Gender Classification Using Human Gait Based on Skeleton-Model-Based Method Through Joint Angle Estimation and Model-Free Method with Gait Energy Motion Derived from Wavelet

September 2013

Department of Science and Advanced Technology

Graduate School of Science and Engineering

Saga University

ROSA ANDRIE ASMARA

GENDER CLASSIFICATION USING HUMAN GAIT BASED ON SKELETON-MODEL-BASED METHOD THROUGH JOINT ANGLE ESTIMATION AND MODEL-FREE METHOD WITH GAIT ENERGY MOTION DERIVED FROM WAVELET

Dissertation Submitted to the Department of Science and Advanced Technology, Graduate School of Science and Engineering, Saga University in Partial Fulfillment for the Requirements of a Doctorate Degree in Information Science

By

ROSA ANDRIE ASMARA

Nationality	:	Indonesia
Previous degrees	:	Bachelor of Electrical Engineering (Electronic) Faculty of Engineering Brawijaya University, Malang, Indonesia
		M.T. of Electrical Engineering Institute of Sepuluh Nopember, Surabaya, Indonesia

Department of Science and Advanced Technology
Graduate School of Science and Engineering
Saga University,

JAPAN

September 2013

APPROVAL

Graduate School of Science and Engineering
Saga University
1 - Honjomachi, Saga 840-8502, Japan

CERTIFICATE OF APPROVAL

Dr. Eng. Dissertation

This is to certify that the Dr. Eng. Dissertation of

ROSA ANDRIE ASMARA

Has been approved by the Examining Committee for the dissertation requirements for the Doctor of Engineering degree in Information Science in September 2013.

Dissertation Committee :

Supervisor, Prof. Kohei Arai
Department of Science and Advanced Technology

Member, Prof. Shinichi Tadaki
Department of Science and Advanced Technology

Member, Associate Prof. Hiroshi Okumura
Department of Science and Advanced Technology

Member, Associate Prof. Koichi Nakayama
Department of Science and Advanced Technology

DEDICATION

Kupersembahkan tesis ini untuk istriku dan anak-anakku tercinta

Anik Nur Handayani,

Muhammad Putra Anandra, dan

Naurisya Asri Anindri.

Acknowledgment

I wishes to express my profound and sincere gratitude to my supervisor, Prof. Kohei Arai, for his understanding, encouraging, excellent guidance and constant support throughout. His wide knowledge and his logical way of thinking have been great value for the author. This dissertation work could not complete without his unstinted help in both academic and personal concerns.

I also would like to thanks to Prof. Shinichi Tadaki, Associate Prof. Hiroshi Okumura, and Associate Prof. Koichi Nakayama for their help, suggestion and serving as members of my doctoral program committees.

The author would like to acknowledge with appreciation the scholarship grant provided by DIKTI from Indonesian Department of Education and Cultural and also State Polytechnics of Malang which made it possible to pursue his doctoral studies at Saga University, Saga Japan.

Terima kasih juga saya ucapkan untuk bapak-bapak *senpai* yang turut memberikan masukan ilmunya sehingga buku tesis ini dapat diselesaikan dengan baik. Pak Basuki, Pak Tri, Pak Herman, Pak Lipur, dan Pak Ronny dan Sang-san, semoga sukses and I love you all. Untuk rekan-rekan satu lab seperjuangan Pak Steven dan Pak Cahya, terima kasih atas bantuan2 kecil yang jika dikumpulkan benar2 jadi tak terhitung jumlahnya. Untuk teman-teman Indonesia di kampung Saga yang tidak mungkin saya sebutkan namanya satu persatu, terima kasih atas dukungan sosial, moral, dan nihongo-nya, semoga teman-teman dapat menyelesaikan studinya tepat waktu dan dapat berkumpul kembali dengan keluarga di Indonesia dengan selamat.

Terakhir, untuk istriku tercinta Anik Nur Handayani. Terima kasih untuk kasih sayang dan kesabaranmu selama ini, semoga perjuangan kita di Saga tercinta selesai dengan indah sesuai rencana kita. Untuk anak-anakku Muhammad Putra Anandra dan Naurisya Asri Anindri, terima kasih atas pengorbanan kalian ya nak, bersedia ikut ke Jepang tanpa tahu maksud kami yang sebenarnya, yaitu ingin agar kalian dapat membantu kami berdiri jika jatuh ditengah perjalanan kami menyelesaikan tesis. Semoga Allah SWT menjadikan kalian anak-anak yang terus berbakti pada kami, menuntun kalian agar dapat menuruti perintah agama dan menjalani larangan-Nya dengan mudah, dan memberikan jalan hidup yang terbaik bagi kalian. Doa kami akan selalu

