correlations between patches. As the local representation is independent on the viewpoint, noise and occlusion,this method thus is usually used in the difficult environment by moving a camera. However, the approaches are complexity and expensive in computation.

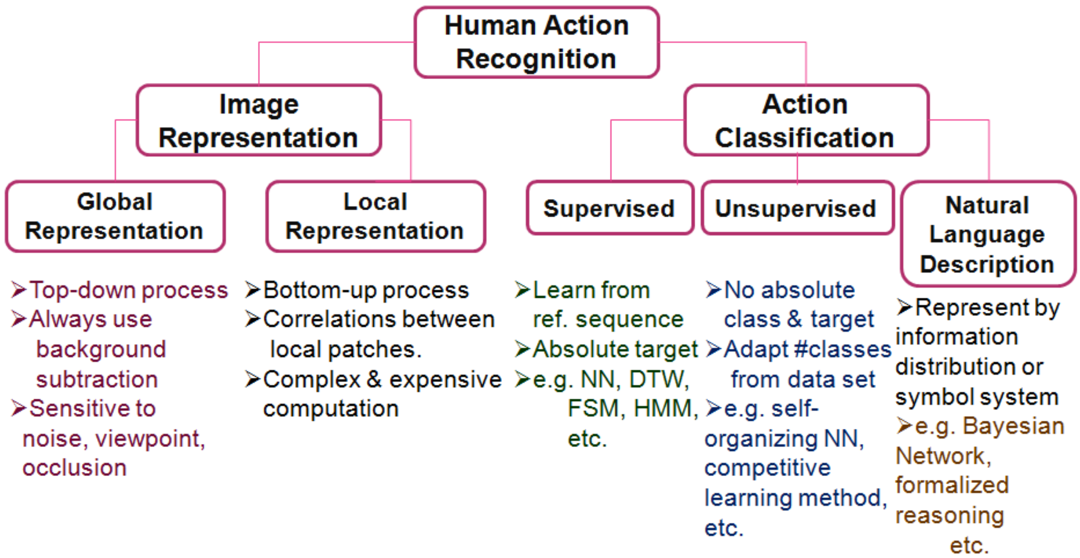


Figure 2.2: Human action recognition main process.

The classification model for human behavior analysis [55] has three types such as supervised, unsupervised and natural language description. The supervised classification model is based on learning the reference sequences or absolute target from training samples to group the testing data set to each class. For example, Dynamic Time Warping (DTW), Finite-State Machine (FSM), Hidden Macov Model (HMM), Time-Delay Neural Network (TDNN) etc. On the contrary, the unsupervised classification method has no absolute target classes. The classifier learns and adapts number of suitable classes from the samples data set, for instance, self-organizing neural network, competitive learning method etc. Recently, the natural language description use the information distribution and the symbol system to represent the behavior patterns, such as the statistical models or Bayesian network models and formalized reasoning model, respectively.

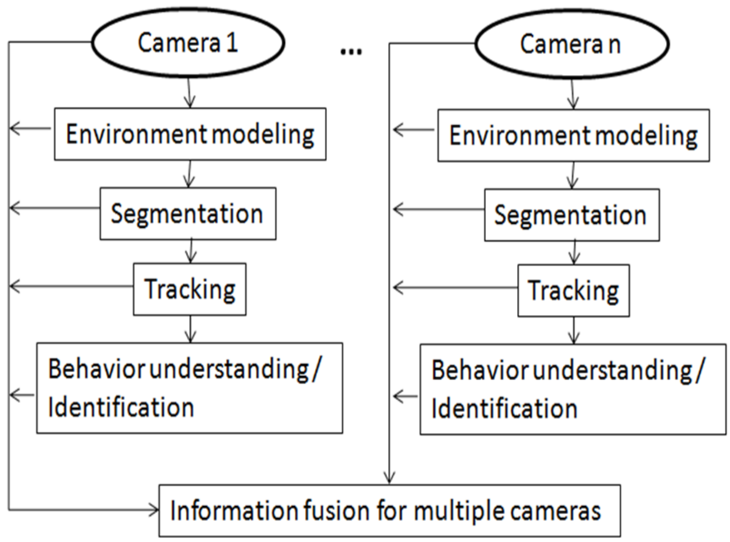


Figure 2.3: General framework of visual surveillance

The main process of the visual surveillance system begins with each fixed camera comprising three important techniques : segmentation, tracking and recognition. Then, the information of all cameras is fused together and analysed for the whole recognition in a multi-cameras system.

The general workflow of a visual surveillance system, which considers the object motion and human behavior [55], is shown in Figure 2.3:

For the segmentation, the most researches use generally three techniques such as background subtraction, temporal differencing, and optical flow. The background subtraction is the popular method used in some literatures in [10, 20, 56]. because the image capturing condition in the experimental room is simply to set up. The assumption of their approaches limits the static background, constant lighting and unchanging environment. The background subtraction technique is developed by Chien et al. [57], called background registration technique, which help improving to remove the noise and shadow effect more efficiently. Second, the temporal differencing technique finds the frame-difference in two or three frames and uses the threshold for selecting the ROI. This method obtains the poor segmented region address in [58, 59]. Lastly, the optical flow computes the displacement vector field called active rays. This technique is the most complex computation and sensitive to noise, which is suitable for complicated environment as moving objects or moving camera [60, 61, 62, 63]. Figure 2.4 presents the comparison techniques in each segmentation type.

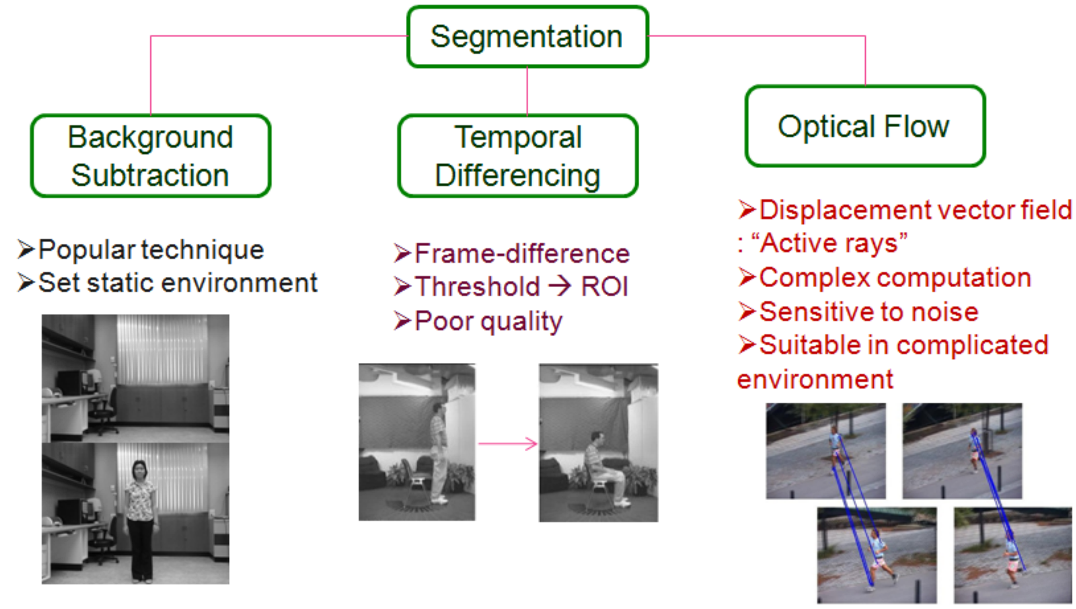


Figure 2.4: Types of segmentation techniques.

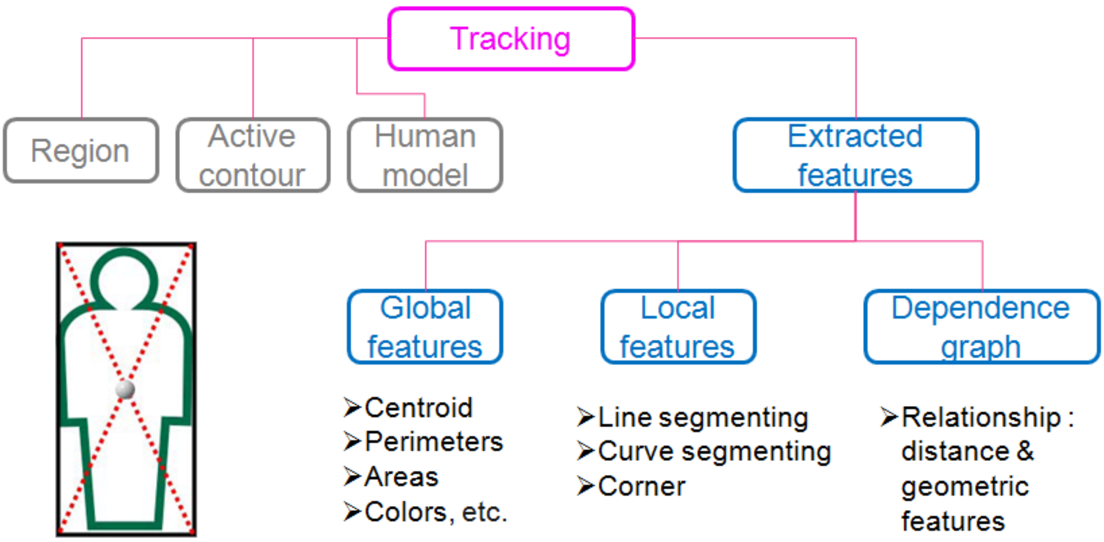


Figure 2.5: Types of segmentation techniques.

After the segmentation step, the ROI is tracked by many methods based on tracking algorithms, such as region-based tracking algorithm, active contour-based tracking algorithm, human model-based tracking algorithm and extracted features-based tracking algorithm, which presented in Figure 2.5. The features-based tracking method is divided into three main types : global features, local features and dependence graph based algorithm. The global features based algorithm extracts and analyses the following information such as centroid, perimeters, areas and colors, etc. from the ROI rectangular boundary box [11, 13]. The local features based algorithm extracts the information such as the line segmenting, curve segmenting and corner [64, 56, 65]. The dependent graph based algorithm uses the information of distances and geometric relationship between features for classification, for instance in [10, 66]. The tracked ROIs are likewise to the representation of the object in image sequence which will be analysed and interpreted with the classification methods for recognizing in the different purposed applications, e.g. behavior understanding and identification. Until the present, the effective and fast human detection and tracking still be the challenged problem, as these several researches [67, 68, 69, 70, 71, 72].

In conclusion, all of these systems which can recognize and detect the human movement efficiently is based on image processing and computer vision that have been developed continuously. Such techniques are segmentation, people localization and tracking, feature extraction, posture estimation and classification.

2.2 Literatures Review

Recognizing human actions from an image sequence is still a challenging problem. Recently, researchers have developed the techniques for recognizing human motion by using only a continuous image sequence to gain some human action information as the important features. Hence, the truly relevant features can improve the efficacy of the overall system. In the recognition part, several techniques can be applied to classify, identify, track, or reconstruct the human actions in different purposes. However, most existing techniques deployed costly computation, especially during feature extraction. To make a system practical to real situation such as homecare environments, simplest features but yet fully informative must be discovered. Besides, the recognizing process must be faster than the scenario change of human actions such as surveillance system.

During ten years, there were many literatures about human behavior understanding, action motion analysis, and human 3D reconstruction and recognition as follows;

Focusing on the human posture recognition field has begun since twelve years ago. In A.D.

2000, Haritaoglu et al. [73] introduced the W^4 technique for a real-time outdoor surveillance system. This was an early technique which studies about action recognition directly, without the additional equipments on the body. The W^4 technique was a simply real-time outdoor surveillance system method which corresponds to four main questions about people in the frames, such as what, when, where, and who are acting in the frames. These main research questions leaded to the several related technical aspects such as detecting and tracking people and luggage, locating body parts and estimating human posture for representing the human regions and their behaviours in the frame. The histogram of the Region Of Interesting (ROI) was used for people detection and counting number of people in the crowd. The periodic function and frequency analysis was adapted for people and luggage tracking. The convex and concave analysis was applied to locating the body part and estimating the human posture. However, their research used only one video clip for testing the method in the part of posture estimation. The robustness of this system was quite low.

The present researches always interpret the human postures analysis as the three dimensions human performing reconstruction and tracking body parts on human region in video. The literatures in the human 3D model reconstruction field are still in the trend, addressed in some researches. The 3D human shape or body is reconstructed from the multi-camera as demonstrated in multiple view angle. The different techniques are used for modelling and deformable the human pose.

In research of Wu et al. [65], the gesture information was fused across the space, time, dimension and feature levels which using Particle Swarm Optimization (PSO). For study of Mukasa [74], the 3D model was reconstructed from the two components features as Geometric configuration and color or texture of the body part. The sample researches in the tracking of human pose, perhaps called *posture estimate* appeared in [56, 75], and [76].

Wu and Aghajan [65] introduced 3D human body model by ellipse fitting embodying in multi-camera network. The gesture information was fused across the space, time, dimension and feature levels using Particle Swarm Optimization (PSO). The 3D model was reconstructed from the two components features as Geometric configuration and color or texture of the body part.

Mukasa et al. [74] presented the 3D human shape reconstruction of the complex posture estimation from multiple-views of yoga video image. The algorithm was based on visibility-based model fitting. This method could reconstruct the invisible surface shaded by body part.

Lao et al. [20] used a consumer camera captured the surveillance video mimic the robbery situation. The system can analyse the human behavior and deform the 3D body model in different views. In addition, the system also estimated the activity of the robber and suspicious object trajectory. The recognition algorithm using Continuous Hidden Markov Model (CHMM) based on temporal modeling and HV-PCA descriptor achieves high percentage of posture recognition.

Juang et al. [64] presented the posture estimation method which located five points on the human body contour, i.e., a head, two hands and two feet with a fixed camera. The algorithm selected the suitable angle of convex points computed by the principle and minor axes contribute to place the five points of the human body contour. In addition, using the skin-color information could prevent the mistaken points. However, this algorithm was limited by a complex posture which contains the overlapping of body parts, caused the limitation of the algorithm.

Gupta et al. [76] adapted the potential function for kinetic and appearance constrained of nodes represented by the graphical structures and used the probability distribution to estimate the 3D pose in clustered scenes.

Bodor et al. [56] introduced the human gait recognition method based on image rendering for classifying the sequential human-posture images in three main actions such as walking, marching and running. The virtual image from a virtual camera was constructed from the whole images captured from the several cameras at the same time. The image of each camera will be matched with the training data set recorded from a single camera at the orthogonal motion position in eight actions, such as walk, run, march, skip, hop, walk sideways, skip sideways and walk a line. These actions were combined and then the orthogonal view angle image was created. For reducing the diverse feature dimensions, the principle component analysis (PCA) was used as an action classifier. These actions were combined and, then, the orthogonal view angle image was created.

From the recent survey [1], the action recognition composes of two important parts which are action representation and recognition.

In the different literatures, several different techniques were applied for the distinct purposes. The example of research that aimed to present the approach for actions representation. For instance, Nazli and Pinar [67] presented the human posture representation using the Histogram of Oriented Rectangles (HOR) which was the rectangle patches.

As the recent survey of human action recognition [2], the most demanding task was the lack of the amount of sample size and variation movement. However, the flexible classifier for adding or removing new action classes was also required. Moreover, the online application was also the challenge problem which had to use the least resource for benefit in real-time information transfer. For the street surveillance system that detected the moving objects on the street, the researches often discriminated the pedestrians from the vehicles by the periodic function from the walking rhythm, for instance:

Lin and Wei [77] presented the real-time vision system for street scene surveillance. The system aimed to recognize moving objects on the street and classified the pedestrians by walking rhythm based on height and width ratio.

Some studies recognized the human actions in during time as the predicted postures. However, most literatures concerning human action recognition at present only concentrated on either motion recognition or stationary posture recognition. The motion recognition expected to analyze the actions or postures that moved all times such as walking, running, jumping and waving hands while the stationary posture recognition intended to analyze the actions that hold still momentarily such as standing, lying, sitting and crouching or bending etc.

The human movement recognition research of Bobick and Davis [58] was an example of the global representations. The motion-history image (MHI) and the motion-energy image (MEI) were constructed for representing each action in the aerobic exercises video clip. The two templates were compared and matched with the image moment. This method was a scaling and translation-invariant algorithm but restricted by the motion speed.

Blank et al. [78] presented the space-time shape feature analysis. While Li et al. [79] presented using the Gaussian Mixture Models (GMMs) to represent the salient postures. The effectiveness of these features was evaluated by similarity measuring between templates classified by basic classification methods such as nearest neighbour, Dynamic Time Warping (DTW), Support Vector Machine (SVM) classification, Finite State Machine (FSM) [79], and Hidden Markov Model (HMM) [80].

The idea of periodicity detection in motion period came from Laptev and his team [60]. The advantage of this method was able to apply to complex scenes such as moving camera and motion parallax immediately with no camera stabilization nor tracking. Nevertheless, the limitation of the method was under the assumption of constant translation as shown in time-linear matrix function and the complex computation.

Presently, there were several researches that analysed the human behaviour from image sequences [1] and applied to the field of home care surveillance application. Most researches intend to focus on human falling detection, for instance;

Cucchiara et al. [81] presented the human classification method using the probability based on Bayesian classifier. The probabilistic projection maps (PPMs) based on Bayes's rule as the classifier and the state transition graph (STG) were adapted in order to increase the efficiency of recognition with the suitable accommodation for each room. The weighted STG corresponding to the relationship among postures in specific room was applied to detect the falling people. They suggested that the human posture can be categorized into four main postures such as standing, sitting, lying, and crouching, presented by the posture stat-transition graph. However, this system cannot distinguish the intentional lying posture from the accidental falling case.

Girondel et al. [82] also classified the stationary human body posture into four main classes. Their hypothesis was to explain the posture classification in a fuzzy set form.

Zhao and Liu [83] suggested the Radii Vector and Normalize Length Vector as the important features for training in SVM and classifying 10 postures but the accuracy was not satisfactory.

Vishwakarma et al. [13] presented the fall detection approach extracting the features from the minimum bounding box such as aspect ratio, horizontal and vertical gradients, and fall angle. Finally, these features was implemented on Finite State Machine (FSM) for detecting the falling human in several situations in sample videos i.e. indoor and outdoor situation, single and multiple person, omni-video. However, their research could not discriminate the normal lying down from the falling by accident.

In the same year, the home care system presented by Juang and Chang [10] used the length-width ratio and DFT coefficients as the important features for training a Self-Constructing Neural Fuzzy Network (SONFIN) [84]. The four main postures such as standing, sitting, bending and lying were classified. The behaviour of falling-down always means the posture immediately changes from standing to lying. With this reason, Juang and Chang [10] suggested that the warning signal should be different in three conditions as emergency accident, normal accident, and sleeping by using the different length of posture changing period. They chose the length-width ratio and 40 DFT-coefficients, extracted from horizontal and vertical projected histogram, as the features and constructed the rules for posture classification by a neuro-fuzzy network. The disadvantage of this system was the limitation of capability to estimate and predict only stationary postures in a single view angle. In the situation of the continuous and complex postures, it gave poor performances. Furthermore, it is noticed that the features which were used for training the system were abundant. The prominent point of this research was the lying posture discrimination between emergency case, accidental case, and normal sleeping case. This system measured the quickness rate of changing posture from other postures to the lying posture. The different signal was sent out from the system calling someone able to help the patient in time. Although the recognition rate of the method was the highest during this time, the number of samples in the experiment was too low and not sufficient to guarantee that the method could cover every cases.

Lin and Ling [11] suggested more simply features for falling detection such as centroid of the body, the maximum vertical projection histogram value, and the duration of an identification event to falling event. However, the posture was categorized into two classes such as standing and lying.

Li et al. [85] used the Gyroscopes and Accelerometer for collecting the posture information and then recognizing the fast accident action such as falling from the stairs or sitting down fast.

For the sleep monitoring application in a home-care system, this application monitors and collects the sleeping information overnight and initially evaluates the quality of sleeping in every night. In case of sleep disorder diagnosis, the doctors use the sleeping information for estimating the symptom of the disease and planning the treatment in order to change the sleep routine. The purpose of the treatment is to avoid the sleep disorder disease as Obstructive Sleep Apnea (OSA), insomnia, parasomnias, and circadian rhythm disorder, etc. that bring to be the terrible disease such as heart disease, untreated OSA, high blood pressure, cerebrovascular (CVD), and death from heart attack, etc. [33]. For the technical approach of the sleep monitoring application as mentioned, the system usually stores the human behaviour information during sleeping night long, as the history of the surveillance system. In general, the personal information for sleep quality diagnosis during presents composed of several biomedical parameters such as data of heart rate, oxygen level, snoring rate, air flowing, body movement and electrical brain signal [31]. These statistical sleeping information in every night of a week should be the initial important personal information for the sleep disorder diagnosis and treatment for the future. Bsoul et. al. suggested to use the signal of electrocardiogram (ECG) channel as the data for training in Support Vector Classifier (SVC) to detect apnea episodes in sleep monitoring application, developed on Android operating system [19]. However, since equipment set-up was rather complicated, the patients always spent a few nights at the hospital for testing the sleep behaviour and it was not convenient for the patients to do so. Sometimes the patients also felt the stress which affected to the diagnose result. Nevertheless, the easier instrument setting can help the patient test the sleep quality at their home. It was not necessary to spend a few nights in the laboratory of the hospital for testing the sleep behaviour.

Occasionally, to collect the important personal information for human behaviour analysis, the additional equipments as sensor was attached to people at all times, such as special suit, wrist sensor [86], or pressure sensor [87]. For example, the accelerometer [15, 88] was the sensor that measures the direction and dynamic of the body. The received signals were applied to classify the human postures to a lot of postures. Signal of center of pressure (COP) was used to classify the standing posture [87]. Spranger et al. used both the signals from the sensors, acceleration sensor and proprioceptive sensor, and extracted visual features of direction changing for falling detection of the humanoid robots [89]. These sensors detected some important data for analysis in the system. However, it was inconvenient to be taken everywhere. So, it was better if the system can recognize the human performance from the image sequence directly.

Peng et al. [23] presented a long-term sleep quality monitoring using the low-cost system by the information from a sensor wearable watch, night-vision video, and PIR Sensor. The pressure sensor under the bed was used in many researches [24, 25, 26, 27, 29, 90], for the lying positions estimation during sleep. The sleep posture recognition from the image directly was presented in Wang et al. [18] study. The 2D-deformed skeleton that represented the sleeping position on the bed was reconstructed. The interesting point of this research was that this system could recognize the postures under the blanket covering the whole body. However, this system could not predict the class of sleep postures and sleep-wake stage. The important part of the home-care system was the human posture classification because the system needed the correctly predicted posture before it used these pieces of information to recognize in other applications.

However, the methods of human posture classification at present were still complicated by the additional instruments for collecting the personal information, complex feature extraction and involving classification method. In addition, most studies were limited by people factors and camera factor, such as human size, appeal, distance from the camera and viewing angle. So, the challenging goal for human posture classification method was how to ease and enhance the correctness and robustness in order to make the system achieve high accuracy and fast processing applicable to real-time systems. The method must be applicable to various situations and everyone independent from the camera and people factors. In conclusion, our research aimed to study the human posture classification method workable in real time, achievable high accuracy, and robust to the people and camera factors for possible applications to a home-care system.

In this paper, we presented a real-time human posture classification method from the indoor video which is adaptable to the home-care system with sleeping monitoring application. Any equipment for collecting the human information such as accelerometer or pressure sensor bed is unnecessary. The proposed method can detect both stationary postures with clear posture and continuous situation and complex postures of the individual behaviour from the human silhouette directly. The human posture is categorized to five main classes such as walking, standing, sitting, bending, and lying. In the case of lying main posture, it is further decomposed into five lying posture sub-classes which are supine, prone, log, fetal, and reclining postures.

CHAPTER III

PROPOSED METHOD

The main framework of the proposed posture recognition method, to distinguish five main human postures such as walking, standing, lying, sitting and bending, is illustrated in the Figure 3.1. The proposed method is composed of three main parts: (1) human region segmentation, (2) feature extraction and (3) posture estimation and classification. The proposed method can classify actions both motion and stationary postures by the effective features based on length and width of body silhouette parts. In segmentation, the human region in the video scene, region of interesting (ROI), is identified by both segmentation techniques such as background subtraction technique and temporal differencing techniques. The background subtraction technique extracts the ROI in the static period whereas the temporal differencing technique extracts ROI from the movement period. Both ROIs are marked by drawing the envelops covering. After that, the important features, geometric information, are extracted from these envelops. Generally, five features, composed of Length-Width ratio(LW), the Difference between the Length and Width (DLW), the Area of human envelope (A), the Width of Upper part (WU) and Lower part (WL), are enough for discriminating five main postures. By the way, WU and WL are extracted from the sub-posture envelop that divide from the posture envelop. The numbers of features that extract from the sub-posture envelop can be adjusted for more suitable to the numbers of required classes. For example, in this study, the proposed posture classification algorithm is also improved to classify ten posture classes, five classes of five main postures and five classes of sub-posture of lying. The considered five sub-postures of lying are comprised of supine, prone, log, fetal, and reclining, respectively. Owing to the confusing of primary feature values overlapping between more concerned postures in different viewing angle, two sub-envelop was not enough for classify ten classes. The compensated features for modify ten posture classification method are added, which are the normalized difference between the longest width and medium width of three sub-envelopes $(\widehat{DLM})$, the normalized difference between the longest width and shortest width of three sub-envelopes $(\widehat{DLS})$, and the length from the centroid to the lower line of posture envelop (l_c). The proposed posture classification method is based on the decision tree. The criteria and all thresholds in algorithm are received from the experimental observation. The details of each step are explained as follow:

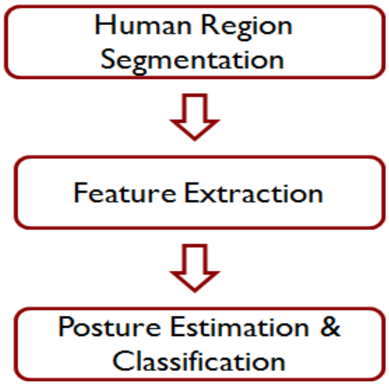


Figure 3.1: The main framework of proposed posture recognition method.

