

行政院國家科學委員會專題研究計畫 成果報告

智慧型居家看護影像監控系統(3/3) 研究成果報告(完整版)

計畫類別：個別型
計畫編號：NSC 95-2221-E-002-234-
執行期間：95年08月01日至96年10月31日
執行單位：國立臺灣大學電機工程學系暨研究所

計畫主持人：陳永耀

計畫參與人員：碩士班研究生-兼任助理：馮孝天、黃璿、劉致廷、江易道

報告附件：出席國際會議研究心得報告及發表論文

處理方式：本計畫可公開查詢

中華民國 97 年 01 月 31 日

行政院國家科學委員會補助專題研究計畫 成果報告
 期中進度報告

智慧型居家看護影像監控系統

Intelligent Vision-based Home Care System

計畫類別： 個別型計畫 整合型計畫

計畫編號：NSC-95-2221-E-002-234

執行期間：95年8月1日至96年10月31日

計畫主持人：陳永耀 教授 (Chen, Yung-Yaw)

共同主持人：

計畫參與人員：馮孝天、黃璿、劉致廷、江易道

成果報告類型(依經費核定清單規定繳交)： 精簡報告 完整報告

本成果報告包括以下應繳交之附件：

- 赴國外出差或研習心得報告一份
- 赴大陸地區出差或研習心得報告一份
- 出席國際學術會議心得報告及發表之論文各一份
- 國際合作研究計畫國外研究報告書一份，

處理方式：除產學合作研究計畫、提升產業技術及人才培育研究計畫、
列管計畫及下列情形者外，得立即公開查詢

涉及專利或其他智慧財產權， 一年 二年後可公開查詢

執行單位：國立台灣大學 電機工程系(所)

中 華 民 國 九 十 七 年 一 月 三 十 一 日

Abstract

Intelligent video surveillance system is discussed for years and applied in many areas, like home care systems, security in the public place and so on. Due to this technique development, it can lower down the product cost and decrease the error judgment by human.

In this NSC project, we focused on the home care system which can analyze the human activities automatically. As the growing number of the elder population, more and more people cannot take care of their parents or the elderly all the time. There is definitely a demand to build a system with the ability to watch the elderly and alarm people when they have abnormal behaviors, such as falling down to the ground. In our concept we first want to recognize the human postures, so we set a CCD camera to grab the image containing the elderly and environment. And then we separate the person from the environment using the background subtraction method. After the object is subtracted, some parameters from the silhouette will be extracted and viewed as the input of our classification system. As for the classification system, we use the adaptive fuzzy rule-based system. It can generate the fuzzy rules automatically according to the feature parameters we extract, and give a good performance in classifying the postures.

Keywords: Human Activities, Background Subtraction, Adaptive Fuzzy-Rule Based Classified System

中文摘要

智慧型監控系統在未來的科學研究領域當中，將會是一個值得令人討論的技術。其應用層面相當之廣泛，不管在居家看護、公共場所安全檢測...等，經由這類的技術發展，不僅可以降低成本還可以進而改善人為疏失的問題。

在本國科會計劃當中，將會著重在居家看護系統分析與建立的部分。而在居家看護當中我們所關切的是人類動作行為異常的判斷，也就是說，隨著老年化的社會成長，越來越多的老年人會待在家中，而當這些老人跌倒甚至發生危險時，我們將會透過這樣的系統提醒正在遠方工作的兒女，讓他們可以及時的做出解救的動作。而在計劃中，除了異常姿態的辨識，我們也將辨識一般日常行為中的人體姿態。我們首先利用 CCD 攝影機擷取環境影像，並且利用背景去除的影像處理方式，將移動中的物體擷取出來，並對此物體作簡單的特徵擷取。藉由得取的這些參數，利用本篇提及的適應性自我學習模糊規則分類系統，產生模糊規則並且利用這些規則判斷人體姿態。本系統建立了一個常態的模糊規則庫，並將此應用在人體姿態的分類，最後達成智慧型辨識的能力。

關鍵字：居家看護系統、人體姿態辨識、模糊適應性、模糊規則、背景去除

Table of Contents

Abstract I	
中文摘要 II	
Table of Contents.....	III
List of Figures.....	V
List of Tables.....	VII
1 INTRODUCTION.....	1
1.1 Motivation	1
1.2 Purpose of Research	2
1.3 Overview of System Architecture	3
2 PRELIMINARIES	5
2.1 Introduction of Human Activities Identification.....	5
2.2 Modeling of Human Body	5
2.2.1 Model Based Approaches	6
2.2.2 Non-Model Based Approaches	10
2.3 Human Activities Recognition Approaches	12
2.3.1 State-Space Approaches	12
2.3.2 Template Matching Approaches.....	16
2.3.3 Silhouette Based Method	17
3 IMAGE PROCESSING AND FEATURE EXTRACTION.....	23
3.1 Moving Object Segmentation	24
3.1.1 Block Diagram for Background Subtraction.....	24
3.1.2 Auto-Adjust Threshold Algorithm	25
3.2 Features Extraction.....	27
3.2.1 Definition of Postures	28
3.2.2 Definition of Feature Parameters	30
4 ADAPTIVE FUZZY RULE-BASED CLASSIFICATION SYSTEMS	37
4.1 Fuzzy Rule-Based Classification Systems	37
4.1.1 Fuzzy Rule Generation Procedure	38
4.1.2 Fuzzy reasoning	40
4.1.3 Learning Algorithm.....	42
4.2 Apply the AFRBCS in Posture Recognition	43
5 EXPERIMENT AND RESULTS	55

5.1 Hardware Architecture and Environment.....	55
5.2 Experimental Section	56
6 DISCUSSIONS AND CONCLUSIONS	67
Reference List	68
計畫成果自評.....	73
可供推廣之研發成果資料表.....	74

List of Figures

Figure 1.1 Procedure of human posture recognition.....	3
Figure 1.2 Procedure of human postures recognition.....	4
Figure 2.1 Classification of human motion analysis [3].....	6
Figure 2.2 A stick-figure human model [12].....	7
Figure 2.3 A 2D contour human model [13].....	8
Figure 2.4 A volumetric human model [23]	10
Figure 2.5 Detecting and Tracking Human “blobs” with The PFINDER System [30]	11
Figure2.6 Hidden Markov Model (HMM)	14
Figure2.7 Example of MEI and MHI [43].....	17
Figure 2.8 Segmentation of a grey-level image through the traditional-snake (a) and GVF-snake (b).Comparison between the final shape of the GVF-snake and the real human shape (c).....	18
Figure 2.9 Left image shows the active contour at the end of the deformation. Right graphics represent the relative Euclidean distance and histogram of radial projections.	19
Figure 2.10 Target pre-processing. A moving target region is morphologically dilated (twice) then eroded. Then its border is extracted.	20
Figure 2.11 The boundary is “unwrapped” as a distance function from the centroid. This function is then smoothed and external points are extracted.....	21
Figure 3.1 Moving object segmentation diagram.....	24
Figure 3.2 (a) Background Image (b) Object Moving Image (c) Segmentation Result	25
Figure 3.3 Histogram of AD values, x axis is the AD value of total image and y axis is the counter number	26
Figure 3.6 Stand Postures	28
Figure 3.7 Sit Posture.....	29

Figure 3.8 Kneel Posture.....	29
Figure 3.9 Bend Posture.....	29
Figure 3.10 Lie Posture	30
Figure 3.11 Bounding box of the human silhouette and its parameters.....	30
Figure 3.12 (a) Segmentation (b) Column Index vs. Counter Numbers.....	31
Figure 3.13 Human Hand Spread Plot	32
Figure 3.14 Modified bounding box of human silhouette.....	32
Figure 3.15 Histogram of the human to modify bounding box boundary	33
Figure 3.16 Definition of neck line plot	34
Figure 3.17 Parameters of kneel-pose	35
Figure 3.18 Parameters of Bend-pose.....	35
Figure 3.19 All of the Feature Parameters Plot.	36
Figure 4.1 Plot of membership function of K=5 subsets.....	38
Figure 4.2 Two dimension membership function with don't-care term.....	39
Figure 4.3 Training vectors of five postures distribution in the two dimensions (Xo,Xm) plot.	44
Figure 4.4 Training vectors of five postures distribution in the two dimensions (Xo,Xk) plot.	45
Figure 4.5 Training Vectors of Five Postures Distribution in The 2-D (Xm,Xb) Plot.	45
Figure 4.6 AFRCs Architecture	51
Figure 4.7 Standing Training Patterns Images.	52
Figure 4.8 Sitting Training Patterns Images.....	52
Figure 4.9 Lying Training Patterns Images	53
Figure 4.10 Kneeling Training Patterns Images.....	53
Figure 4.11 Bending Training Patterns Images	53
Figure 5.1 General Postures Recognition Demo Images.....	61
Figure 5.2 Ill-postures recognition demo images.....	65

List of Tables

Table 4.1 Generating Rules Index Form.....	47
Table 4.2 Training Patterns for Stand Postures.....	48
Table 4.3 Rules for Recognizing Standing Postures (the first 15 terms).....	49
Table 4.4 Each β_{class} Value of Rule 2060.	49
Table 4.5 Five postures and its corresponding feature parameters.	50
Table 4.6 Training Vectors vs. $\max(\mu(\mathbf{x}_p) * CF_{class})$ Values.....	51
Table 5.1 List of experiment equipment.	55
Table 5.2 Hit rate for the demo images.	66

1 INTRODUCTION

1.1 Motivation

The concept of e-home is developed for years, and it involves many aspects of technology. The main goal for an e-home system is always to find an automatic control mechanism that involves watching if there exist some dangers in the home environment and switching on the power of the electric equipments before people coming home and so on. There are many researches in solving these problems, in our research we will focus on the real time intelligent video surveillance systems aspect.

Intelligent video surveillance systems have been mentioned in recent years. We can find many applications in many areas such as airports monitoring systems, home care systems, bank monitoring systems...etc. Especially in home care systems, it is more important in this time as the growing numbers of the elder people. As we introduce before the development of video surveillance systems is so fast cause of that we do not need a lot of security personnel to monitor the human actions. Computer vision is usually used in this kind of intelligent video surveillance systems. It is convenient cause of the information of human body is gathered by using the image processing method without touching or putting the sensor directly on the human body.

As the growth of the elder people, the elderly have more and more time to stay at home without the children's care. The fast solution is to hire a person nursing beside, but it always has the cost problem. As we know that the cost of the consuming electronics in this time is cheaper and cheaper, to establish a system with some simple electrical products is more convenient and cost down. That is people just need to buy a display and set it up on its own office. Then the parents living situations will be transferred to the display and tell what if there happens some

dangerous.

The home care systems are expected to provide the functions that can monitoring the people inside the house and alarm the watcher when people in some dangerous situations. To give more applications even the human postures can be classified is also needed added. In our approach we want to establish a simple system with a PC and a CCD camera only to get the goal. More application it may be used in the other regions such as the robotics or some artificial science.

1.2 Purpose of Research

In the human action recognition part there are many tasks that are hard to solve. There are two main problems, one is the image processing part and the other is how to define the human action.

In discussing the image processing, the images shot by CCD camera are usual including the moving object and the background. So how to subtract the object we want is a main task dealing with in our system. After getting the object the human action is defined in what descriptions is the next problem. As for the human action there are many words expressed in the similar way like action, behavior, motion, gesture, posture, deed, and movement. For simply dividing, one is the individual static frames like gesture and postures, and the other is continuous dynamic frames like action and behavior.

Our research provides a solution to establish a home care system with the images processing and human action recognition. The image processing part will use a easy concept of background subtraction to complete, and the human action recognition here is focus on the static frame discussion. That is we mainly discuss on the postures recognition parts. The postures part is also can be view as the basic of the human behavior recognition, as you can recognize the human postures the human behavior will be more easily defined.

1.3 Overview of System Architecture

As mentioned before for the easy setup, in our system a camera and a PC are needed and show as figure 1.1. In figure it shows the whole system architecture, the CCD camera is fixed in the corner of the indoor environment and using the image grabbing card to get images, and a PC will process the images and tell the situation happened in the house.

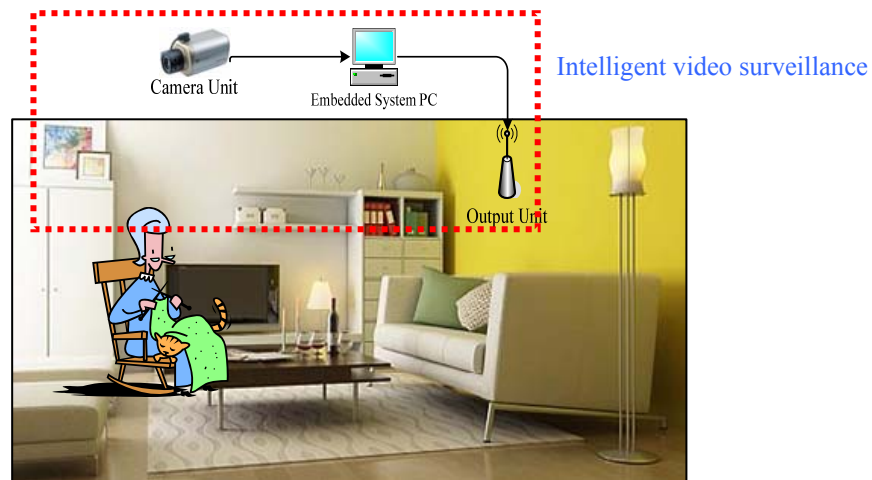


Figure 1.1 Procedure of human posture recognition.

The procedure of imaging processing subtracts the human silhouette from the background image. The silhouette helps us to define the human parameters for the next postures recognition procedure. Of course the parameters we get from the silhouette are usually some characteristic of the human postures. In the postures recognition, it can use the figure 1.2 to explain how it works precisely.

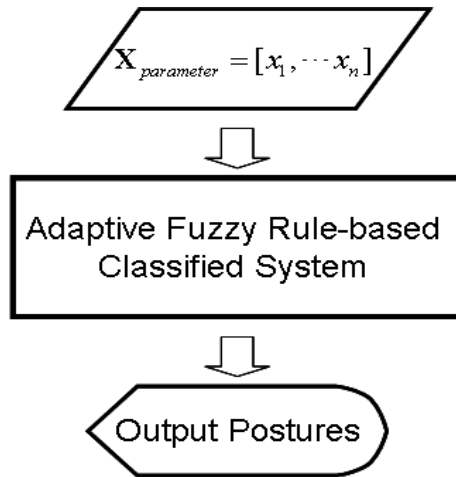


Figure 1.2 Procedure of human postures recognition.

In figure 1.2, the vectors X_p are composed of the parameters we get from the procedure before, and they are viewed as the systems input. In the adaptive fuzzy rule based classified system, it is based on the original fuzzy rule based method with a little change in the generating fuzzy rules. The detail will be introduced in the following sections.

2 PRELIMINARIES

2.1 Introduction of Human Activities Identification

Human activities identification is a quite challenge topic in many kinds of respects. It is complicated in that it can be divided into motion estimation, behavior understanding, and postures recognition. As the various applications of the identification of human activities such as video surveillance, artificial intelligence, medical diagnosis and so on, it is more and more discussed in the computer vision science. In analyzing this problem it usually can be separated from the processing procedure parts. That is in the high level processing tasks usually deal with the body modeling and the whole action identification. As for the low level processing part, they are usually taking the video tracking, segmentation, and pose recovery.

In other words before discussing the problem we need to figure out how the different processing parts works. So in this chapter, section 2.2 will show how to model the human body in two kind of respects, one is based on the pre-build modeling based method to recover a human model and the other is build a human model without pre-model. The section 2.3 will show the key method in modifying the human actions including the HMM method, template matching method and so on. Finally we will give a summary in section 2.4 for the method introduced in this chapter.

2.2 Modeling of Human Body

Human motion analysis usually includes human body part segmentation and the recovery of 3D structure from the 2D projections in an image sequence. There are two typical approaches related motion analysis. One is model based method and the other is non-model based approach.

The difference between the two methodologies is in the process of establishing feature correspondence between consecutive frames. Methods which assume a priori shape models match the 2D image sequences to the model data. Feature correspondence is automatically achieved once matching between the image and the model data is established. As no a priori shape models are available, correspondence between successive frames is based upon prediction or estimation of features related to position, velocity, shape, texture, and shape. Conventionally, the representation of human body maybe uses the stick figures, 2D contours, or volumetric model. Section 2.2.1 and 2.2.2 would respectively discuss model based and non-model based approaches. Figure 2.1 shows the classification of human motion analysis.

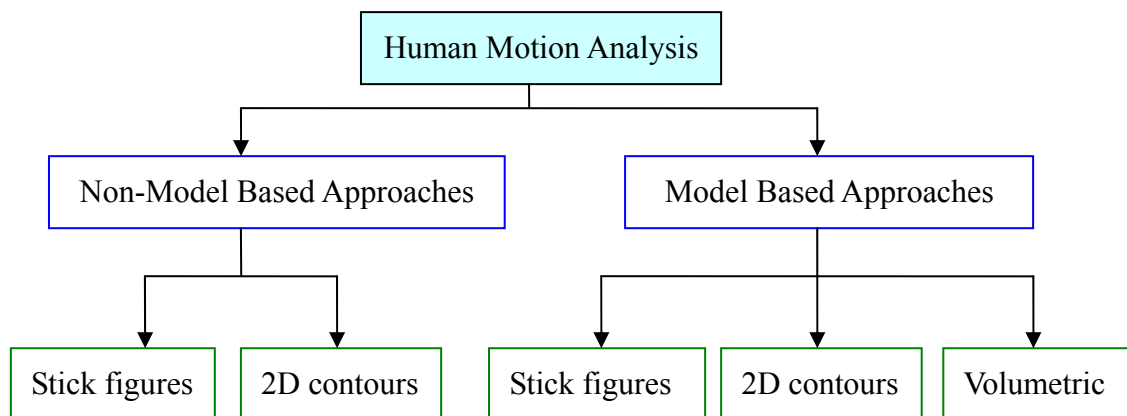


Figure 2.1 Classification of human motion analysis [3]

2.2.1 Model Based Approaches

Model-based approaches build the body representation by fitting to the image data the predefined parameter values of a parametric body model. This approach can efficiently integrate shape knowledge and visual input, and are better for high-level identification of complicated motions. However, model-based approaches usually require additional processing steps of model selection and parameter estimation to fit the model to a given visual input. Addition of a new

activity or motion may require significant complexity to model-based techniques. In model-based approaches, the fitting process involves either an optimization scheme such as the least square method of Gavrilu and Davis [10] or a stochastic sampling scheme such as the particle filtering method by Isard and Blake [11].

In general, human bodies could be represented as stick figures, 2D contours, or volumetric models. Thus, body segments can be approximated as lines, 2D ribbons, and elliptical cylinders, accordingly. Bharatkumar *et al.* [29] used stick figures to model the lower limbs of the moving human body. They aimed at constructing a general model for gait analysis in human walking. Chen and Lee [12] recovered the 3D configuration of a moving subject according to its projected 2D image. Their model used 17 line segments and 14 joints to represent the features of the head, torso, hip, arms, and legs. The model is shown in figure 2.2.