teriring untuk kalian. Untuk ayah, ibu, dan saudara-saudaraku di Surabaya, Dik Ratih dan Om Ricky, Dik Aris dan Dik Ronny. Aku kehabisan kata-kata untuk mengungkapkan terima kasihku. Semoga Allah membalas dengan kebaikan yang setimpal. Untuk bapak, ibu, dan saudara-saudara beserta keponakan-keponakanku di Malang, terima kasih dukungan doa, pengertian dan kelapangan waktunya berskype ria dengan kami, juga kiriman snack dan bahan-bahan makanan khas Malang yang benar-benar mengobati rindu kami akan tempat tinggal kami.

Saga, 2013-June-02

ROSA ANDRIE ASMARA

Abstract

Biometrics as an identification system had been famous recently due to their wide implementation in security, identification and recognition. Application of biometrics such as gender, ages, pregnancy, disable classification and human recognition. Many researcher had focus in finding best feature for biometrics. One of the topics is finding best feature for gait biometrics. Due to the high technology improvement, also improve the image processing technology. Using image processing technology to be used as a feature extractor is a promising application in near future because of the small and cheap existence camera in the market today. Gait is very unique compare to other biometrics which is used video as the input rather than image or frequency signal. Gait also can only be detected in further distance compare to other biometrics creating an obtrusiveness. Spatial and temporal information is hiding behind and hope to be found.

This research provide the experimental results of some feature extracted from gait using Chinese Academy of Sciences (GAIT) Database. The first feature is extracted the kinematics feature, one of the simple and tradition feature extractor in biometrics. We used skeleton model to extract this kind of feature based on Nixon model, then we compare our kinematics feature to their feature. To improve the skeleton model, we proposed 3d skeleton model using depth sensor Kinect Camera and implement it in disable gait classification. The third step is experiment on gait images and analyze the images itself if they can or cannot be used as feature. Some of the researcher also shown that gait image feature extraction giving promising results. We use Image motion sequences as the image feature and 2D Discrete Wavelet Transform (2D DWT) Energy as the feature extractor. The fourth experiment is creating best gait image to be used as a feature. The purpose is to improve the image motion sequences and we call our image "Gait Energy Motion (GEM)".

From the experimental results shown that our kinematics feature from skeleton model giving 81.6% classification accuracy compare to Nixon kinematics. Using this model also we should consider that the hip angle is the most discriminant feature for gender classification. 3d skeleton model also giving good results around 86.7% classification accuracy. This results cannot be

compare because there are not such of this kind of research before. Our image feature extractor giving 92.9% classification accuracy using motion image sequences and 97.63% using GEM.

Keyword: *Gait gender classification, gait skeleton model, gait image feature extraction, SVM classification, morphological image, 3d gait skeleton model, Kinect skeleton,*

Contents

List of Figures

List of Tables

1. Introduction

This chapter presents the overview, the motivation, the objectives, and thesis writing layout of the work presented in this dissertation.

1.1 Overview

The topic of this research is a study about Human Gait as one of Biometrics System and its implementations in Gender and Disable Person Classification.

Biometrics is used in computer science as a form of identification and access control. It is also used to identify individuals in groups that are under surveillance. Biometric identifiers are the distinctive, measurable characteristics used to label and describe individuals. Biometric identifiers are often categorized as physiological versus behavioral characteristics. A physiological biometric would identify by one's voice, DNA, hand print or behavior. Behavioral biometric are related to the behavior of a person, including but not limited to typing rhythm, gait, and hand gesture. Some researchers have coined the term behavior-metrics to describe the latter class of biometrics.

More traditional means of access control include token-based identification systems, such as a driver's license or passport, and knowledge-based identification systems, such as a password or personal identification number. Since biometric identifiers are unique to individuals, they are more reliable in verifying identity than token and knowledge-based methods; however, the collection of biometric identifiers raises privacy concerns about the ultimate use of this information. The most popular biometric researches in the computer science subject are analyzing face, fingerprint, palm print, iris, gait or a combination of these traits [1].

The topic in this research covering biometrics application for gender classification. Not all biometrics can be used for gender classification. Gesture (gait), voice, and face in some references show promising results in gender classification [2–6]. This information will not only enhance existing biometrics systems but can also serve as a basis for passive surveillance and control in “smart area” (e.g., restricting access to certain areas based on gender) and collecting valuable demographics (e.g., the number of women entering a retail store, airports, post office, or

public smoking area etc. on a given day) [5]. Gender classification along with age and race classification can also be used to help user perceptual if human recognition is failed.