3.1 Human Silhouette Segmentation

Segmentation is an important step to extract the Region Of Interest (ROI) from the image. In this research, the ROI is the human performing different actions composing of stationary and moving postures. We use two methods, i.e. *Background Subtraction* and *Temporal Differencing* for segmenting the human silhouette and movement from the background. Background subtraction is the main method for segmenting the outline in order to extract the important features. Meanwhile, Temporal Differencing is another method for determining the moment of moving and steady postures. The result of this method is also adapted to evaluate the period of walking posture. The details of both methods are described as follows.

3.1.1 Background Subtraction

In this study, the objective of the background subtraction is to remove the registered background from a current video frame and maintain the actual human silhouette as accurate as possible. The human silhouette is considered as the foreground object and used further for posture analysis. Several techniques of background subtraction were reviewed in [91]. However, background subtraction is not a new problem, several existing methods were adapted in our work. Similar to the method proposed by Chien and his colleagues [57], we assume that the first frame video image is the background image and no human appears in this image. The context in the background image will not be altered by any means throughout the analysis and experiments. Since the amount of relevant information of the actual human silhouette is defined only by the intensity, not by its colors, the RGB color space is transformed to YIQ color space by the following computation:

$$\begin{bmatrix} Y \\ I \\ Q \end{bmatrix} = \begin{bmatrix} 0.299000 & 0.587000 & 0.114000 \\ 0.595716 & -0.274453 & -0.321263 \\ 0.211456 & -0.522591 & 0.311135 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \tag{3.1}$$

Note that, In YIQ color space, Y represent the luma information and I and Q represent the chrominance information.

Only the luminescence channel(Y) was considered. To eliminate noise of a considered frame, a median filter was used before the background subtraction. At coordinate (x, y), let $I^{(b)}_{(x,y)}$ be the value of Y channel of the background image and $I^{(c)}_{(x,y)}$ be the value of Y channel of the considered image. Then, the difference between these two images, denoted by $d(x, y)$, was obtained from $I^{(b)}_{(x,y)}$ and $I^{(c)}_{(x,y)}$ as follows:

$$d(x, y) = |I^{(b)}_{(x,y)} - I^{(c)}_{(x,y)}| \tag{3.2}$$

However, the extracted human region might be still noisy due to the frequency of the fluorescent light, vibration of the camera, and some irrelevant regions around the interested human regions. To eliminate these factors, a threshold filter similar to that of Grest, Frahm, and Koch [92] was applied. Assume that each video image frame was of size $M \times N$. Let $\bar{d}$ be the average difference defined as $\bar{d} = \frac{1}{MN} \sum_{(x,y)} d(x, y)$. At coordinate (x, y), if the following condition is satisfied.

$$|d(x, y) - \bar{d}| > \frac{\lambda\sigma}{\eta} \tag{3.3}$$

then the pixel at this location is considered as a foreground pixel. The value σ^2 is the variance of $d(x, y)$ defined as follows:

$$\sigma^2 = \frac{1}{MN} \sum_{i=0}^{M-1} \sum_{i=0}^{N-1} (d(x, y) - \bar{d})^2 \tag{3.4}$$

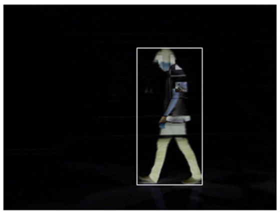
In addition, the η value is calculated by

$$\eta = \begin{cases} 1 & \text{for } |\bar{I}^{(b)} - \bar{I}^{(c)}| \leq \beta \\ 2 & \text{for } \beta < |\bar{I}^{(b)} - \bar{I}^{(c)}| \leq 2\beta \\ 3 & \text{for } 2\beta < |\bar{I}^{(b)} - \bar{I}^{(c)}| \leq 3\beta \end{cases} \tag{3.5}$$

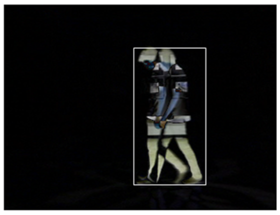
where $\bar{I}^{(b)} = \frac{1}{MN} \sum_{(x,y)} I^{(b)}_{(x,y)}$ and $\bar{I}^{(c)} = \frac{1}{MN} \sum_{(x,y)} I^{(c)}_{(x,y)}$ were the mean intensity of background image and the considered image, respectively. The values λ and β were empirically set to achieve the best performance for all test sets.

The segmenting threshold values were set between σ and 3σ and automatically adjusted on the condition of light intensity. In our experiment, the values λ and β were adjusted to 3 and 20,

(a)



(b)



(c)

Figure 3.2: Human silhouette segmentation and movement detection. (a) From left to right, registered background, walking in frame 32, and frame 35, respectively. (b) Human silhouette segmenting after applying background subtraction technique and its posture envelope. (c) Movement of human silhouette after applying temporal differencing technique and its temporal posture envelope, respectively.

respectively, which were more suitable for our video sample data sets and others as well. This approach helped to solve the auto-white balance problem from the camera as well. The problem happened when the actor changes the distance far from the camera, making the intensity of backgrounds in considered image and registered image significantly different. This adaptive threshold effectively enhanced the actual human silhouette in ROI. After the human region was obtained, the dilation was applied to increase the contour and to improve the important feature extraction.

3.1.2 Temporal Difference

To detect any moving postures such as walking in a considered image frame, it was necessary to detect the value difference of a pixel at least two consecutive times. Bobick and Davis [58] presented a method using a temporal template for the human movement recognition. The idea of this method was to subtract the intensity of the same pixel at two or three adjacent frames. Then, the result was determined by a threshold whether there was a movement or not.

Let $I^{(c)}_{(x,y)}(t)$ be the intensity of pixel located at coordinate (x, y) of a considered image frame at time t. The temporal difference of a pixel at coordinate (x, y), denoted by $TD(x, y)$, can be simply defined as follows:

$$TD(x,y) = |I^{(c)}_{(x,y})(t) - I^{(c)}_{(x,y)}(t-1)| \tag{3.6}$$

Although this method was limited by the low speed movement, it worked very well for periodic movement detection. For this reason, this method was adapted for detecting the moving action in each posture, especially walking posture. When $TD(x, y)$ was zero, it implied that there is no movement. Thus, the walking posture was impossible to appear in the frame. For this reason, the walking posture could be completely eliminated from the other stationary postures.

Our features were based on the observation that the outline of human silhouette or human region can be bounded by an envelope. Each posture provides rather unique width and length of this bounding envelope. Hence, before extracting the features from the human region, a bounding envelope called *posture envelope* was drawn to cover all pixels of human region. This posture envelope was created after the background subtraction process. Similar to posture envelope, another envelope called *temporal posture envelope* could also be created after the temporal difference process.

3.2 Feature Extraction

The features used in our recognizing process were based on the geometrical information of both posture envelop and temporal posture envelop [93]. Obviously, five postures could be distinguished by the related features extracted from the lengths and widths of the posture envelope and temporal posture envelope. In general case, six related features were enough for five main posture discrimination. Two of them were extracted from the width of two sub-envelops covering upper and lower parts. However, to enhance the accuracy of recognition, each envelop was decomposed into three sub-envelops covering the head, torso, and legs parts. The sizes of these sub-envelops would be discussed in the experiment section. The value of each feature was given in the following definitions.

Definition 1. Let $(x_u^{(t)}, y_u^{(t)})$ and $(x_l^{(t)}, y_l^{(t)})$ be the coordinates of the upper left corner and the lower right corner of the posture envelope of the frame at time t, respectively. At frame t, the length, denoted by $l^{(t)}$, and the width, denoted by $w^{(t)}$, of the posture envelope are defined as follows.

$$l^{(t)} = |y_l^{(t)} - y_u^{(t)}| \tag{3.7}$$

$$w^{(t)} = |x_l^{(t)} - x_u^{(t)}| \tag{3.8}$$

(a)

(b)

Figure 3.3: An example of sitting posture, bending posture, and the both upper part envelope and lower part envelope, respectively. (a) The sitting posture (*left image*) and the corresponding envelops (*right image*). (b) The bending posture (*left image*) and the corresponding envelops (*right image*).

Definition 2. The length-width ratio of frame t, denoted by $LW^{(t)}$, is equal to

$$LW^{(t)} = \frac{l^{(t)}}{w^{(t)}} \tag{3.9}$$

Definition 3. The normalized length, denoted by $\widehat{l}^{(t)}$ and the normalized width, denoted by $\widehat{w}^{(t)}$, of frame t are defined as follows.

$$\widehat{l}^{(t)} = \frac{l^{(t)}}{\max_{\forall i}\{l^{(i)}\}} \tag{3.10}$$

$$\widehat{w}^{(t)} = \frac{w^{(t)}}{\max_{\forall i}\{w^{(i)}\}} \tag{3.11}$$

Definition 4. The difference between the length and width of a posture envelope at frame t, denoted by $DLW^{(t)}$, is defined as follows.

$$DLW^{(t)} = |\widehat{l}^{(t)} - \widehat{w}^{(t)}| \tag{3.12}$$

A posture envelope may not be effective when being applied to sitting and bending postures because the sizes of both posture envelopes may be the same in length and width. Hence, the following two additional features must be introduced.

Definition 5. The upper part width of frame t, denoted by $WU^{(t)}$, is defined as the width of a smallest envelope, whose y-coordinate of the upper left corner is at $y_u^{(t)}$ and the length is equal to $\frac{1}{3}l^{(t)}$, that can fit the human region. This envelope is called *upper part envelope*.

(a)

(b)

Figure 3.4: An example of sitting posture, bending posture, and the envelops of their heads, torsos, and legs, respectively. (a) The sitting posture (*left image*) and the corresponding envelops (*right image*). (b) The bending posture (*left image*) and the corresponding envelops (*right image*).

Definition 6. The lower part width of frame t, denoted by $WL^{(t)}$, is defined as the width of a smallest envelope, whose y-coordinate of the lower right corner is at $y_l^{(t)}$ and the length is equal to $\frac{2}{3}l^{(t)}$, that can fit the human region. This envelope is called *lower part envelope*.

Two examples of $WU^{(t)}$ and $WL^{(t)}$ are shown in Figures 3.3. The left image shows the posture and the right image shows the envelopes of $WU^{(t)}$ and $WL^{(t)}$.

To enhance the posture classification method to distinguish five main postures and additional five sub-postures of lying, three sub-envelops covering the head, torso, and legs parts were defined for solving the overlap between sitting and bending postures and sub-lying postures in different viewing angles as the examples in Figure 3.15 and 3.16.

Definition 7. The *head part envelope* was defined as the width of a smallest envelope, whose y-coordinate of the upper left corner was at $y_u^{(t)}$ and the length was equal to $\dfrac{1}{6}\ l^{(t)}$, that can fit the human region.

Definition 8. The *torso part envelope* was defined as the width of a smallest envelope, whose y-coordinate of the upper left corner is at $y_u^{(t)}+\dfrac{1}{6}\ l^{(t)}$ and the length was equal to $\dfrac{2}{6}\ l^{(t)}$, that could fit the human region.

Definition 9. The *legs part envelope* was defined as the width of a smallest envelope, whose y-coordinate of the lower right corner was at $y_l^{(t)}$ and the length was equal to $\dfrac{3}{6}\ l^{(t)}$, that could fit the human region.

Definition 10. Let W_{head}, W_{torso}, and W_{legs} be the widths of head, torso, and legs parts of the envelops of head, torso, and legs part. The normalized difference between the longest width and medium width of three sub-envelopes, denoted by $\widehat{DLM}^{(t)}$, of frame t was defined as follows.

$$\widehat{DLM}^{(t)} = \frac{|max\{W_{head}, W_{torso}, W_{legs}\} - med\{W_{head}, W_{torso}, W_{legs}\}|}{max\{W_{head}, W_{torso}, W_{legs}\}} \tag{3.13}$$

Here, *med* was the operator for finding the medium value from a 3-element set and *max* was the operator for finding the maximum value from a 3-element set.

Definition 11. The normalized difference between the longest width and shortest width of three sub-envelopes, denoted by $\widehat{DLS}^{(t)}$, of frame t was defined as follows.

$$\widehat{DLS}^{(t)} = \frac{|max\{W_{head}, W_{torso}, W_{legs}\} - min\{W_{head}, W_{torso}, W_{legs}\}|}{max\{W_{head}, W_{torso}, W_{legs}\}} \tag{3.14}$$

Here, *min* was the operator for finding the minimum value from a 3-element set.

Definition 12. The area of a posture envelope at frame t, denoted by $A^{(t)}$, was defined as follows.

$$A^{(t)} = l^{(t)} \times w^{(t)} \tag{3.15}$$

Definition 13. The normalized area of a posture envelope at frame t, denoted by $\hat{A}^{(t)}$, was defined as follows.

$$\hat{A}^{(t)} = \frac{A^{(t)}}{\max_{\forall i}\{A^{(i)}\}} \tag{3.16}$$

Definition 14. Let $(x_c^{(t)}, y_c^{(t)})$ be the coordinate of the centroid of the posture envelope of the frame at time t. At frame t, the length from the centroid to the lower line of posture envelop, denoted by $l_c^{(t)}$, was defined as follows.

$$l_c^{(t)} = |y_l^{(t)} - y_c^{(t)}| \tag{3.17}$$

Definition 15. Let $(x_{\tau u}^{(t)}, y_{\tau u}^{(t)})$ and $(x_{\tau l}^{(t)}, y_{\tau l}^{(t)})$ be the coordinates of the upper left corner and the lower right corner of the *temporal posture envelope* at frame t, respectively. The temporal width at frame t, denoted by $\tau w^{(t)}$, of the template posture envelope was defined as follows.

$$\tau w^{(t)} = |x_{\tau l}^{(t)} - x_{\tau u}^{(t)}| \tag{3.18}$$

Definition 16. The temporal difference width of a temporal posture envelope at frame t, denoted by $TDW^{(t)}$, was defined as follows.

$$TDW^{(t)} = |\tau w^{(t)} - \tau w^{(t-1)}| \tag{3.19}$$

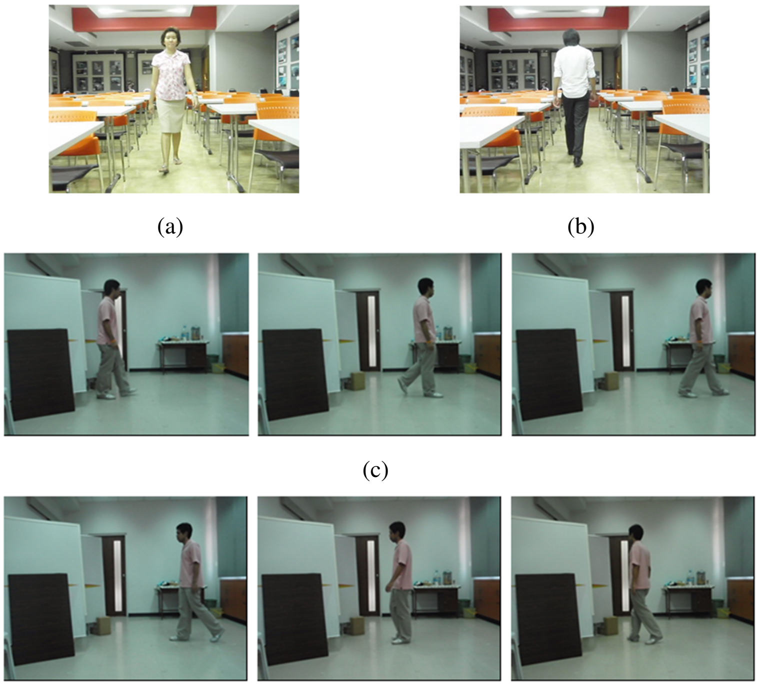
3.3 Posture Recognition

A motion posture as walking and four stationary postures such as standing, lying, sitting, and bending are concerned in this research. The 5-posture classification algorithm based on a decision tree is proposed. The criteria and all thresholds in the algorithm is based on the following observation in the experiments. To demonstrate and testing the efficiency of the proposed algorithm, many short video clips consisting of a mixture of continuous postures were constructed as a new human activity video database based on four main conditions, such as 1) clear posture condition 2) complex posture condition 3)walking testing condition and 4)various sleeping postures testing condition. The details of our database in each condition will be described in the next chapter. All of conditions, the general five postures classification algorithm is achieved to classify the posture into five main classes in high percentage of recognition. Moreover, the Modification of proposed algorithm for classifying ten classes of postures, five main postures and five sub-postures of lying, is also proposed as an example of modifying the general method. The classification steps are described as follow.

3.3.1 General Proposed 5-Posture Classification Algorithm

This method is developed to classify the human posture in the video scene into five main postures such as *walking, standing, lying, sitting,* and *bending*. To classify the movement period from the stationary period, the first posture that we concern is *walking*. The walking postures are limited to *walking towards* the camera, *walking away* from the camera, *left-to-right walking*,and *right-to-left walking*. To make it easy to refer the types of walking, the *walking towards* the camera and *walking away* from the camera will be named *type-1 walking*, whereas the *left-to-right walking* and *right-to-left walking* will be named *type-2 walking*. Figure 3.5 illustrate an example of the studied walking posture in both types.

When a person is walking, the $TDW^{(t)}$ and the area of posture envelope $(A^{(t)})$ will be changed according to time frames due to the distance measure with respect to the camera. But, for the other postures, the $TDW^{(t)}$ is rather constant. With this observation, walking posture must be firstly analyzed before the other postures. However, the area of a posture envelope in the case of walking posture of each person is different. The problem is how to determine whether a given sequence of frames belongs to walking posture. To obtain a simple decision rules for this problem, the areas of posture envelope for walking posture of several experimental persons are plotted against the time frames. Each line in Figure 3.6, denoted by a series of "$\times$" symbols, represents the curve of area of posture envelope (vertical axis) versus each time frame (horizontal axis) for each experimental person when walking towards the camera. Similar to Figure 3.6, Figure 3.7 shows the curves of area of posture envelope when walking away from the camera. Notice that all curves of the same walking posture have a similar shape. All these curves can be combined and mathematically represented by another curve called *standard behaviour curve* (SBC). This SBC will be used for determining whether a person is walking or acting the other postures. Based on the data from 10 walking persons, the following high degree polynomial is created as a representative of these curves.



(a) (b) (c) (d)

Figure 3.5: An example of walking postures of both types. (a) Type-1: walk towards the camera, (b) Type-1: walk away from the camera, (c) Type-2: left-to-right walking, (d) Type-2: right-to-left walking.

Here, x is the frame number, a_i is the coefficient of term i, and y is the area of posture envelope in frame x. Degree of order 20 is most empirically suitable for these experimental data. The thick lines in Figures 3.6 and 3.7 are the SBCs. For the other postures besides walking, their SBCs are just a straight line since there is no area change due to those postures. Hence, three SBCs are created, one for walking towards, one for walking away, and one for the other postures.

Figure 3.8 illustrates an example of a test curve as denoted by a thick curve for determining its posture and the other three standard behaviour curves denoted by a series of "*", "+", and "·" for walking towards, walking away, and the other postures, respectively. It is obvious that the shape of the test curve is similar to the curve of walking away.

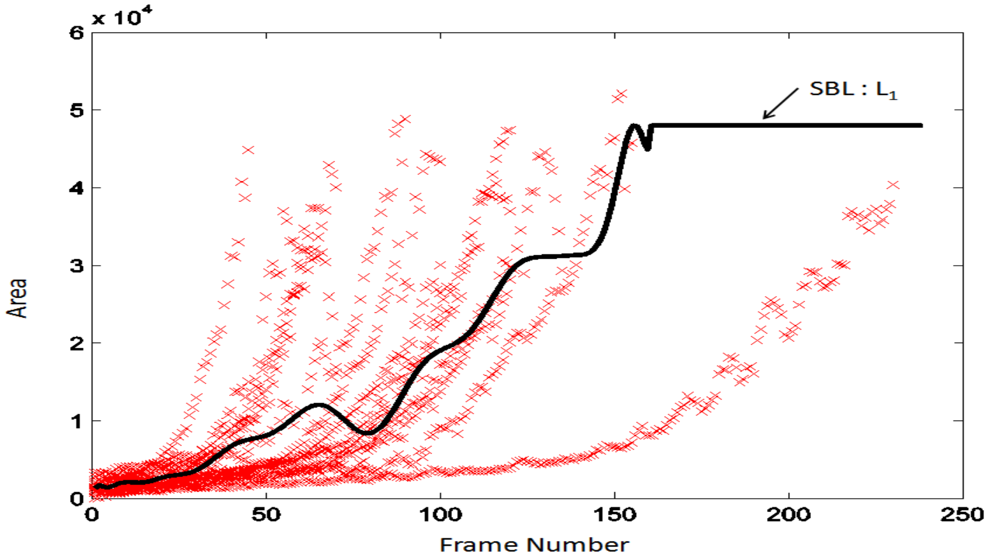


Figure 3.6: Curves, denoted by a series of ”×” symbols, of area of posture envelope versus time frames of the experimental persons for walking towards the camera and the Standard Behaviour Curve denoted by a thick curve.

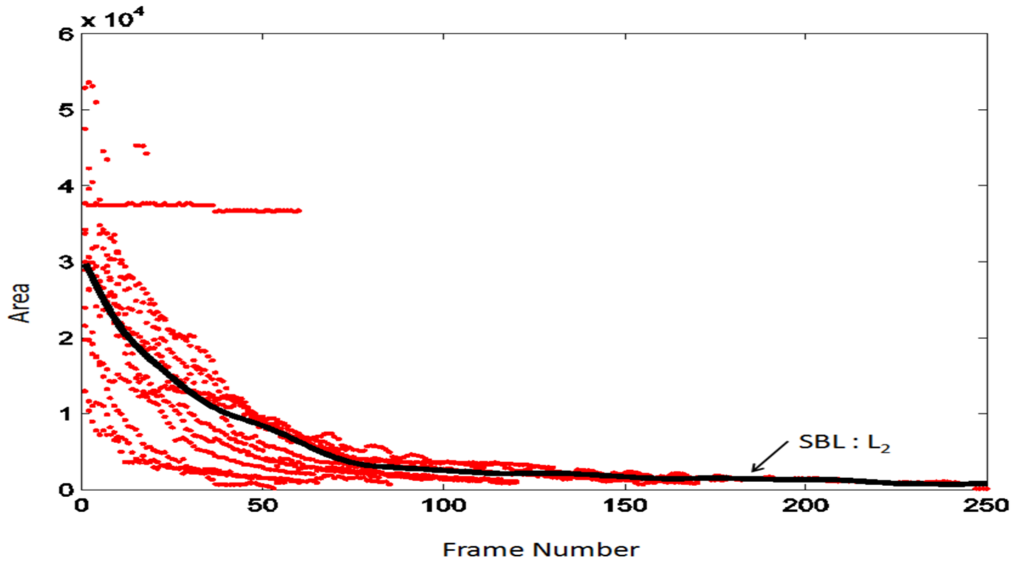


Figure 3.7: Curves, denoted by a series of ”×” symbols, of area of posture envelope versus time frames of the experimental persons for walking away from the camera and the Standard Behaviour Curve denoted by a thick curve.