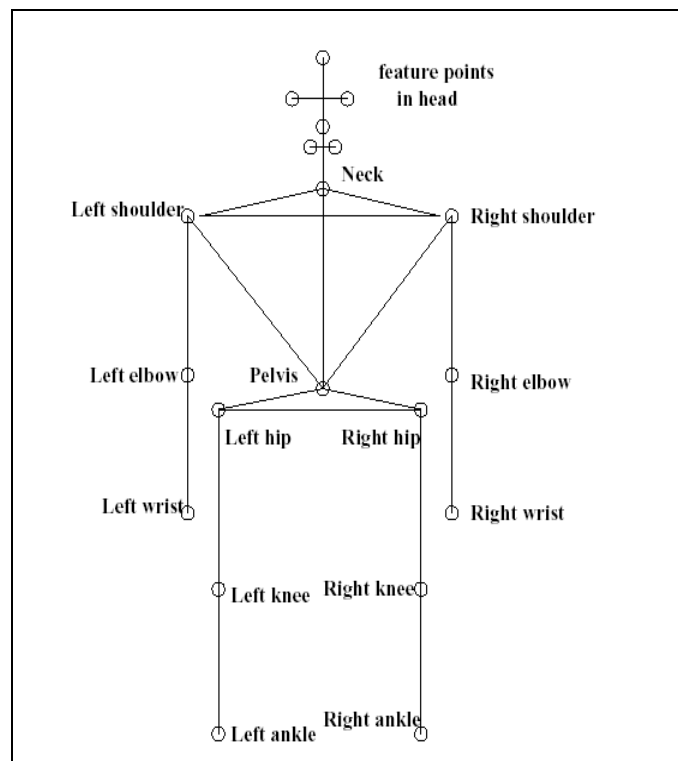


Figure 2.2 A stick-figure human model [12]

Various constraints were imposed for the basic analysis of the gait. The method was computational expensively, as they use the full search to find the possible combinations of 3D

configurations, given the known 2D projection, and required accurate extraction of 2D stick figures.

The pioneer work of O'Rourke and Badler [27] used a volumetric model consisting of 24 rigid segments and 25 joints. The surface of each segment is defined as a collection of overlapping sphere primitives. Sato *et al.* [28] developed a different approach that treats a moving human as a combination of various blobs of its body parts. Leung and Yang [13] applied a 2D ribbon model to recognize poses of human performing gymnastic movements. The emphasis of their work is to estimate motion solely from the outline of a moving human subject. The system consists of two major processes: extraction of human outline and interpretation of human motion. The 2D ribbon model is comprised of two components, the "basic" body model and the "extended" body model. The basic body model outlines the structural and shape relationships between the body parts, shown in Figure 2.3.

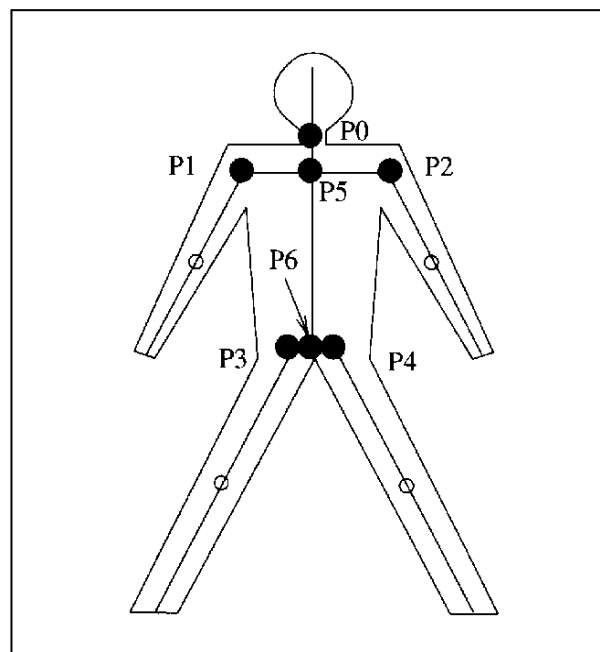


Figure 2.3 A 2D contour human model [13]

The extended model consists of three patterns: the support posture model, the side-view kneeling model, and side horse motion model. A modified edge detection technique was

developed based on the work in Jain and Nagel [22] to generate a complete outline of the moving object images, A spatial-temporal relaxation process was proposed to determine which side of the moving edge belongs to the moving object. Two sets of 2D ribbons on each side of the moving edge, either a part of the body or that of the background, are identified according to their shape changes over time. The body parts are labeled according to the human body model. Then, a description of the body parts and the appropriate body joints is obtained.

Elliptical cylinders are one of the commonly used volumetric models for modeling human forms. Hogg [23] and Rohr [24] used the cylinder model originated by Marr and Nishihara [25], in which the human body is represented by 14 elliptical cylinders. Each cylinder is described by three parameters: the length of the axis and the major and minor axes of the ellipse cross section. The origin of the coordinate system is located at the center of the torso. Hogg attempted to generate 3D descriptions of a human walking by modeling. Hogg [23] presented a computer program (WLKER) which attempted to recover the 3D structure of a walking person. Rohr [26] found the initial pose of the human first and then incremented it from frame to frame. He used the Kalman filter, edge segmentation, and a motion model tuned to the walking image object for a more robust result. He applied eigenvector line fitting to outline the human image, and then fitted the 2D projections into the 3D human model using a distance measure similar to Hogg [23]. Besides, Rohr [26]'s method identified the straight edges for tracking the restricted movement of walking human parallel to the image plane. The representation of volumetric human model is shown in figure 2.4.

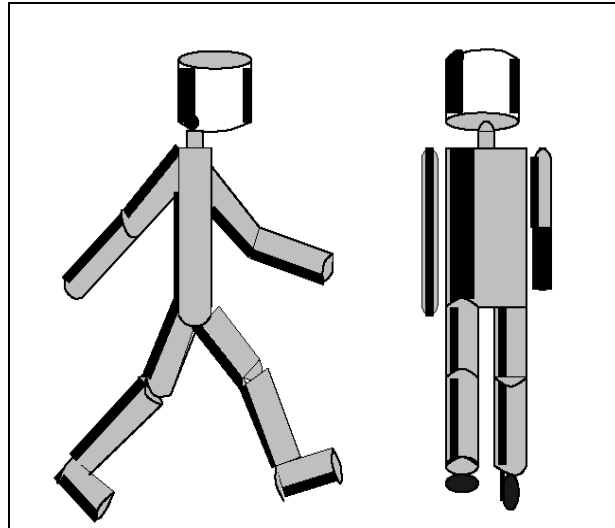


Figure 2.4 A volumetric human model [23]

All of these approaches must match each real image frame to the corresponding model, which represents the human body structure at an abstract level. This procedure is itself non-trivial. The complexity of the matching process is governed by the number of model parameters and the efficiency of human body segmentation. When fewer model parameters are used, it is easier to match the feature to the model, but more difficult to extract the feature. For example, the stick figure is the simplest way to represent a human body, and thus it is relatively easier to fit the extracted lines into the corresponding body segments. However, extracting a stick figure from real images needs more care than searching for 2D blobs or 3D volumes.

2.2.2 Non-Model Based Approaches

As no predefined shape models are assumed, heuristic assumptions, which impose constraints on feature correspondence and decreasing search space, are usually used to establish the correspondence of joints between successive frames. The approaches of non-model based human motion analysis are applicable to more diverse situations. However, these approaches are sensitive to noise in general, because they lack any mechanism to distinguish noise from signal in

visual input.

The approaches of non-model based human motion analysis are also called as view-based approaches. They build a body representation in a bottom-up fashion by first detecting appropriate features in an image. The methods do, however, build some sort of model to represent the pose. These pose representations are points, simple shapes, and stick-figures. The pose of the subject may be represented by a set of points and is widely utilized as markers are attached to the subject. Without the markers, the hands and the head may be estimated from the image represented by just three points [30] as figure 2.5. This is a very compact representation which suffices for a variety of applications. By utilizing color segmentation [31] or blob segmentations [30], these three points are usually found.

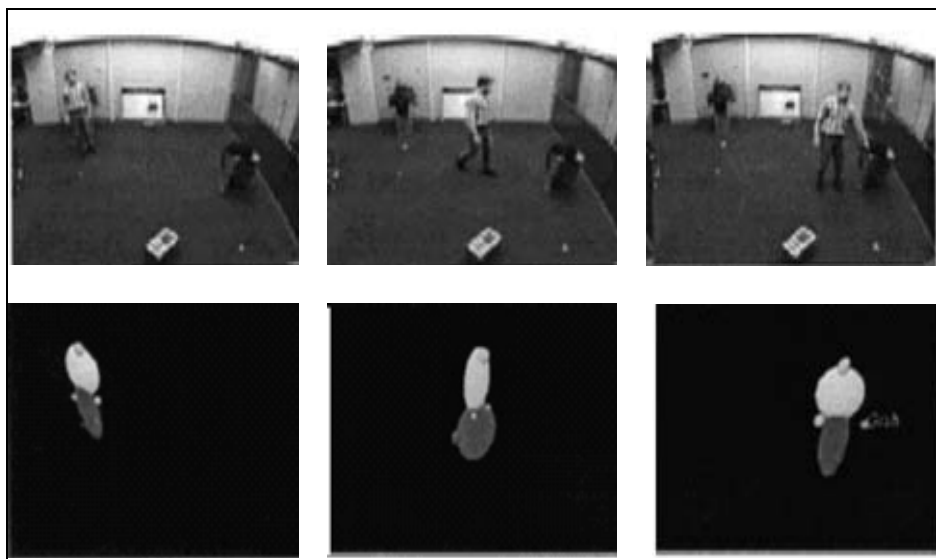


Figure 2.5 Detecting and Tracking Human “blobs” with The PFINDER System [30]

A subject in an image can be simply represented into some boxes by the boundary of subject [32]. However, the boundary box representation is mainly utilized as an intermediate representation during processing and not as the final representation. Instead, shapes which are more human-like, such as ellipses [33], may be used as a final representation.

A rather different approach of view-based model is to learn a mapping directly from image

features to posture data. In the work proposed by Brand [34], the paths through the 3D state space obtained through training are modeled using an HMM where the states are linear paths modeled by multivariate Gaussians. As in [35] the moments (central) are found by synthesizing various poses. The moments are associated to the HMM. Altogether a sequence of moments is mapped to the most likely sequence of 3D poses. Obviously this approach is either offline or includes a significant time-lag, but it has the ability to resolve ambiguities using hindsight and foresight.

2.3 Human Activities Recognition Approaches

For human activity or behavior recognition, most efforts have been concentrated on using state-space approaches [36] to understand the human motion sequence [37, 38, 39, 40]. Another approach is to use the template matching technique [41, 42, 43] to compare the feature which is extracted from the given image sequence to the pre-stored patterns during the recognition process. There are still some researches in discussing how to get the human feature from the silhouette image, like image skeletonization [5, 29], active contours based [1], and so on.

The approaches which are related to human action recognition will be surveyed and introduced in this section. Besides, the recognition of human behavior is our future desired research. Some approaches of human behavior recognition will be presented below.

2.3.1 State-Space Approaches

Approaches using state-space models define each static posture as a state. These states are connected by certain probabilities. Any motion sequence as a composition of these static poses is considered a tour going through various states. Joint probabilities are computed through these tours, and the maximum value is selected as the criterion for classification of activities [3]. Most

studies using state-space model have applied the methods of ‘Dynamic Time Warping’ (DTW) [33] or ‘Hidden Markov Model’ (HMM) [30].

DTW is a method of sequence comparison used in various applications such as DNA comparison in microbiology, comparison of strings of symbols in signal transmission, and analysis of bird songs and human speech. DTW deals with differences between sequences by operations of deletion-insertion, compression-expansion, and substitution, of subsequences. By defining a metric of how much the sequences differ before and after these operations, DTW classifies the sequences. DTW can also be applied to image sequences. However, DTW lacks the consideration of interactions between nearby subsequences occurring in time. In many actual situations, a sequence has higher correlation between closer subsequences than between distant subsequences [44].

State-space models have been widely used in the recognition of human activities. Hidden Markov Model (HMM) is a finite set of states, each of which is associated with a probability distribution. HMM has been successfully applied in speech recognition [10, 35]. Recently, it has been applied in the research field of image processing and computer vision such as handwritten character recognition, human action recognition [30, 45], sign language recognition, and facial expression identification [12]. The conception of HMM comes from the Markov chain. The most difference between hidden Markov model and Markov chain is that HMM can not acquire the states of event directly. HMM must utilize the observation function to infer the event states indirectly. The observation is a probabilistic function of the state. The basic HMM model can see the Figure 2.6.

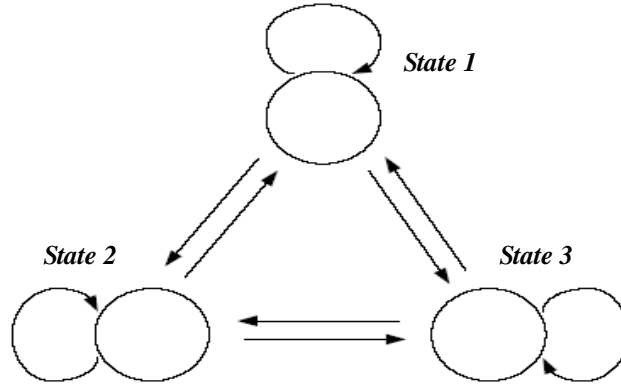


Figure2.6 Hidden Markov Model (HMM)

In order to define an HMM completely, following elements are needed.

- N : the number of states in the posture graph. The model state at time t is denoted as q_t , $1 \leq q_t \leq N$, $1 \leq t \leq T$.
- T : length of the observation sequence.
- $S = \{ S_i \mid 1 \leq i \leq N \}$: a set of model-states; if the state is k at time t , it is denoted as $q_t = S_k$.
- M : the number of observation model-states.
- $V = \{ v_i \mid 1 \leq i \leq M \}$: a set of observation model-states.
- $A = \{ a_{ij} \mid 1 \leq i, j \leq N \}$: a state transition probability distribution, where $a_{ij} = P(q_j \text{ at } n+1 \mid q_i \text{ at } n)$.
- $B = \{ b_j(k) \mid 1 \leq j \leq N \text{ and } 1 \leq k \leq M \}$: an observed state probability distribution where $b_j(k) = P(v_k \text{ at } t \mid q_t = S_j)$.
- $\Pi = \{ \pi_i \mid 1 \leq i \leq N \}$: an initial state probability distribution, where $\pi_i = P(q_1 = S_i)$.
- $\lambda = (A, B, \pi)$: HMM parameter.
- $O = \{ O_1, O_2, \dots, O_T \}$: observable symbol sequence.

The complete parameter set λ of the discrete HMM is represented by vector π and matrices A

and B . In order to accurately describe a real world process such as posture with a HMM, the HMM parameters have to be selected accurately. The parameter selection process is called the HMM training process. This parameter set λ can be used to evaluate the probability $P(O | \lambda)$, that is to measure the maximum likelihood performance of an output observable symbol sequence [46].

There are three basic problems in HMM design:

1. Evaluation problem

Given the observable symbol sequence $O=\{O_1, O_2, \dots, O_T\}$ and the HMM parameter set $\lambda=(A, B, \pi)$, how do we efficiently compute the probability $P(O | \lambda)$?

2. Estimation problem

Given the observable symbol sequence $O=\{O_1, O_2, \dots, O_T\}$ and the HMM parameter set $\lambda=(A, B, \pi)$, how do we choose an optimal corresponding state sequence $q = (q_1, q_2, \dots, q_T)$?

3. Training problem

How do we adjust an HMM parameter set $\lambda=(A, B, \pi)$ in order to maximize the output probability $P(O | \lambda)$ of generating the output observable symbol sequence $O=\{O_1, O_2, \dots, O_T\}$?

These problems have had complete and effective theory to solve. Problem 1 utilizes the forward-backward algorithm to reduce computation [49, 50]. Problem 2 uses the Viterbi algorithm to find the state which has a maximal probability at a given time [47, 48]. Problem 3 utilizes the Baum-Welch algorithm to increase training efficiency [46].

A good tutorial for details of HMM is proposed by Rabiner [48]. HMM has been more popular than DTW for dynamic representation of action because of its ability to handle uncertainty in its stochastic framework. For instance, hand motions are classified by HMM [49]. Min *et al.* [50] used both static representation and dynamic representation for hand gesture recognition. Yang *et al.* [51] applied HMM to classify human action intent and to learn human skills.

2.3.2 Template Matching Approaches

The template matching approaches analyze the individual frames first and then combine the results. The motion magnitude in each cell is summed, forming a high dimensional feature vector used for recognition. To normalize the duration of the movement, they assume that human motion is periodic and divide the entire sequence into a number of cycles of the activity. Motion in a single cycle is averaged throughout the number of cycles and differentiated into it fixed number of temporal divisions. Finally, activity recognition is processed to use the nearest neighbor algorithm.

In an early research, Herman [52] utilized the template matching of stick-figures of a given frame to analyze different poses of a person. The emotions and actions at a given frame based on the person's pose are inferred. His stick figure was built by manually locating body parts, and he analyzed individual frames separately, without considering the interrelations between the frames in a sequence. Akita [53] used template matching of silhouette images to recognize different motions of a person in tennis play. Recent work by Bobick and Davis [41], they follows the same vein, but extracts the motion feature differently. They interpret human motion in an image sequence by using motion-energy images (MEI) and motion-history images (MHI). The example can see the figure 2.7. The MEI is binary images containing motion blobs. The motion images in a sequence are calculated via differencing between successive frames and then threshold into binary values. The MEI ($E_T(x, y, t)$) is combined by these motion images in time and is defined:

$$E_T(x, y, t) = \bigcup_{i=0}^{\tau-1} D(x, y, t-i) \quad (2.1)$$

where $D(x, y, t)$ is a binary image sequence indicating regions of motion, τ is critical in defining the temporal extent of a movement.

MHI is to represent how motion the image is moving. The each pixel value is proportional to the duration of motion at that position. The MHI ($H_{\tau}(x, y, t)$) is defined:

$$H_{\tau}(x, y, t) = \begin{cases} \tau & \text{if } D(x, y, t) = 1 \\ \max(0, H_{\tau}(x, y, t-1) - 1) & \text{otherwise} \end{cases} \quad (2.2)$$

The MEI can be generated by threshold the MHI above zero. Moment-based featured are extracted from MEI and MHI and employed for recognition using template matching.

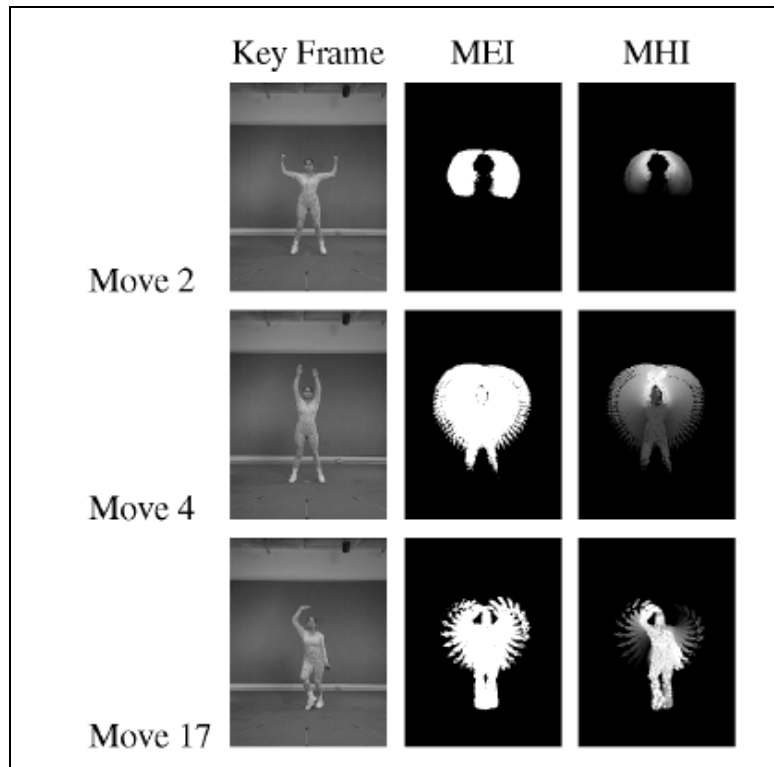


Figure 2.7 Example of MEI and MHI [43]

2.3.3 Silhouette Based Method

In this kind of method, they all focus on how to get the human characteristic features to do

the next activities identifications. It means first the image subtractions for the moving object is all needed to be done. As the silhouette getting the next procedure is divided into some different method. [1] efforts the active contours features to represent the human postures, [2] take the Fourier transform concept to develop an image Fourier descriptor method. It will be introduced one by one.