Human gait is the way locomotion is achieved using human limbs. Human gait is defined as bipedal, biphasic forward propulsion of center of gravity of human body, in which there is alternate sinuous movements of different segments of the body with least expenditure of energy. Different gaits are characterized by differences in limb movement patterns, overall velocity, forces, kinetic and potential energy cycles, and changes in the contact with the surface (ground, floor, etc.).

Compared to other biometric methods, gait recognition offers several unique characteristics. The most attractive characteristic is its unobtrusiveness, which does not require observed subjects' attention and cooperation. Also, human gait can be captured at a far distance without requiring physical information from subjects. This favorable characteristic has great advantages, especially when individual information such as face image is confidential [7][8]. Moreover, gait recognition offers great potential for recognition of low-resolution videos, where other biometrics technologies may be invalid because of insufficient pixels to identify the human subjects [9]. Several review articles provide a general overview of gait recognition [7][8]. However, a comprehensive survey of recent development of gait recognition can be rarely found.

Gait biometrics for identification have disadvantages compare to other biometrics. Some of the disadvantages are walking style of a person can be easily change due some factors like the change of a body weight, gait accident, aging etc.

The general framework of automatic gait recognition consists of subject detection, silhouette extraction, feature extraction, feature selection, and classification. Once moving subjects are captured, individuals will be detected and separated from the image background. The most widely used method is background subtraction, which attempts to separate objects from the difference between the modeled background and the current frame [10]. The initial detection of humans within images and the consequent separation from the background and silhouette extraction can be considered as a preprocessing step of gait recognition.

After individuals have been separated from the background, features that can be used for recognition are extracted from these segmented walking persons. There are mainly two kinds of

gait features, i.e., model-based features and model-free features. Model-based features employ static and dynamic body parameters and are generally view and scale invariant [11][12][13]. On the other hand, model-free features usually only use binary silhouettes and do not need construction of a model for walking persons [14][15]. This research also considered using combination of model and free model based [16].

Features extracted from segmented video sequences are commonly not effective for classification and require too many training samples because of high dimensionality. Many dimensionality reduction methods have been proposed to solve this problem. Among them, principal component analysis (PCA) [14][17] and linear discrimination analysis (LDA) [17] are widely used. The proposed research also considered using 2D Discrete Wavelet Transform (2D DWT) to reduce the dimensionality of the feature [18][19]. Some researcher also using their feature reduction method used special for gait silhouettes like Gait Energy Motion (GEM) [16].

The last step is to classify the test sequence to a particular individual based on the extracted features. The classification of gait features is mainly based on three categories of methods, namely, direct classification, similarity of temporal sequences and state-space model. The direct classification is usually used after single representation or key frames are extracted from a temporal sequence of gait frames. While the similarity of temporal sequences is used to measure the distance between two temporal sequences of gait, the state-space model such as Hidden Markov Model (HMM) [20–22]focuses on the pattern of state related to succession of stance. This stochastic approach explicitly employs both the similarity information between test and reference sequences, and probability of shapes appearance [7].

1.2 Motivation

Biometrics system is a human identification technology using human physiology or behavior. Finger print, iris and face are some examples of the famous biometrics based on human physiology. Signature and gait are an examples of behavioral biometrics. Most popular and already implemented until today is fingerprint biometric. India is one of the country using fingerprint biometrics to use as an ID for citizen. Currently, India is the biggest biometric identification program in the world with more than 200 million people enrolled as of January 2012.

One biometrics have a superiority among the other ones. For example fingerprint recognition method (*minutiae*) is better computation time compare to face recognition (*Gabor filter*). In contrary, face recognition giving more feature rather than fingerprint. In the application, face recognition can identify more people rather than fingerprint. To overcome the weakness of one biometrics, some application use the combination of more than one biometrics namely emerging biometrics. The research group at University of Wolverhampton lead by Ramaswamy Palaniappan has shown that people have certain distinct brain and heart patterns that are specific for each individual. The advantage of such 'futuristic' technology is that it is more fraud resistant compared to conventional biometrics like fingerprints. However, such technology is generally more cumbersome and still has issues such as lower accuracy and poor reproducibility over time.