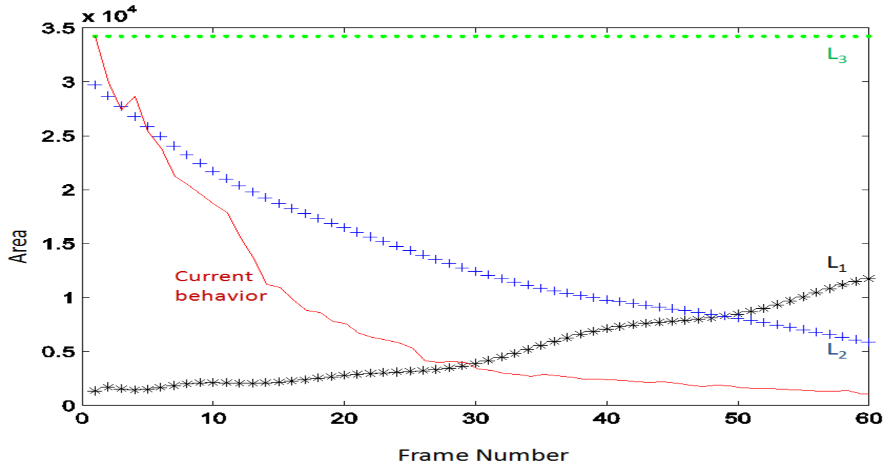


Figure 3.8: Three SBCs and a sample testing curve comparison : L_1(black line, "*"), L_2(blue line, "+"), L_3(green line, "·") and current behavior (red line), for this case, the current curve is similar to L_2

To determine the similarity of both curves, each curve is captured by a posture vector whose elements are the y-coordinates. Then, the angle between the test vector and the SBC vector is computed. Let $\mathbf{L}_1$, $\mathbf{L}_2$, $\mathbf{L}_3$, and $\mathbf{T}$ be the posture vectors for walking towards, walking away, other postures, and the considered posture. The angles between $\mathbf{T}$ and other posture vectors are defined as follows:

$$\theta_1 = arccos\left(\frac{\mathbf{L}_1 \cdot \mathbf{T}}{|\mathbf{L}_1||\mathbf{T}|}\right) \tag{3.20}$$

$$\theta_2 = arccos\left(\frac{\mathbf{L}_2 \cdot \mathbf{T}}{|\mathbf{L}_2||\mathbf{T}|}\right) \tag{3.21}$$

$$\theta_3 = arccos\left(\frac{\mathbf{L}_3 \cdot \mathbf{T}}{|\mathbf{L}_3||\mathbf{T}|}\right) \tag{3.22}$$

The correct posture is finally determined by $\arg(\min_{1 \leq i \leq 3}\{\theta_i\})$.

Now, we noticed that the area of a posture envelope in the case of walking away and walking towards are clearly different. Some examples of the area versus frame number curves in both walking types were shown as Figure 3.8. In conclusion, the shapes of *walking curve* (SBL), that observed from 10 experimental person are alike. In case of walking towards the camera, the area of posture envelope polynomially increases as a function of frame number. On the other hand, the area of the posture envelope decreases rapidly in case of walking away from the camera. Experimentally, it is noticeable that the variance of envelop lengths computed from all frames in the walking curve is a very large number. Therefore, the difference value can be used to filter *type-2* walking.

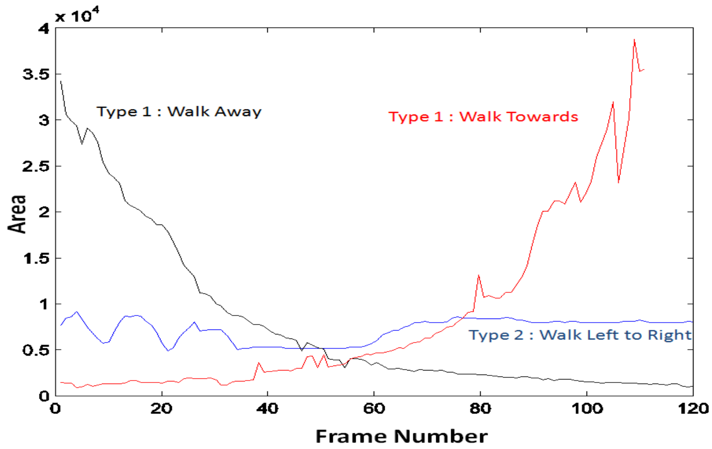


Figure 3.9: Comparison of three walking curves of both walking types : *(right-line)* Type-1: walk towards; *(left-line)* Type-1: walk away; and Type-2: walk left to right. The x-axis denotes the time frame and the y-axis denotes the area of posture envelope.

When a person is walking from left to right or from right to left, the swinging of person's legs and arms makes the width of the posture area and its length-width ratio periodically changed, obviously illustrated in Figure 3.10. Therefore, to identify this type of walking, the periodic change of length-width ratios of a series of frames must be detected. The periodic detection part discussed by Haritaoglu, Harwood, and Davis [73] was adapted and modified to solve this problem for the effective walking posture recognition in the current frame. The periodic detection part in our method is applied by counting the recurrent frequency as the periodicity of the current *length-width ratio (LW)*curve. If the current *LW* curve is similar to the periodic function then the current posture is classified as *type-2 walking*. The detail of periodic change detection was given in **Algorithm 1**.

The *start* and *stop* points in *Algorithm 1* are determined by the condition of $TDW^{(t)} > 0$.

After filtering the walking postures, the stationary postures, i.e. standing, lying, sitting, and bending, are identified by comparing their corresponding $LW^{(t)}$, $DLW^{(t)}$, $WU^{(t)}$, and $WL^{(t)}$ with thresholds. The threshold values in Algorithm 2 are empirically obtained from the postures of ten individual persons. The postures are orderly filtered and identified according to the decreasing sequence of threshold values. In **Algorithm 2**, the following filtering order is executed: standing, lying, sitting, and bending.

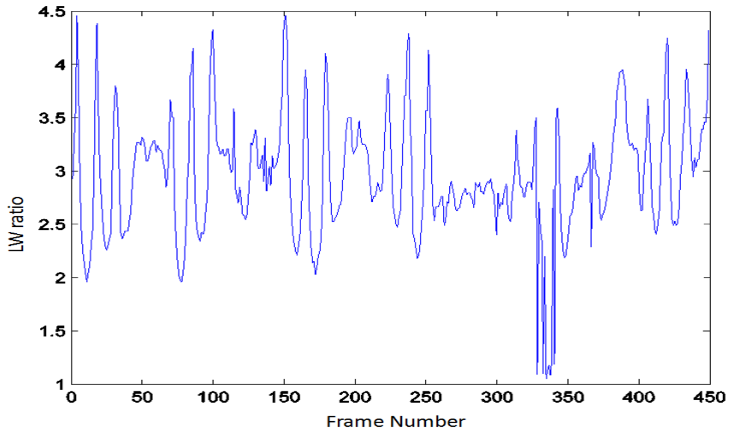


Figure 3.10: Length-width ratio graph of sample video performing walking in Type-2. The x-axis denotes the time frame and the y-axis denotes the length-width ratio.

The overall framework of the general proposed 5-posture classification method is concluded as Figure 3.11. For motion part discrimination, the feature as Temporal Difference Width (TDW) is applied. The temporal differencing segmentation method is adapted for extracting the TDW by the temporal posture envelope because of the advantage in sensitivity movement between adjacent frames. When a moving period is available, the TDW is non-zero. So the *walking* posture is possible to appear in that time. To identify the current image is motion posture; the current period graph is compared with the three Standard Behaviour Curves (SBCs), such as walking towards, walking away, and other postures like the global determination. If the walking takes too short or too long period, the local determination using periodically function is adapt to confirm the motion posture at current frame. In this experiment, we assume that the background is known. Thus the background subtraction technique is applied to segment the human silhouette from the background. Then, the important features are extracted from the posture envelope is constructed from the human outline. The important features that use for postures recognition is composed of the Length-Width ratio(LW), the Difference between the Length and Width (DLW), the Width of Upper part (WU) and Lower part (WL) and the Area of human envelope (A). The Area graph is used for representing the current period graph which is compared to three SBCs. If the structure of the area graph is similar to other postures, the four stationary postures are considered instead. The four stationary postures comprise of *standing*, *lying*, *sitting*, and *bending*. Only LW feature is enough for *lying* posture discrimination. The *standing* posture is examined from LW and DLW. Whilst the LW, WU and WL features are considered together for distinguishing between the *sitting* and *bending* postures.

Algorithm 1: Periodic Detection for *Type-2 Walking* Postures

begin

1. Set the $start$ and $stop$ points of the *movement period.*
2. Let $cutframe = start - stop$.
3. Let $mean = \frac{1}{cutframe} \sum_{t \in [start, stop]} LW^{(t)}$. The values of $LW^{(t)}$ are from the training data set.
4. Let $\Delta^{(t)} = LW^{(t)} - mean$.
5. Let $count = 0$.
6. **For** $i = 1$ to $cutframe - 1$
7. Let $u = \Delta^{(i)} \times \Delta^{(i+1)}$.
8. **If** $u \leq 0$, **Then** $count = count + 1$.
9. **Endfor**
10. Let $period = \frac{count}{2}$.
11. **If** $period > 1$ **then** identify as $walking$.

end

Algorithm 2: Posture Recognition

begin

1. **If** $TDW \neq 0$ **then** identify as the *movement period*.
2. **If** *movement period* ≥ 40 frames **then** compute the angles according to equations (17), (18), and (19).
3. **If** the test curve is similar to Standard behaviour curve of L_1 or L_2
4. **then** identify the posture as *walking*.
5. **else** detect whether it is types 3 or 4 walking posture.
6. **If** $LW^{(t)} \geq 0.5$ **and** $DLW^{(t)} > 0.4$
7. **then** identify the posture as *standing*.
8. **else If** $LW^{(t)} < 0.5$
9. **then** identify asthe posture as *lying*.
10. **else If** $WU^{(t)} < WL^{(t)}$ **and** $DLW^{(t)} < 0.2$
11. **then** identify the posture as *sitting*.
12. **else**identify the posture as *bending*.

end

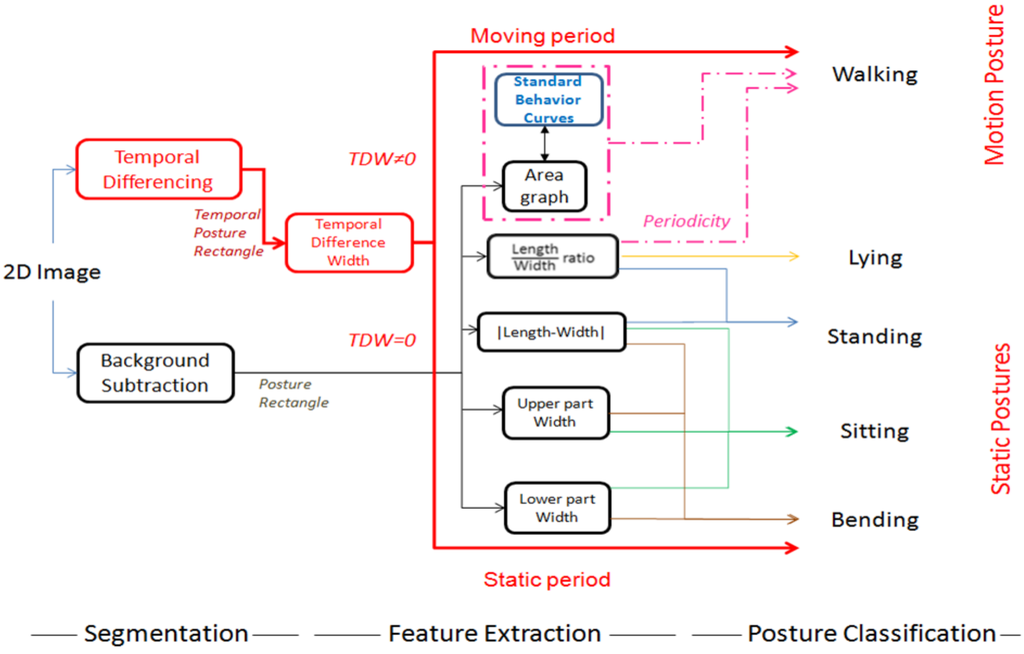


Figure 3.11: The framework of our proposed method.

3.3.2 Modification of Proposed Posture Classification Algorithm

The proposed method can be enhanced to recognized the specific posture in five main postures. In this study, the sleep-positions detection is an example. The considered sub-postures of lying or sleep-positions are comprised of supine, prone, log, fetal, and reclining, respectively. Figure 3.13a shows some examples of length-width ratio of body envelop of five considered postures and Figure 3.13b shows the length-width ratio of body envelop for different lying postures. If both Figures are combined as shown in Figure 3.12, the length-width ratio (*LW ratio*) of the envelop of each time frame is plotted against the frame number. A curve is represented the characteristics of all acting postures. Notice that, a challenge problem is the confusing of the same *LW* ratio between some lying postures and some non-lying postures. An important factor is caused from the various viewing angle of capturing the postures. To improve the performance of the posture classification of our method, the modify features, such as $\widehat{DLM}^{(t)}$, $\widehat{DLS}^{(t)}$,l_c, are used instead of WU, WL. The thresholds and decision tree in the general algorithm are also adjusted from the observation in the experiment training in condition 4 dataset. Each challenge problem is discussed as follows.

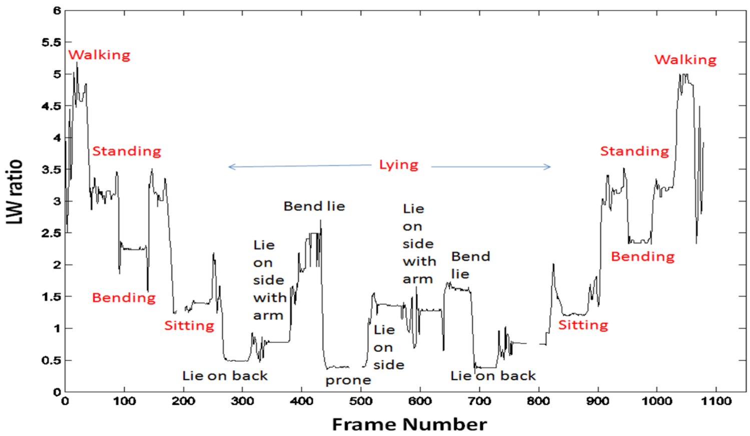


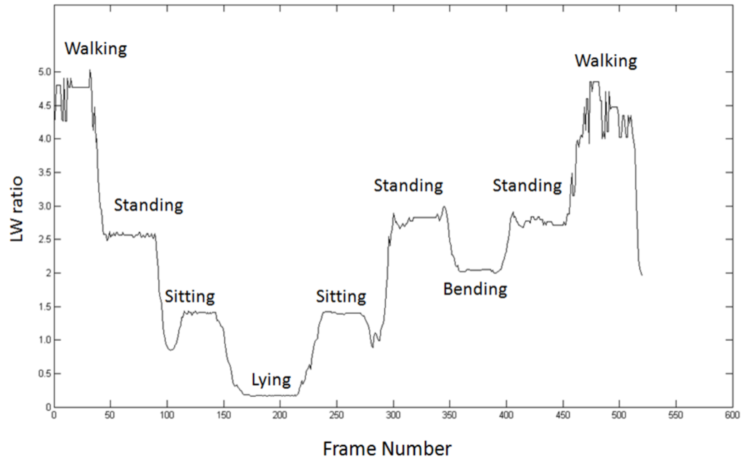
Figure 3.12: An example of length-width ratio graph of complex postures, mixed with five main postures and five lying postures at 90° viewing angle. The x-axis denotes the time frame and the y-axis denotes the length-width ratio.

Standing versus Bending Postures

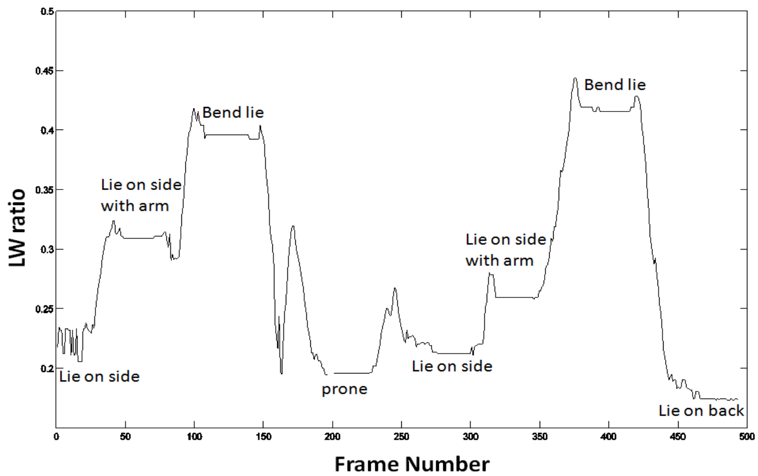
In general case of standing posture, the values of $LW^{(t)}$, $DLW^{(t)}$, and $l_c^{(t)}$ are enough for discriminating the standing posture from bending posture. Considering the $LWratio$ graph of changing posture in Figure 3.13a, The standing and bending can be identified by some thresholds. The threshold values of LW and DLW in **Algorithm 3** are empirically obtained from the postures of ten individuals. However, in some situations shown in Figure 3.14, the standing posture in a far distance resembles the bending posture in 90° viewing angle in a near distance. The sizes of both posture envelopes are almost the same. The dependent features are not enough to distinguish both postures. So, the normalized feature such as $\hat{A}^{(t)}$, is more suitable to consider for discriminating the standing posture than $l^{(t)}$, $w^{(t)}$, and $LW^{(t)}$ because it compares $A^{(t)}$ with $\max_{\forall i} A^{(i)}$ value of that video clip. The normalized area value of standing posture is larger than that of the bending posture and then, both postures can be identified with some thresholds. The threshold value was presented in **Algorithm 3**.

Bending versus Lying Postures

Similar to the previous case, the bending and lying postures at 0° to 81° of viewing angles can be clearly discriminated by the threshold of LW. However, in 90° viewing angle as shown in Figure 3.12. The $LW^{(t)}$ of bending posture and fetal lying posture are approximate the same value. To discriminate both ambiguous postures, the silhouette is divided to three part envelopes. Figure 3.15 presents the examples of three sub-envelopes of bending and fetal lying postures which are head, torso, and legs part envelope. It can be seen that the way these three



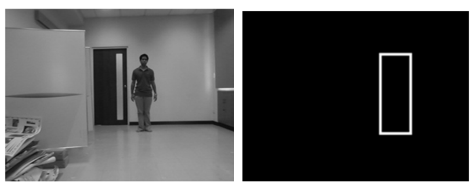
(a)



(b)

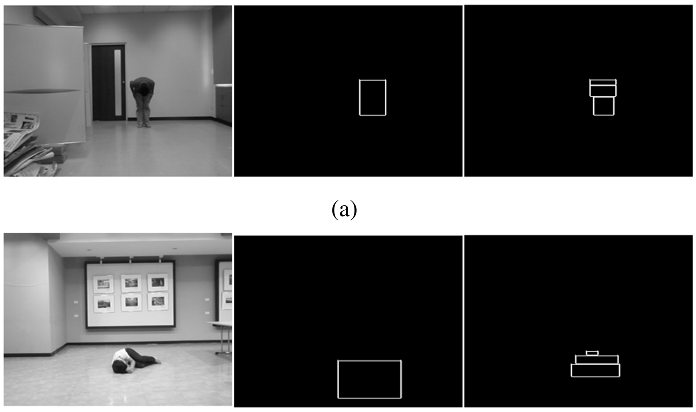
Figure 3.13: An example of length-width ratio graph at 0° viewing angle. The x-axis denotes the time frame and the y-axis denotes the length-width ratio. (a) Five clear main postures. (b) Five clear lying postures.

(a)



(b)

Figure 3.14: An example of bending and standing in a far distance postures with their corresponding posture envelopes. (a) Bending in 90° viewing angle in near distance and its posture envelope. (b) Standing in far distance and its posture envelope.



(a)

(b)

Figure 3.15: An example of bending and fetal lying postures with their corresponding posture envelopes and three sub-envelopes. (a) Bending in far distance at $90°$ viewing angle. (b) Fetal lying at in the middle distance at $45°$ viewing angle.

envelops are stacked for bending posture is opposite to that of fetal lying posture. Hence, the values of $\hat{A}^{(t)}$ and $l_c^{(t)}$ must be concerned to distinguish the bending posture from fetal lying posture. The detail is given in **Algorithm 3**.

Sitting versus Lying Postures

In Figure 3.12, the $LW^{(t)}$ value of sitting posture is similar to that of reclining lying and fetal lying postures. This problem happens in 90° viewing angle. To distinguish the resembled postures as sitting and lying postures in 90°, the silhouette is divided to three part envelopes as illustrated in Figure 3.16 to capture the local features of the postures. The sitting and lying postures are identified by the thresholds of features $\hat{A}^{(t)}$, $\widehat{DLM}^{(t)}$and $\widehat{DLS}^{(t)}$ as written in **Algorithm 3**.

In *Algorithm 3*, a decision tree for classifying five main postures was constructed from the training data set and the concepts previously discussed. Each posture is identified by comparing their corresponding $LW^{(t)}$, $DLW^{(t)}$, $\widehat{DLM}^{(t)}$, $\widehat{DLS}^{(t)}$, $\hat{A}^{(t)}$, $l_c^{(t)}$ with thresholds. The postures are orderly filtered and identified according to the decreasing sequence of threshold values. The following filtering order is executed: walking, standing, lying, sitting, bending, and lying in five sleep-positions, i.e. supine, prone, log, fetal, and reclining. The detail of the algorithm is as follows.

The modify algorithm is adapted to the homecare system application in sleep monitoring application. The overall framework of Modification of proposed method, adapted to the sleep monitoring application, is presented as Figure 3.17. The main idea of the posture classification

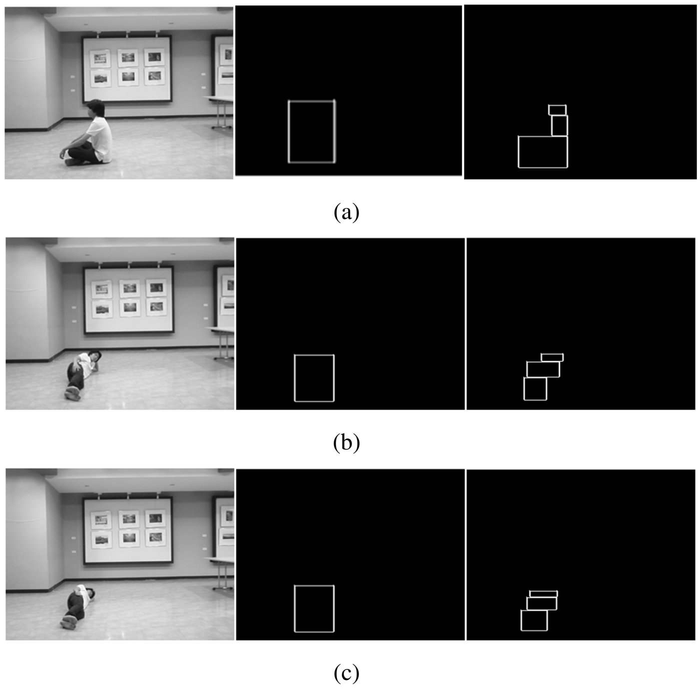


Figure 3.16: An example of sitting, reclining lying, and log lying postures with their corresponding posture envelope and three sub-envelopes at 90° viewing angle in the same distance. (a) Sitting posture. (b) Reclining lying posture. (c) Log lying posture.

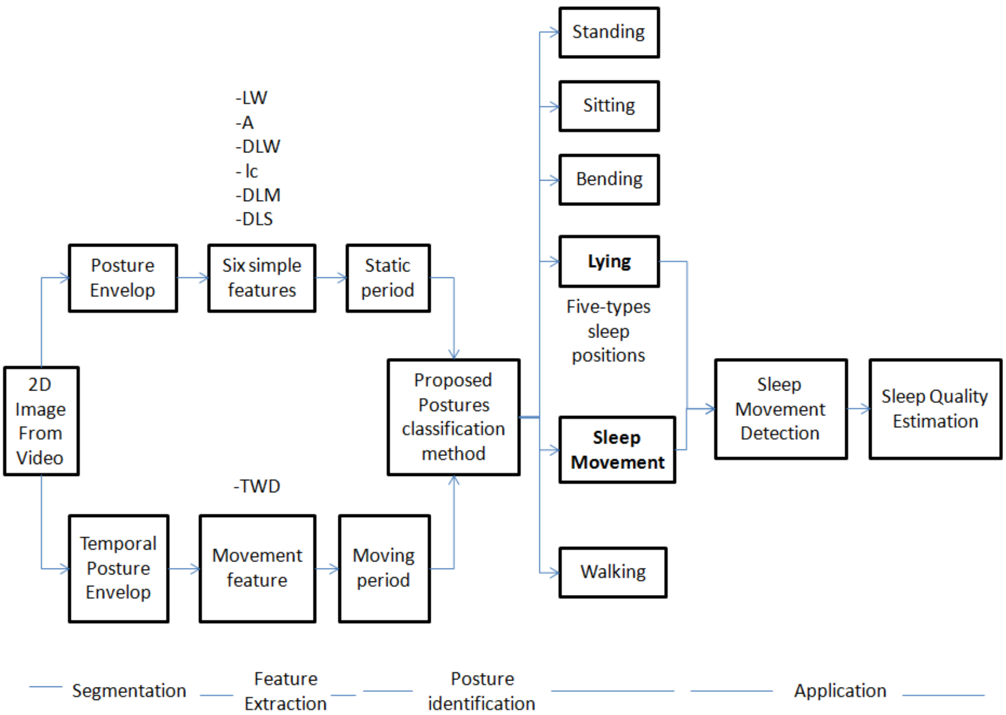


Figure 3.17: The framework of our modified proposed system.