A. Active Contour Based Method

In the research mentioned, Buccolieri [1] give a method using the active contours as the image features and take this as the NN input to model a systems in recognition the postures. The system architecture consists of five sequential modules that include the moving target detection process, two levels of segmentation process for interested element localization, features extraction of the object shape and a human posture classification system based on the Radial Basis Functions Neural Network. Moving objects are detected by using an adaptive background subtraction method with an automatic background adaptation speed parameter and a new fast Gradient Vector Flow snake algorithm for the elements segmentation is proposed.

The GVF-snake method can separate the object as the figure 2.8, its detail derivation is not mentioned here.

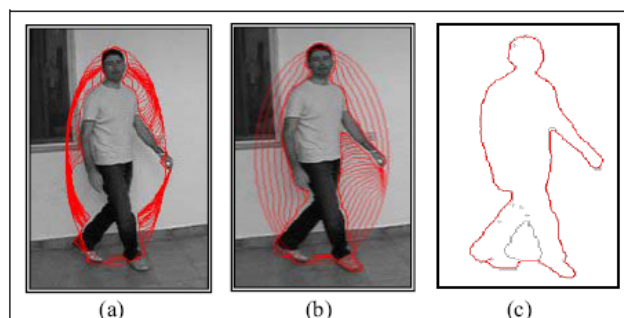


Figure 2.8 Segmentation of a grey-level image through the traditional-snake (a) and GVF-snake (b). Comparison between the final shape of the GVF-snake and the real human shape (c).

In its research, it use the three layers Radial basis function neural network, and as for the input of these system is the Euclidian distance form the snake center that is shown in the figure 2.9. The Euclidian distance is calculated by the equation 2.X expressed as bellow.

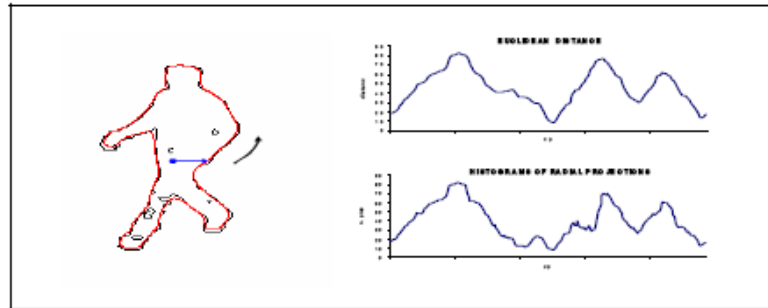


Figure 2.9 Left image shows the active contour at the end of the deformation. Right graphics represent the relative Euclidean distance and histogram of radial projections.

$$d_i = \sqrt{(x_i - x_c)^2 + (y_i - y_c)^2} \quad i = 1 \dots n \quad \dots \dots \dots (2.3)$$

$$x_c = \frac{1}{n} \sum_{i=1}^n x_i, \quad y_c = \frac{1}{n} \sum_{i=1}^n y_i \quad \dots \dots \dots (2.4)$$

Where (x_c, y_c) means the center of the contours and (x_i, y_i) is the coordinate of each point on the contour, and n is the number of the contour points. As the features can be extracted the input is chosen from these data. Although these features are easily evaluated, a high number of them is necessary (depending of the object’s size) and, moreover, they are too sensitive to excess parts (noise and shadow) and to lacking parts.

There is another extension active contours method for the analysis of human motions. [5, 29] take a more deepgoing process to represent a human model. From these, a “star” skeleton is produced. Two motion cues are determined from this skeletonization: body posture, and cyclic motion of skeleton segments. These cues are used to determine human activities such as walking or

running, and even potentially, the target S gait. Unlike other methods, this does not require an a priori human model, or a large number of ‘pixels on target’’. Furthermore, it is computationally inexpensive, and thus ideal for real-world video applications such as outdoor video surveillance.

For getting the ‘star’ shape of a tracking human model it is using some image process from segmentation of the moving object and using morphology process like dilation and erosion it shows in the figure 2.10. The figure shows the whole preprocess procedure step by step. The process dilation is first fitting the hole in the original image then using erosion the make the image thinner. Finally grab the contour for the image and finish the pre-process.

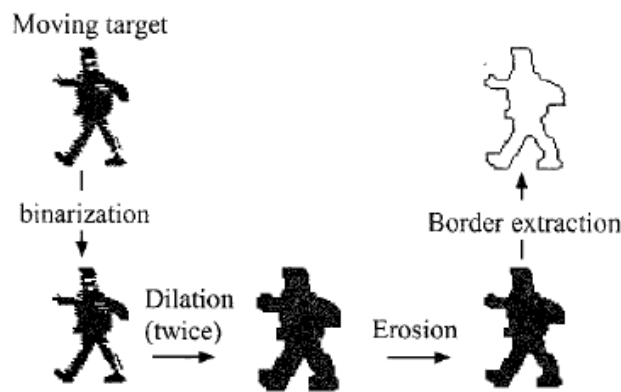


Figure 2.10 Target pre-processing. A moving target region is morphologically dilated (twice) then eroded.

Then its border is extracted.

As the pre-process finished, the data process part is then continued. It is shown in the figure 2.11. The figure shows the main procedure in getting the star skeleton, that is first calculate the active contour like before and the take the DFT for the data curve, and by using low pass filter to get the low frequency characteristic. Finally do the inverse DFT can get another smooth curve and find the characteristic point to represent the apex of the star shape. And the whole diagram is shown in the figure 2.11.

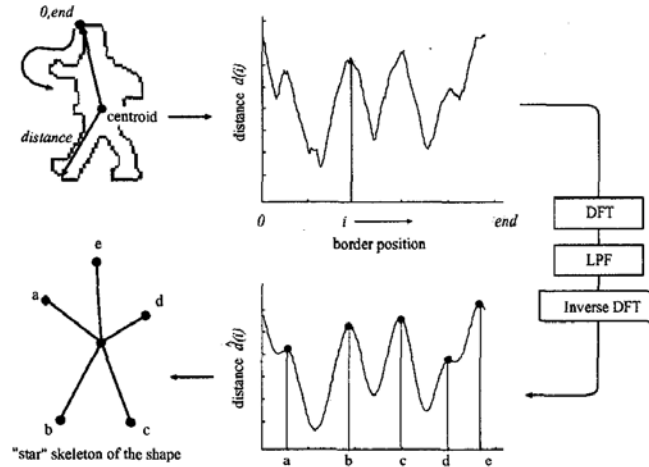


Figure 2.11 The boundary is “unwrapped” as a distance function from the centroid. This function is then smoothed and external points are extracted.

As a reason that the star shape is somewhat like our hand and foot, the angle from the shape can be view as the motion analysis key point. Besides the angle of the star apex the period of the skeleton changing is also another human motion analysis property. The system then applies the concept to model a running person motion and the detail here will not to be focused.

B. Fourier Descriptor

In the paper, it assumes that a human silhouette can be extracted from monocular visual input. The idea behind Fourier descriptors is to describe a silhouette by a fixed number n of sample points $\{(x_0, y_0), \dots, (x_{n-1}, y_{n-1})\}$ on the boundary. Since Fourier descriptors can describe only a single closed-curve shape, we only sample along the external boundary. Sampling can be done randomly but in practice, equidistant sampling is preferred to make sure the sampling is more uniform. The sample points are transformed into complex coordinates $\{z_0, \dots, z_{n-1}\}$ with $z_i = x_i + y_i * \sqrt{-1}$ and are further transformed to the frequency domain using a Discrete Fourier Transform (DFT). The results of this transformation are called

the Fourier coefficients, denoted by $\{f_0, \dots, f_{n-1}\}$.

The coefficients with low index contain information on the general form of the shape and the ones with high index contain information on the finer details of the shape. The first coefficient depends only on the position of the shape and setting it to zero makes the representation position invariant. Rotation invariance is obtained by ignoring the phase information and scale invariance is obtained by dividing the magnitude values of all coefficients by the magnitude of the second coefficient f_1 . Since after normalization f_0 is always zero and f_1 is always one, we have $n-2$ unique coefficients given by:

$$FD = \left(\left| \frac{f_2}{f_1} \right|, \dots, \left| \frac{f_{n-1}}{f_1} \right| \right) \quad (2.5)$$

This descriptor is a shape signature and can be used as a basis for similarity and for retrieval. When we deal with Fourier descriptors of length n , we actually use only $n - 2$ coefficients. It is clear that Fourier descriptors describe a shape globally. The finer details are filtered out. Now consider two shapes indexed by Fourier descriptors $FD1$ and $FD2$. Since both Fourier descriptors are $(n - 2)$ - dimensional vectors, we can use the Euclidian distance d as a similarity measure between the two shapes:

$$d = \sqrt{\left(\sum_{j=0}^{n-3} |FD1_j - FD2_j|^2 \right)} \quad (2.6)$$

3 IMAGE PROCESSING AND FEATURE EXTRACTION

A real time man-less monitoring system requires three main process procedures. One is moving object segmentation, one is feature extraction and the final is the smart sensing procedure. The first of all is to subtract the foreground image like the elderly from the background image. Many segmentation techniques have been developed which can be classified into five types :

- (1) Temporal Segmentation,
- (2) Spatial Segmentation,
- (3) Spatial-Temporal Segmentation,
- (4) Motion Estimation Segmentation,
- (5) Background Subtraction Segmentation.

In our research we use the fifth kind of method to develop our home care system, we will introduce in the next paragraph. As the foreground object has been captured the features characteristic will be extracted from the object. These features will help us in classifying the human motions. At the final stage we use the Adaptive Fuzzy Rule-based Classification method to model the system. The extracted features are the system input vectors and the output is the classified postures. The whole system will be introduced in the next paragraph step by step.

3.1 Moving Object Segmentation

In this plan, the method we applied in moving object segmentation is the background subtraction. Its main concept is to subtract the current frame from the background frame to separate the foreground object. More precisely, we take the difference operation to see the pixel value change from these two frames. So the problem now is to determine the suitable threshold value to separate the foreground and background. In this plan, the histogram method is used to count the difference pixel value and then the threshold value will be decided. More detail will be introduced in the next sections.

3.1.1 Block Diagram for Background Subtraction

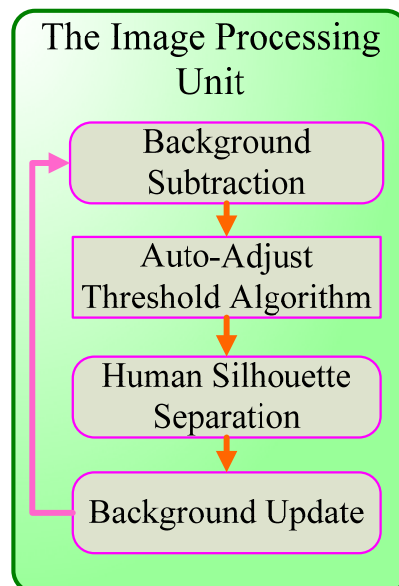


Figure 3.1 Moving object segmentation diagram.

In figure 3.1, it shows the segmentation procedure. In this method the main goal is to get the background and foreground image, so at first the CCD camera will automatically take a quick

shot as the background image. Next step it takes the next frame as the foreground image. Then the system will put these two image into auto-adjust threshold algorithm to get the moving object silhouette, i.e. the gray level value between foreground image and background image changing too large to exceed the threshold the output image gray level value will be assigned as 0. In the final we will get a two gray level value image which gray level value is 0 or 255. In figure 3.2, it shows the result of this applied subtract method.

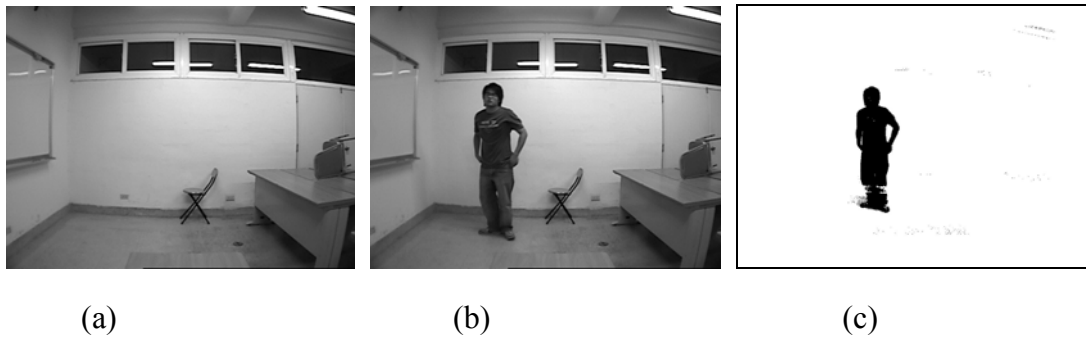


Figure 3.2 (a) Background Image (b) Object Moving Image (c) Segmentation Result

3.1.2 Auto-Adjust Threshold Algorithm

In this section, the plan will introduce how to decide a good threshold to subtract the object. In the traditional method is determined by the following equations :

$$D(x, y, t) = |F(x, y, t) - B(x, y)| \quad (3.1)$$

$$I(x, y, t) = \begin{cases} F(x, y, t) & , \text{if } D(x, y, t) \geq Th \\ 0 & , \text{if } D(x, y, t) \leq Th \end{cases} \quad (3.2)$$

In these equations, $F(x, y, t)$ is the pixel value of current frame. $B(x, y)$ is the pixel value of background frame Th is the global threshold value and $I(x, y, t)$ is the segmentation result. $D(x, y, t)$ is the different value between current frame and background frame. So in these two equations above the different threshold value we choose the segmentation will also have different result.

In order to get the proper threshold value, in this system we will first calculate the AD

(absolute difference), which is defined as :

$$AD_{ij} = |F(i, j, t) - B(i, j)|, \text{ where } i = 1 \sim dx, j = 1 \sim dy \quad (3.3)$$

It means that we first need to calculate the difference value for each pixel in the capture image, where dx means the resolution of the x dimension of the image and dy means the resolution of the y dimension of the image. After getting the AD value we make a histogram to count this AD value. More precisely, we take each pixel absolute difference from (dx , dy)=(1 , 1) to (dx , dy)=(480 , 640), and assign a row vector, HISTOGRAM(1, 260), to save this value. The length of this row vector is 260 in order to make sure that every pixel's gray level difference from 0~255 can be saved. For example, if the AD value is 23 we will put one count in the row vector (1, 23), i.e. the value of the HISTOGRAM (1, 23) will be added one. Figure 3.3 shows the histogram plot.

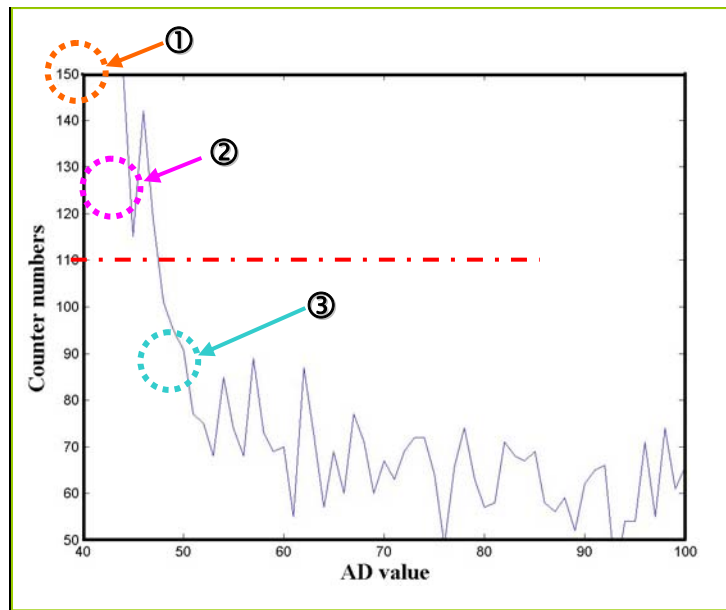


Figure 3.3 Histogram of AD values, x axis is the AD value of total image and y axis is the counter number

As long as we got the histogram, the Rules of Auto-Adjust Threshold will be realized in the next four steps.

1. Find the globally maximal counter number.
2. Search the first locally minimal counter number that is located at right side of globally maximal counter number

3. If the first locally minimal counter number is greater than 100, find the next locally minimal counter number as step 2, otherwise the threshold value is not changed and go to step four
4. Check the index of globally maximal counter number. If the index of global maximum equals zero, the system will decrease the threshold value. The new threshold value is decided from the point where the SAD variation greater than 150. Otherwise the system does not change the threshold.

The first step assumes the SAD value of background information is lower than foreground information. The second step assumes that there is a border between moving object and background so we find the first local minimum to determine the threshold value. The third step confirmed that the threshold value would not be influenced by false histogram of SAD different value. The fourth step checks the index of globally maximal counter number to monitor the camera noise to apply on different video sequence. If the index of globally maximal counter number equals zero that means the camera noise is very small. If camera noise is small, we will decrease the threshold value to adapt different video sequence.

3.2 Features Extraction

Like the figure 3.2 (c), we got the binary value image i.e. the black pixel gray value is 0 and the white pixel gray value is 1. There are many kind of method in discussing how to extract the human silhouette features, [2] transfer the image into the FD (Fourier descriptor), [1, 5, 29] calculates the image counter called active contour...etc. In the FD method, it is often viewed the 2-D image as the complex plane. Assume the close plane image curve can be got, it take the Fourier transform of this curve, and take the dominate term to represent the image feature, It has the space-invariant characteristic, i.e. the rotation will just differ in phase. But in our goal to

determine a man which is lying is not properly used here because the lying pose is the rotation of the standing pose. The other method, Active Contour, is a kind of contour based method. It first to find out the contour mass center and next to calculate the Euclidean norm between each point and the mass center. Then the distance data will be regard as the feature of the image. Above methods are quite useful for some special situation, but it usually takes time to process this algorithm. So in order to deal with the real time problem, we should take a quick process to find the proper feature. In the next paragraph, the method we use will be discussed.

3.2.1 Definition of Postures

Actually the obvious human pose in this plan is divided into five kinds of postures such as stand, sit, lie, kneel and bend. This section will show this kind of pictures and explain some definition. First we show the stand pose as the figure 3.6.