There are some reason to motivate the conduction of gait research analysis. Many researchers had focus in gait biometrics today [2], [7–9], [11–16], [18], [19], [21–42]. One advantage of gait biometrics compared to other is it does not need observed subjects attention and cooperation. The subject even does not realized that they have been investigate from the distance. The research proposed can classify object about 3 meter distance from camera. Gait biometrics can also implemented for age classification, gender classification, pregnant classification, gait disabilities and human recognition. This research will focus on gender classification and gait disabilities. Gender classification will help us analyze what kind of method and feature extraction can be used in gait disabilities and human recognition. Gender gait will classify people between male and female from the way their walk. Gender classification is choose in the implementation due to its simplicity because this system only classify 2 class. In gait disabilities, the system will classify 5 class. In implementation, gait disabilities will help therapist determine how far their therapy improve their patient’s gait disease. There are also some dataset already provide by gait researcher. The datasets are USF (belong to *University of South Florida*) gait, SOTON (belong to *University of Southampton, UK*) gait, and CASIA (belong to *Chinese Academy of Sciences*) gait. Compared to other gait dataset, CASIA provide some advantages. CASIA dataset is a multi-view dataset. The object taken using 11 different angles of camera, have 124 different object and every object taken 6 times. The proposed research used CASIA gait dataset in normal condition using 90 degree angles of single-view camera. The object also have different condition like clothing, normal, and carrying condition. Further work of this research is for human recognition [25], [26], [29], [43].

1.3 Objective

The main objective of this research is to analyze the ability of human gait for biometrics classification and identification. Particularly, the proposed method implement in gender classification, analyze the dimensional reduction and gait disabilities classification. More specifically, this dissertation deals with the following objectives:

1. Analyze the capabilities of CASIA gait as a Human Gait Recognition Dataset [18].
2. Using 2D Discrete Wavelet Transforms (2D DWT) Energy for dimensionality reduction in gait dataset and implemented in Human Gait Gender Classification [19].
3. Create free-model based feature extraction and implemented in gender classification [16], [19].
4. Create a combination of free-model based feature extraction as a spatial information and some temporal information and implemented in gender classification [16].
5. Create model based feature extraction and implemented in gait disabilities classification [42].

1.4 Layout of Thesis

This dissertation is organized as five chapter. The structure and relation among chapters are shown in Figure 1-1: Flowchart of the thesis continuity.

Chapter 1 presents the introduction, motivation, objectives and relationship between chapters of this dissertation.

Chapter 2 reviews the background information regarding image processing for preprocessing subsystem creation, specifically in background subtraction, create image silhouettes, and creation of image skeleton as a part of gait motion tracking. This chapter also explain about dimensional reduction used such as creation of Gait Energy Motion (GEM), Principal Component Analysis (PCA), and 2D Discrete Wavelet Transforms (2D DWT) Energy. Feature extraction used in this research such as kinematics feature (spatial domain), statistical feature and time frame feature (temporal domain). The classification will be explained in chapter 2 is the usage of Decision Tree and Support Vector Machine (SVM). Decision tree used to remove the feature which is decreasing the classification accuracy.

Chapter 3 describe the design and implementation of the system which is describe in 1.3.

- First task is to analyze the CASIA gait dataset. The purpose of this task is to understand the capabilities of this dataset to be implemented in the next stage of the research. The analysis will be use 2D Discrete Wavelet Transform (2D DWT) Energy as a dimensional reduction also as a feature extracted [18].
- Second task is the usage of 2D Discrete Wavelet Transforms (2D DWT) Energy for dimensionality reduction in gait dataset and implemented in Human Gait Gender Classification [19].
- Fourth task is to create a combination of free-model based feature extraction as a spatial information with temporal information and implemented in gender classification [16].
- Third task is to create free-model based feature extraction and implemented in gender classification [16], [19]. Free-model based created is an image combination from silhouettes frame operation namely Gait Energy Motion [GEM].
- Fifth task is to create model based feature extraction and implemented in gait disabilities classification [42]. Model based created is called the skeleton model.

Chapter 4 describe the experimentation results from all the task done.

Chapter 5 provides the conclusion, summary and future research works.