Algorithm 3: Five Main Postures and Five Lying Postures Recognition

begin

1. **If** $TDW\neg 0$ **then** identify as the *movement period.*
2. **If** *movement period* ≥ 40 frames **and** $var(A) > 10^7$.
3. **then** identify the posture as *Type 1 walking*.
4. **else** detect whether it is *Types 2 walking* posture.
5. **If** $LW^{(t)} > 2.4244$ **and** $DLW^{(t)} > 0.4$ **and** $\hat{w}^{(t)} > \hat{l}^{(t)}$
6. **If** $\hat{A}^{(t)} > 0.5$ **or** $l_c^{(t)} > 55$.
7. **then** identify the posture as *standing.*
8. **else** identify the posture as *bending*.
9. **else If** $LW^{(t)} > 0.85$
10. **then** identify the posture as *lying (supine or prone)*
11. **else If** $W_head^{(t)}$ is the biggest envelop **or** $W_legs^{(t)}$ is the smallest envelop.
12. **If** $\hat{A}^{(t)} > 0.5$ **or** $l_c^{(t)} > 40$
13. **then** identify the posture as *bending*.
14. **else** identify the posture as *lying (fetal)*.
15. **else If** $WL^{(t)}$ is the biggest envelop **or** $WU^{(t)}$ is the smallest envelop
16. **If** $\widehat{DLM}^{(t)} > 0.25$ **or** $\widehat{DLS}^{(t)} > 0.6$
17. **then** identify the posture as *sitting*.
18. **else** identify the posture as *lying(log or reclining) end*

is still keeping but increased some features of three sub-envelop for classifying more sleep-position classes. The sleep quality in each night will be estimated from detecting the sleep-positions all night. This application monitors and collects the sleeping information overnight and initially evaluates the quality of sleeping in every night. In the sleep disorder diagnosis, the doctors use the sleeping information for estimating the symptom of the disease and planning the treatment in order to change the sleep routine. The purpose of the treatment is to avoid the sleep disorder disease as Obstructive Sleep Apnea (OSA), insomnia, parasomnias, and circadian rhythm disorder, etc. that bring to be the terrible disease such as heart disease, untreated OSA, high blood pressure, cerebrovascular (CVD), and death from heart attack, etc. [33]. For the technical approach of the sleep monitoring application as mentioned, the system usually stores the human behaviour information during sleeping night long, as the history of the surveillance system. It is the benefit that the patients can live in their home whereas they will be in the care of doctor in the hospital. This application helps to improve the patient's quality of life. In present, many equipments such as the accelerometer or pressure sensor bed, are used for collecting the human information but our method is developed to collect the information directly from the video surveillance. The more information of the sleep monitoring application is explained briefly in Appendic.

Chapter IV

Experimental

In this dissertation, there were three main experiments which test: (1)the efficacy of proposed features, (2)the applicability of our algorithm to various data sets, and (3)the processing time of our algorithm. The first experiment was to demonstrate how feasible the proposed features is when they are applied to classify five main postures and the additional lying postures. Our features were compared with DFT-coefficients classified by ANFIS model. For the second experiment, we aimed to test the performance and the robustness of our classification method in various human action data sets. The following four main data sets: Our data set, Weizmann data set [78], UIUC data set [94], and CMU data set[1], were selected in our experiment. In the third experiment, the time complexity of our algorithm was computed and proved by implementing on the real processing. The details of each experiment are explained as follows.

4.1 Experimental Set-Up

The experimental procedures of proposed method are composed of three steps:

1. programming the codes of algorithm
2. testing the code on the database
3. evaluating performance of algorithm

The efficiency of each method was evaluated by the percentage of accuracy, calculated as

$$\%\text{accuracy} = \frac{\text{correctpredictedposture}}{\text{correctposture}} \times 100 \tag{4.1}$$

4.1.1 Hardware and Software system

1. In the experiment, every algorithm was implemented on the MATLAB© programme.
2. processing on Desktop PC with Intel(R) Core (TM) processor; 7-2600 CPU @ 3.40 GHz, 16 GB. of RAM with 64-bit operating system.
3. The video camera recorder was a compact camera, Olympus µ830 , settled on a tripod with 90-110 centimeters height from the floor (as the eye-level).

[1] Available on http://mocap.cs.cmu.edu/

4. The main captured viewing angle was 0, 45, 90 degrees.

5. The captured video file was set at resolution 320x240 pixels at the rate 15 frames per second and captured in real RGB color space.

4.1.2 Locations

All human actions video was captured at Imaging and Colour Technology Laboratory, at the 6^{th} floor, Museum of Imaging and Printing Technology Building, Department of Imaging and Printing Technology, Faculty of Science, Chulalongkorn University. In every testing conditions, the most contents of background in each clip were not altered during capturing. The luminescence of every room was between 300-800 lux.

Nine sample backgrounds in six rooms are presented as shown in Figure 4.1.

Figure 4.1: The sample captured background of our dataset : Nine backgrounds of six rooms.

4.1.3 Subject

The volunteers were healthy and involved having different physical characteristics. They were informed the consent as to be the participants of the experiment in order to meet ethical requirement. Figure 4.2 shows the fifteen volunteers in different sizes and cloths and in the conditions as follows:

1. wore the normal appeal (did not hold any sensor on the body)

2. performed the four main postures in individual style vary

3. 3-6 meters distance and 0-90 degrees of action direction.
4. number of volunteers : 8 males, 7 females
5. age: 23-35 years old
6. height: 145-185 cm. , weight: 45-85 kg.
7. single actor in a frame

4.1.4 Postures Representative

Definition for difficulty levels:

Basic

A clear posture means the pause posture which easy to decide to posture class immediately. The definition of each posture is defined as follow:

1. *standing* : pull up, standing strength: two legs is closed and two arms are enclosed with the body.
2. *lying* : lie down on the back, two legs is closed and two arms are enclosed with the body.
3. *sitting* : sit down on the floor, two arms are enclosed with the body.
4. *bending* : standing with closed two legs and bend down with enclosed of two arms with the body.

Advance

A complex posture is comprised of 3 types:

1. the posture which in transition state during changing between posture.
2. the posture which has some additional posture, i.e., raise a hand, raise a leg, wave a hand, crawling, various sleep positions, etc.
3. specific individual styles posture.

Figure 4.3 presents the sample posture of the clear postures, main posture, and the complex postures in 3 types.

Determination of defined postures: The most complex postures are ambiguous and difficult to classify to an actual class. So, in the constructing database process, after the actor/actress

Figure 4.2: fifteen volunteers perform actions in different physical characteristics, various size and the texture of cloths.

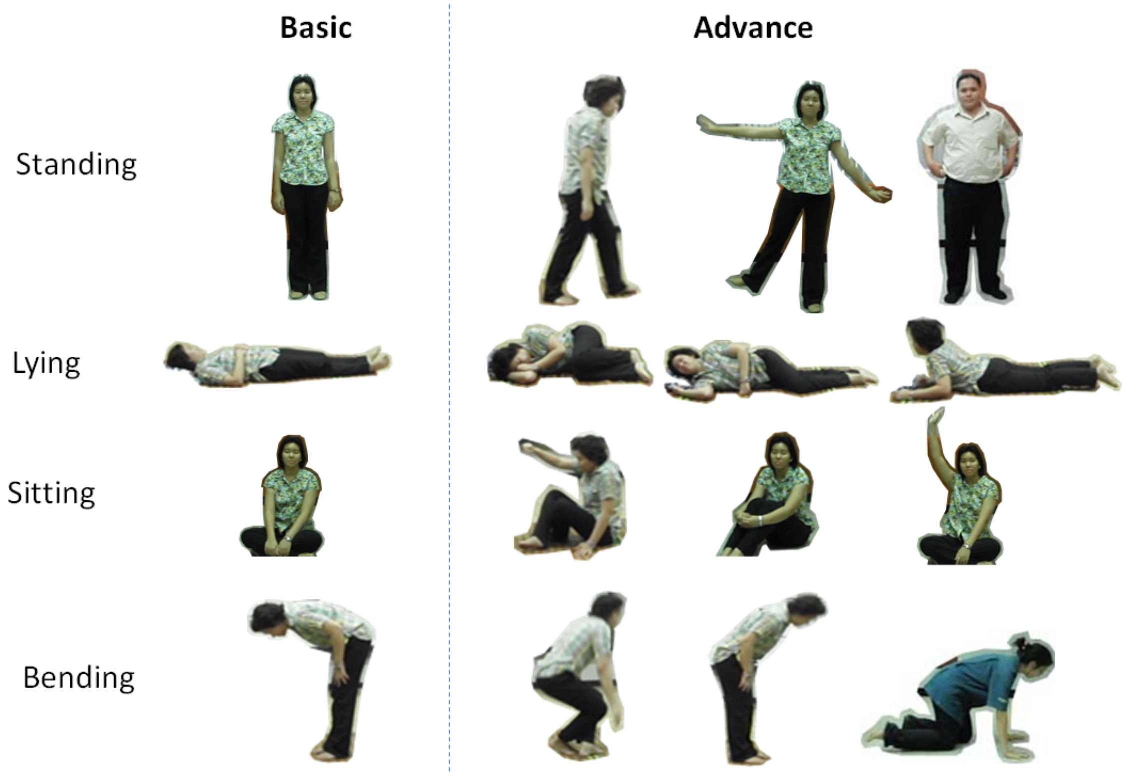


Figure 4.3: presents some sample recognized postures (a) basic posture, (b) advance postures.

had finished performing in the video clip, he/she would judge his/her own postures from the rewind performance video clip. Hence, all answer of postures in the action video came from the confirmation of actor/actress who performed those postures. All advanced postures determined by the actors.

4.2 Human Action Databases

Every experiment was mainly trained and tested on our human actions database. In addition, three other published datasets as Wiezman, UIUC, and CMU, are also the testing data set in second experiment.

There were four databases, selected as the testing set for proposed algorithm. Some clips of our own database was chosen as the training set. The training and testing set were separated from each other. The details of each database is described in next subsections.

4.2.1 New Human Action Database Construction

In our study, we constructed our own data set of human action in five main postures and various styles of five lying postures under different conditions defined in Chapter I for training and testing the capability of proposed algorithms. Our data set consists of 141 clips of 68,500 frames shot from 15 individuals and 9 different indoor background. The background scene was not altered throughout the experiment. The sample nine backgrounds in six rooms are presented as Figure 4.1. To test the applicability of the proposed method, the body size and the texture of the clothes of each person were not controlled. Figure 4.2 shows the fifteen volunteers in different sizes and cloths.

Figure 4.4: The sample captured images different in distances, (left to right) far(7-9 meters), middle(5-7 meters), and near (3-5 meters) distances, respectively.

In this research, we aimed to develop the human posture recognition method to be independent for the size and clothing of subject and the perspective problem, such as distances and viewing angles, from the camera. Hence, the distance of the actors and the camera in each location was not fixed. The video clips were captured in 3 distances, namely near, middle, and far from the camera, within 3 to 7 meters at $0°$, $45°$, and $90°$ viewing angles, respectively.

The video clips for testing the camera distance effect in far(7-9 meters), middle(5-7 meters), and near (3-5 meters) distances are illustrated in Figure 4.4. In addition, the effect of view angles variation was also concerned in three main conditions of viewing angles, i.e. $0°$, $45°$, $90°$. Some images of video clips for testing the robustness of view angle effect of our method is presented in Figure 4.5.

Figure 4.5: The sample captured images different in viewing angles, (left to right) $0°$, $45°$, $90°$, respectively.

Moreover, the following constraints are set as the capturing conditions for our new database construction. Four human actions videos datasets are recording under these conditions. The details of how our data set were collected as follows;

1. A subject performing the postures did not wear any special suit or sensor instruments.

2. The postures can be varied by each individual's styles and viewing angle of the camera.
3. Viewing angles ranged from 0° to 90°.
4. The camera was settled in a tripod with 90-110 centimetres height from the floor.
5. For testing the effect of perspective, a volunteer performed the individual postures with distances of 3 to 9 meters apart from the camera and captured on the 0, 45 and 90 degrees of viewing angles, depending on different purposes of testing set.
6. Each volunteer has different body sizes and can wear in different cloth textures in varied apparel.
7. The video file was captured at resolution 320x240 pixels at the rate 15 frames per second. Each training video was about 1-2 minutes length and each image file was captured in true colors of RGB color space.
8. The most contexts in the background image were not be altered by any means throughout the analysis and experiments.

Our constructed human action database was composed of four dataset for testing with different objectives; (1) clear posture testing data set, (2) complex posture testing data set, (3) walking testing data set, and (4) sleep position testing data set. The details of each data set are explained and some samples are also presented as follow:

Human Action Data Set1: Clear Posture Testing Data Set

There are 8 clips performed by 8 actors in different clothes and backgrounds. The actors performed 4 main stationary postures such as walking, standing, lying, sitting and bending with clear posture. A clear posture means that each actor arbitrarily starts acting at any posture. Once the actor fully performs the posture, he or she pauses for a short moment, approximately 40-60 frames, before changing to the next posture.

Figure 4.6 presents some sample images of video clip in data set of clear posture testing and its length-width ratio graph is shown as Figure 4.7. For example, the clear sitting posture occurred at two frame periods. The first period occurred at frames 110 to 150 and another period was at frames 240 to 300.

Human Action Data Set2: Complex Posture Testing Data Set

There are 21 clips performed by 12 actors in different clothes and different backgrounds. The actors performed five main postures, which were walking, standing, lying, sitting, and bend-

Figure 4.6: Sample images of a video clip from *clear posture testing data set*,at frame (*left to right*)30, 80, 130, 200, 250, 330, 380, 420, 485, respectively.

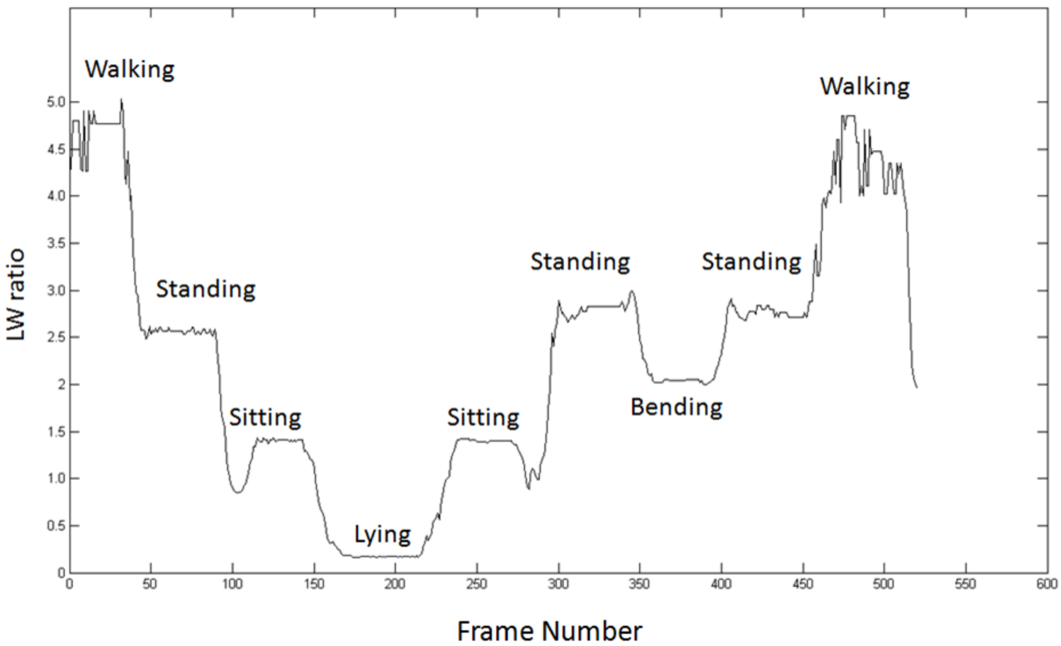


Figure 4.7: An example of length-width ratio graph of clear postures. The x-axis denotes the time frame and the y-axis denotes the length-width ratio.

ing with complex postures. The complex postures are the set of continuous acting of different postures under different distances from the camera.

Figure 4.8: Sample images of a video clip from *complex posture testing data set*,at frame (*left to right*)40,65,80,160,280,335,400,450, respectively.

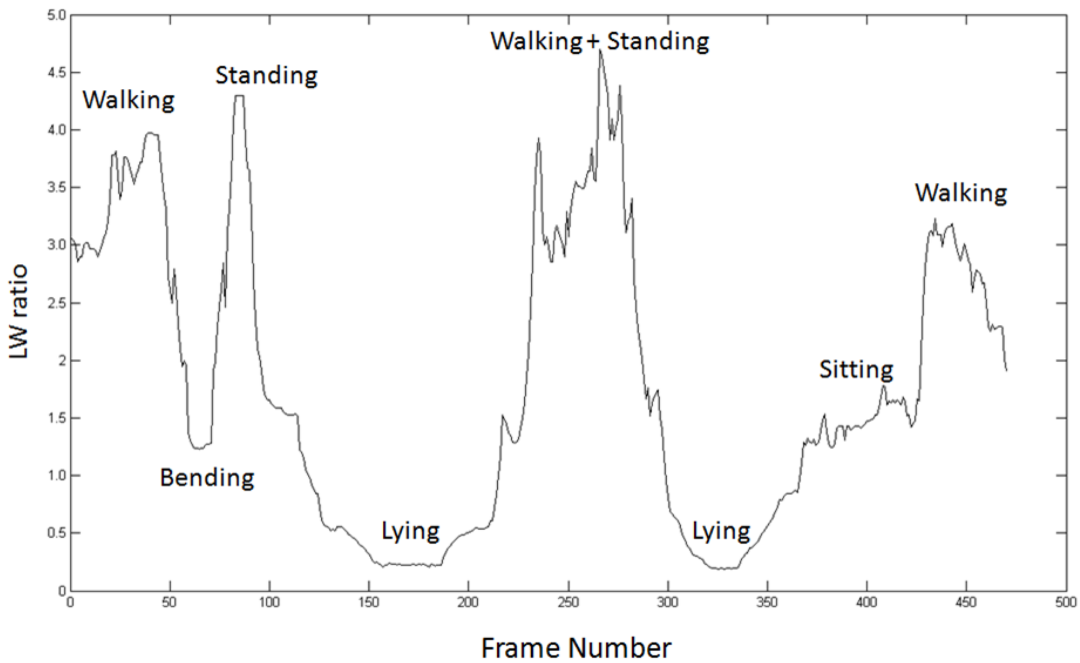


Figure 4.9: An example of length-width ratio graph of complex postures.

Figure 4.8 presents some sample images of video clip in data set of complex posture testing and its length-width ratio graph which is presented in Figure 4.9. the continuous standing, bending, and sitting postures occurred from frames 100 to 200. Obviously, these complex postures are quite difficult to distinguish.

Human Action Data Set3: Walking Testing Data Set

There are 35 clips performed by 8 actors in different cloths and difference stationary backgrounds. The actor performed 3 walking types such as walking around the room with frequent changing distances and viewing angles, walking away from the camera, and walking towards the camera in 90° viewing angle with various distances and speeds of walking.

This data set can be divided into three data subsets: (1)Walking away from the camera, (2)Walking towards the camera, and (3)Complex Walking. The first and second data subsets aims to testing the *Walking Type1* and *Walking Type2* was tested in the third data subset. The sample captured images from a clip of walking away and walking towards as Type1 and their area graphs are illustrated in Figure 4.10 and 4.11.

- **Walking away from the camera.** There were sixteen clips performed by six actors in different clothes and different stationary backgrounds. The actor walked away from the camera in 90 degrees of view angle with various distances and speeds of walking. Figure 4.10 shows a sample image of walking away from the camera.

- **Walking towards the camera.** There were fourteen clips performed by six actors in different clothes and different backgrounds. Each actor walked towards the camera in 90 degrees of view angle with various distances and speeds of walking. Figure 4.11 shows a sample image of walking towards the camera.

- **Complex Walking.** There were five clips performed by three actors in different clothes and different backgrounds. Each actor walked around the room or frequently changes distances and viewing angles from the camera and with various speeds of walking. Figure 4.14 shows the area graph of an example of complex walking.

Figures 4.12shows some images from some sample clips of walking type2 and their length-width ratio graphs plotting against the time.

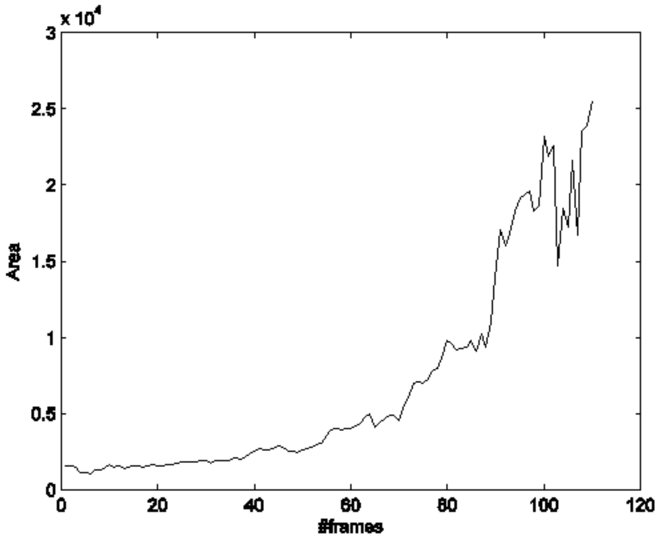
The comparison of area graph of walking in both type is illustrated in Figure 4.13.

The other videos in complex walking data subset are the mixing of walking postures in both 2 types and also some main postures. An example of various walking video clip is presented as LW graph as Figure 4.14

Human Action Data Set 4: Sleeping Position Data Set

There are 100 clips performed by 10 actors in different cloths and difference stationary backgrounds. The actor performed 5 lying postures, i.e. supine lying, prone lying, log lying,

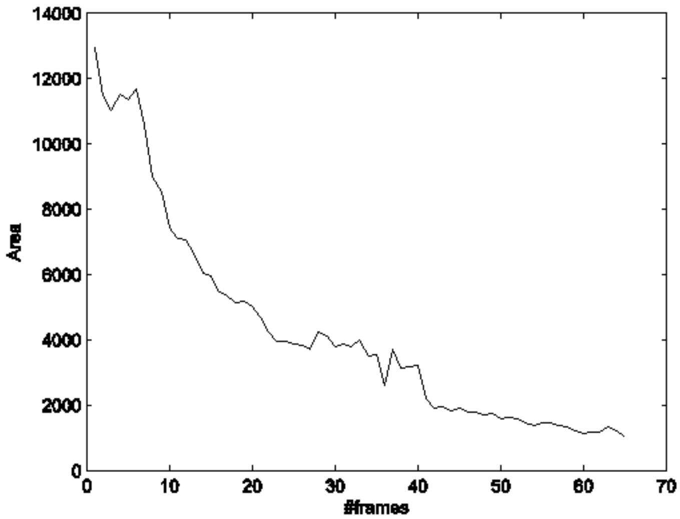
(a)



(b)

Figure 4.10: Sample captured images of two walking Type 1 video clips from *Walking testing data set*: (a)walking away clip at frame 20,60,100, and (b)area graph of walking away clip.

(a)



(b)

Figure 4.11: Sample captured images of two walking Type 1 video clips from *Walking testing data set*: (a)walking towards clip at frame 10,20,60, and (b)area graph of walking towards clip.

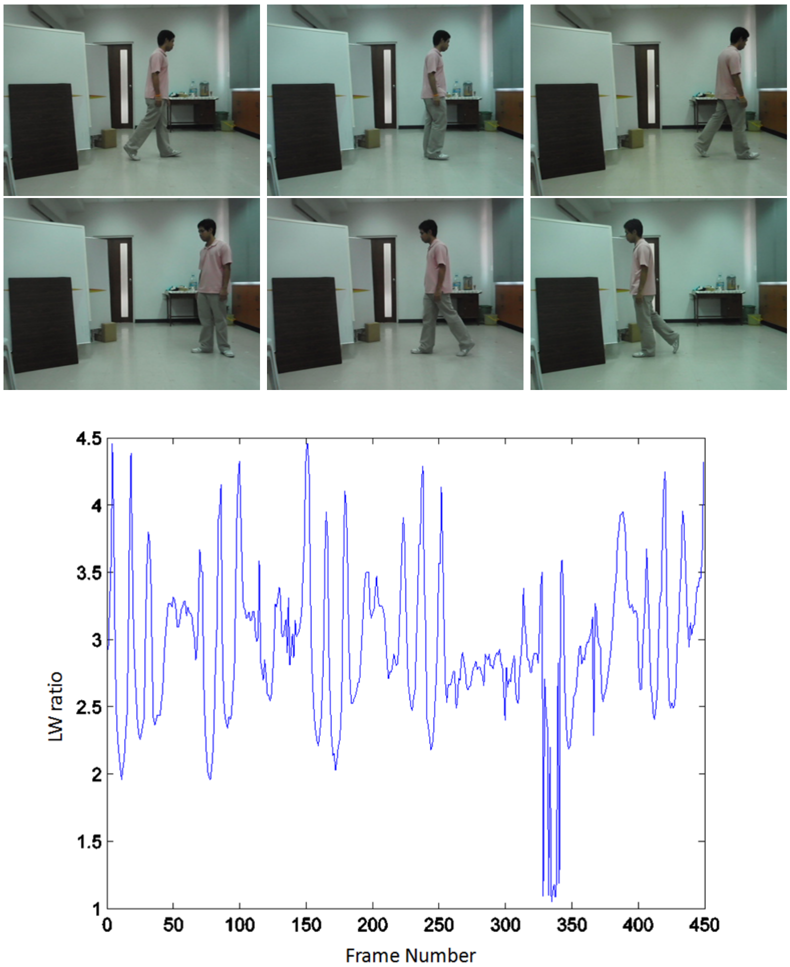


Figure 4.12: Some sample images of walking around the room, a video in *complex walking video data subset* and its LW graph.