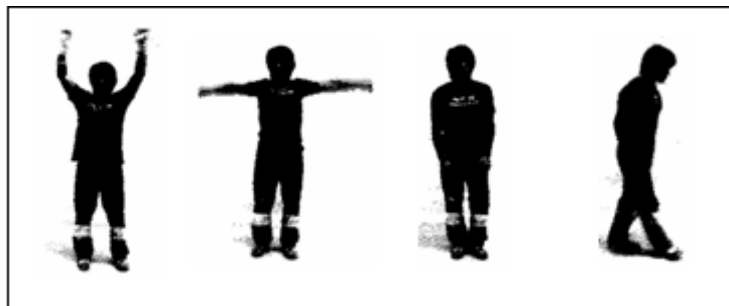


Figure 3.6 Stand Postures

In the standing poses the hands motions may be an effect of the parameter we define in the next section. So in this plan the most frequent poses people will act are included and discussed. For the purpose to get the real time postures recognition system, besides the standing postures we also consider the other four kinds of postures, such as sit, kneel, bend and lie. They are all shown

as below. In the kneeling and bending postures for the convenience of feature extraction, the side view images are used.



Figure 3.7 Sit Posture



Figure 3.8 Kneel Posture



Figure 3.9 Bend Posture

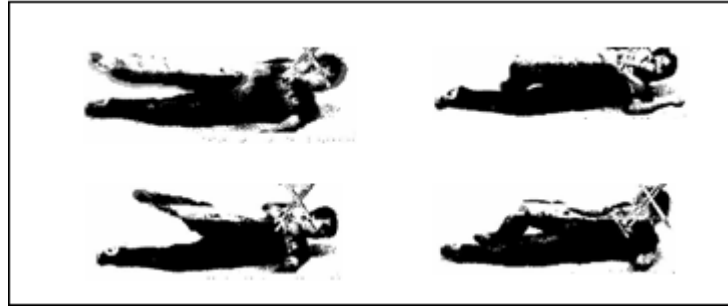


Figure 3.10 Lie Posture

As getting these postures definition, it is now important to find the different feature expression for these different postures. In the next section the feature extraction methods are discussed.

3.2.2 Definition of Feature Parameters

In fact, in this plan we take the bounding box of human contour, see as the figure 3.11.

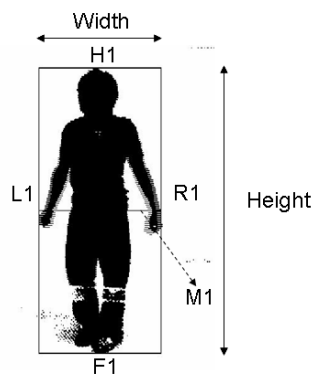


Figure 3.11 Bounding box of the human silhouette and its parameters

In the figure 3.11, L1 means the left boundary of the bounding box, R1 means the right boundary of the bounding box, H1 means the upper boundary of the boundary box and F1 means the lower boundary of the bounding box. M1 is calculated by $(F1+H1)/2$, it means the middle line of the bounding box.

Intuitively, the most difference between the human pose is the ratio of height and width, and

the bounding box will help us to obtain the value we need. Cause of the imperfect of the image segmentation the subtraction image we got usually has some noise. It is shown in figure 3.2 (c), the up-right corner has some noise disturbances. The noise will be viewed as the moving object if we take the boundary by finding the first black pixel we met form left to right, top to down. To solve this problem, we count the black pixel number column by column. As we known the moving object will be assigned the black pixels, and it will occupy the large image space. Again we use the histogram method to record this counter values, figure 3.12 shows the plot. In figure 3.12 the x axis of the lower plot means the x coordinate of the image i.e. it ranges from 1 to 640, and the y axis means the counter numbers of each coordinate. The max counter line here is point out the maximum black pixel counter numbers.

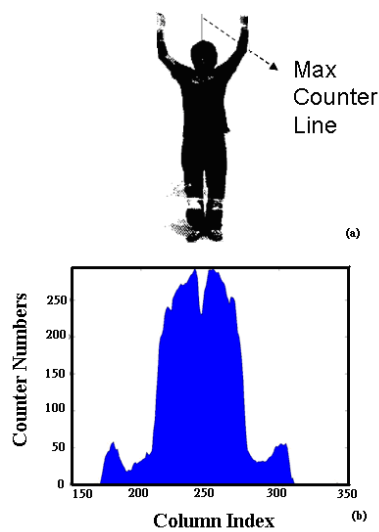


Figure 3.12 (a) Segmentation (b) Column Index vs. Counter Numbers

For finding the outer bounding box we check the black pixel counter numbers from the index of the maximum counter line to the boundary of the image. If the counter numbers of index (i) and (i+1) are not equal to 0, we set the index to be the bounding box boundary. For example in figure 3.12, the index of maximum counter line is 239. From column index 239 to column index 640 the first two zero count numbers are column index 311 and 312. Then the right bounding box boundary, i.e. the parameter R1 in the figure 3.9 is the column index 313. In the same way we can

find the outer bounding box H1, F1 and L1.

In fact, the hand is rising or not will make the bounding box more complicated. It means that when the hand is spread like the figure 3.13, the width of the outer bounding box will become larger than the original width, i.e. the ratio also becomes smaller than the original stand pose ratio. To accurately determine the real human body boundary in this case we should modify the bounding box boundary.

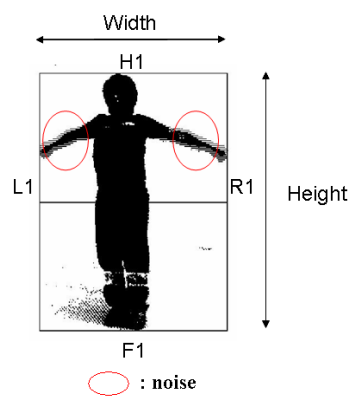


Figure 3.13 Human Hand Spread Plot

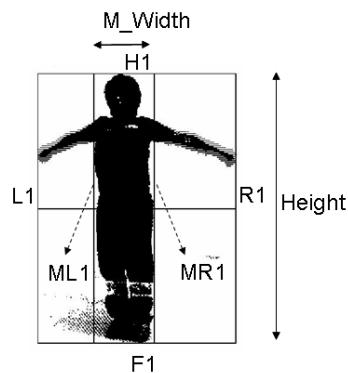


Figure 3.14 Modified bounding box of human silhouette

In further discussion, we modify the figure into figure 3.14. The ML1 and MR1 mean the left and right boundary after modifying. The M_Width means the distance between MR1 and ML1. The way we find the modified parameters is also using the same histogram way discussed

before.

As mentioned before, we take the figure 3.15 for the example. We take some mark for the figure 3.12 (b) and it becomes figure 3.15. The symbol B is the main body part of human, symbol A and symbol C are the hand part. It can be obviously found from the figure 3.12 (a) that the black pixel numbers of hand parts are much smaller than the body part, so the boundary of part A and part B is chosen by this concept.

First, after finding the maximum counter numbers of part B, we move to the right part to get the counter numbers of each column. As the process going we pick the column index as the boundary of part B and part C if the counter numbers of the right part divide by the maximum counter numbers is larger than the threshold value. The boundary of part A and part B is also chosen as the same way. It can be seen as the figure, at column index 280 the counter numbers suddenly drop to a small value so we take the 280 to boundary of the modified bounding box.

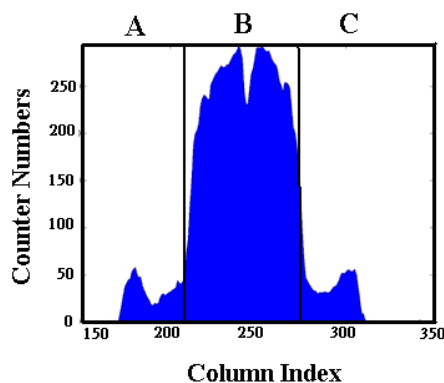


Figure 3.15 Histogram of the human to modify bounding box boundary

Once again the hand effect is not only in spreading out case, but the raising effect. It means that if the hand is rising over the head the ratio of width-length will be larger than the original case. By solving this kind of problem we take a quick process to define the head line it is shown below. The symbol of N1 in the plot indicates the line which is modified as the neck line. The definition of the line index is calculated by dividing the upper bounding box into three part and we take the 1/3 length for its coordinate. M_Height here represents the length form N1 to F1.

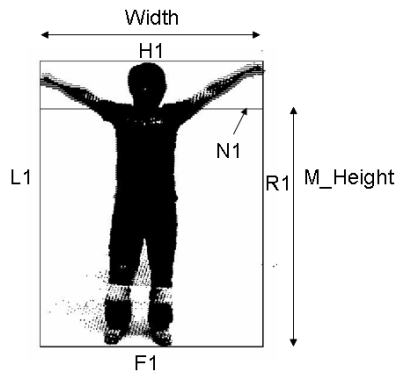


Figure 3.16 Definition of neck line plot

In the next paragraph the two postures, kneel and bend will be discussed. Cause of the specialty of these two postures, that is the length-width ratio of these two postures are easily confused with the sit postures, it is necessary for finding another proper parameters.

The main concepts of our feature extraction are based on the shape method. That is the formation of the postures shape is a kind of index to point out the different postures. Take the kneel postures for example, that is show a below in figure 3.17. In the figure the symbol KN_1 means the parameter to define the kneel-pose line, and the BN_1 means the parameter to define the bend-pose line. More concisely, the other two parameters are defined to calculate the features. The line index of KN_1 is found by finding the maximum black pixel numbers from the half line of lower bounding box to the bottom line of the lower bounding box. It can be drawn as the fig. 3.17 shown. The meaning of BN_1 line is also very clear seen as the fig. 3.18. In the bend-pose, the number of black pixels in the line BN_1 are obviously large than others. For the quick process procedure the line of BN_1 are chosen by take the half of the upper bounding box. For recoding the two parameters we use the symbol of Pixel(BN_1) and Pixel(KN_1) to stand for the pixel numbers of the KN_1 line and BN_1 line.

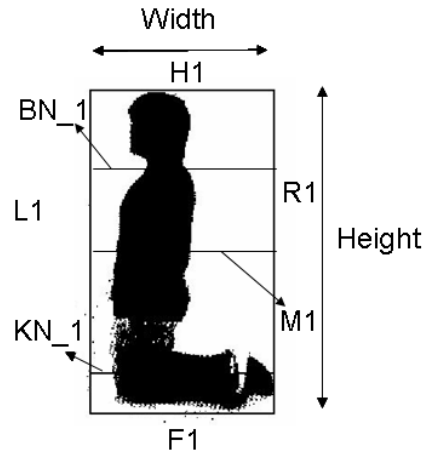


Figure 3.17 Parameters of kneel-pose

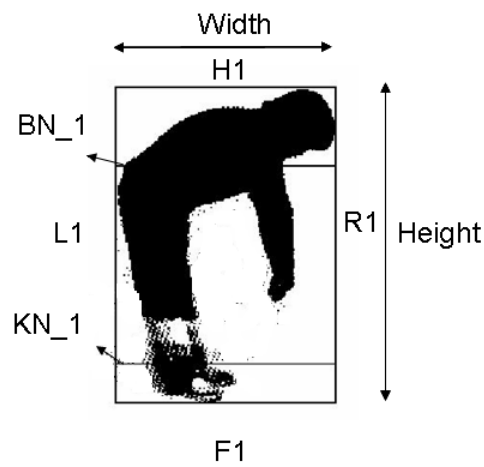


Figure 3.18 Parameters of Bend-pose

1. $Original_length-width_ratio = (Height / Width) \equiv x_o$
2. $Modify_length-width_ratio(1) = (Height / M_Width) \equiv x_{m1}$
3. $Modify_length-width_ratio(2) = (M_Height / M_Width) \equiv x_{m2}$
4. $Kneel_ratio = (Pixel(KN_1) / Pixel(BN_1)) = x_k$
5. $Bend_ratio = (Pixel(BN_1) / Pixel(KN_1)) \equiv x_b$

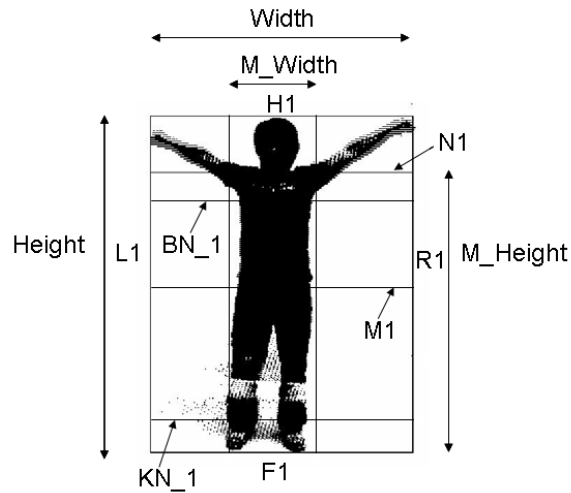


Figure 3.19 All of the Feature Parameters Plot.

By summing all above up, we can derive five main feature parameters for our classifier system. It can be summarized as the five equations shown below. Figure 3.19 shows the parameters plot for the system. All the five parameters can be derived from the plot, and the plot in other words is the summary of all the plots shown before. So in our processing, every image taken by the CCD camera will be through the image subtraction and feature extraction. Then we can get the feature extraction vector which has the five components, like $\mathbf{x} = (x_o, x_{m1}, x_{m2}, x_k, x_b)$ represents.

4 ADAPTIVE FUZZY RULE-BASED CLASSIFICATION SYSTEMS

This paper proposes an adaptive method to construct a fuzzy rule-based classification system with high performance for pattern classification problems. The proposed method consists of two procedures: an error correction-based learning procedure and an additional learning procedure. The error correction-based learning procedure adjusts the grade of certainty of each fuzzy rule by its classification performance. That is, when a pattern is misclassified by a particular fuzzy rule, the grade of certainty of that rule is decreased. On the contrary, when a pattern is correctly classified, the grade of certainty is increased. Because the error correction-based learning procedure is not meaningful after all the given patterns are correctly classified, we cannot adjust a classification boundary in such a case. To acquire a more intuitively acceptable boundary, we propose an additional learning procedure. We also propose a method for selecting significant fuzzy rules by pruning unnecessary fuzzy rules, which consists of the error correction-based learning procedure and the concept of forgetting. We can construct a compact fuzzy rule-based classification system with high performance. Finally, we test the performance of the proposed two methods on the pose recognition system.

4.1 Fuzzy Rule-Based Classification Systems

In this section, a method for auto generating fuzzy if-then rules systems will be introduced. First step is generating the fuzzy if-then rules by using the data partition in the fuzzy subsets. Next we define the fuzzy classification procedure, and at last we introduce an adaptive learning method to correct the error classified postures. As the three main processes are done the system is

constructed.

4.1.1 Fuzzy Rule Generation Procedure

Let us consider the training patterns are normalized between the axis [0,1]. For example, if the training vectors have the dimension of $X_1 \times X_2$ then the dimension is $[0,1] \times [0,1]$. Now suppose that we have n training vectors pair $\mathbf{x}_p = (x_{p1}, x_{p2}, \dots, x_{pm})$ $p = 1, 2, \dots, n$ which is divided into M classes : class 1(C1), class 2(C2)... class M(CM). In our research, the training vectors have five components like length-width ratio, modified length-width ratio... , and the classified categories are stand, sit, kneel, bend and lie. The classification problem here is to generate fuzzy rules that divide the pattern space into M disjoint decision areas. For this problem, we employ fuzzy rules of the following type :

$$\begin{aligned}
 &R_{ij \dots u}^K : \\
 &\text{if } x_{p1} \text{ is in } A_i^K \text{ and } x_{p2} \text{ is in } A_j^K \dots \text{and } x_{pm} \text{ is in } A_u^K \\
 &\text{then classify } \mathbf{x}_p \text{ as class } Cm \text{ with } CF = CF_{ij \dots u}^K
 \end{aligned}
 \tag{4.1}$$

$$i = 1 \dots K, j = 1 \dots K, \dots, u = 1 \dots K, m = 1 \dots M$$

Where the $R_{ij \dots u}^K$ the rule number, $A_i^K, A_j^K, \dots, A_u^K$ are fuzzy sets on the unit interval [0,1], Cm means the consequent class (in this plan is five). The K means the numbers of the fuzzy subsets in the axis of universe [0,1]. In the plan we set the K equals to 5 like figure 4.4.

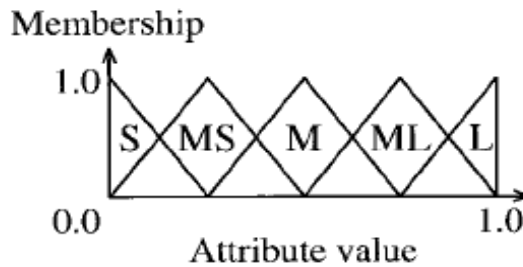


Figure 4.1 Plot of membership function of K=5 subsets.

In fuzzy systems, the membership functions have various kind of type, like triangular, bell and trapezoid. The membership function we choose here is the triangular membership function. It can be described as the following equation :

$$\mu_i^K(x) = \max \left\{ \left(1 - \frac{x - a_i^K}{b^K} \right), 0 \right\}, \quad i = 1 \dots K$$

$$\text{Where } a_i^K = \frac{i-1}{K-1}, \quad i = 1 \dots K \quad b^K = \frac{1}{K-1} \quad (4.2)$$

For more detail discussions, a don't care term is added for the consideration, the don't-care term membership function is plot as figure 4.5. In the fig. 4.5, it shows the membership function shape of the don't-care term for a two dimensional training vector. The generating rules method are calculated by $X_1 \times X_2 \times \dots \times X_m$ so the numbers of rules generated by the method are $(K+1)^m$, where m is the numbers of the training vector, in the plan m is equal to 5. It means that we have 6^5 rules before eliminating the dummy or unnecessary rules. After the procedure the CF values are calculated by the method below.

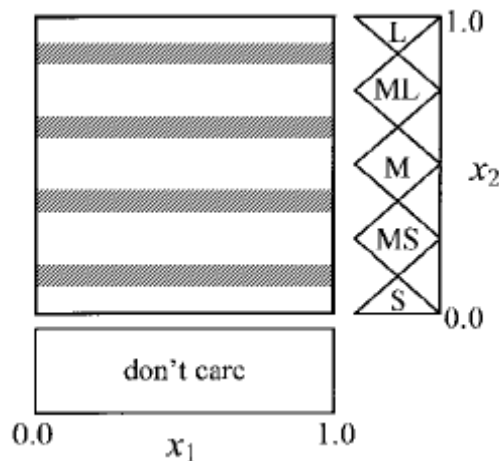


Figure 4.2 Two dimension membership function with don't-care term

1. Calculate the β_{CT} for $T = 1 \dots M$ as

$$\beta_{CT} = \sum_{p \in CT} \mu_i^K(x_{p1}) \square \mu_j^K(x_{p2}) \cdots \mu_u^K(x_{pm}) \quad (4.3)$$

2. Find Class X (CX) by

$$\beta_{CX} = \max(\beta_{C1}, \beta_{C2}, \dots, \beta_{CM}) \text{ In the plan which is represented as}$$

$$\beta_{CX} = \max(\beta_{C_sit}, \beta_{C_stand}, \beta_{C_kneel}, \beta_{C_bend}, \beta_{C_lie}) \quad (4.4)$$

If multiple classes take the maximum value in (4.4), the consequent CX of the fuzzy rule corresponding to the fuzzy subspace cannot be determined uniquely. Thus, let CX be dummy class and the procedure is terminated. Otherwise, CT is determined as CX in (4.4).

3. $CF_{ij \dots u}^K$ is calculated by the following equations

$$CF_{ij \dots u}^K = \frac{\beta_{CX} - \bar{\beta}}{\sum_{T=1}^M \beta_{CT}}, \text{ where } \bar{\beta} = \sum_{CT \neq CX} \frac{\beta_{CT}}{M-1} \quad (4.5)$$

Here if we take an example for a three class i.e. we just classified the pose into pose-sit, pose-stand and pose-lie. The CF value will be calculated as below.

$$CF_{ij \dots u}^K - sit = \frac{\beta_{C_sit} - \bar{\beta}}{(\beta_{C_sit} + \beta_{C_stand} + \beta_{C_lie})}, \text{ where } \bar{\beta} = \frac{(\beta_{C_stand} + \beta_{C_lie})}{2} \quad (4.6)$$

Take more consideration, the meaning of $\beta_{CT} = \sum_{p \in CT} \mu_i^K(x_{p1}) \square \mu_j^K(x_{p2}) \cdots \mu_u^K(x_{pm})$ is to see the partitions of each pose in the fuzzy subsets. Then compare and find the maximum value to decide the rules which is guiding the decision of this pose. It is described as the equation (4.4). CF value are the index of fuzzy if-then rules, it represents how certain the rules in classifying the poses. After the parameters are getting by us the next step is to the main classified process i.e. the fuzzy reasoning parts.

4.1.2 Fuzzy reasoning

When the antecedent fuzzy sets of each fuzzy if-then rule are given, we can determine the consequent class and the grade of certainty by the heuristic rule generation procedure in the previous section. Here we assume that we have already generated a set of fuzzy if-then rules for a pattern classification problem. In our fuzzy classifier system, we perform the fuzzy reasoning via the single winner rule. The winner rule $R_{ij\dots u}^K$ for the input pattern \mathbf{x}_p is determined as :

$$\mu'(\mathbf{x}_p) \square CF' = \max(\mu(\mathbf{x}_p) \square CF_{ij\dots u}^K \mid R_{ij\dots u} \in S) \quad (4.7)$$

where $\mu(\mathbf{x}_p) = \mu_i^K(x_{p1}) \square \mu_j^K(x_{p2}) \cdots \mu_u^K(x_{pm})$, S is the generating rule sets. That is, the winner rule has the maximum product of the compatibility $\mu(\mathbf{x}_p)$ and the grade of certainty $CF_{ij\dots u}^K$. In other words, assume that we got a training pattern, $\mathbf{x}_p = (x_{p1}, x_{p2}, x_{p3})$, once the membership function is needed to be considered, we divide the five linguistic variables are small, medium small, medium, medium large and large, so the fuzzy if-then reasoning rules is described as below.

$$R_{ij\dots u}^K : \\ \text{If } x_{p1} \text{ is small and if } x_{p2} \text{ is large and if } x_{p3} \text{ is medium small} \\ \text{then Class } C_j \text{ with } CF = CF_{ij\dots u}^K$$

As shown in the previous, if we have 100 if-then rules for the classifier problem, we will calculate 100 times the CF values and find the maximum CF value to see which class the CF represents. The mathematical representations are shown below.

1. Calculate α_{CT} for class $T = 1 \cdots M$ as

$$\alpha_{CT} = \max \left\{ \mu_i^K(x_{p1}) * \mu_j^K(x_{p2}) * \mu_k^K(x_{p3}) * CF_{ijk}^K \mid R_{ijk} \in S \right\} \quad (4.8)$$

2. Find Class X (CX) maximizing α_{CT} by

$$\alpha_{cX} = \max \{ \alpha_{c1}, \alpha_{c1}, \dots, \alpha_{cT} \mid T = 1 \dots M \} \quad (4.9)$$

By combining these equations the classified problem can be easily done. When multiple classes take the maximum value of α_{cT} , the classification of the unknown pattern will be rejected in this procedure.

4.1.3 Learning Algorithm

So far the classified system is almost finished except the miss classified part. The fuzzy reasoning part although can almost solve the classified problem but the blurred data. It means that it can be misclassified in the follow steps if we don not add the learning algorithm. To adjust the grades of certainty of the fuzzy rules, we use the following error correction-based learning procedure.

1. When \mathbf{x}_p is correctly classified by $R_{ij \dots u}^K$.

$$CF_{ij \dots u}^K \text{ - NEW} = CF_{ij \dots u}^K \text{ - OLD} + \eta 1 (1 - CF_{ij \dots u}^K \text{ - OLD}) \quad (4.10)$$

2. When \mathbf{x}_p is misclassified by $R_{ij \dots u}^K$.

$$CF_{ij \dots u}^K \text{ - NEW} = CF_{ij \dots u}^K \text{ - OLD} - \eta 2 * CF_{ij \dots u}^K \text{ - OLD} \quad (4.11)$$

Where $\eta 1$ and $\eta 2$ are learning constants. Generally, because the number of correctly classified patterns is much larger than that of misclassified patterns, the grade of certainty of each fuzzy rule tends to be increased to its upper limit (i.e., $CF_{ij \dots u}^K = 1$) by (4.10) if we choose the proper $\eta 1$. In other way, the misclassified training patterns will be narrow down by choosing the proper $\eta 2$. By this assumption, we take the value for these tow constants $0 < \eta 1 \leq \eta 2 < 1$. A

learning method with Procedure C for generating a fuzzy rule-based classification system can be written as the following algorithm.

1. Initialization :
 - a. Set the stop iteration numbers J_{\max} and desirable classification rate.
 - b. Let K be the number of the fuzzy subsets in each axis of the training vectors.
 - c. Generate the fuzzy if-then rules by using procedure A.
2. Classification :
 - a. Classify all the training patterns with the generated fuzzy rules and calculate the classification rate.
 - b. Stop the algorithm if $J = J_{\max}$ or the classification rate is reached.
3. Tuning the grades of certainty (CF values) :
 - (a.) Stop algorithm when the classification rate is reached.
 - (b.) Set $J = J + 1$
 - (c.) Perform the learning procedure in C for each pattern.
 - (d.) Go back to step 2.

4.2 Apply the AFRBCS in Posture Recognition

In our plan the, each subtracted image has the five feature parameters as in the section mentioned above. That is the training pattern vector for each image is expressed as $\mathbf{x} = (x_o, x_{m1}, x_{m2}, x_k, x_b)$, each element in the vector is introduced before. As mentioned before we have the five postures needed to be classified, and it will show the feature parameter distribution for them. Figure 4.6 shows the five postures feature vector distribution plot. In our training patterns the vectors have five dimensions it exceeds the three dimensions expression, so

in the plot it will show just only two dimensions to explain the concepts. In figure 4.6 x axis means the first element of the training vectors that is the original length-width ratio, and y axis means the second element of the training vectors that is the modified length-width ratio. In this figure the circle points represents stand postures, star points represents sit postures, diamond points represents lie postures, triangular points represents kneel postures and the pentagram means the bend postures. It is easily found that the boundary of these data is not well-defined.

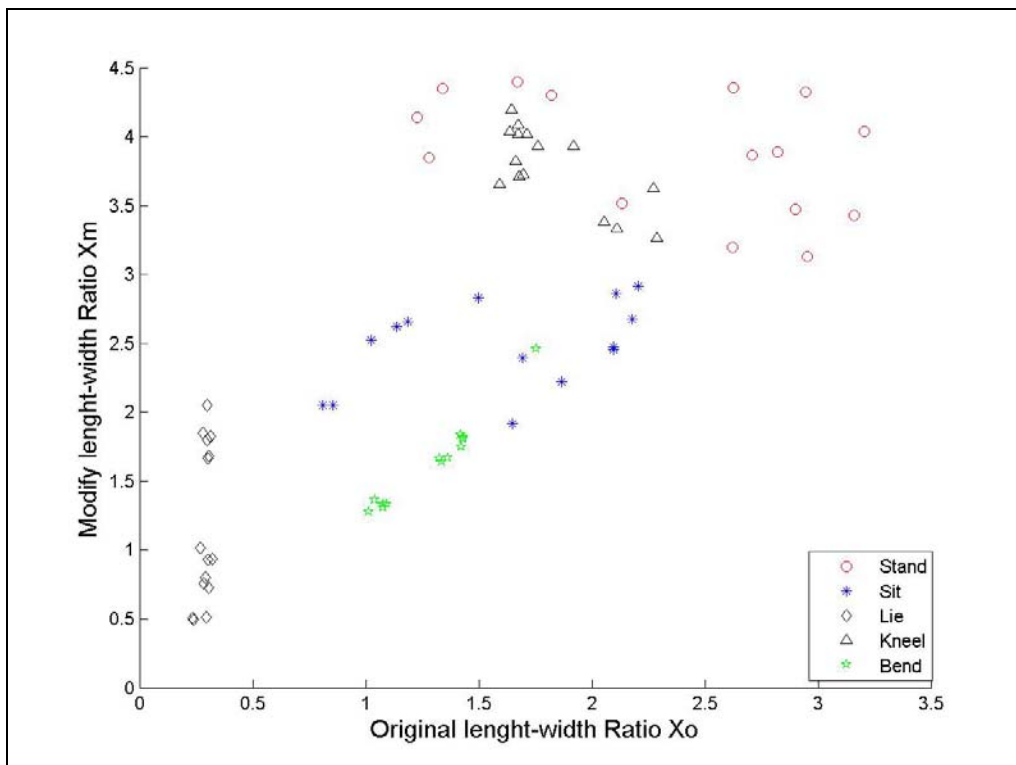


Figure 4.3 Training vectors of five postures distribution in the two dimensions (Xo,Xm) plot.

For the kneel postures, parts of them have almost the same partitions with stand postures. It means that the tow dimensions can not character all of these five postures, so that's why the kneel ratio and bend ratio are needed. Figure 4.7 will show another two different dimensions plot for differing kneel and stand postures. It can be found that kneel ratio can separate kneel posture from the other postures.

So dose the bend posture once again we plot it as the figure 4.8. In the figure we use a

different x axis as the modified length-width ratio for explaining more conveniently, and the y axis means the bend ratio.

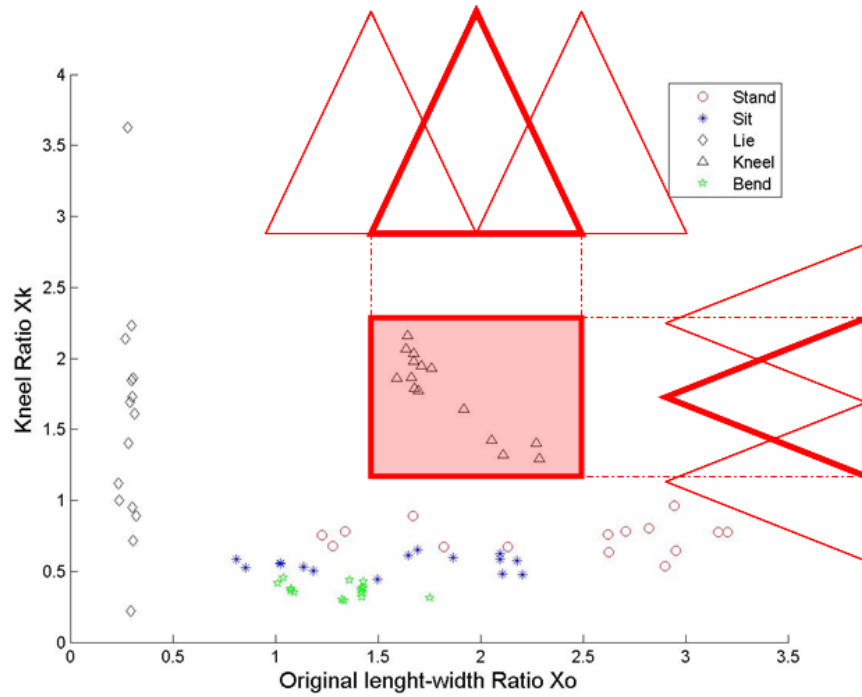


Figure 4.4 Training vectors of five postures distribution in the two dimensions (X_o, X_k) plot.

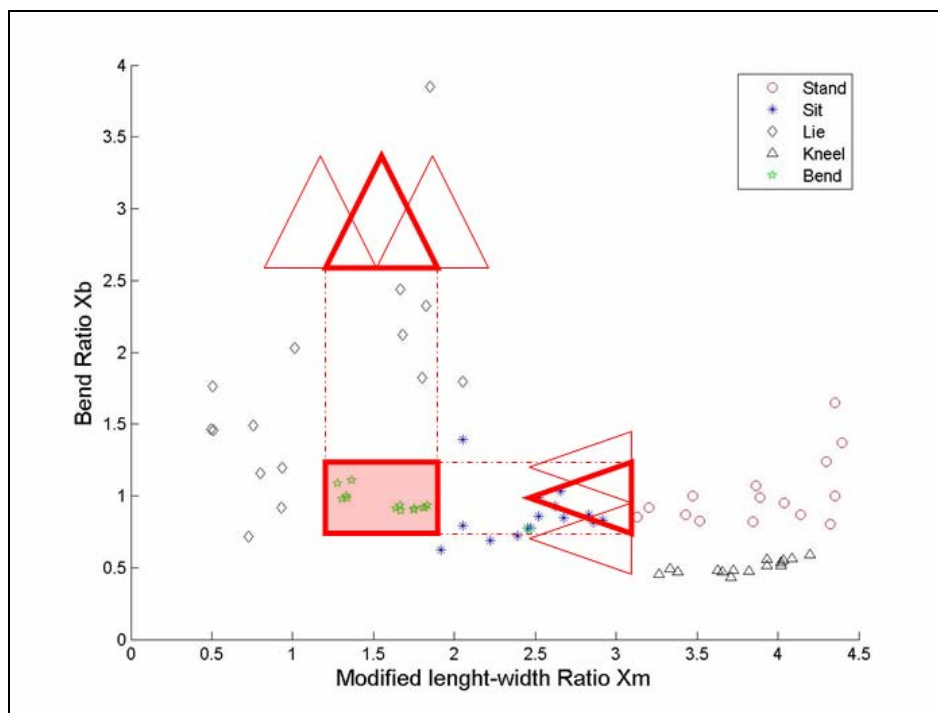


Figure 4.5 Training Vectors of Five Postures Distribution in The 2-D (X_m, X_b) Plot.

For generating rules, the data partitions in the fuzzy subsets are calculate. Take the standing postures for example, we will show how to generate the fuzzy rules for recognition the standing postures. At first we choose 15 training vectors for each pose, and each vector has the 5 elements as shown before. In the fuzzy membership functions we defined are distribution in the x axis [0,1], so we have to normalize the data into the interval [0,1]. The normalized equation is stated as below.

$$\mathbf{x}_{normalize} = (\mathbf{x} - \mathbf{x}_{min}) / (\mathbf{x}_{max} - \mathbf{x}_{min}) \quad (4.12)$$

First of all as we get the training vectors the generating rules procedure is to calculate the β_{CT} by using equation (4.3). For convenience to show the β_{CT} values, the membership functions are assigned some index for them. That is the membership function of “small” we assign it as number 1, the membership function of “medium small” we assign it as number 2, and so on. The “don’t care” membership function is assigned as number 6. As mentioned before there are $6^5 = 7776$ rules before vanishing unnecessary dummy rules.

In table 4.1 the rule 1 means the fuzzy if-then rule which represents as

Rule 1:

If x_1 is small and if x_2 is small and if x_3 is small

and if x_4 is small and if x_5 is small then Class C_j with $CF = CF_{ij...u}^K$

For the same inference the other rules are the same form as shown. As we define the index for these membership functions the next is to calculate β_{stand} so that we can use the β_{stand} value to design the rules for understanding the stand postures.

And the β_{stand} can be calculated from the membership function rule by rule, that is we will fit the standing training patterns into these membership functions and see how they distribute. In the final results we have 675 if-then rules for standing postures, and it is too trivial to explain how to generate each rule by the β_{stand} values. So we just take some of these rules for example.

	x_1	x_2	x_3	x_4	x_5
Rule 1	1	1	1	1	1
Rule 2	1	1	1	1	2
Rule 3	1	1	1	1	3
Rule 4	1	1	1	1	4
Rule 5	1	1	1	1	5
Rule 6	1	1	1	1	6
Rule 7	1	1	1	2	1
⋮	⋮	⋮	⋮	⋮	⋮
Rule 7775	6	6	6	6	5
Rule 7776	6	6	6	6	6

Table 4.1 Generating Rules Index Form.

In table 4.2 it shows the 15 standing postures training patterns, each element of the patterns are normalized into interval (0,1). As we put the 15 input standing training vectors in the table 4.2 into each rules in the table 4.1, the first nonzero β_{stand} term is the Rule 2053 that is the index of membership functions 24411. The nonzero β_{stand} value is 0.0037, and it is a very small value to compare with other β_{class} values. It can be considered as that the Rule 2053 (index 24411) is not so fitted for recognizing standing postures.

	x_1	x_2	x_3	x_4	x_5
Stand 1	0.912843	0.981392	0.981469	0.36904	0.184669
Stand 2	1	0.907653	0.9077	0.277973	0.257307
Stand 3	0.803819	0.69307	0.692588	0.269139	0.242014
Stand 4	0.871147	0.870021	0.868772	0.29144	0.275882
Stand 5	0.832914	0.863768	0.865014	0.27932	0.318522
Stand 6	0.89761	0.762438	0.764133	0.159426	0.281974
Stand 7	0.80421	0.988638	0.989521	0.207169	0.281974
Stand 8	0.534074	0.975375	0.975586	0.226751	0.402334

Stand 9	0.352448	0.858483	0.859735	0.228675	0.192778
Stand 10	0.334406	0.934312	0.935649	0.267355	0.217936
Stand 11	0.372614	0.988231	0.988933	0.280958	0.605972
Stand 12	0.483742	1	1	0.333943	0.466719
Stand 13	0.639042	0.774204	0.773661	0.224579	0.19415
Stand 14	0.915131	0.674471	0.673494	0.213016	0.208518
Stand 15	0.984026	0.75206	0.750975	0.276192	0.217523

Table 4.2 Training Patterns for Stand Postures

In the table 4.3 it shows parts of the nonzero β_{stand} value terms. Cause of too large number of rules generated, table 4.3 just shows the first 15 rules. In the table the bolded sign means the values of β_{stand} are larger than the other β_{class} values. Although part of the β_{stand} values are seems very small like Rule 2062, Rule 2065 Rule 2067 and Rule 2068 and so on. It means the other β_{class} values may be zero, so the zero β_{class} value-rules may not act in recognizing standing postures.

So in such a system, how to decide which rule is the postures precognition rule is decided by the β_{class} values. Once we take the table 4.3 for example, as we mentioned before the bolded sign means the rules' β_{class} are larger than other postures. Like Rule 2060 the β_{stand} value is equal to 0.168557, and its β_{sit} value is equal to 0.0551, and the rest β_{class} values are equal to 0, so the β_{stand} value is the biggest of all. Then the Rule 2060 will be defined as the rule to recognize the standing postures, that is the rule can be written as

Rule 2060 :

If x_1 is MS and if x_2 is ML and if x_3 is ML

and if x_4 is MS and if x_5 is MS

then the posture is Class C_{stand} with $CF_{stand} = CF_{2060}$

Here the MS and ML are represented as the membership functions for the linguistic variable of medium-small and medium-large. In the above rule the CF value is calculated by the equation (4.5), and it will be introduced in the next paragraph.

	x_1	x_2	x_3	x_4	x_5	β_{stand}
Rule 2053	2	4	4	1	1	0.00366
Rule 2054	2	4	4	1	2	0.01233
Rule 2058	2	4	4	1	6	0.015989
Rule 2059	2	4	4	2	1	0.044592
Rule 2060	2	4	4	2	2	0.168557
Rule 2061	2	4	4	2	3	0.000536
Rule 2062	2	4	4	2	4	0.000394
Rule 2064	2	4	4	2	6	0.214079
Rule 2065	2	4	4	3	1	0.000399
Rule 2066	2	4	4	3	2	0.002711
Rule 2067	2	4	4	3	3	7.58E-05
Rule 2068	2	4	4	3	4	5.57E-05
Rule 2070	2	4	4	3	6	0.003241
Rule 2083	2	4	4	6	1	0.048651
Rule 2084	2	4	4	6	2	0.183598

Table 4.3 Rules for Recognizing Standing Postures (the first 15 terms).