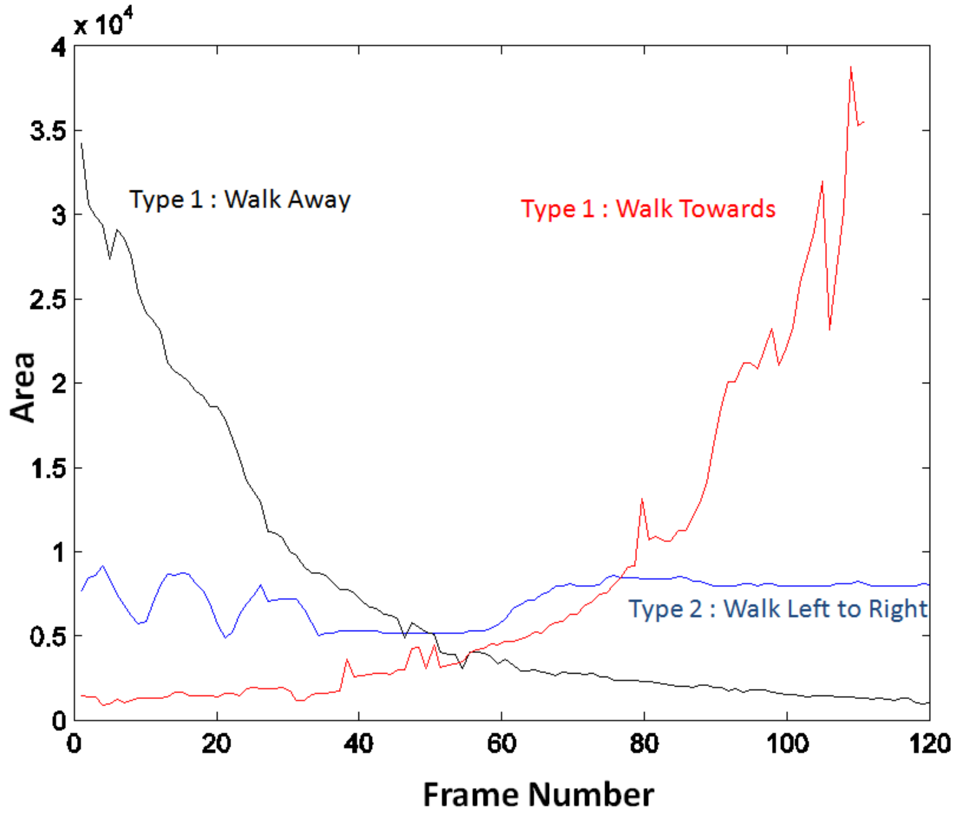


Figure 4.13: Comparison of area graph of both types of walking.

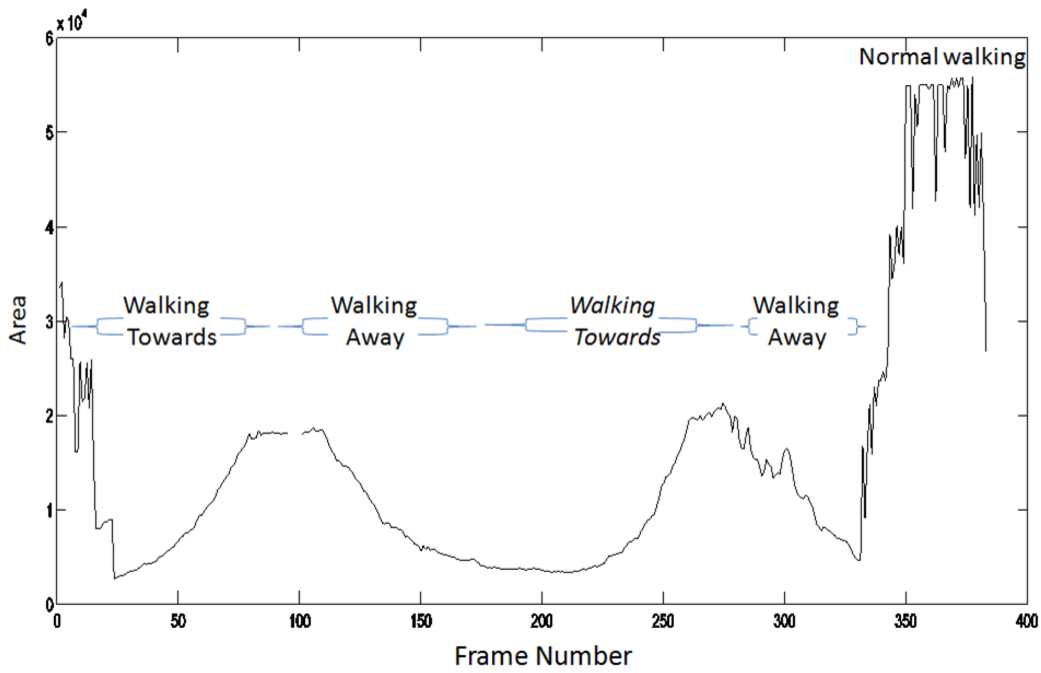


Figure 4.14: an example of area graph of complex walking.

fetal lying, and reclining lying, respectively and some clips including 5 main postures. A sample set of 5 lying postures and 5 main postures performing in each video clip is shown in Figure 4.15 and its $LW^{(t)}$ is presented in Figure 4.16.

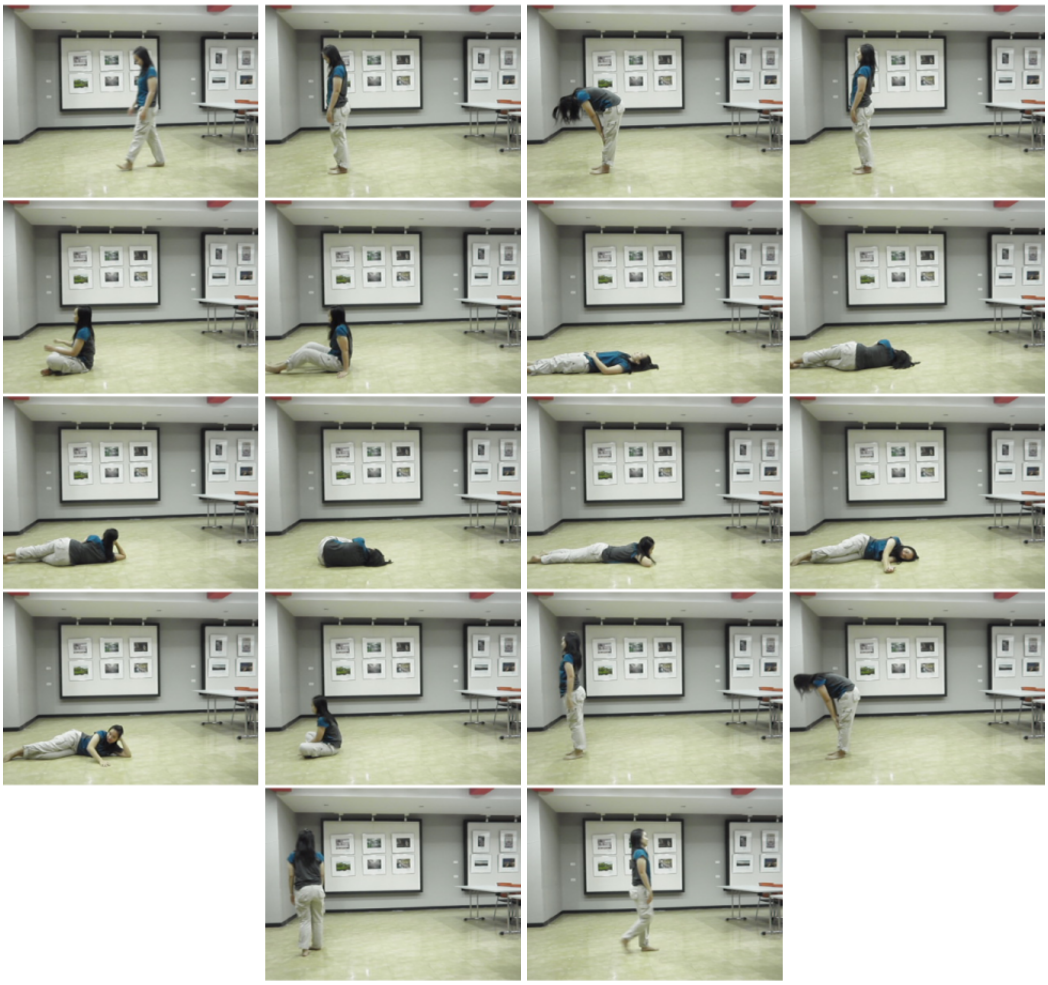


Figure 4.15: A set of actions in video clips in *Sleeping position data set*, comprised of varied styles of five main postures and five types of concerned lying postures, i.e. supine, prone, log, fetal, and reclining at frame 35,75,100,135,185,235,285,335,385,435,535,585,635,835,885,935,1035,1085 respectively.

The challenge problems cause from various distances and viewing angles factors of our own data set are described in Chapter III. These problems cause from the overlap of some complex postures between five lying positions and five main postures. Only using length-width ratio cannot distinguish these postures. The effective features are required.

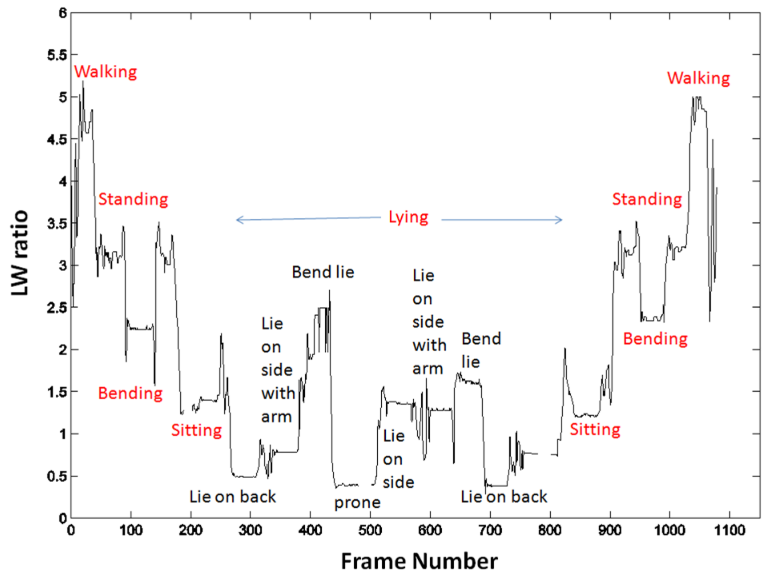


Figure 4.16: The length-with ratio of varied styles of five types of concerned lying postures supine, prone, log, fetal, and reclining and five main postures at 90 degrees.

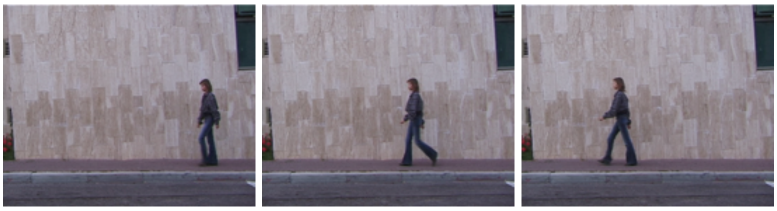
4.2.2 Other Human Actions Databases

We aimed to test the robustness of our method for the various styles of human actions. Thus, to signify the competency of proposed method, the human postures in the other 3 available data sets, i.e Weizmann human posture data set, UIUC Complex Active data set, and CMU Graphics Lab Motion Capture database, were tested with our algorithms. Since there are various postures in these selected data sets, only those postures beyond our scope of study were filtered out from the experiment. The details of other data sets are as follows.

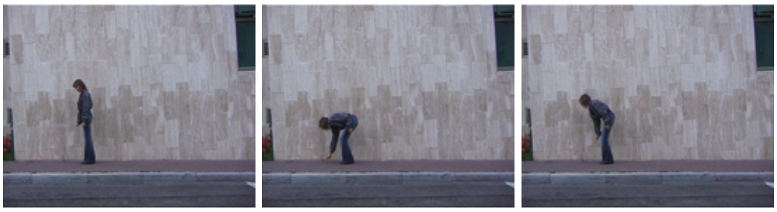
Weizmann Human Posture Data Set

The human posture data set was constructed by Blank et.al. at Weizmann institute. This is the new publicly available database [50] containing 10 postures of walking, running, jumping, galloping, walking side-way, bending, one-hand waving, two-hands waving, jumping in place, jumping jack, and skipping postures. Each posture was performed by 10 persons with a stationary background. Walking and bending postures were chosen to be the testing data set. The sample video clips are presented as Figure 4.17.

Besides this database, the testing of robustness concerning deformable postures and various viewpoints was included for walking postures as well. For robustness evaluation, testing with 10 captured viewpoint angles from 0° to 81° and 10 deformation styles of walking posture such as normal, wearing skirt, carrying briefcase, swinging a bag, limping, being occluded legs, being



(a)



(b)

Figure 4.17: An example video clips in Weizmann Human Posture Data Set (a)walking video clip and (b)bending video clip of Daria actress, at frame 12, 32, 52, respectively.

occluded with a pole, kneeing up, sleepwalking, and walking with a dog were involved. Some examples of the data set for testing robustness are presented in Figure 4.18 and 4.19.

Consequently, the number of total video clips used for testing was 40 sequences of 3,410 frames.

UIUC Complex Active Data Set

The motion data set constructed by Ikizler and Forsyth [94] captured the snap shot images of 3 persons in 5 clothing performing complex postures, differing in length and individual activity. Only 29 sample video clips from the total of 73 clips were selected to be the testing data set for our method. The selected clips comprise of the several postures such as crouch-run, stand-pick up, walk-carry, walk-pick up-walk, run-backwards-wave, run-backwards, run-carry, run-wave, run-turn back-run, walk-turn back-walk, walk-run, walk-stop-run. The samples of image frames were presented in Figure 5.5.

All clips of 1,347 frames contain some complex postures which are not in the scope of our study but rather similar to our experimental postures. Thus, in this experiment, some postures in the clips must be categorized to conform with our studied postures as follows:

- *running* and *walking* postures were categorized in the same posture class as *walking*.
- *standing* and *stopping* postures were categorized in the same posture class as *standing*.
- *picking-up* and *crouching* postures were categorized in the same posture class as *bending*.

Figure 4.18: An example video clips in Weizmann Human Posture Data Set for the deformable robustness testing : (left to right)swinging a bag, carrying briefcase, walking with a dog, kneeing up, limping, sleepwalking, being occluded legs, normal, being occluded with a pole, and wearing skirt.

Figure 4.19: An example video clips in Weizmann Human Posture Data Set for the view-angle robustness testing : 0°-81°.

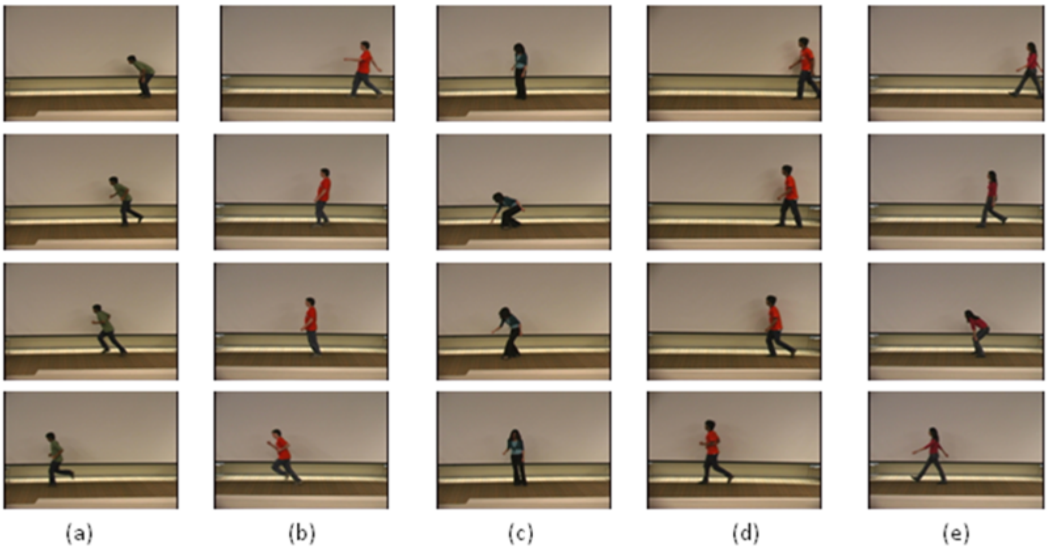


Figure 4.20: Sample image postures from the UIUC Complex Active data set. In each colum from top to buttom rows, the postures and their corresponding are the following. (a) *crouch-run* from image frames 27, 35, 39, and 47. (b) *walk-stop-run* from image frames 31, 43, 53, and 60. (c) *stand-pickup* from image frames 41, 61, 66, and 73. (d) *walk-run* from image frames at 20, 29, 31, and 42. (e) *walk-crouch-walk* from image frames 25, 40, 65, and 93.

CMU Graphic Lab Motion Capture Database.

The database form the Carnegie Mellon University was published on website CMU[2]. Each clip of this data set is a 3D deformed skeleton of human acting with no background. Only walking and bending postures of 134 clips were selected from the locomotion data set performing in various time frames, speeds, viewpoints, and styles. With the same reason as the UIUC data set, some complex postures must be categorized as follows:

- *lifting-up* posture was categorized as *standing* posture.
- *picking-up* posture was categorized as *bending* posture.

Some examples of image frames in this data set were shown in Figure 18. The total number of tested frames is 57,380.

4.3 Experiments on Efficacy of Proposed Features

The following six features were proposed: $LW^{(t)}$ ratio, $DLW^{(t)}$, $\widehat{DLM}^{(t)}$, $\widehat{DLS}^{(t)}$, $\hat{A}^{(t)}$, and $l_c^{(t)}$. Our features were compared with the features proposed by Juang and Chang [10]. In

[2]http://mocap.cs.cmu.edu/

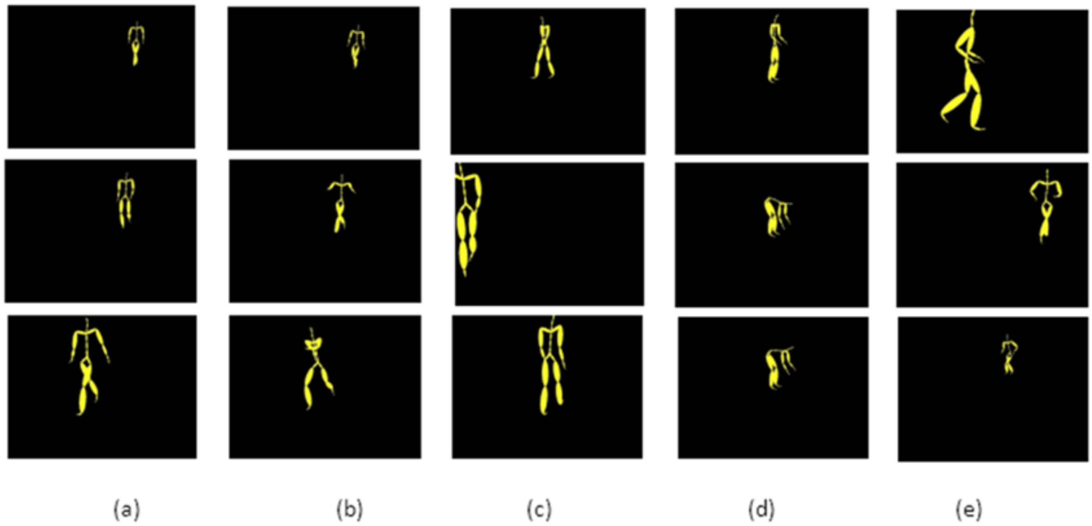


Figure 4.21: Sample video clips of CMU Graphics Lab Motion Capture Database. The images in each column from top to bottom rows are the postures from different frames. (a) *brisk walking* from frames 5, 25, and 60. (b) *walking/exaggerated stride* from frames 5, 35, and 60. (c) *walking forward/ backward/ sideways* from frames 35, 105, and 171. (d) *bending/lift up* from frames 15, 45, and 75. (e) *walking around/ frequent turning/ cyclic walking* from frames 1, 31, and 95.

their study, the *LW* ratio and the normalized coefficients of Discrete Fourier Transform (DFT) of the projected horizontal and vertical histograms obtained from the coordinates of the outline of human shape were used to train a neuro-fuzzy network.

The neuro-fuzzy network is a model combining between Neural Network (NN) and Fuzzy Logic. The model can learn from the data in training set and predict the answer as the fuzzy logic period [95]. The neuro-fuzzy network were applied in several researches as [10, 96, 97, 98].

In this study, the Adaptive Neuro-Fuzzy Inference System (ANFIS) [99] was chosen to be the neuro-fuzzy network model because of its superior performance. Table 4.1 presents the results of Abrahams's research [95] reporting the comparison of ANFIS performance with other models such as NEFPROX, EFuNN, dmEFuNN, and SONFIN on the Mackey-Glass chaotic time series [100] data set, with 500 training data sets and 500 testing data sets.

The architecture of ANFIS model is illustrated as Figure 4.22. ANFIS composes of 5 layers; 1) fuzzification layer, 2) rule layer, 3) normalized layer, 4) defuzzification layer, and 5) output layer. The input signal will go through the fuzzification layer to perform the bell activation function. Then the Takgi-Sugeno fuzzy rule type is adapted to each neuron for calculate the firing strength of the rule it represents. The normalized neuron is defuzzied before the whole signal is summarized to be the output signal of the system. The least-squares estimator and the

Table 4.1: The performance of some nuero-fuzzy inference network.

System	RMSE
ANFIS	0.0017
NEFPROX	0.0332
EFuNN	0.0140
dmEFuNN	0.0042
SONFIN	0.0180

gradient descent method are combined as the learning algorithms for ANFIS. A Fuzzy Inference System (FIS) is constructed by learning the training sample with the backpropagation method. FIS is the model adapting for predicting class of the testing sample, in this study, the human posture. In the experiment, we set the FIS model as the Takgi-Sugeno fuzzy rule type, using two memberships of Gaussian function, and training with 20 epochs. After that, the answer was predicted in four and five main posture classes, respectively.

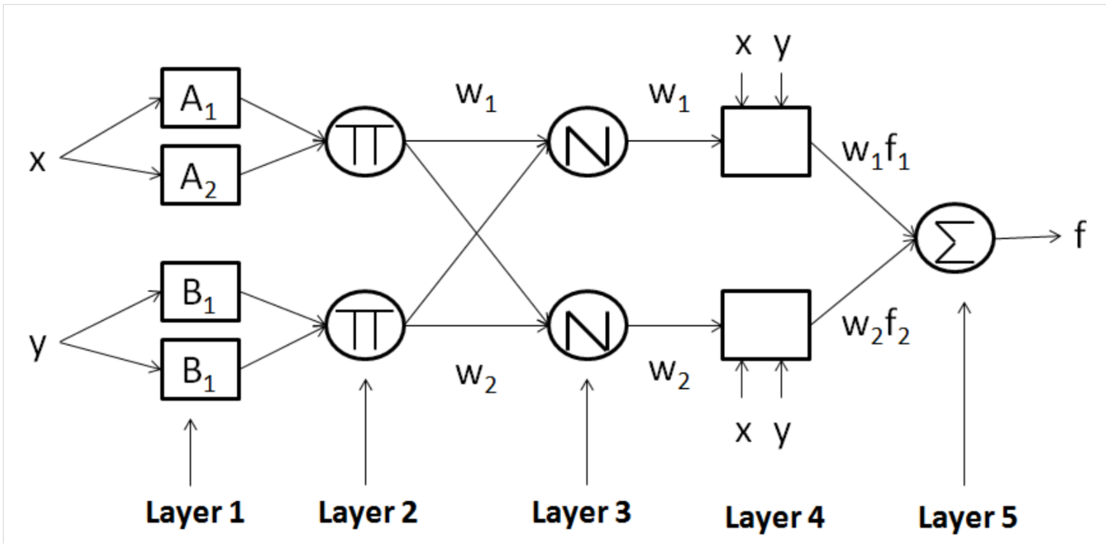


Figure 4.22: Architecture of ANFIS Model

For the features that used in ANFIS model, we select our features compare with the Juang and Changs features [10] which are LW-ratio and DFT-coefficients. The DFT transformation is calculated as follow;

$$F_H[u] = \frac{1}{N}\sum_{y=1}^{n} H(y)exp(-\frac{j2\pi yu}{N}), u = 0, \ldots, N-1 \quad (4.2)$$

$$F_V[v] = \frac{1}{M}\sum_{x=1}^{n} H(y)exp(-\frac{j2\pi xv}{M}), v = 0, \ldots, M-1 \quad (4.3)$$

Where the image size is $N \times M$ and $j = \sqrt{(-1)}$. The histograms of horizontal and vertical projections are denoted by $H(y)$ and $V(x)$, respectively. In Juang and Changs research, the 40 DFT-coefficients, 20 from the horizontal, and 20 from the vertical histograms are used and normalized by

$$NF_H[u] = \frac{|F[u]|}{|F_H[1]|}, u = 1, \ldots, 20 \quad (4.4)$$

$$NF_V[v] = \frac{|F[v]|}{|F_V[1]|}, v = 1, \ldots, 20 \quad (4.5)$$

An example of how the DFT features were extracted in Juang and Chang's work is demonstrated in Figure 4.23. Because of the limitation of our memory capacity when running the ANFIS model, we only tested 10 large DFT coefficients.