	β_{stand}	β_{sit}	β_{lie}	β_{kneel}	β_{bend}
Rule 2060	0.168557	0.0551	0	0	0

Table 4.4 Each β_{class} Value of Rule 2060.

For calculating the CF_{2060} value for the Rule 2060, we list each β_{class} value to explain how to get the CF_{2060} value. According to the equation (4.5) we put each β_{class} value into the equation and we can write as the following form

$$CF_{2060} = \frac{\beta_{stand} - (\beta_{sit} + \beta_{lie} + \beta_{kneel} + \beta_{bend})/4}{(\beta_{stand} + \beta_{sit} + \beta_{lie} + \beta_{kneel} + \beta_{bend})} = \frac{0.169 - (0.055/4)}{(0.169 + 0.055)} = 0.919.$$

For this example it can be more easily to explain the meaning of CF values. That is the meaning of the value is to see how certain the rule is in recognizing a posture. From the above equation we know that if the value of $\beta_{sit} + \beta_{lie} + \beta_{kneel} + \beta_{bend}$ is bigger than the β_{stand} value, the CF value will become much smaller that is the rule is not efficient enough. For the same method the other β_{class} values can be determined. As the same procedure all the rules' CF values can be done, and based on these CF values we can do the fuzzy reasoning parts as the section 4.2.2.

For reasoning part the value $\mu'(\mathbf{x}_p) \square CF'$ of each rule is needed to be calculated that we can use the values and find the $\max(\mu(\mathbf{x}_p) \square CF_{ij \dots u}^K | R_{ij \dots u} \in S)$ value to decide the training vector is which posture class. To explain the reasoning procedure we list some tables for it, the table 4.5 shows the training vectors versus the postures. It is said that the first column shows the postures and each row shows the postures' parameters. In the table 4.6 it shows the values of $\mu(\mathbf{x}_p) \square CF_{ij \dots u}^K | R_{ij \dots u} \in S$ for the five postures, and the bolded signs mean the maximum values that can also represent as $\max(\mu(\mathbf{x}_p) \square CF_{ij \dots u}^K | R_{ij \dots u} \in S)$.

	x_1	x_2	x_3	x_4	x_5
Standing	0.9128	0.9814	0.9815	0.369	0.1847
Sitting	0.4761	0.3645	0.3634	0.1957	0.0934
Lying	0.0223	0.3	0.2966	0.7507	1
Kneeling	0.4741	0.9485	0.9459	0.9652	0.0768
Bending	0.262	0.199	0.1986	0.0975	0.3266

Table 4.5 Five postures and its corresponding feature parameters.

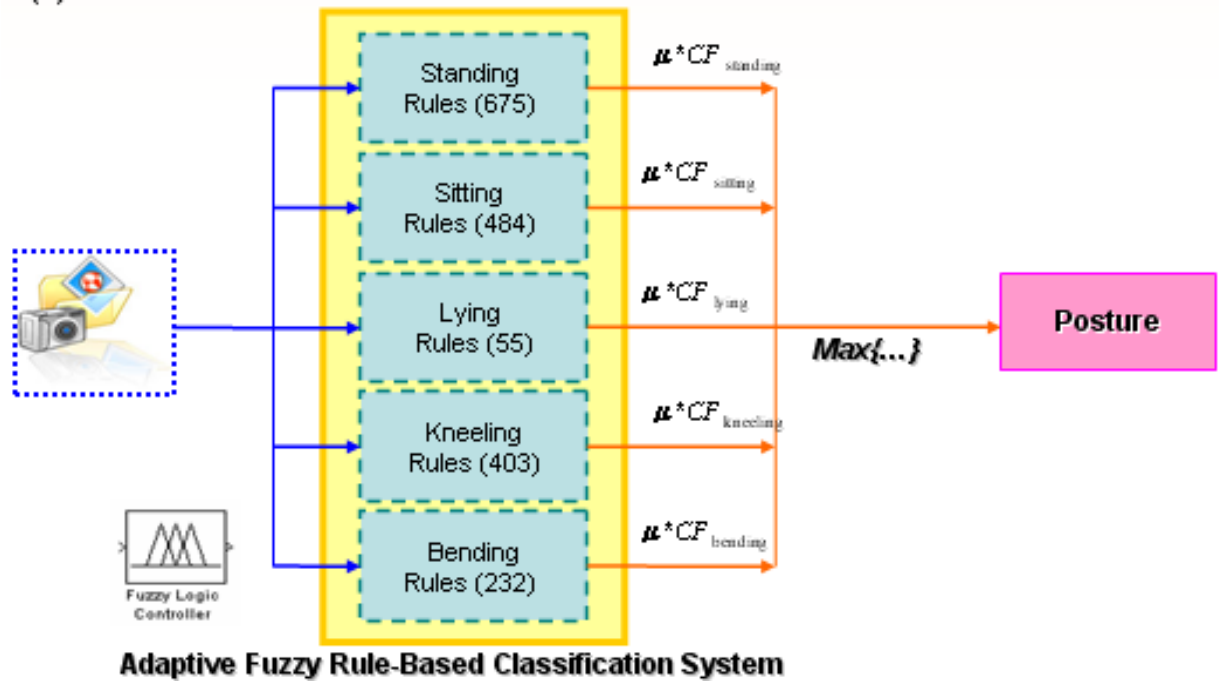


Figure 4.6 AFRCS Architecture

	Standing	Sitting	Lying	Kneeling	Bending
$\max(\mu(\mathbf{x}_p) \square CF_{stand})$	0.6514	0.0506	0	0.3349	0.056
$\max(\mu(\mathbf{x}_p) \square CF_{sit})$	0.0787	0.3379	0.1528	0.0041	0.0971
$\max(\mu(\mathbf{x}_p) \square CF_{lie})$	0.1186	0	0.1851	0.0059	0.1826
$\max(\mu(\mathbf{x}_p) \square CF_{kneel})$	0.2135	0.1388	0.0133	0.7717	0.0041
$\max(\mu(\mathbf{x}_p) \square CF_{bend})$	0.1131	0.3197	0.0593	0.0739	0.5929

Table 4.6 Training Vectors vs. $\max(\mu(\mathbf{x}_p) * CF_{class})$ Values.

In our system, each posture we have 15 training patterns that is we have 15 standing training patterns, 15 sitting training patterns, 15 lying training patterns, 15 kneeling training patterns and 15 bending training patterns. In table 4.5 the first column is one of each posture training pattern. Before we talk about this, in the generating rules procedure it generates 675 rules for standing

postures, 484 rules for sitting postures, 55 rules for lying postures, 403 rules for kneeling postures and 232 rules for bending postures. Figure 4.8 to figure 4.12 shows parts of the training postures pictures.



Figure 4.7 Standing Training Patterns Images.



Figure 4.8 Sitting Training Patterns Images





Figure 4.9 Lying Training Patterns Images



Figure 4.10 Kneeling Training Patterns Images



Figure 4.11 Bending Training Patterns Images

In order to obtain the table 4.6, the $\mu(\mathbf{x}_p) \square CF_{class}$ values are needed to be calculated rule by rule first. Once we take the standing training pattern in table 4.5 for example, we put the vector into all the fuzzy if-then rules and calculate the $\mu(\mathbf{x}_p) \square CF_{class}$ values then we list

the $\max(\mu(\mathbf{x}_p) \square CF_{class})$ values in the table 4.6. More precisely speaking, the value 0.6514 is the maximum value of these standing rules for the standing vector, which is the result of the equation (4.8). In other words here the values we calculate in the table 4.6 are defined in the equation (4.8) that is $\alpha_{stand} = 0.6514$, $\alpha_{sit} = 0.0787$, $\alpha_{lie} = 0.1186$, $\alpha_{kneel} = 0.2135$ and $\alpha_{bend} = 0.1131$. By using the equation (4.9), the largest α_{CT} value is $\alpha_{stand} = 0.6514$ so the output pose is standing. In the same reason the other four postures can be judged.

Although most of the postures can be classified by using the first generated rules, there still exist few error classifying conditions. Like our classified system, in the first iteration (that is without the learning algorithm) there are 2 errors in classifying standing postures, 2 errors in classifying sitting postures, 4 errors in classifying lying postures, 1 error in classifying kneeling postures and 2 errors in classifying bending postures. So adding the learning algorithm is necessary.

By using the learning algorithm shown in the above section 4.2.3, we set the learning constants $\eta_1 = 0.001$ and $\eta_2 = 0.1$. As the learning procedure it iterates about 20 times to get the 100% classified rate. In the next chapter we will introduce the whole system architecture and the testing demo pictures also be introduced.

5 EXPERIMENT AND RESULTS

The real time monitor system in postures recognition in our research is constructed. To deal with the moving object subtraction we use the auto-threshold background segmentation method, and a human silhouette will be got. Then the feature extraction will be determined in the next step, and apply the AFRBCS system in recognition the postures parts. The whole system will be realized in the previous method we mentioned. In the next section we will show the experiment results and shows the classification rate.

5.1 Hardware Architecture and Environment

In the system process diagram as shown in the figure 1.X, in our system we use the PC platform to set up this experiment environment. The equipment are listed in the below table.

1. PC (Intel Pentium 4 with CPU 3.2G Hz 1.96G RAM)
2. CCD camera (EverFocus EQ500/N)
3. Image restoration card (Matrox Meter II)
4. Simulation Software (MATLAB® 7.1 R14)

Table 5.1 List of experiment equipment.

The system can take 30 FPS like a usual real time monitor system. In general situation it is not necessary to take so many frames to monitor an environment, so we take 3 frames per second to deal with the general monitoring system.

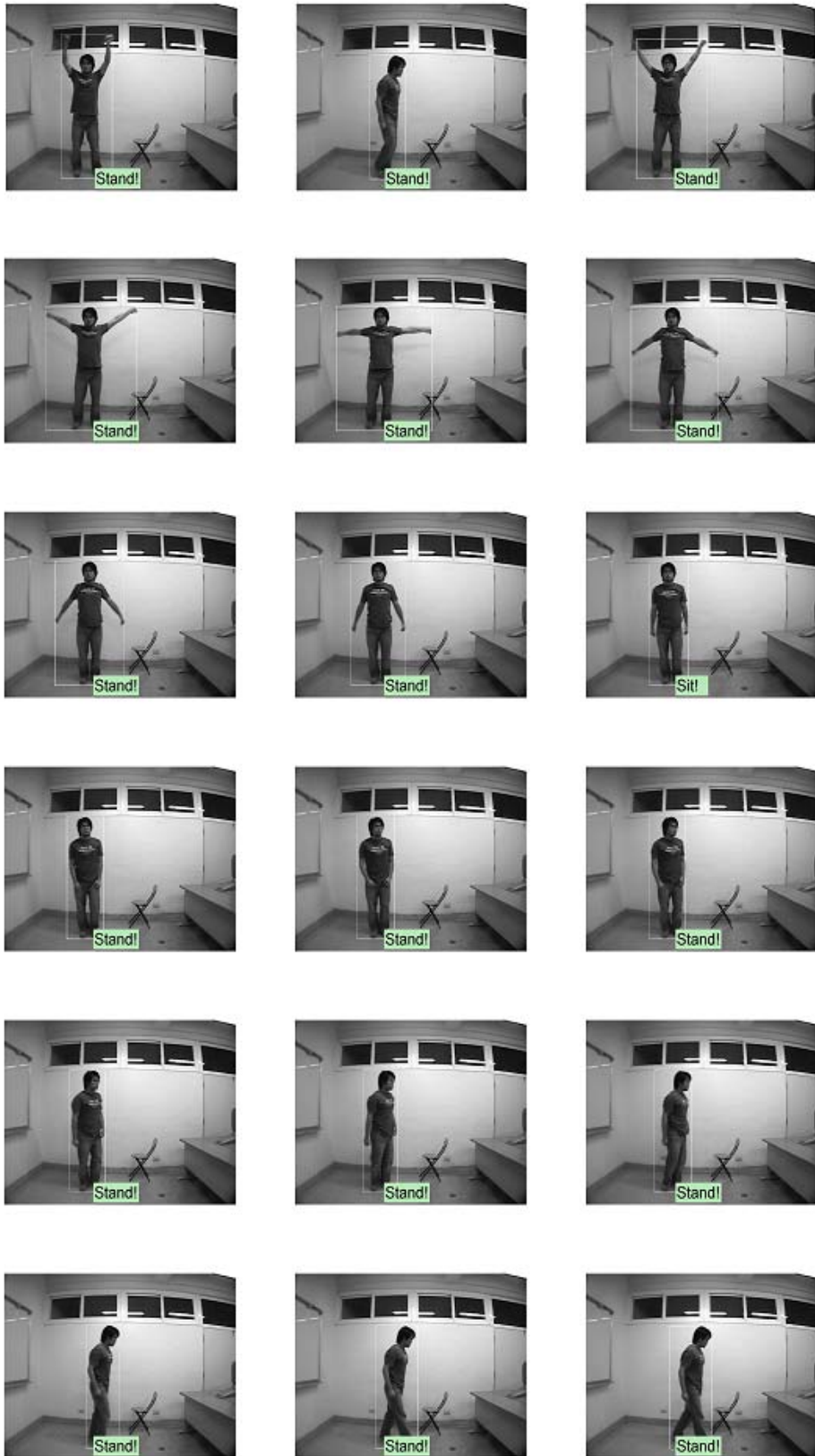
As a reason for the assumption of the home care system, so the experiment is taken in the indoor environment. For the background subtraction method often we will take a more monotonic scene and with only one moving object. In the next section it will show the experiment results.

5.2 Experimental Section

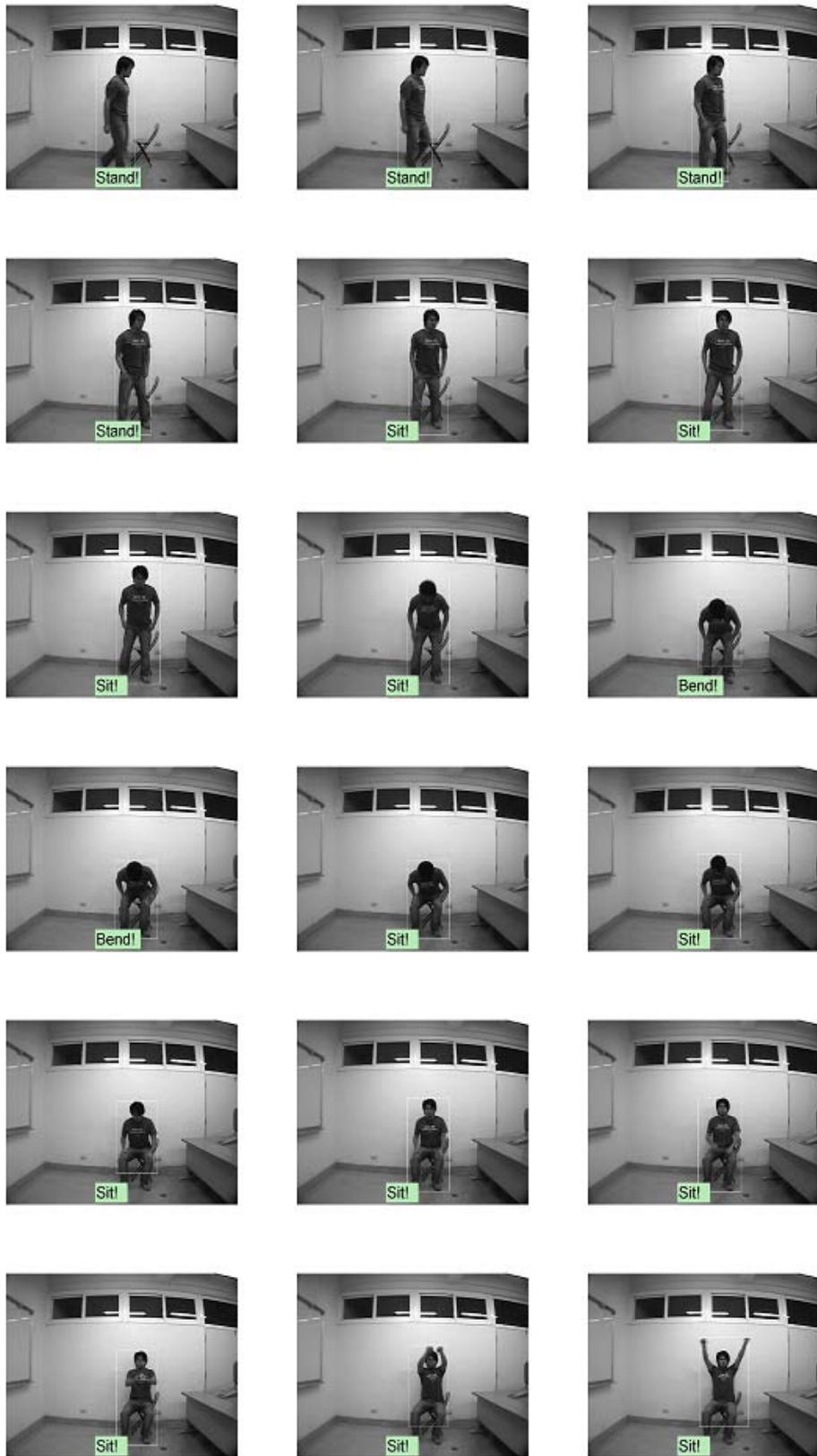
For two kind of situation we take two different demo video to explain the result, one is the different postures recognition the other is ill-postures recognition. That is in the first demo we will show the image and output in recognizing which posture the image is. The second demo part will show the ill-posture, when a people is lying down in the floor.

A. General Posture Recognition Demo

In this part the demo will show the output of each image frame. In the output images, the white line index out the outer bounding box and the bottom label shows the postures recognition results. The figure 5.1 shows the demo frames one by one, the first 25 frames are the standing postures and the next 30 frames are sitting postures and the kneeling postures and bending postures. In the figure we just only show parts of the demo images, we cut the main postures parts to show the results.