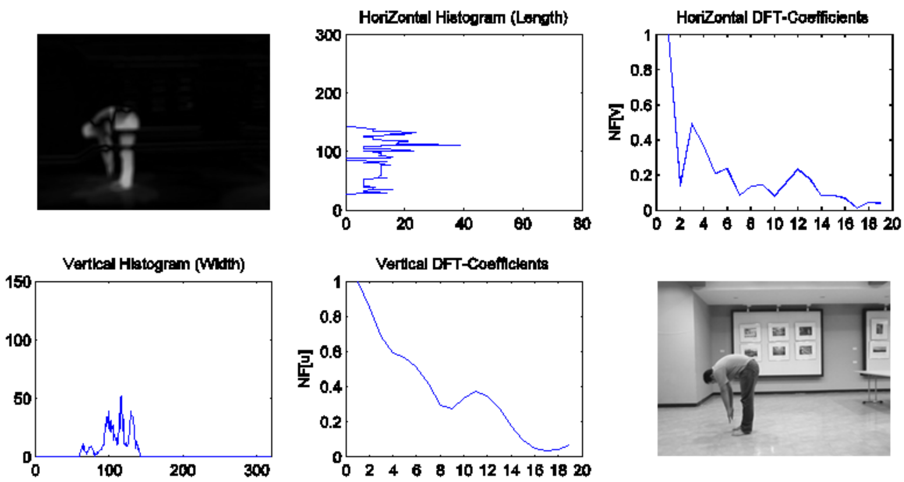


Figure 4.23: An example of projected horizontal and vertical histograms and their 20 DFT-coefficients of bending in 0° viewing angle, respectively.

Two sets of postures were used to test the efficacy of the features proposed by Juang and Chang [10] and ours. The first set consists of standing, sitting, lying, and bending postures. The second set has one additional posture which is walking. The details of the numbers of training and testing set for implementing on ANFIS model in testing 4 postures classification and 5 postures classification are reported on Table 4.2.

Table 4.2: The numbers of training and testing set for each sub-experiment in experiments on efficacy of proposed features .

Experiment	Training (frames)	Testing (frames)
4 classes	4,531	18,826
5 classes	13,686	54,814

4.4 Experiment on Testing Applicability of Our Algorithm

The objective of applicability testing experiment is to test the robustness of our algorithm to various human action video datasets. This experiment examined whether the proposed method is suitable for the real situation which some uncontrollable videos are possible occurred.

Four datasets are composed of our own database, Wiezman dataset, UIUC dataset, and CMU dataset, selected as the testing set of our proposed method. In addition, the ANFIS model is also implemented on our own dataset, in both 4 and 5 postures classification, to compare the average percentage of correct recognition with our proposed method.

The conclusion of each testing data set in all experiment is summarized in Table 4.3. Each number in the table is the number of frames of each posture.

Table 4.3: The number of image frames in each posture in different testing data set.

Data Sets	Postures						Description
	Walking	Standing	Lying	Sitting	Bending	Total	
Ours	6,570	4,908	49,519	4,832	2,671	68,500	Five postures with various styles, camera distances, and viewing angles in clear complex postures.
Weizmann [50]	2,859	170	-	-	381	3,410	normal walking, bending with deformable actions and various view points
UIUC [94]	1,162	69	-	-	116	1,347	complex postures
CMU	56,920	262	-	-	198	57,380	various styles of walking, complex postures

4.5 Experiment on Time Complexity Analysis of Our Algorithms

This experiment aimed to prove the time complexity in processing process of our algorithm. The time complicity shows how long the method takes the time for processing the results. If the complexity of algorithm shows that the method take more or less time to process the answer. It implies to possible to be the real-time method and more suitable to applied to the actual application. The approach to compute the time complexity is presented in next chapter. Furthermore, the confirmed experiment, which actually run on the real computer in same database, is examined. The actual times comparing between proposed algorithm and ANFIS model are reported in next Chapter.

CHAPTER V

RESULTS AND DISCUSSIONS

We constructed the user interface from the GUI interface tools in MATLAB© to facilitate the testing and evaluating process. Sample of captured image from the proposed method is illustrated in Figure 5.1 and the user manual is explained in Appendix. Every video data set was evaluated by the posture prediction from the program in frame by frame. In addition, the efficiency of recognition in each posture was concluded using the confusing matrix, reported by the percentage of accuracy.

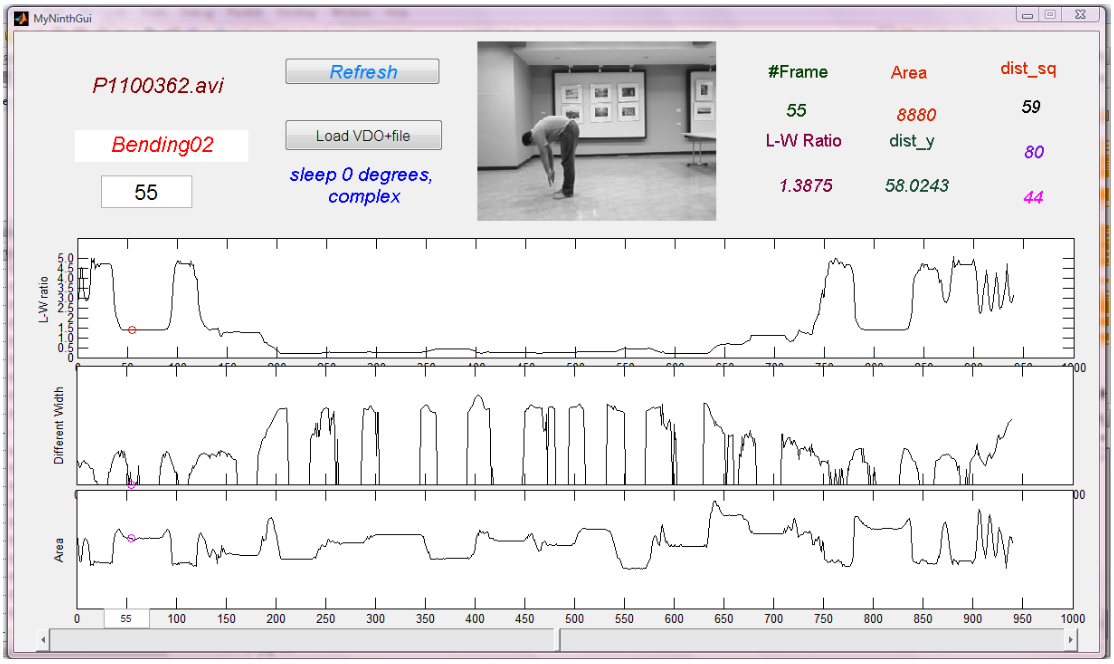


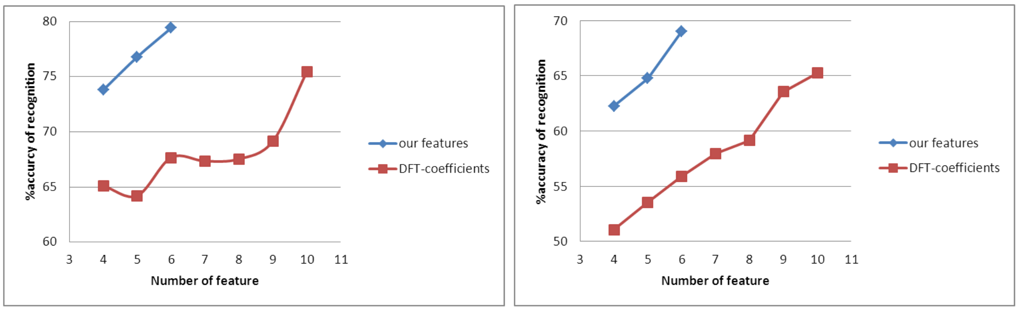
Figure 5.1: Captured image of the GUI user interface using proposed method.

5.1 Results of Proposed Features Efficacy Experiments

The experiment of features efficiency was divided into 2 main sections followed by the amount of frames in our data set. For the first section, we intended to set the same viewing angle condition as that in Juang and Chang's work [10]. We tested the features of postures in 0° viewing angle from 23,357 frames which decomposed into 4,531 frames for training and 18,826 frames for testing. For the second section, the whole data set of 68,500 frames was separated into 13,686 frames for training and 54,814 frames for testing, with the condition of camera distance ranging from 3 to 7 meters and three viewing angles of 0°, 45°, and 90°. Both experiments were implemented on the ANFIS model for four and five posture sets as previously mentioned.

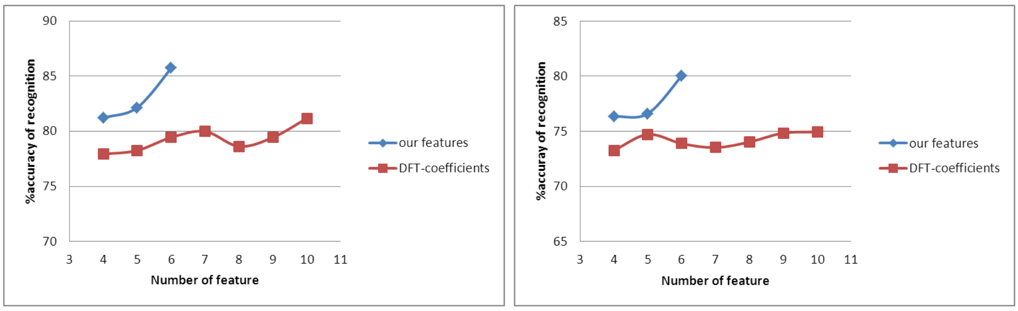
The results of both experiments were summarized in Figure 5.2 and Figure 5.3, respectively. In both experiments, our features were selected for implementing on ANFIS model with 4, 5 and 6 features, respectively. The important features of our 4 testing features were $LW^{(t)}$, $DLW^{(t)}$, $\widehat{DLM}^{(t)}$, and $\widehat{DLS}^{(t)}$. Moreover, the $\hat{A}^{(t)}$ and $l_c^{(t)}$ were the additional features for 5 and 6 features of our feature testing, respectively. The sequence of these features was obtained by the measuring of the highest correct percentage of recognition.

The percentage of accuracy was plotted against the number of features for presenting the trend. Due to the limitation of ANFIS model, all 40 DFT-coefficients cannot be implemented. Table 5.1 summarizes the comparison of average percentage of accuracy. Thus, the efficiency of our features were compared to the DFT-coefficients in 2 times.



(a) (b)

Figure 5.2: Efficacy of feature testing on 0 degree of viewing angle dataset : the percentage of accuracy of the proposed features and DFT-coefficients implementing on ANFIS model: (a) in four-posture classification and (b) five-posture classification.



(a) (b)

Figure 5.3: Efficacy of feature testing on whole dataset (3 to 7 meters of distances and 0, 45, 90 degrees of viewing angle) : the percentage of accuracy of the proposed features and DFT-coefficients with ANFIS model : (a) in four-posture classification and (b) five-posture classification.

Note that, regardless of four and five posture sets, our proposed features achieved much higher accuracy than the accuracy of Juang and Chang's features for all three considered degrees.

Table 5.1: Comparison of average percentage of accuracy for all postures using proposed features and DFT-coefficients implemented on ANFIS models in our data set at only $0°$ viewing angle and at all $0°$, $45°$, $90°$ viewing angles for 4 and 5 main postures, respectively.

Kind of Database	**Classes**	**Our features 6 feats.**	**Juang/Chang's feats. [10] 6 feats.**	**10 feats.**
Only View angle at 0°	4	**79.4061**	67.6617	75.4488
	5	**69.0269**	55.8802	65.2449
Whole Data set at 0°,45°,90°	4	**85.7216**	79.4380	81.1688
	5	**80.0368**	73.8780	74.9399

From Table 5.1, although the numbers of DFT-coefficients were almost twice as many as our features, the results showed that our simply features achieved the accuracy more than DFT-coefficients in both experiments and both 4 and 5 postures classification. Therefore, the efficiency of our features was sufficiently enough to be the important features for representing the human posture in the video clips. However, the percentage of recognition was still low when we used the ANFIS model as the classification method.

Table 5.2: Comparison average percentage of accuracy for all postures using our features implemented on ANFIS Model and our proposed method with the data from our database.

Data set	**Average percentage of correct recognition**	
	ANFIS Model	**Proposed Method**
4 postures	85.7216	**93.6390**
5 postures	80.0368	**93.2700**

The overall recognition results between ANFIS model and our proposed method when implemented with our six important features in both 4 and 5 postures classifications are presented in Table 5.2. From the Table 5.2, the average percentage of recognition of our proposed method is higher than the ANFIS model for both four and five postures.

5.2 Results of Applicability Testing of Our Algorithm Experiment

The objective of this experiment was to test the robustness of the proposed posture recognition that could be applied to the vary situation. The experiment was set for testing our 2 proposed method, such as general algorithm and modify algorithm. General algorithm was designed to recognize the five main postures. Meanwhile, the modified algorithm was developed for lying recognition for sleeping monitor application.

In this experiment, our proposed method was tested on various dataset to signify the competency. One from our datadset and other three data sets are the publicized data set, such as 1)Weizmann human posture data set, 2)UIUC complex active data set, and 3)CMU graphics lab motion capture database. Since there are various postures in these selected data sets, only those postures beyond our scope of study will be filtered out from the experiment. The details of results in each experiment are as follows.

5.2.1 Results on General Algorithm

The General algorithm of our proposed method is designed to recognize 5 main postures such as, walking, standing, lying, sitting, and bending, respectively. Various data sets were implemented on the algorithm for testing the robustness. The evaluation were presented as the percentage of accuracy of recognition in confusion matrix and concluded the average accuracy of overall system.

Our Own Database

The results of implementing the our own video data set on general algorithm of 5 postures recognition are reported as the confusion matrix in Table 5.3.

Table 5.3: Summary of percentage of correctly identified postures with respect to the actual posture and other postures using our data sets.

Actual postures	**Identified postures**				
	Walking	**Standing**	**Lying**	**Sitting**	**Bending**
Walking	**96.09**	3.09	0.05	0.21	0.56
Standing	6.48	**91.98**	0	0.62	0.91
Lying	0	0	**96.68**	1.25	2.08
Sitting	7.48	0	0	**91.54**	0.98
Bending	11.04	6.38	0.93	12.29	**69.36**

According to the recognition results from Table 5.3, the proposed method can correctly recognize 11,090 frames from 12,079 frames of human postures and the average percentage of recognition is 91.812. Lying achieved the highest accuracy with 96.68%. Every posture achieved the percentage of accuracy higher than 90, excepting the bending posture, correct as 69.36%, mostly confusing with sitting.

Weizmann Human Posture Data Set

The human posture data set was constructed by Blank et.al. at Weizmann institute. This is the new publicly available database [78] containing ten postures of walking, running, jumping, galloping,walking sideway, bending, one-hand waving, two-hands waving, jumping in place, jumping jack and skipping postures. Each posture was performed by 10 persons with stationary background. *Walking* and *bending* postures were chosen to be the testing data set. Besides this database, the testing of robustness concerning *deformable posture* and *various viewpoints* was included for *walking* posture as well. For robustness evaluation, testing with 10 capturing viewpoint angles (0 to 81 degrees) and 10 deformation styles of $walking$ posture such as normal, wearing skirt, carrying briefcase, swinging a bag, limping, occluded legs, occluded with a $pole$, kneeing up, sleepwalking, and walking with a dog are involved. Consequently, the number of total video clips used for testing was 40 sequences (3,410 frames). Some examples of the selected testing data set are presented in Figure 5.4.

Figure 5.4: Some sample images from the Weizmann human posture data set. The upper row shows $walking$ posture, and $bending$ posture. The lower row shows the *deformable* and different view points of walking postures, respectively from left to right, for testing the robustness of our algorithms.

The results are shown in Table 5.4. The label in each row of the first column is the name of an posture actually performed. But the label in each column is the name of an posture identified

Table 5.4: Summary of percentage of correctly identified postures with respect to the actual posture and other postures using Weizmann data sets.

Actual postures	**Identified postures**				
	Walking	**Standing**	**Lying**	**Sitting**	**Bending**
Walking	**100.00**	0	0	0	0
Standing	0	**94.71**	0	4.71	0.59
Bending	0	0.52	0	10.24	**89.24**

by the algorithms. Each number in Table 5.4 is the percentage of correctness based on the total number of 3,410 testing frames. For example, in the third row, the algorithms identifies the performed posture as bending posture with 89.24% correctness but it also identifies this posture as standing posture with 0.52% and sitting with 10.24%. The overall accuracy for this data set is 98.534%.

UIUC Complex Active Data Set

The motion data set constructed by Ikizler and Forsyth [94] was captured by snap shot using 3 persons in 5 performing complex postures in length and individual activity. Only 29 movie clips from the total of 73 clips were selected to be the testing data set for our method. The selected clips were comprised of several postures such as crouch-run, stand-pick up, walk-carry, walk-pick up-walk, run-backwards-wave, run-backwards, run-carry, run-wave, run-turn back-run, walk-turn back-walk, walk-run, walk-stop-run. The samples of image frames are presented in Figure 5.5.

As all clips of 1,347 frames contain some complex postures which were not in the scope of our study but rather similar to our experimental postures, thus, in this experiment, some postures in the clips must be categorized to conform with our studied postures as follows:

- *running* and *walking* postures are categorized in the same posture class as *walking*.
- *standing* and *stopping* postures are categorized in the same posture class as *standing*.
- *picking-up* and *crouching* postures are categorized in the same posture class as *bending*.

The recognition results are summarized in Table 5.5. The overall accuracy for this data set is 95.834%. The satisfied results are walking and standing posture, higher than 90%. Meanwhile, bending can be recognized as 70.69%, confusing as walking 22.41% and sitting 6.9%, respectively.

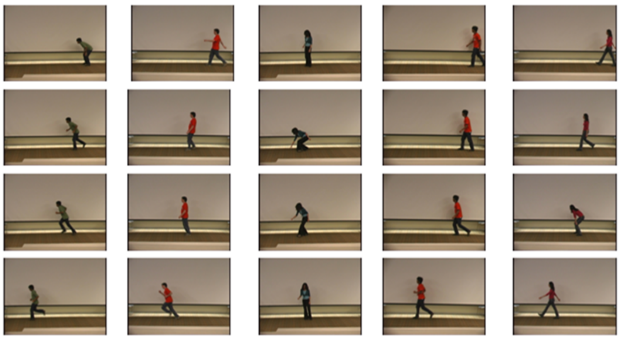


Figure 5.5: Sample image postures from the UIUC Complex Active data set. In each colum from top to buttom rows, the postures and their corresponding are the following. (a) *crouch-run* from image frames 27, 35, 39, and 47. (b) *walk-stop-run* from image frames 31, 43, 53, and 60. (c) *stand-pickup* from image frames 41, 61, 66, and 73. (d) *walk-run* from image frames at 20, 29, 31, and 42. (e) *walk-crouch-walk* from image frames 25, 40, 65, and 93.

Table 5.5: Summary of percentage of correctly identified postures with respect to the actual posture and other postures using UIUC data sets.

Actual postures	**Identified postures**				
	Walking	**Standing**	**Lying**	**Sitting**	**Bending**
Walking	**98.45**	0	0	0.95	0.60
Standing	0	**94.20**	0	1.45	4.35
Bending	22.41	0	0	6.90	**70.69**

CMU Graphics Lab Motion Capture Database.

The database form the Carnegie Mellon University is published on website CMU. Each clip of this data set was a 3D deformed skeleton of human acting with no background. Only *walking* and *bending* postures of 134 clips were selected from the locomotion data set performing in various time frames, speeds, viewpoints, and styles . With the same reason as the UIUC data set, some complex postures must be categorized as follows:

- *lifting-up* posture is categorized as *standing* posture.
- *picking-up* posture is categorized as *bending* posture.

Some image frames in this data set are shown in Figure 5.6. The total number of tested frames is 56,509. The recognition result is summarized in Table 5.6.

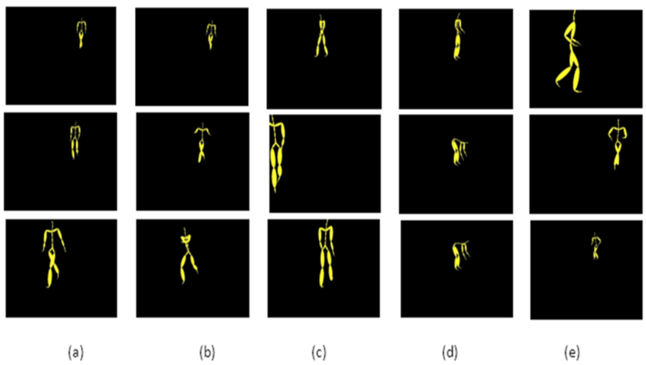


Figure 5.6: Sample video clips of CMU Graphics Lab Motion Capture Database. In each colum from top to buttom rows, the postures and their corresponding are the following. (a) *brisk walking* from frames 5, 25, and 60. (b) *walking/exaggerated stride* from frames 5, 35, and 60. (c) *walking forward/ backward/ sideways* from frames 35, 105, and 171. (d) *bending/lift up* from frames 15, 45, and 75. (e) *walking around/ frequent turning/ cyclic walking* from frames 1, 31, and 95.

The overall accuracy for this data set is 99.892%. The satisfied results are *walking* and *standing* posture which are 99.77% and 100%, respectively. Meanwhile, *bending* can be recognized as 70.69%, confusing as *walking* 22.41% and *sitting* 6.9%, respectively.

To signify the recognition accuracy of our algorithms with respect to the other techniques, only walking and bending posture images from Weizmann database were carefully selected. The reasons of this selection were because this database contains the mixture of walking, bending, and some other different postures from our studied postures and the comparison of recognition performance must be fairly conducted.

Table 5.6: Summary of percentage of correctly identified postures with respect to the actual posture and other postures using CMU data sets.

Actual postures	Identified postures				
	Walking	Standing	Lying	Sitting	Bending
Walking	**99.97**	0	0	0.03	0.48
Standing	0	**100.00**	0	0	0
Bending	22.41	0	0	6.90	**70.69**

Table 5.7 summarized the recognition accuracy of our algorithms with respect to the techniques of Ikizier [49] and Blank [50]. The other two databases, i.e. UIUC and CMU, were not considered in this comparison. It was because the other techniques using these two databases have different research objectives. The study of Ikizier and Forsyth [94] based on UIUC database concentrated on retrieving the most correct and matched image sequence defined by a query of a sequence posture words, e.g. walk-sit-walk. For CMU database, the images were artificial 3D deformable human animation created for different objective from our objective. Currently existing techniques do not use the test data from this database. Nevertheless, this database was considered in our test to signify the robustness of our algorithm.

Table 5.7: Performance comparison of our algorithms with respect to the other algorithms for walking and bending postures selected from Weizmann database.

Algorithms	Average percentage of accuracy
Ours	100.00
Ikizier [94]	100.00
Blank [50]	99.64

The average percentage of recognition of our proposed method, implemented on 4 data set was summarized in Table 5.8

5.2.2 Results on Modified Algorithm

This experiment applied the various data sets on the proposed modified algorithm, 5 main postures recognitions and 5 sleeping positions identification. This algorithm was applied to the sleep-monitoring application. The results of recognition are reported as follows;

Table 5.8: Summary of average percentage of recognition results based on our algorithms and tested with different databases.

Data set	Average percentage of correct recognition
Ours	**91.812**
Weizmann [50]	**98.534**
UIUC [94]	**95.834**
CMU	**99.892**

Our Own Database

According to the recognition results from Table 5.9, the proposed method can correctly recognize 63,890 frames from 68,500 frames of human postures and the average percentage of recognition is 93.270.

Table 5.9: Summary of percentage of correctly identified postures with respect to the actual posture and other postures using our data sets.

Actual postures	Identified postures				
	Walking	Standing	Lying	Sitting	Bending
Walking	**97.25**	0.03	1.64	0.02	1.07
Standing	5.11	**91.75**	0.77	0.45	1.92
Lying	2.28	0	**96.17**	0.92	0.63
Sitting	7.22	0.02	18.63	**71.52**	2.61
Bending	9.47	3.44	12.80	2.32	**71.96**

Weizmann Databases

Results are shown in Table 5.10. The label in each row of the first column is the name of a posture actually performed. However, the label in each column is the name of a posture identified by the algorithms. Each number in Table 5.10 is the percentage of correctness based on the total number of 3,410 testing frames. For example, in the third row, the algorithms identify the performed posture as bending posture with 86.88% of correctness but it also identifies this posture as standing posture with 0.79% and sitting with 12.34%. In conclusion, the average percentage of recognition is 98.328.

Table 5.10: Summary of percentage of correctly identified postures with respect to the actual posture and other postures using Weizmann data sets.

Actual postures	Identified postures				
	Walking	Standing	Lying	Sitting	Bending
Walking	**100.00**	0	0	0	0
Standing	0	**95.88**	0.59	0	3.53
Bending	0	0.79	12.34	0	**86.88**

UIUC Databases

The recognition results are summarized in Table 5.11. The overall accuracy is 95.100%.

Table 5.11: Summary of percentage of correctly identified postures with respect to the actual posture and other postures using UIUC data sets.

Actual postures	Identified postures				
	Walking	Standing	Lying	Sitting	Bending
Walking	**98.46**	0	0.77	0.77	0
Standing	8.70	**85.51**	0	5.80	0
Bending	22.41	0	10.34	0	**67.24**

CMU Databases

The recognition result is summarized in Table 5.12. To signify the recognition accuracy of our algorithms with respect to the other techniques, only walking and bending posture images from Weizmann database were carefully selected. The average percentage of recognition is 99.711. The reasons of this selection are because this database contains the mixture of walking, bending, and some other different postures from our studied postures and the comparison of recognition performance must be fairly conducted.

The only wrong recognized posture was bending. It was recognized as sitting and walking in some frames because the length-width ratio of some persons when they were bending is also the same as when they were sitting and walking. Usually, it occurred to chubby and rather fat persons. In addition, there may be a variety of postures under the same name. For example, in Figure 5.7, the bending posture could be easily recognized as sitting posture.

Table 5.12: Summary of percentage of correctly identified postures with respect to the actual posture and other postures using CMU data sets.

Actual postures	Identified postures				
	Walking	Standing	Lying	Sitting	Bending
Walking	**99.87**	0.01	0.08	0.03	0.02
Standing	0	**100.00**	0	0	0
Bending	0	0	45.45	0	**54.55**

Figure 5.7: Images of posture *run-pickup-run* from UICU database obtained from frames (left to right) 16, 22, 27, 31, and 35.

Our method achieved the satisfied performance for each data set with our algorithms as presented in the average percentage of correct recognition in Table 5.13. Furthermore, our proposed method is sufficient for adapting to a sleep monitoring application.

5.3 Time Complexity Analysis of Our Algorithms

Suppose a considered image size is $m \times n$ and the value of variable *cutframe* in *Algorithm 3* is equal to c. Our recognition consists of the following main steps: (1) image segmentation, (2) envelop construction, (3) feature extraction, and (4) posture determination based on decision tree. The time complexity of each step is as follows.

1. **Image segmentation.** Since every pixel must be involved, the time for subtracting two images is $O(m \times n)$. However, there are c consecutive frames included in this step. Thus, the total time is $O(c \times m \times n)$.

2. **Envelop construction.** To obtain the length and width of the envelop of silhouette, in the worst case, every pixel of the image must be tested. Thus, the time complexity of one frame is $O(m \times n)$. But there are c frames. Therefore, the total time is $O(c \times m \times n)$.

Table 5.13: Summary of average percentage of recognition results based on our algorithms and tested with different databases.

Data set	**Average percentage of correct recognition**
Ours	**93.270**
Weizmann [50]	**98.328**
UIUC [94]	**95.100**
CMU [?]	**99.711**

3. **Feature extraction.** For all c frames, the total time spent in this step is the time to compute equations (7) to (19). Thus, the time complexity is $O(c)$.

4. **Posture determination.** Since *Algorithm 2* determines the posture types based on the structure of a decision tree, the time of information flows through the decision tree must be a constant time or $O(1)$.

From the above analysis, the total time complexity of our algorithms is $O(c \times m \times n) + O(c \times m \times n) + O(c) + O(1) = O(c \times m \times n)$. This time complexity is much less than the time complexity of the others' works, which deployed complex features such as coefficients Fourier transform. The actual average time per frame was also measured during the experiments on the efficacy of our proposed features. Table 5.14 shows the processing times in terms of second per frame for different methods, experimented on Desktop PC with Intel(R) Core (TM) processor; 7-2600 CPU @ 3.40 GHz, 16 GB. of RAM with 64-bit operating system.

Table 5.14: Average processing time in terms of second per frame of posture recognition using the proposed method and ANFIS models with different features.

Method	**Features**	**Second/frame**
Proposed Method	**Our 6 features**	**0.0626**
ANFIS Model	Our 6 features	0.2223
	Juang/Chang's 6 features.	0.3333
	Juang/Chang's 10 features.	8.7777

CHAPTER VI
CONCLUSION

6.1 General Discussion of Our Research

In this research, the five-postures recognition method based on decision tree which were independent from background, camera distance and apparel were developed. The new features which extracted from the morphological of the human silhouette and tested the effect of these features in the first experiment were proposed. The results showed that our features achieved the higher percentage of recognition that the features extracted from DFT-coefficients with the numbers of features lesser than twice. Our features can be used in both cases of 4 and 5 postures recognition.

With the satisfied results from first experiment, we developed the five-postures classification method based on decision tree for posture identifying. Firstly, our proposed general algorithm for five-postures classification was tested the robustness on our new constructed data set and three published human action video data sets, such as 1)Weizmann human posture data set, 2)UIUC complex active data set, and 3)CMU graphics lab motion capture database. The average percentages of overall recognition in every data set were more than 90%. The posture that had the recognition problem was bending. The misclassification became from the advanced postures which were ambiguous postures. For example, the transition state from a pose to another pose, with the addition part of body such as rising hands or legs, and the individual styles. These postures were also the difficult to classify to an actual class for subject. However, the answers of the actual postures in this experiment came from the subjects who performed the actions by themselves.

In case of lying, the modify algorithm to identify additional 5 sleep-positions were developed. It is because this posture was the best recognized posture in the general algorithm. The modified algorithm was also tested the applicability. It was found to achieve the average accuracy in every dataset more than 93%. It is implied that our proposed method is effective, robust and suitable for posture classification.

Finally, we presented the time complexity of our algorithm and confirmed the computational results by real implementing with MATLAB© on our data set. Our algorithm took only 0.0626 second per frame. Based on the proposed system which the video was captured in rate 15 frames per second, our algorithm could be applied as the condition in real time processing.

6.2 Conclusion

A new set of effective features and recognition algorithms for human posture recognition were proposed. Our features are composed of Length-Width ratio (LW), Difference between the Length and Width of a posture envelope (DLW), normalized Difference between the Longest width and Medium width of three sub-envelopes ($\widehat{DLM}$), normalized Difference between the Longest width and Shortest width of three sub-envelopes ($\widehat{DLS}$), normalized Area of a posture envelope ($\hat{A}$), Length from the Centroid to the lower line of posture envelop ($l_c^{(t)}$), and Temporal Difference Width of a temporal posture envelope (TDW). Our features are related to the morphological of the human body. Thus, our proposed features are simple, understanding, and meaningful when comparng with DFT-coefficients features.

Unlike the other studies, our studied postures are based on real situation constraints, involving human's factors, viewing angles, and camera distances. 5 common daily postures, including walking, standing, lying, sitting, and bending postures, and additional 5 lying postures, i.e. supine, prone, log, fetal, and reclining postures, were concentrated in our study. The recognition concerns both stationary and moving postures acting in any arbitrary sequence.

The larger human actions video data set and user-interface was constructed for testing in covered concern conditions and facilitating in evaluation process. The proposed features were tested against the features used in the other works. The results confirmed that our features are superior than other features. Our algorithm is also robust and efficiently applicable to several testing postures from various benchmarked databases with average accuracy of 96%. Since our recognizing algorithms deployed the concept of decision tree, its processing speed is fast enough to apply to sleep disorder detection in a home-care system.

6.3 Suggestion for Future Research

The human posture recognition is very useful to adapt in many fields such as, surveillance system, home-care system, smart-home, security system and tele-medicine. In this research, we suggested to apply our proposed method to a sleep-monitoring application. The sleep-monitoring application is used for collecting the sleep position information during the patient was sleeping in each night. These information were very useful for sleep disease analysis to diagnose the sleeping disorder, such as insomnia, obstructive sleep apnea (OSA), parasomnias, and circadian rhythm disorder. The details of this application is described in Appendix.

However, It was always a problem with confusing with other postures of our algorithm in the bending posture, appeared in the second experiment. Moreover, based on the image size, resolution, and capturing rate of image capturing per second of video, our proposed algorithm cannot process the video in conditions such as, large size, high resolution, and high caricaturing rate per second, in the real time condition. Since the trend of recording the video file will be high definition (HD) system in near future, the developed method should also be concerned in these topics.

REFERENCES

[1] Poppe, R. W. A survey on vision-based human action recognition, Image and Vision Computing, vol. 28, pp. 976–990, June 2010.

[2] Moeslund, T. B., Hilton, A. and Krüger, V. A survey of advances in vision-based human motion capture and analysis, Comput. Vis. Image Underst, vol. 104, pp. 90–126, Nov. 2006.

[3] Howe, N.R. Silhouette lookup for monocular 3d pose tracking, ImageandVisionCom- *puting*, vol. 25, 2006.

[4] Rajkumar, K. and Ramadoss, B. Semantic modeling and retrieval of dance video anno- tations, CoRR, vol. abs/1001.0441, 2010.

[5] Rashidi, P. and Cook, D. J. Keeping the resident in the loop: Adapting the smart home to the user, IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans, vol. 39, pp. 949–959, Sept. 2009.

[6] Mihailidis, A., Carmichael, B. and Boger J., The use of computer vision in an intelligent environment to support aging-in-place, safety and independence in the home, Digital Object Identifier, pp. 238-247, Sep. 2004

[7] Ni, H., Abdulrazak, B., Zhang, D. and Wu, S. Unobtrusive sleep posture detection for elder-care in smart home, in Proceedings of the Aging friendly technology for health and independence, and 8th international conference on Smart homes and health telematics, ICOST'10, (Berlin, Heidelberg), pp. 67–75, Springer-Verlag, 2010.

[8] De Bruijn, H. and Herder, P. M. System and actor perspectives on sociotechnical sys- tems, Trans. Sys. Man Cyber. Part A, vol. 39, pp. 981–992, Sept. 2009.

[9] Rehunathan, D., Bhatti, S., Chandran, O. and Hui, P. vnurse: Using virtualisation on mobile phones for remote health monitoring, in e-Health Networking Applications and Services

(Healthcom), 2011 13th IEEE International Conference on, pp. 82 –85, june 2011.

[10] Juang, C.-F. and Chang, C.-M. Human body posture classification by a neural fuzzy network and home care system application, Trans. Sys. Man Cyber. Part A, vol. 37, pp. 984–994, Nov. 2007.

[11] Lin, C.-W. and Ling, Z.H. Automatic fall incident detection in compressed video for intelligent homecare, in Computer Communications and Networks, 2007. ICCCN 2007. Proceedings of 16th International Conference on, pp. 1172 –1177, aug. 2007.

[12] Miaou, S. G., Sung, P.H. and Huang, C.Y., A customized human fall detection system using Omni-Camera images and personal information, Distributed Di- agnosis and Home Healthcare, vol. 0, pp. 39–42, 2006.

[13] Vishwakarma, V., Mandal, C., and Sural, S. Automatic detection of human fall in video, in Proceedings of the 2nd international conference on Pattern recognition and machine intelligence, PReMI'07, (Berlin, Heidelberg), pp. 616–623, Springer-Verlag, 2007.

[14] Abbate, S., Avvenuti, M., Cola, G., Corsini, P., Light, J. V. and Vecchio, A. Recognition of false alarms in fall detection systems, in 1st IEEE International Workshop on Consumer eHealth Platforms, Services and Applications (CCNC'2011 Workshop CeHPSA), (Las Vegas, NV, USA), pp. 538–543, 1 2011.

[15] Lustrek, M. and Kaluza, B. Fall detection and activity recognition with machine learn- ing, Informatica (Slovenia), vol. 33, no. 2, pp. 197–204, 2009.

[16] Corchado, J. M., Bajo, J., Tapia, D. I. and Abraham, A. Using heterogeneous wireless sensor networks in a telemonitoring system for healthcare, Trans. Info. Tech. Biomed, vol. 14, pp. 234–240, Mar. 2010.

[17] Wang, C.W. Hunter, A. Gravill, N. and Matusiewicz, S. Real time pose recognition of covered human for diagnosis of sleep apnoea, Comp. Med. Imag. and Graph, vol. 34, no. 6, pp. 523–533, 2010.

[18] Wang, C.W. and Hunter, A. Robust pose recognition of the obscured human body, Int. J. Comput. Vision, vol. 90, pp. 313–330, Dec. 2010.

[19] Bsoul, M., Minn, H. and Tamil, L. Apnea medassist: Real-time sleep apnea monitor using single-lead ecg, Information Technology in Biomedicine, IEEE Transactions on, vol. 15, pp. 416 –427, may 2011.

[20] Lao, W., Han, J. and De With, P. H. Automatic video-based human motion analyzer for consumer surveillance system, IEEE Trans. on Consum. Electron, vol. 55, pp. 591–598, May 2009.

[21] Harms, H., Amft, O., and Tröster, G. Estimating posture-recognition performance in sens- ing garments using geometric wrinkle modeling, Trans. Info. Tech. Biomed, vol. 14, pp. 1436–1445, Nov. 2010.

[22] Gallagher, A. T., Matsuoka, Y. and Ang, W.T. An efficient real-time human posture tracking algorithm using low-cost inertial and magnetic sensors, in Proceedings of the 2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS '04), vol. 3, pp. 2967 – 2972, October 2004.

[23] Peng, Y.T., Lin, C.Y., Sun, M.T. and Landis, C. A. Multimodality sensor system for Long-Term sleep quality monitoring, IEEE Transactions on Biomedical Circuits and Systems, vol. 1, pp. 217–227, Sept. 2007.

[24] Huang, W., Wai, A. A. P., Fook, F. S., Biswas, J., Hsia, C.-C. and Liou, K. Multimodal sleeping posture classification, in ICPR, pp. 4336–4339, IEEE, 2010.

[25] Verhaert, V., Haex, B., Wilde, T. D., Berckmans, D., Vandekerckhove, M., Verbraecken, J. and Sloten, J. V. Unobtrusive assessment of motor patterns during sleep based on mattress indentation measurements, IEEE Transactions on Information Technology in Biomedicine, vol. 15, no. 5, pp. 787–794, 2011.

[26] Adami, A., Hayes, T. and Pavel, M. Unobtrusive monitoring of sleep patterns, in Engi- neering

in Medicine and Biology Society, 2003. Proceedings of the 25th Annual Interna- tional Conference of the IEEE, vol. 2, pp. 1360 – 1363 Vol.2, sept. 2003.

[27] Hsia, C.C., Liou, K.J., Aung, A.W., Foo, V., Huang, W., and Biswas, J. Analysisandcom- parison of sleeping posture classification methods using pressure sensitive bed system, Conf Proc IEEE Eng Med Biol Soc, vol. 1, pp. 6131–4, 2009.

[28] Boughorbel, S., Bruekers, F. and Breebaart, J., Baby-postureclassificationfrompressure- sensor data, Pattern Recognition, International Conference on, vol. 0, pp. 556–559, 2010.

[29] Yousefi, R., Ostadabbas, S., Faezipour, M., Farshbaf, M., Nourani, M., Tamil, L. and Pompeo, M. Bed posture classification for pressure ulcer prevention, Conf Proc IEEE Eng Med Biol Soc, vol. 2011, pp. 7175–8, 2011.

[30] Yang, C.C. and Hsu, Y.L. A review of accelerometry-based wearable motion detectors for physical activity monitoring, Sensors, vol. 10, no. 8, pp. 7772–7788, 2010.

[31] Morillo, D., Ojeda, J., Foix, L. and Jimenez, A. An accelerometer-based device for sleep apnea screening, Information Technology in Biomedicine, IEEE Transactions on, vol. 14, pp. 491 – 499, march 2010.

[32] Li, Q., Stankovic, J., Hanson, M., Barth, A., Lach, J. and Zhou, G. Accurate, fast fall detection using gyroscopes and accelerometer-derived posture information, in Wearable and Implantable Body Sensor Networks, 2009. BSN 2009. Sixth International Workshop on, pp. 138 –143, june 2009.

[33] Collop, N. A., Anderson, W. M., Boehlecke, B., Claman, D., Goldberg, R., Gottlieb, D. J., Hudgel, D., Sateia, M. and Schwab, R. A. Clinical guidelines for the use of unattended portable monitors in the diagnosis of obstructive sleep apnea in adult patients. portable monitoring task force of the american academy of sleep medicine, J Clin Sleep Med, vol. 3, no. 7, pp. 737–47, 2007.

[34] Sinha, R. K. Artificial neural network and wavelet based automated detection of sleep spindles,

rem sleep and wake states, J. Med. Syst., vol. 32, pp. 291–299, Aug. 2008.

[35] Wang, C.W. and Hunter, A. A novel approach to detect the obscured upper body in ap- plication to diagnosis of obstructive sleep apnea, IAENG International Journal of Com- puter Science, vol. 35, pp. 110–118, Feb. 2008.

[36] Yang, F.C., Kuo, C.H., Tsai, M.Y. and Huang, S.C. Image-based sleep motion recog- nition using artificial neural networks, in Machine Learning and Cybernetics, 2003 In- ternational Conference on, vol. 5, pp. 2775 – 2780, nov. 2003.

[37] Vangeneugden, J., De Maziegravere, P. A., Van Hulle, M. M., Jaeggli, T., Van Gool, L. and Vogels, R., Distinct mechanisms for coding of visual actions in macaque temporal cortex, J Neurosci, vol. 31, pp. 385-401, 2011.

[38] Cheng, S. and Trivedi, M. Human posture estimation using voxel data for "smart" airbag systems: Issues and framework, in IEEE Intelligent Vehicles Symposium, pp. 84–89, 2004.

[39] Choi, H. and Baraniuk, R. Image segmentation using wavelet-domain classification, in Proceedings of SPIE technical conference on Mathematical Modeling, Bayesian Estimation, and Inverse Problems, pp. 306–320, 1999.

[40] Ishtiaq, M., Jaffar, A., Hussain, A., Basit, A. and Mirza, A. Wavelet based video seg- mentation using self organizing map neural network, in Computer Science and Informa- tion Technology - Spring Conference, 2009. IACSITSC '09. International Association of, pp. 122 –125, april 2009.

[41] Schoenemann, T. and Cremers, D. A combinatorial solution for model-based image seg- mentation and real-time tracking, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, pp. 1153–1164, 2010.

[42] Brox, T. and Malik, J. Large displacement optical flow: Descriptor matching in varia- tional motion estimation, IEEE Transactions on Pattern Analysis and Machine Intelli- gence, vol. 33, pp. 500–513, 2011.

[43] Werghi, N., Xiao, Y. and Siebert, J. A functional-based segmentation of human body scans in arbitrary postures, IEEE Transactions on Systems, Man, and Cybernetics, Part B, vol. 36, pp. 153–165, February 2006.

[44] Chaumette, F. Image moments: A general and useful set of features for visual servoing, 2004.

[45] Jenkins, O. C., Serrano, G. G. and Loper, M. M. Interactive Human Pose and Action Recognition Using Dynamical Motion Primitives, International Journal of Humanoid Robotics, vol. 4, pp. 365–385, 2007.

[46] Kodi, R. T., Madhavi, K. V. and Sudha, K. J. Segmentation of image using watershed and wavelet transform, in Proceedings of the International Conference and Workshop on Emerging Trends in Technology, ICWET '10, (New York, NY, USA), pp. 156–158, ACM, 2010.

[47] Alonso, M. T., Lpez-Martnez, C., Mallorqu, J. J. and Salembier, P. Edge enhancement algorithm based on the wavelet transform for automatic edge detection in sar images, IEEE T. Geoscience and Remote Sensing, vol. 49, no. 1-1, pp. 222–235, 2011.

[48] Liu, C. and Yuen, P. C. Human action recognition using boosted eigenactions, Image Vision Comput, vol. 28, pp. 825–835, May 2010.

[49] Ikizler, N. and Duygulu, P. Histogram of oriented rectangles: A new pose descriptor for human action recognition, Image Vision Comput, vol. 27, pp. 1515–1526, Sept. 2009.

[50] Gorelick, L., Blank, M. Shechtman, E., Irani, M. and Basri, R. Actions as space-time shapes, Transactions on Pattern Analysis and Machine Intelligence, vol. 29, pp. 2247– 2253, December 2007.

[51] Hu, W., Tan, T., Wang, L. and Maybank, S. A survey on visual surveillance of object motion and behaviors, Trans. Sys. Man Cyber Part C, vol. 34, pp. 334–352, Aug. 2004.

[52] Goffredo, M., Bouchrika, I., Carter, J. and Nixon, M. Performance analysis for auto- mated gait extraction and recognition in multi-camera surveillance, Journal of Multime- dia Tools and

Application, 2009.

[53] Lee, K.D., Nam, M. Y., Chung, K.Y., Lee, Y.H. and Kang, U.G. Context and profile based cascade classifier for efficient people detection and safety care system, Multimedia Tools and Applications, Apr. 2012.

[54] Corchado, J. M., Bajo, J., Tapia, D. I. and Abraham, A. Using heterogeneous wireless sensor networks in a telemonitoring system for healthcare, Trans. Info. Tech. Biomed, vol. 14, pp. 234–240, Mar. 2010.

[55] Moeslund, T. B. and Granum, E. A survey of computer vision-based human motion cap- ture, Comput. Vis. Image Underst, vol. 81, pp. 231–268, Mar. 2001.

[56] Bodor, R., Drenner, A., Fehr, D., Masoud, O. and Papanikolopoulos, N. View-independent human motion classification using image-based reconstruction, Image Vision Comput, vol. 27, pp. 1194–1206, July 2009.

[57] Yi Chien, S., Yih Ma, S. and Chen, L.G. Efficient moving object segmentation algo- rithm using background registration technique, IEEE Trans. Circuits Syst. Video Tech- nol, vol. 12, pp. 577–586, 2002.

[58] Bobick, A. F.and Davis, J. W., The recognition of human movement using temporal templates, IEEE Transactions on Pattern Analysis and Ma- chine Intelligence, vol. 23, pp. 257–267, 2001.

[59] Tai, Y.W., Du, H., Brown, M. S. and Lin, S. Correction of spatially varying image and video motion blur using a hybrid camera, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 32, pp. 1012–1028, 2010.

[60] Laptev, I., Belongie, S. J., Prez, P. and Wills, J. Periodic motion detection and segmen- tation via approximate sequence alignment, in In Proceedings of the International Con- ference on Computer Vision, 2005.

[61] Babu, R.V. and Ramakrishnan, K.R. Recognition of human actions using motion history

information extracted from the compressed video, Image and Vision Computing, vol. 22, no. 8, pp. 597–607, 2004.

[62] Ali, S. and Shah, M. Human action recognition in videos using kinematic features and multiple instance learning, IEEE Transactions on Pattern Analysis and Machine Intelli- gence, vol. 32, pp. 288–303, 2010.

[63] Min, J. and Kasturi, R. Extraction and temporal segmentation of multiple motion tra- jectories in human motion, in Proceedings of the 2004 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'04)(Washington, DC, USA), vol. 7, pp. 118, IEEE Computer Society, 2004.

[64] Juang, C.F., Chang, C.M., Wu, J.-R. and Lee, D. Computer vision-based human body segmentation and posture estimation, Trans. Sys. Man Cyber. Part A, vol. 39, pp. 119– 133, Jan. 2009.

[65] Wu, C. and Aghajan, H. Model-based human posture estimation for gesture analysis in an opportunistic fusion smart camera network, in Proceedings of the 2007 IEEE Conference on Advanced Video and Signal Based Surveillance, AVSS '07, (Washington, DC, USA), pp. 453–458, IEEE Computer Society, 2007.

[66] Fathi, A. and Mori, G. Human pose estimation using motion exemplars.

[67] Zhu, Q. and et al. Fast human detection using a cascade of histograms of oriented gradi- ents, in In CVPR06, pp. 1491–1498, 2006.

[68] Zhou, J. and Hoang, J. Real time robust human detection and tracking system, in Pro- ceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Workshops - Volume 03, CVPR '05, (Washington, DC, USA), pp. 149–, IEEE Computer Society, 2005.

[69] Ning, H., Tan, T., Wang, L. and Hu, W. Kinematics-based tracking of human walking in monocular video sequences, Image Vision Comput, vol. 22, no. 5, pp. 429–441, 2004.

[70] Takahashi, K. and Naemura, M. Remarks on human body posture estimation using neural network and kalman filter, in Systems, Man and Cybernetics, 2005 IEEE International Conference on, vol. 3, pp. 2495 – 2500 Vol. 3, oct. 2005.

[71] Zhao, T., Nevatia, R. and Wu, B. Segmentation and tracking of multiple humans in crowded environments, IEEE Transactions on Pattern Analysis and Machine Intelli- gence, vol. 30, pp. 1198–1211, 2008.

[72] Montabone, S. and Soto, A. Human detection using a mobile platform and novel features derived from a visual saliency mechanism, Image Vision Comput, vol. 28, pp. 391–402, Mar. 2010.

[73] Haritaoglu, I., Harwood, D. and Davis, L. S. W4: Real-time surveillance of people and their activities, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, pp. 809–830, 2000.

[74] Mukasa, T., Miyamoto, A., Nobuhara, S., Maki, A. and Matsuyama, T. Complex human motion estimation using visibility. in FG, pp. 1–6, IEEE, 2008.

[75] Elgammal A. and Lee, C.S. Tracking people on a torus, IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, pp. 520–538, 2009.

[76] Gupta, A., Mittal, A. and Davis, L. Constraint integration for efficient multiview pose estimation with self-occlusions, Pattern Analysis and Machine Intelligence, IEEE Transactions on, vol. 30, pp. 493 –506, march 2008.