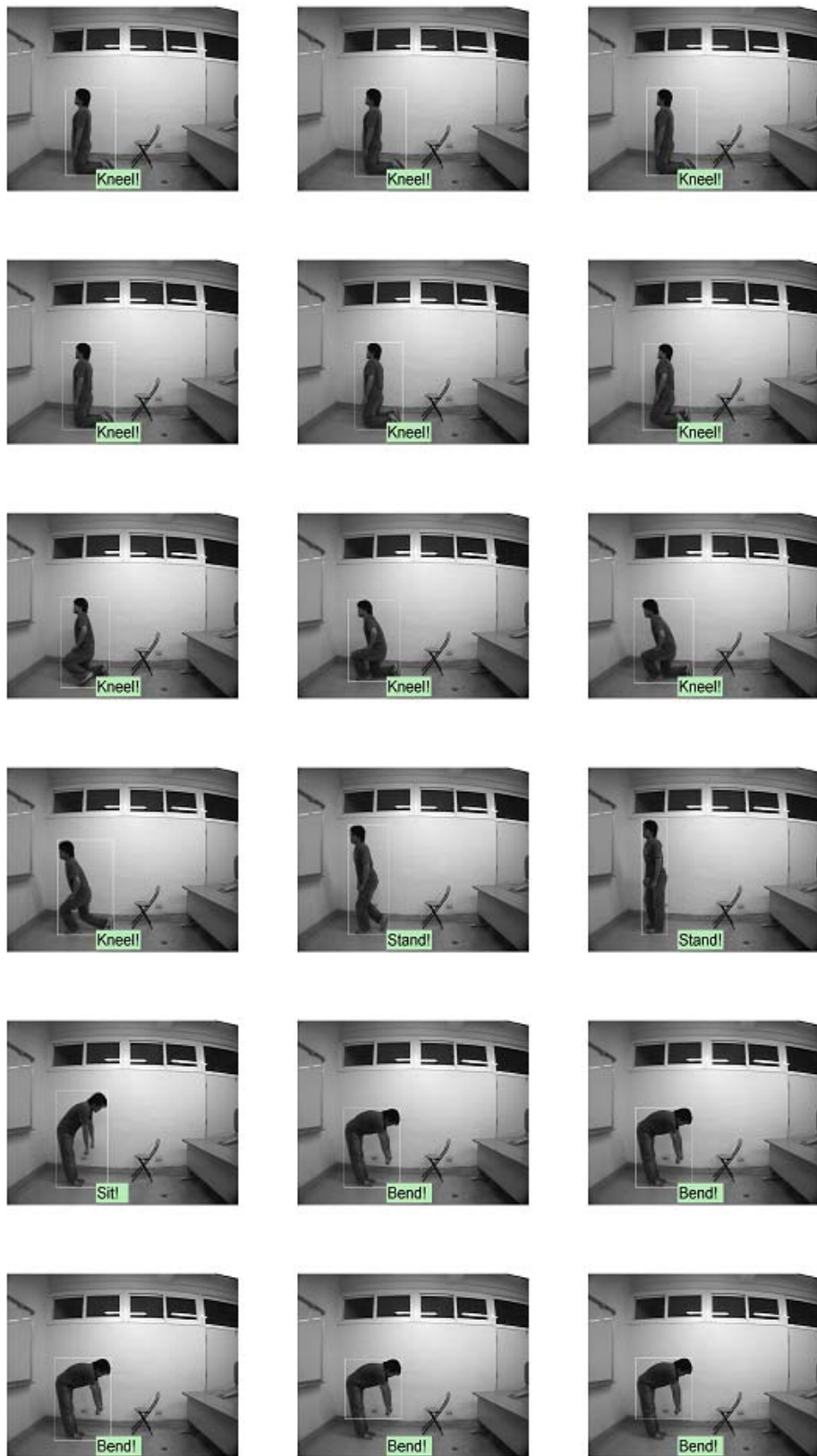
(a) Frames of Standing Postures



(b) Frames of Sitting and Standing Postures



(c) Frames of Sitting and Kneeling Postures



(d) Frames of Kneeling and Bending Postures.

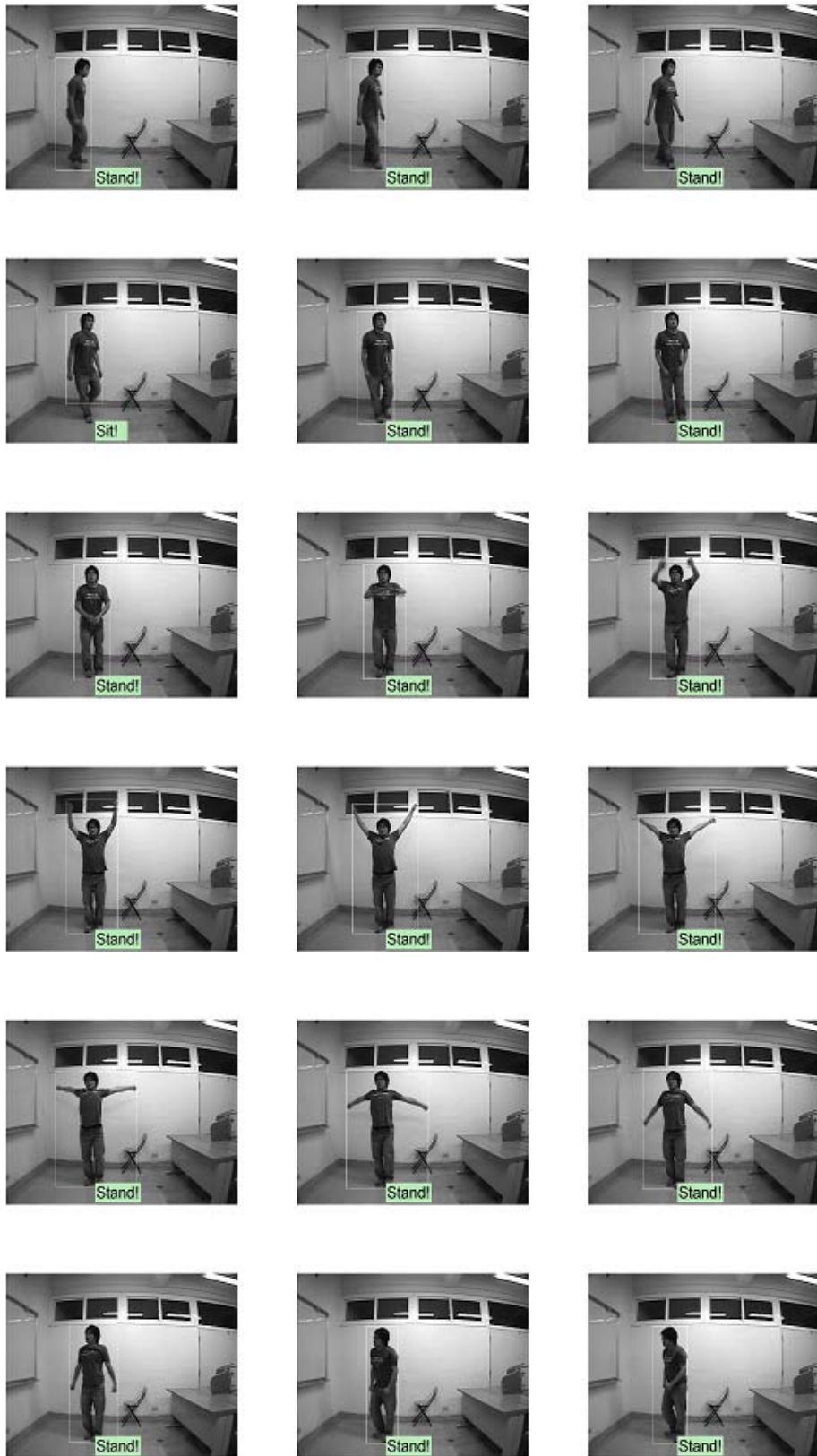


(e) Frames of Bending Postures.

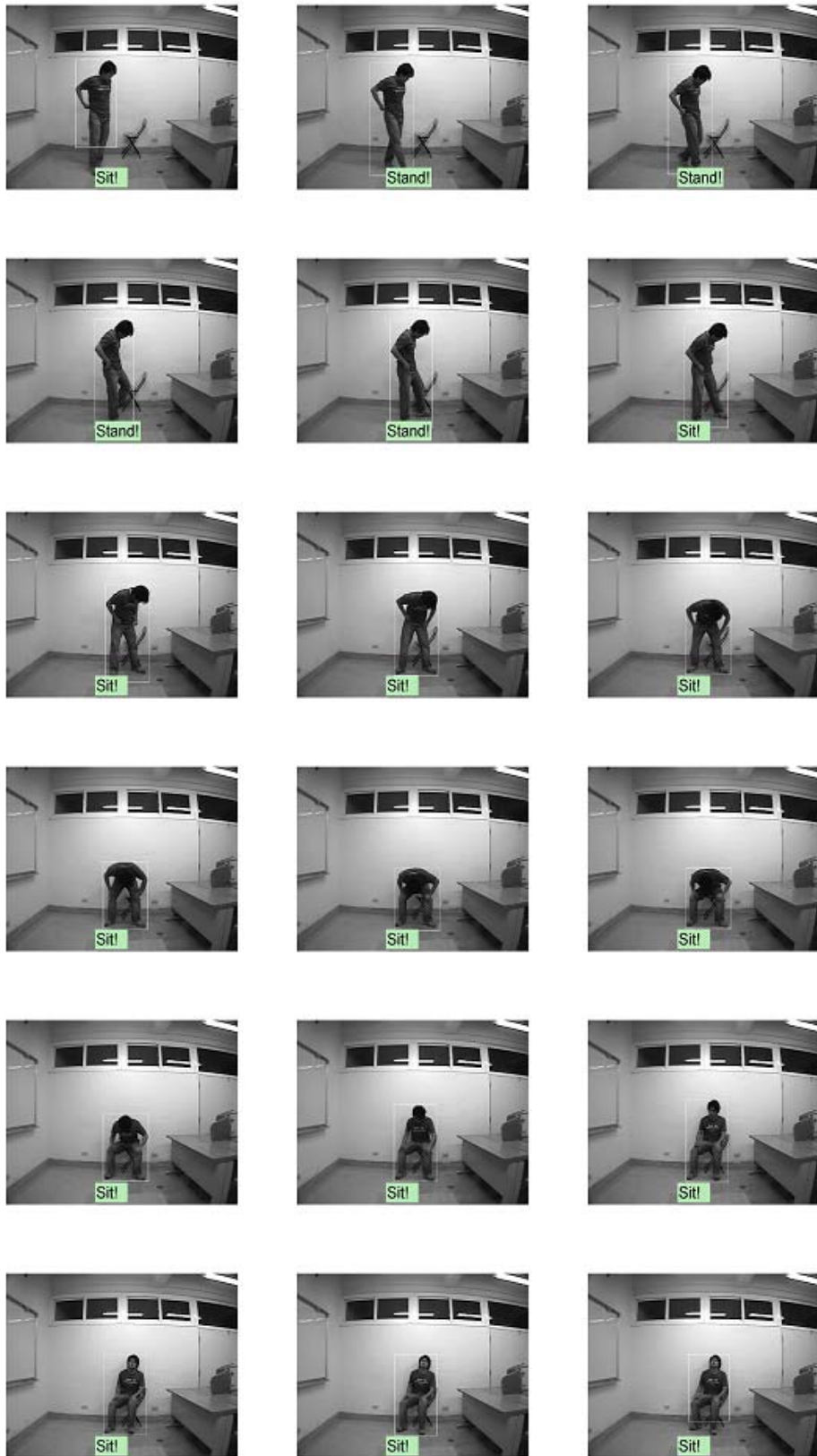
Figure 5.1 General Postures Recognition Demo Images

B. Ill-postures Recognition Demo

In the paragraph we assume a situation that if a people walk inside the room and then sit on a chair finally when the people get up from chair and suddenly fall down on the floor. In our system the main goal is to find out the people fall down and alarm the watcher to see what happen in the room. Figure 5.2 will show the demo frame by frame and the output image expressions are the same meaning as the figure 5.1. In the figure 5.2 (c) and (d) the people are falling down and our system can alarm immediately to the output window.



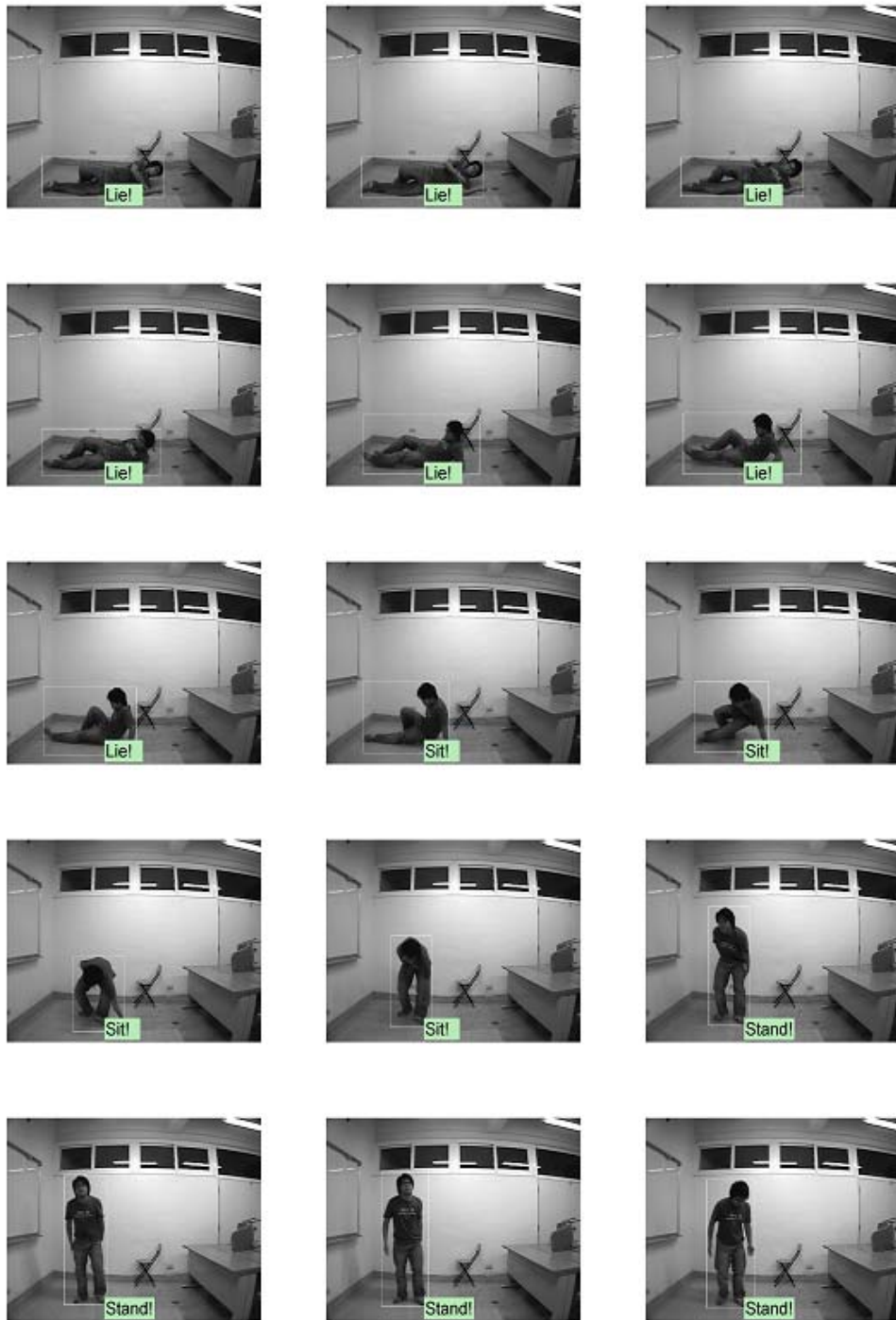
(a) Frames of standing postures.



(b) Frames of standing and sitting postures.



(c) Frames of sitting and lying postures.



(d) Frames of lying postures.

Figure 5.2 Ill-postures recognition demo images.

For above two demo images we show a real time monitoring home care system, no matter in the various postures recognition or in the ill-postures recognition it works with little errors. In

table 5.2 it shows the hit rate for our classified system. In the figure 5.1 and figure 5.2 we just show the part frames from the original demo frames, so in the table 5.2 we will show the total frames of these two demo images and the hit rate. From table 5.2 the two demos, when the postures are transforming the classified rate will decrease that is the transition of two postures act a poor appearance.

Postures	Stand	Sit	Kneel	Bend	Total
Demo(1)	47/50=94%	30/33=91%	27/27=100%	15/15=100%	119/125=95.2%
Postures	Stand	Sit	Lie	Total	
Demo(2)	50/56=90%	34/34=100%	10/10=100%	94/100=94%	

Table 5.2 Hit rate for the demo images.

In this section we show the total system architecture and the demo results for our real time human postures recognition. Although cause of the processing time it just take three frames per second, the classified rate is also can hold on the 90 percentage. In the general situation that people action is not so variable in one second the system will work well.

6 DISCUSSIONS AND CONCLUSIONS

In the plan, we present a real time postures recognition monitoring system. It has the potential to recognize the general postures and gives a good classified hit rate. In the research, we first use some simple concept to grab the human bounding box to get the feature parameters, like some ratio of the body length and width. The feature extraction may not so difficult to get but it may happen some ambiguous situation in data distribution. So we then present a method based on the fuzzy if-then rules systems to modify how the class distribute.

In our method, the data with close characteristic may still have some boundary to separate them. That is with our method the auto generating rules can apply to this kind of classified problem and solve it well. The adaptive fuzzy rule based method can automatically adjust the grades of certainty of fuzzy rules. The method also has the learning ability to improve the error situation that is the ability to correct the errors. Finally we establish a real time system for recognizing the postures three frames per second.

For our future work, we may build an embedded system to make the whole system in a chip. As the chip is created, everyone can set it on any camera and show a real time postures recognition results. It will be a benefit to the development for the home care systems.

Reference List

- [1]. F. Buccolieri, C. Distance and A. Leone, “Human Posture Recognition Using Active Contours and Radial Basis Function Neural Network”, IEEE conference on Advanced Video and Signal Based Surveillance, 2005. AVSS 2005, 15-16 Sept. 2005 Page(s):213–218 Digital Object Identifier 10.1109/AVSS.
- [2]. Ronald Poppe and Mannes Poel, “Comparison of Silhouette Shape Descriptors for Example-based Human Pose Recovery”, Proceedings of the 7th International Conference on Automatic Face and Gesture Recognition (FGR’06), 2006.
- [3]. J.K. Aggarwal and Q. Cai, ”Human motion analysis: a review”, *Nonrigid and Articulated Motion Workshop*, 1997. Proceedings, IEEE, Page(s):90 – 102,16 June 1997.
- [4]. G. Johansson, “Visual motion perception”, *Sci. Am.* 232(6), 1975, 76-88.
- [5]. Hironobu Fujiyoshi, Alan J. Lipton and Takeo Kanade, “Real-Time Human Motion Analysis by Image Skeletonization”, *IEICE TRANS. INF. & SYST.*, Vol. E87-D, No.1 January 2004.
- [6]. R. F. Rashid, “Toward a system for the interpretation of moving light display”, *IEEE Trans. PAMI* 2(6), 574-581, November 1980.
- [7]. J. A. Webb and J. K. Aggarwal, “Visually interpreting the motion of objects in space”, *IEEE Comput.* August 1981, 40-46.
- [8]. J. A. Webb and J. K. Aggarwal, “Structure from motion of rigid and jointed objects”, *in Artif. Intell.* 19, 1982, 107-130.
- [9]. S Kurakake and R. Nevatia, “Description and tracking of moving articulated objects”, *in 11th Intl. Conf. on Pattern Recognition*, Hague, Netherlands, 1992, Vol. 1, pp. 491-495.
- [10]. D. Gavrilu and L. Davis, “3D model-based tracking of human upper body movement: a multi-view approach”, *In Proceedings of Int’l Symposium on Computer Vision*, pages: 253-258, 1995.