[77] Lin, H.Y. and Wei, J.Y. A street scene surveillance system for moving object detection, tracking and classification, in Intelligent Vehicles Symposium, 2007 IEEE, pp. 1077 – 1082, june 2007.

[78] Gorelick, L., Blank, M., Shechtman, E., Irani, M. and Basri, R. Actions as space-time shapes, in In ICCV, pp. 1395–1402, 2005.

[79] Li, W., Zhang, Z. and Liu, Z. Expandable Data-Driven graphical modeling of human ac- tions

based on salient postures, Circuits and Systems for Video Technology, IEEE Trans- actions on, vol. 18, no. 11, pp. 1499–1510, 2008.

[80] Liang, Y.M., Shih, S.W., Shih, A. C.C., Liao, H.Y. M. and Lin, C.C. Learning atomic human actions using variable-length markov models, Trans. Sys. Man Cyber. Part B, vol. 39, pp. 268–280, Feb. 2009.

[81] Cucchiara, R., Grana, C., Prati, A. and Vezzani, R. Probabilistic posture classification for human-behavior analysis, Trans. Sys. Man Cyber. Part A, vol. 35, pp. 42–54, Jan. 2005.

[82] Girondel, V. and et al. Belief theory-based classifiers comparison for static human body postures recognition in video, 2005.

[83] Haiyong, Z. and Zhijing, L. Human action recognition based on non-linear svm decision tree, Journal of Computational Information Systems (JCIS), vol. 7, no. 7, pp. 2461– 2468, 2011.

[84] Juang, C.F. and Lin, C.T. An online self-constructing neural fuzzy inference network and its applications, IEEE Transactions on Fuzzy Systems, vol. 6, no. 1, pp. 12–32, 1998.

[85] Li, Q., Stankovic, J. A., Hanson, M. A., Barth, A. T., Lach, J. and Zhou, G. Accurate, fast fall detection using gyroscopes and accelerometer-derived posture information, in 2009 International Conference on Body Sensor Networks, vol. 9, p. 138–143, 2009.

[86] Karantonis, D., Narayanan, M., Mathie, M., Lovell, N. and Celler, B. Implementation of a real-time human movement classifier using a triaxial accelerometer for ambulatory monitoring, Information Technology in Biomedicine, IEEE Transactions on, vol. 10, pp. 156 –167, Jan. 2006.

[87] Ramdani, S., Seigle, B., Lagarde, J., Bouchara, F. and Bernard, P. L. On the use of sample entropy to analyze human postural sway data, Med Eng Phys, vol. 31, no. 8, pp. 1023– 31, 2009.

[88] Rednic, R., Gaura, E. and Brusey, J. Classact: Accelerometer-based real-time activity classifier,

in Sensors & Instrumentation KTN: Wireless Sensing Demonstrator Showcase (WiSIG), 2009.

[89] Spranger, M., Hoefer, S. and Hild, M. Biologically inspired posture recognition and posture change detection for humanoid robots, in Proceedings of ROBIO'09: IEEE In- ternational Conference on Robotics and Biomimetics, IEEE, 2009.

[90] Boughorbel, S., Bruekers, F. and Breebaart, J. Baby-postureclassificationfrompressure- sensor data. in ICPR, pp. 556–559, IEEE, 2010.

[91] Piccardi, M. Background subtraction techniques: a review, in IEEE International Con- ference on Systems, Man and Cybernetics, vol. 4, pp. 3099–3104, IEEE, Oct. 2004.

[92] Grest, D., Frahm, J. M. and Koch, R. A color similarity measure for robust shadow removal in real time, *in In Vision, Modeling and Visualization*, pp. 253–260, 2003.

[93] Silapasuphakornwong, P., Phimoltares, S., Lursinsap, C. and Hansuebsai, A. Posture recognition invariant to background, cloth textures, body size, and camera distance using morphological geometry. in ICMLC, pp. 1130–1135, IEEE, 2010.

[94] Ikizler, N. and Forsyth, D. A. Searching for complex human activities with no visual examples, Int. J. Comput. Vision, vol. 80, pp. 337–357, Dec. 2008.

[95] Abraham, A. Neuro fuzzy systems: State-of-the-art modeling techniques, in AND AR- TIFICIAL INTELLIGENCE, SPRINGER-VERLAG GERMANY, JOSE MIRA AND AL- BERTO PRIETO (EDS), pp. 269–276, Springer Verlag, 2001.

[96] Wachs, J. P., Goshorn, D. and Kölsch, M. Recognizing human postures and poses in monocular still images, in IPCV, pp. 665–671, 2009.

[97] Hazelhoff, L., Han, J. and With, P. H. Video-based fall detection in the home using principal component analysis, in Proceedings of the 10th International Conference on Advanced Concepts for Intelligent Vision Systems, ACIVS '08, (Berlin, Heidelberg), pp. 298–309, Springer-Verlag, 2008.

[98] Gkalelis, N., Tefas, A. and Pitas, I. Combining fuzzy vector quantization with linear discriminant analysis for continuous human movement recognition, Circuits and Systems for Video Technology, IEEE Transactions on, vol. 18, no. 11, pp. 1511–1521, 2008.

[99] Jang, J.S. R., Anfis: Adaptive-network-based fuzzy inference system, IEEE Transactions on Systems, Man, and Cybernetics, vol. 23, pp. 665–685, 1993.

[100] Mackey, M. C. and Glass, L. Oscillation and Chaos in Physiological Control Systems, Science, vol. 197, pp. 287–289, July 1977

APPENDICES

APPENDIX A

ANFIS MODEL IN MATLAB©

In this section we discuss the use of the function $anfis$ in the Fuzzy Logic toolbox. These tools apply fuzzy inference techniques to data modeling. As you have seen from the other fuzzy inference GUIs, the shape of the membership functions depends on parameters, and changing these parameters change the shape of the membership function. Instead of just looking at the data to choose the membership function parameters, you choose membership function parameters automatically using these $FuzzyLogicToolbox$ applications.

The basic structure of the type of fuzzy inference system seen thus far is a model that maps input characteristics to input membership functions, input membership function to rules, rules to a set of output characteristics, output characteristics to output membership functions, and the output membership function to a single-valued output or a decision associated with the output. You have considered only fixed membership functions that were chosen arbitrarily. You have applied fuzzy inference to only modeling systems whose rule structure is essentially predetermined by the user's interpretation of the characteristics of the variables in the model.

***anfis* function)**

Training routine for Sugeno-type Fuzzy Inference System (MEX only)

A.1 Syntax

[fis,error,stepsize,chkFis,chkErr] = ...
anfis(trnData,numMFs,trnOpt,dispOpt,chkData,optMethod)

A.2 Description

This syntax is the major training routine for Sugeno-type fuzzy inference systems. anfis uses a hybrid learning algorithm to identify parameters of Sugeno-type fuzzy inference systems. It applies a combination of the least-squares method and the backpropagation gradient descent method for training FIS membership function parameters to emulate a given training data set. anfis can also be invoked using an optional argument for model validation. The type of model validation that takes place with this option is a checking for model overfitting, and the argument is a data set called the checking data set.

The arguments in the description for $anfis$ are as follows. Note that you can specify the arguments $trnOpt$, $dispOpt$, $chkData$, and $optMethod$ as empty, [], when necessary:

- *trnData*: the name of a training data set. This matrix contains data input in all but the last column. The last column contains a single vector of output data.

- *initFis*: the name of a fuzzy inference system (FIS) used to provide $anfis$ with an initial set of membership functions for training. Without this option, $anfis$ uses $genfis1$ to implement a default initial FIS for training. This default FIS has two membership functions of the Gaussian type, when it is invoked with only one argument. If $initFis$ is provided as a single number (or a vector), it is taken as the number of membership functions (or the vector) whose entries are the respective numbers of membership functions associated with each respective input when these numbers differ for each input). In this case, both arguments of $anfis$ are passed to $genfis1$ to generate a valid FIS structure before starting the training process.

- *numMFs*: the number of membership functions. Use numMFs, an integer scalar value, as the second argument to anfis when you do not already have a FIS to train and you want anfis to build a default initial FIS using your data. Each input and output to this FIS is characterized by one or more membership functions. Specify the number of membership functions in numMFs.

- *trnOpt*: a vector of training options. When a training option is entered as NaN, the default options is in force. These options are as follows:

  - trnOpt(1): training epoch number (default: 10)
  - trnOpt(2): training error goal (default: 0)
  - trnOpt(3): initial step size (default: 0.01)
  - trnOpt(4): step size decrease rate (default: 0.9)
  - trnOpt(5): step size increase rate (default: 1.1)

- *dispOpt*: a vector of display options that specify what message to display in the MATLAB Command Window during training. The default value for a display option is 1, which means that the corresponding information is displayed. A 0 means the corresponding information is not displayed. When a display option is entered as NaN, the default options will be in force. These options are as follows:

  - dispOpt(1): ANFIS information, such as numbers of input and output membership functions, and so on (default: 1)
  - dispOpt(2): error (default: 1)

  - dispOpt(3): step size at each parameter update (default: 1)
  - dispOpt(4): final results (default: 1)

- *chkData*: the name of an optional checking data set for overfitting model validation. This data set is a matrix in the same format as the training data set. When you supply chkData as an input argument, you must also supply chkFis and chkErr as output arguments. *

- *optMethod*: an optional optimization method used in membership function parameter training: either 1 for the hybrid method or 0 for the backpropagation method. The default method is the hybrid method, which is a combination of least-squares estimation with backpropagation. The default method is invoked whenever the entry for this argument is anything but 0.

The training process stops whenever the designated epoch number is reached or the training error goal is achieved.

Note When anfis is invoked with two or more arguments, optional arguments take on their default values if they are entered as NaNs or empty matrices. Default values can be changed directly by modifying the file anfis.m. Either NaNs or empty matrices must be used as placeholders for variables if you do not want to specify them, but do want to specify succeeding arguments, for example, when you implement the checking data option of anfis.

The range variables in the previous description for anfis are as follows:

- *fis* is the FIS structure whose parameters are set according to a minimum training error criterion.

- error or *chkErr* is an array of root mean squared errors representing the training data error signal and the checking data error signal, respectively. The function only returns chkErr when you supply chkData as an input argument.

- *stepsize* is an array of step sizes. The step size is decreased (by multiplying it with the component of the training option corresponding to the step size decrease rate) if the error measure undergoes two consecutive combinations of an increase followed by a decrease. The step size is increased (by multiplying it with the increase rate) if the error measure undergoes four consecutive decreases.

- *chkFis* is the FIS structure whose parameters are set according to a minimum checking error criterion. The function only returns chkFis when you supply chkData as an input argument.

anfis has certain restrictions (see Constraints of anfis for more information).

A.3 Examples

```
x = (0:0.1:10)';
y = sin(2*x)./exp(x/5);
trnData = [x y];
numMFs = 5;
mfType = 'gbellmf';
epochn = 20;
infis = genfis1(trnData,numMFs,mfType);
outfis = anfis(trnData,infis,20);
plot(x,y,x,evalfis(x,outfis));
legend('Training Data','ANFIS Output');
```

APPENDIX B

OUR USER INTERFACE CONSTRUCTED BY MATLAB©

In the experiment, we constructed the user interface of GUI interface of MATLAB© to facilitate for testing and evaluating the correct posture. The capture image of user-interface, implemented of our proposed algorithms, shown as Figure B.1.

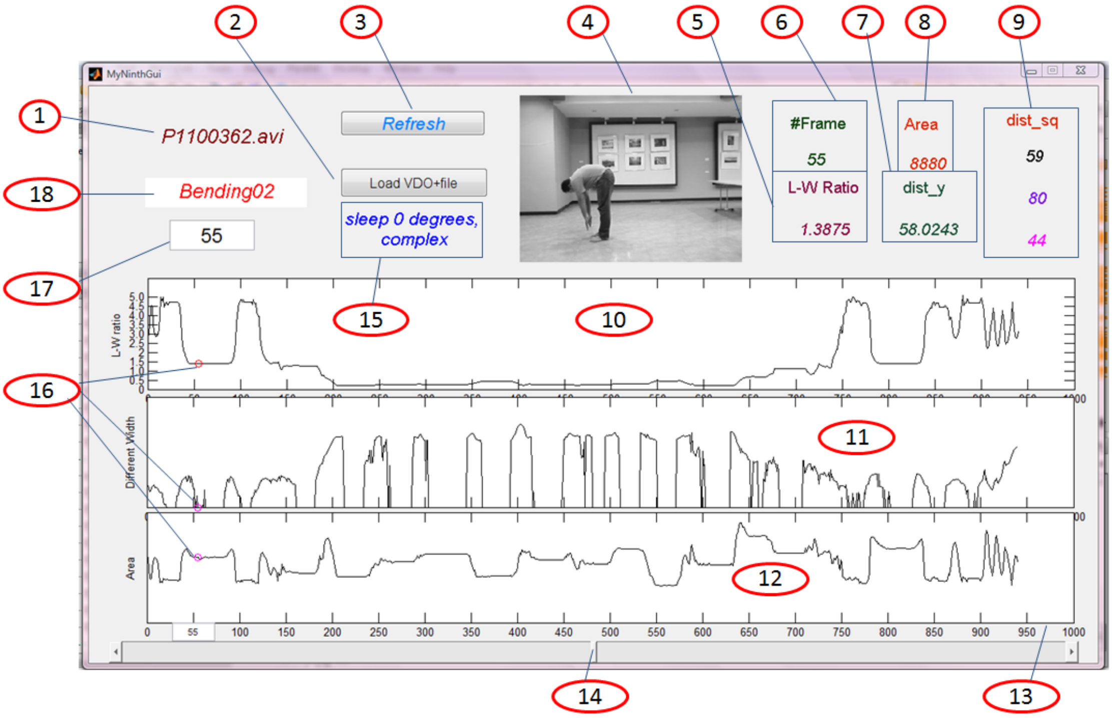


Figure B.1: Captured image of the GUI user interface using proposed method.

The user guide description of each part on the user-interface is as follow;

1. Video file name
2. Button of loading video clip
3. Refresh button for clear last current point on every graph
4. Captured image from current frame
5. Length-With ratio (LW) value of posture envelop in current frame
6. Number of the current frame
7. Length of centroid (l_c) value of posture envelop in current frame
8. Area (A) value of posture envelop in current frame

9. Width of 3 sub-posture envelops, head,torso,leg, in current frame
10. Length-With ratio (LW) graph
11. Different between Length and Width (DLW) graph
12. Area (A) graph
13. X-axis: Number of the frames
14. Slide bar for selecting the current frame
15. Descriptions of the current video clip
16. Current point : present the current frame
17. Box for get number of current frame from the user
18. predicted posture

APPENDIX C
SLEEP MONITORING APPLICATION FOR SLEEPING QUALITY ANALYSIS

A home-care system is very useful when it is adapted to monitor the convalescent and elderly live at home alone. The posture recognition can be applied to interpret the human behaviour video in home-care system. In general, the system mainly works by monitoring the behavior of each person and detecting the period of changing postures when the important situation is happening. For the sleep monitoring application, the system detects the movement during sleep which causes the crisis of sleep disorder. The sleep information during each night is very useful for sleep disease and diagnosis such as insomnia, obstructive sleep apnea (OSA), parasomnias, and circadian rhythm disorder. In this paper, we suggested the sleep monitoring application as one of the real applications of our human posture recognition process.

The sleep movement at night is an evaluation of sleep disorder diagnosis. A huge number of changing sleeping positions per a night implies the bad sleep quality of a person because the movement causes people to wake up during sleeping. If the person wakes up for many times, the good sleeping period is not enough for some rest. For a long term period, it is a cause of sleeping disease such as insomnia, obstructive sleep apnea (OSA), and parasomnias. Moreover, the disease will become a problem as hard as circadian rhythm disorder and death from the heart failure. The proposed sleep monitoring application works by counting the number of sleeping position changes during a night. Then, the system will conclude the result in each night and estimate the state of sleep quality.

From the flow chart of the system in Figure C.1, the human behavior in video is predicted firstly as a posture among 5 main postures. If the recent main posture is lying, the system will detect the changing posture period, determined by the condition of $TDW^{(t)} > 0$. After that, the system will count the number of sleeping position changes, defined by the variable *SM*. *SM* value is counted continuously all night when a person changes the sleeping positions until the person rises up. This means that the posture is changed from lying to bending or sitting in the morning. In general, the suggested sleeping period which is suitable for adults is about eight to twelve hours per night. After the person wakes up in the morning, the system will evaluate the sleep quality from the *SM* value in 3 levels such as good, risky, and poor sleep quality. The *good* sleep quality means the normal sleeping with the numbers of movement during sleeping less than 12 times per night (the average is less than one time per hour). The *risky* sleep quality which has the sleeping position changing rate is during 13 to 20 times per night. There are risks

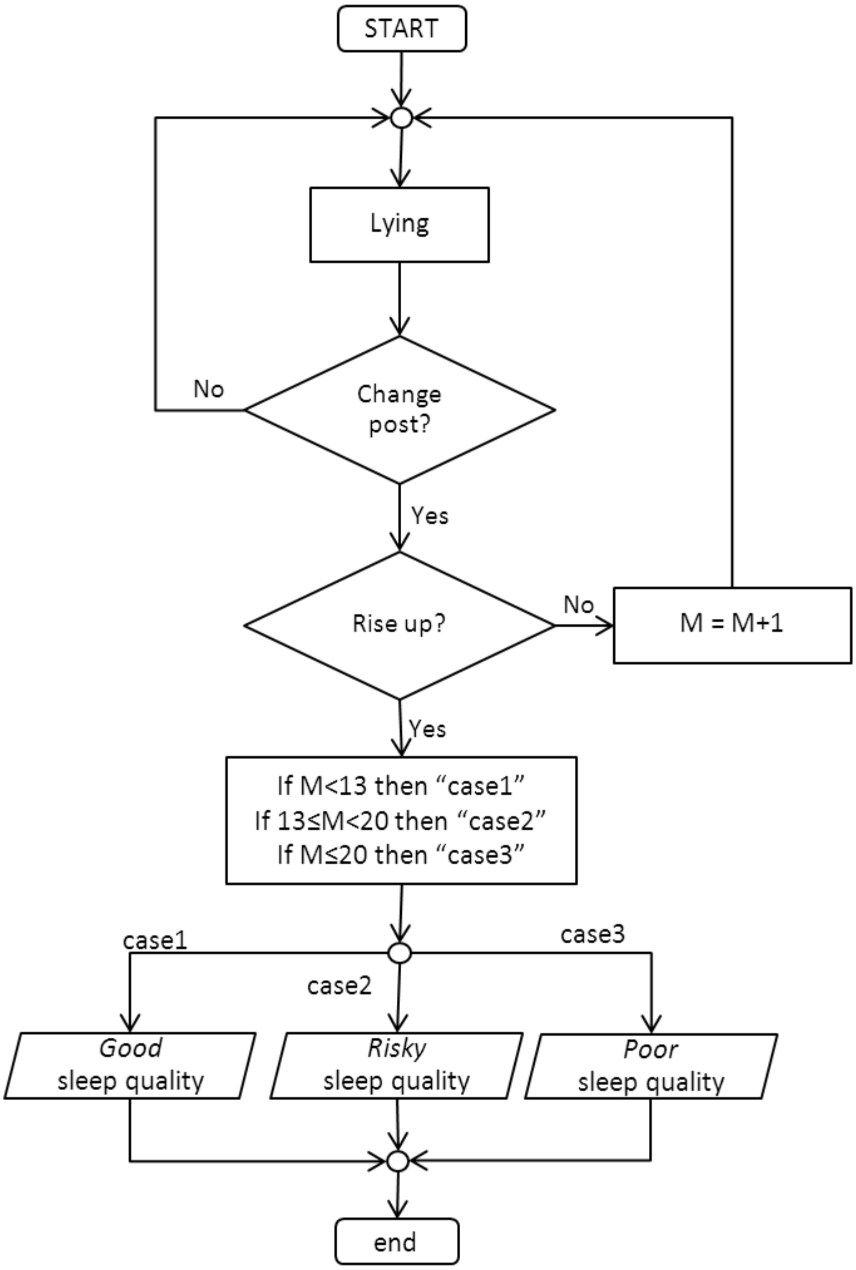


Figure C.1: The flow chart of the sleep monitoring system for sleeping quality analysis.

for sleep disorder disease, if the patient is in this state for a long time. So, the patient should be taken care from a doctor. For the last state, the *poor* sleep quality means the sleeping is trouble. There are more than 21 times of sleep movement per night (average is more than two time per hour), also including the sleep walking. The patient is suggested to see a doctor for treatment planning. Furthermore, the sleep information is able to be collected in a period such as a few days, a week, or a month to be the information for sleep disorder diagnosis and treatment for a long term. The sample of five types of sleeping postures was illustrated in Figure C.2.

Table C.1: The percentage of accuracy in sleep movement detection and its fault positive and fault negative.

%	**Accuracy**	**Fault Positive**	**Fault Negative**
Performance	**96.381**	3.351	3.619

The initial result of sleep movement detection was reported as Table C.1. We tested the application with our data set as correctly counting the number of sleeping position changes. The system achieved the accuracy of sleep movement prediction of 96.381% (719 from 746 times). The fault positive and fault negative of the system are 3.351% (25 times) and 3.619% (27 times), respectively.

Figure C.2: An example of sleep movement situation, captured at 0 degree of viewing angle, which change posture in 5 sleep-positions *(left to right)* log, prone, supine, fetal, and reclining, respectively.

APPENDIX D

SEGMENTATION TECHNIQUE FOR BACKGROUND SUBTRACTION

The ideals of background subtraction technique in this dissertation came from the researches of [92].In their research, they proposed the techniques of foreground segmentation techniques as follow;

”To detect foreground regions we use a background reference image and calculate the difference image with the current one. To acquire a background reference a ground truth image is incorporated as a mean image of a sequence of the interaction area without the user.

The segmentation of the foreground is often obtained by applying a threshold to the difference values. The choice of the threshold parameter varies in each approach, while it usually depends on the camera noise.

The variance of the pixel noise over the whole difference image each frame are measured, leaving the segmented foreground region out.

Let $d(x;y) = r(x;y).\ a(x;y)$ be the difference image value, $r(x;y)$ the background reference image intensity and $a(x;y)$ the current image intensity at image positions $(x;y)$. The variance of noise in an image with size MxN is,

$$\sigma^2 \quad = \quad \frac{1}{MN}\sum_{i=0}^{M-1}\sum_{i=0}^{N-1}(d(i,j)-\bar{d})^2 \tag{D.1}$$

where $\bar{d}$ is the mean value of the difference image.

For efficient computation this equation can be rewritten as

$$\sigma^2 \quad = \quad \frac{1}{MN}\sum_{i=0}^{M-1}\sum_{i=0}^{N-1}(d(i,j))^2 - (\frac{1}{MN}\sum_{i=0}^{M-1}\sum_{i=0}^{N-1}d(i,j))^2 \tag{D.2}$$

Assuming a Gaussian distribution of noise we set the segmentation threshold to $2\sigma^2$. Therefore 97:5% of noise are in the segmentation interval. Each pixel belongs to the foreground,if

$$|d(x,y)-\bar{d}| \quad > \quad 2\sigma^2 \tag{D.3}$$

The main advantage of this approach is the property of fast adaptation to lighting changes, because the new threshold is computed for each frame and is independent from previous frames. Even large changes of the overall brightness can be adapted, as the mean value in the difference image incorporates such intensity variations.

However, it is very important to calculate the new threshold for the background only and not for shadows or parts of the person, because a sufficiently large region of not detected foreground leads to a large variance, which leads to less detected foreground in the next frame. Accumulating even larger variances until almost everything is detected as background. To overcome this problem we assume for initialization that only background is visible. In later steps a bounding box around all thresholded pixel can be spared out to calculate the new threshold. In our environment the camera is mounted such that we can assume nobody is appearing in the lower five percent of the image. Therefore we calculate the threshold only in this part of the image.

Overall changes in the scene are not treated, it is assumed to be static. However it is easily possible to update the background reference if only slow changes in the background scene occur."

From the main ideal from this research, we set the small experiment for finding the suitable threshold to subtract the region of interest from the background. The details have been explained in the chapter III.

BIOGRAPHY

Miss Piyarat Silapasuphakornwong was born on February 2, 1984 in Bangkok, Thailand. She received the second class honor of Bachelor's Degree of Science in Imaging Science and Printing Technology from Chulalongkorn University, Thailand in 2007.

Scholarships and Awards

During she studied in Doctoral's Degree in 2008-2012, she was a recipient of the scholarship from The 72 nd Anniversary of His Majesty King Bhumibol Adulyadej and The 90TH Anniversary of Chulalongkorn University Fund. (Ratchadaphiseksomphot Endowment Fund.) from Chulalongkorn University and the research supporting scholarship of Center of Excellence in Mathematics from Mahidol University during 2010-2012.

In 2010, she participated in Ph.D. Colloquia award section in the International Conference on Machine Learning and Cybernetics (ICMLC) and the International Conference on Wavelet Analysis and Pattern Recognition (ICWAPR) 2010 at Qingtao, China.

Publication

P. Silapasuphakornwong, S. Phimoltares, C. Lursinsap, and A. Hansuebsai, "Posture recognition invariant to background, cloth textures, body size, and camera distance using morphological geometry.," in ICMLC, pp. 1130–1135, IEEE, 2010.

Research interests

Her research interests include image processing, computer vision, and human-computer interaction.