- [11]. M. Isard and A. Blake, "Condensation-conditional density propagation for visual tracking", *Int. Journal of Computer Vision*, 29(1): 5-28, 1998.
- [12]. H. J. Lee and Z. Chen, "Knowledge-guided visual perception of 3-D human gait from a single image sequence", *Trans. Systems, Man, Cybernetics* 22(2), 336-342, 1992.
- [13]. M. K. Leung and Y. H. Yang, "First sight: A human body outline labeling system", *Trans Pattern Anal. Mach. Intelligence* 17(4), 359-377, 1995.
- [14]. D. Ayers and M. Shah, "Recognizing human action in a static room", *In Proceedings Computer Vision and Pattern Recognition*, pages 42-46, 1998.
- [15]. James Davis and Aaron Bobick, "The representation and recognition of action using temporal plates", *In Proceedings Computer Vision and Pattern Recognition*, pages 928-934, 1997.
- [16]. Stephen S. Intille, J. Davis, and A. Bobick, "Real time closed world tracking", *In Proceedings Computer Vision and Pattern Recognition*, pages 697-703, 1997.
- [17]. Ken Nozaki, Hisao Ishibuchi and Hideo Tanaka, "Adaptive Fuzzy Rule-Based Classified Systems", *IEEE transaction on Fuzzy Systems*, Vol. 4, No. 3, August 1996.
- [18]. Hisao Ishibuchi and Tomoharu Nakashima, "Improving the Performance of Fuzzy Classifier Systems for Pattern Classification Problems with Continuous Attributes", *IEEE transactions on Industrial Electronics*, Vol. 46, No. 6, December 1999.
- [19]. Hisao Ishibuchi, Tomoharu Nakashima and Tadahiko Murata, "Performance Evaluation of Fuzzy Classifier Systems for Multidimensional Pattern Classification Problems", *IEEE Transactions on Systems, Man and Cybernetics*, Vol. 29, No. 5, October 1999.
- [20]. Hisao Ishibuchi and Tomoharu Nakashima, "Effect of Rule Weights in Fuzzy Rule-Based Classification Systems", *IEEE Transaction on Fuzzy Systems*, Vol. 9, No. 4, August 2001.
- [21]. Hisao Ishibuchi, Tomoharu Nakashima and Gaku Nakai, "Credit Assignment by Fuzzy Rule-based Systems in Fuzzy Classifier Ensembles", *Proceedings 2003 IEEE International Symposium on Computational Intelligence in Robotics and Automation*, July 16-20, 2003.

- [22]. R. Jain and H. H. Nagel, "On the analysis of accumulative difference pictures from image sequences of real scenes", *IEEE Trans. on PAMI*, 1(2): 206-214, 1979.
- [23]. D. Hogg, "Model-based vision: A program to see a walking person", *Image and Vision Computing 1(1)*, 5-20, 1983.
- [24]. K. Rohr, "Towards model-based recognition of human movements in image sequences", *CGVIP: Image Understanding*, 59(1): 94-115, 1994.
- [25]. D. Marr and H. K. Nishihara, "Representation and recognition of the spatial organization of three dimensional shapes", *In Proc. R. Soc. London, volume B*, pages 269-394, 1978.
- [26]. K. Rohr, "Human Movement Analysis Based on Explicit Motion Models", chap 8, pp. 171-198, Kluwer Academic, Dordrecht/Boston, 1997.
- [27]. J. O'Rourke and N. I. Badler, "Model-based image analysis of human motion using constraint propagation", *IEEE Trans. PAMI*, 2: 522-536, 1980.
- [28]. K. Sato, T. Maeda, H. Kato, and S. Inokuchi, "CAD-based object tracking with distributed monocular camera for security monitoring", *Proc. 2nd CAD-Based Vision Workshop*, Champion, PA, February 1994, pp. 291-297.
- [29]. Robert T. Collins, Alan J. Lipton and takeo Kanade, "A System for Video Surveillance and Monitoring".
- [30]. C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland, "Pfinder: real-time tracking of the human body", *Trans. Pattern Anal. Mach. Intelligence 19(7)*, 780-785, 1997.
- [31]. T. B. Moselund and E. Granum, "Multiple cues used in model-based human motion capture", *in The Fourth International Conference on Automatic Face and Gesture Recognition*, Grenoble, France, March 2000.
- [32]. T. Darrell, P. Maes, B. Blumberg, and A. P. Pentland, "A novel environment for situated vision and behavior", *in Workshop for Visual Behaviors at CVPR-94*, 1994.
- [33]. A. Nakazawa, H. Kato, and S. Inokuchi, "Human tracking using distributed video systems", *in International Conference on Pattern Recognition*, 1998.

- [34]. M. Brand, "Shadow puppetry", in *International Conference on Computer Vision*, Corfu, Greece, September 1999.
- [35]. R. Rosales and S. Sclaroff, "Learning and synthesizing human body motion and posture", in *Fourth International Conference on Automatic Face and Gesture Recognition*, Grenoble, France, March 2000.
- [36]. J. Farmer, M. Casdagli, S. Eubank, and J. Gibson, "State-space reconstruction in the presence of noise", *Physics D*, 51D: 52-98, 1991.
- [37]. Nuria Oliver Barbara Rosario and Alex Pentland, "A Bayesian computer vision system for modeling human interactions", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, August, 2000.
- [38]. Jie Yang, Yangsheng Xu, and Chiou S. Chen, "Human action learning via hidden markov model", *IEEE Transactions on Systems, Man and Cybernetics*, A: 34-44, 1997.
- [39]. A. F. Bobick and A. D. Wilson, "State-based technique for the summarization and recognition of gesture", *In Proc. Of 5th Intl. Conf. on Computer Vision*, pages 382-388, 1995.
- [40]. J. Yamato, J. Ohya, and K. Ishii, "Recognizing human action in time-sequential images using Hidden Markov Model", *In Proc. IEEE Conf. CVPR*, pages 379-385, Champaign, IL, June 1992.
- [41]. James Davis and Aaron Bobick, "The representation and recognition of action using temporal plates", *In Proceedings Computer Vision and Pattern Recognition*, pages 928-934, 1997.
- [42]. A. F. Bobick and J Davis, "Real-time recognition of activity using temporal templates", *In Proc. of IEEE Computer Society Workshop Applications on Computer Vision*, pages 39-42, Sarasota, FL, 1996.
- [43]. R. Polana and R Nelson, "Low level recognition of human motion (or how to get your man without finding his body parts)", *In Proc. of IEEE Computer Society Workshop on Motion of Non-Rigid and Articulated Objects*, pages 77-82, Austin, TX, 1994.

- [44]. J. K. Aggarwal and Sangho Park, "Human Motion: Modeling and Recognition of Actions and Interactions", *3DPVT04* (640-647). IEEE Abstract. IEEE Top Reference. 0412 BibRef, 2004.
- [45]. A. Wilson and A. Bobick, "Parameter Hidden Markov Models for Gesture Recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(9), 1999.
- [46]. 張意政, "Non-rigid Motion Analysis for Human Body", 國立清華大學電機工程研究所博士論文, 1999.
- [47]. C. Becchetti and L. P. Ricotti, "Speech Recognition Theory and C++ Implementation", New York: Wiley, 1999.
- [48]. L. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", *Proceeding of the IEEE*, vol. 77, No. 2, February 1989.
- [49]. M. Shah and R. Jain, "Motion-Based Recognition", chapter 9, pages 201-226. Kluwer Academic Publishers, 1997.
- [50]. B. Min, H. Yoon, J. Soh, Y. Yang, and T. Ejima, "Visual recognition of static/dynamic gesture: Gesture-driven editing system", *Journal of Visual Languages and Computing*, 10: 291-309, 1999.
- [51]. J. Yang, Y. Xu, and C. S. Chen, "Human action learning via hidden markov model", *IEEE Transactions on Systems, Man and Cybernetics*, pages 34-44, 1997.
- [52]. M. Herman, "Understanding body postures of human stick figures", PhD plan, University of Maryland, 1979.
- [53]. K. Akita, "Image sequence analysis of real world human motion", *Pattern Recognition*, 17(1), 1984.
- [54]. 李乾丞, "Fast Human Posture Recognition by Heuristic Rules.", 國立台灣大學電機工程研究所碩士論文, 2006.

計畫成果自評

日期：97年1月31日

自評成果項目	自評項目說明
<p>研究內容與原計畫相符程度</p>	<p>原研究計畫主要目標為建構一套「智慧型影像居家看護系統」，當家中的個人發生危險時，可以經由攝影機畫面的取得以及智慧型人體動作辨識軟體的操作，發出警訊至醫院或是通知正在工作的其他家人，避免延遲導致憾事發生。計畫規劃分成影像處理、人體動作與行為辨識、及硬體實現三個主題分年進行研究。而計畫執行方向皆依據初期研究內容，雖在演算法及系統架構的研究上有些微適當調整，就整體來說與原計畫相符且一致的。</p>
<p>達成預期目標情況</p>	<p>就計畫分為三個方面說明。首先影像處理方面，主要達成目標為研究已擷取到的影像資訊建立人體模型，藉由此人體模型上的各種參數來代表受測者肢體動作等參數。其次行為辨識方面，將這些參數交由智慧型行為辨識系統來做為辨識受測者的姿勢、行為的依據，此辨識系統以自適應性模糊演算法來實踐分類器，克服因外界干擾而造成的錯誤判斷。最後在硬體實現上，主要目的在於建立影像處理的硬體平台，為影像辨別方面建立起完整的系統架構，為提升效能將原採用嵌入式架構改為雙核心PC-based系統，以分散式平行運算取代之，實踐即時人體動作辨識與影像擷取演算法實現在所架構的平台之上。</p>
<p>研究成果之學術或應用價值</p>	<p>在本研究中，我們利用特定特徵變數的定義及採用邊界分割處理人形姿態的特徵取樣問題，再配合上自適性模糊控制的演算法成功的將其適用於一般性的姿態辨識，並解決一些較不易分辨姿態的問題，以及在因背景或外界干擾所產生辨識雜訊的情況下更具強健性。藉由系統的學習功能，讓我們可將其提高效能，將影像處理達到即時監控系統的要求。在本計畫中我們成功了實踐一個即時的辨識監控系統。</p>
<p>是否適合在學術期刊發表或申請專利</p>	<p>本計畫研究成果適合發表於訊號及影像處理以及智慧演算法應用領域學術期刊以及用於專利申請上皆極具價值。</p>
<p>主要發現或其他有關價值</p>	<p>可輕易、快速將技術移植於居家監控系統、居家安全及看護系統、金融保全系統、一般消費性電子產品等等領域應用上。</p>
<p>綜合評估</p>	<p>成功的將影像處理技術改良以及智慧型演算法應用，合併不同領域的技術予以整合，並將想法於實際系統實踐。</p>

可供推廣之研發成果資料表

可申請專利

可技術移轉

日期：97年1月31日

國科會補助計畫	計畫名稱：智慧型居家看護影像監控系統 計畫主持人：陳永耀 教授 計畫編號：NSC-95-2221-E-002-234 學門領域：控制學門
技術/創作名稱	即時姿態辨識監控系統 The Real-time Postures Recognition Monitoring System
發明人/創作人	陳永耀教授
技術說明	中文：在本研究中，我們利用特定特徵變數的定義及採用邊界分割處理人形姿態的特徵取樣問題，再配合上自適性模糊控制的演算法成功的將其適用於一般性的姿態辨識，並解決一些較不易分辨姿態的問題，以及在因背景或外界干擾所產生辨識雜訊的情況下更具強健性。藉由系統的學習功能，讓我們可將其提高效能，將影像處理達到即時監控系統的要求。在本計劃中我們成功了實踐一個即時的辨識監控系統。 英文：We present a real time postures recognition monitoring system. It has the potential to recognize the general postures and gives a good classified hit rate. In the plan, we first use some concepts to grab the human bounding box to get the feature parameters, like some ratio of the body length and width. We then present a method based on the fuzzy if-then rules systems to modify how the class distribute. In our method, the data with close characteristic may still have some boundary to separate them. That is with our method the auto generating rules can apply to this kind of classified problem and solve it well. The adaptive fuzzy rule based method can automatically adjust the grades of certainty of fuzzy rules. The method also has the learning ability to improve the error situation that is the ability to correct the errors. Finally we establish a real time system for recognizing the postures.
可利用之產業及可開發之產品	居家監控系統、居家安全及看護系統、金融保全系統、一般消費性電子產品等等。
技術特點	採用一般影像攝影機單機，無須額外感測器（三維加速規、壓電器或是壓力規），系統硬體成本具競爭力。且裝設架構簡易方便，體積小維護容易，且不影響裝設地點原來之規劃及作息。
推廣及運用的價值	可輕易快速將技術移植於居家監控系統、居家安全及看護系統、金融保全系統、一般消費性電子產品等等領域應用上。

※ 1. 每項研發成果請填寫一式二份，一份隨成果報告送繳本會，一份送 貴單位研發成果推廣單位（如技術移轉中心）。

※ 2. 本項研發成果若尚未申請專利，請勿揭露可申請專利之主要內容。

※ 3. 本表若不敷使用，請自行影印使用。

出席國際學術會議心得報告

計畫編號	NSC 95-2221-E-002-234
計畫名稱	智慧型居家看護影像監控系統(3/3)
出國人員姓名 服務機關及職稱	陳永耀 台灣大學電機系教授
會議時間地點	2007/10/28~2007/10/31 美國紐約紐約州紐約市
會議名稱	2007 年 IEEE International Ultrasonics Symposium
發表論文題目	(中文) 1. 分離對焦多方向加熱式超音波於乳房腫瘤熱手術 2. 雙曲線超音波陣列於乳房腫瘤熱治療之探討 (英文) 1. Split-focused ultrasound for breast tumor thermal surgery with multidirectional heating 2. Investigation of dual curved ultrasound phased array for breast tumor thermal therapy

一、參加會議經過

2007 年 IEEE 國際超音波研討會於 10 月 28 日至 31 日在美國紐約州紐約市 Hilton & Towers 舉行。此次大會是由 IEEE 超音波、磁性及頻率控制學會(IEEE Ultrasonics, Ferroelectrics, and Frequency Control Society)所主辦，參與人數相當多，大會收到 1000 份以上的投稿，在四年內成長了 33%，分別來自於美國、歐洲、亞洲等多個國家。



大會的主席為 John A. Kosinski, U.S. Army RDECOM, Fort Monmouth, NJ, USA

會議議程主席為 Mauricio Pereira da Cunha, University of Maine, Orono, ME, USA

本次大會會議相關主題分為五大組別，包括：

1. Group 1: Medical Ultrasonics

Vice Chair: Stanislav Emelianov: University of Texas at Austin, USA

2. Group 2: Sensors, NDE, and Industrial Application

Vice Chair: Jafar Saniie: Department of Electrical & Computer Engineering, Illinois Institute of Technology, USA

3. Group 3: Physical Acoustics

Vice Chair: Yook-Kong Yong, Rutgers University, USA

4. Group 4: Surface Acoustic Waves

Vice Chair: Peter Smith, McMaster University, Canada

5. Group 5: Transducers and Transducer Materials

Vice Chair: Scott Smith, GE Global Research, USA

同時每個組別都安排了多場邀請演講：

GROUP 1:

- Lihong Wang, "High-Resolution Photoacoustic Tomography"
- Bjorn Angelsen, "New Methods of Nonlinear Ultrasound Imaging"
- Evan C. Unger, "Therapeutic Applications of Microbubbles -Sonothrombolysis and Beyond"
- Juin-Jet Hwang, "Portable Echo Imaging System"
- Kjell Kristoffersen, "Real-Time 3D Cardiac Imaging With 2D Array Transducers"
- Kullervo Hynynen, "Ultrasound Potentiated Therapy"
- George Sutherland, "Myocardial Strain Rate Imaging"
- Steven Feinstein, "Contrast-Enhanced, Ultrasound Imaging of Atherosclerosis"
- Zahi A. Fayad, "Multimodality (PET/CT, MR, and CT) Imaging of the Atherosclerotic Plaque and

Novel Nanoparticulate Drug Delivery: Implications for Cardiovascular Drug Design and Discovery"

GROUP 2:

- Michael Thompson and Scott Ballantyne, "Ultra High Frequency Acoustic Wave Detection Of HIV Antibody In Whole Serum"
- Paul D. Wilcox and Bruce W. Drinkwater, "Ultrasonic Arrays for NDE - Historical Perspective and Future Concepts"
- B. T. Khuri-Yakub , K-K Park, J. Lee, and G. Yaralioglu, "The Capacitive Micromachined Ultrasonic Transducer(CMUT) as a Chem/Bio Sensor"

GROUP 3:

- Joel Kent, M. Takeuchi, and G. Laux, "Robert Adler's Touchscreen Inventions & A Memorial Video"
- Sergey Nikitov, "Review of Phononic Crystals, Devices, and Prospects"
- Jyrka Kaitila, "Review of Wave Propagation in BAW Thin Film Devices - Progress and Prospects"
- Abdelkhrim Khelif, "Phononic Crystals: a New Acoustic Material System"

GROUP 4:

- Michio Kadota, "High Performance and Miniature Surface Acoustic Wave Devices with Excellent Temperature Stability Using High Density Metal Electrodes"
- Victor Steel, "Single-Chip Radio Solutions for Mobile Phone Front-Ends"
- Richard Ruby, "Review and Comparison of Bulk Acoustic Wave FBAR and SMR Technology"

GROUP 5:

- Levent Degertekin, "Design Optimization and Integrated Electronics for Dual Electrode

CMUTs"

- Rajiv Chopra, "Integrating Ultrasound Transducers With MRI For Therapeutic And Diagnostic Applications"
- Sung Min Rhim, "Piezoelectric Single Crystal for Medical Ultrasound Transducer"

大會於 10 月 28 日開始，分為 6 個 Track，總共 60 個 Oral Sessions，6 個 Poster Sessions。

第二天早上八點至九點三十分舉行 Keynote 演講，大會邀請了 Columbia University 的 Prof.

V. Mow and Prof. E. Konofagou 擔任本次大會 President's Speaker，演講題目為：

"Cartilage and Osteoarthritis: Biomechanics and Ultrasound"

內容對生物力學及超音波有精闢之說明。

本次參加會議發表之論文排定在 P3C 場次，於 10 月 30 日發表，論文題目為 –

1. Split-focused ultrasound for breast tumor thermal surgery with multidirectional heating (P3C-5)
2. Investigation of dual curved ultrasound phased array for breast tumor thermal therapy (P3C-7)





二、與會心得

大會在前後四天的議程後，圓滿落幕。此次參加 IEEE International Ultrasonics Symposium 能夠見到許多超音波研究相關之專家學者，齊聚一堂，共同討論，個人覺得頗有收穫。廣泛的參與國際學術活動對提升我國之學術地位應有極為正面的意義。

整體而言，IEEE International Ultrasonics Symposium 參加的人數相當多，會場中常常很擁擠，不過參加的學者專家也互動很多。會場之準備適宜，口頭報告的時間掌控良好，poster 的場地及展示也不錯，各項優點值得借鏡。目前國內舉辦國際會議之水準也已相當不錯，我國在經濟發展與爭取國際空間之目標上，應努力爭取主辦國際會議，同時加強軟硬體之設施，鼓勵與各國之學術交流，提高我國學術地位。此次參會，除攜回一些大會資料外，尚有大會論文集光碟。