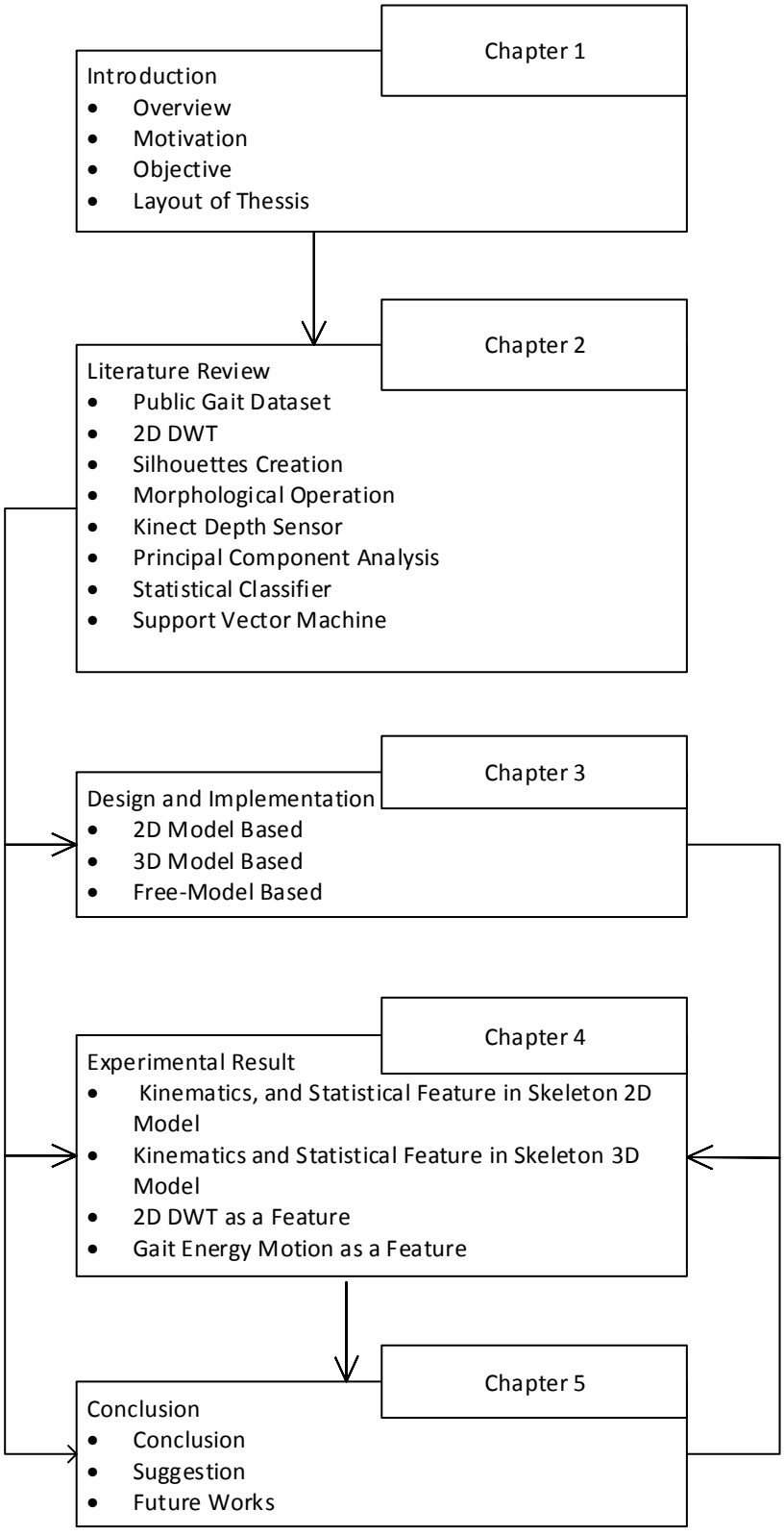


Figure 1-1: Flowchart of the thesis continuity

2. Literature Review

2.1 Public GAIT Dataset

Almost all gait researchers using same gait dataset for their experiments due to its efficiency. Standard publically available gait datasets are needed to fairly compare and evaluate the performance of gait recognition algorithms. Creating own dataset will need more time because of some factors like the amount of people for gait subject. External factor also giving more effect like lighting and big space room for video shooting. Good hardware specification like video camera and should be more than one camera installed. There are a lot of gait dataset but the famous among them are University of South Florida (USF) Gait Dataset from United States of America, University of Southampton (SOTON) Gait Dataset from United Kingdom, and Chinese Academy of Sciences (CASIA) Gait Dataset. This chapter will be detail explained all 3 dataset mentioned. All the experiment conducted in this research using only one dataset which is CASIA gait dataset.

2.1.1 University of South Florida (USF) Gait Dataset

Among two other gait dataset, USF dataset is the oldest one. The file in USF Gait Dataset consists of raw image sequences. The data was collected over four days, May 20-21, 2001 and November 15-16, 2001 at University of South Florida (USF), Tampa, United States of America. This dataset consist of 33 subjects common between the May and November collections. USF IRB approved informed consent was obtained from all the subjects in this dataset. The dataset consists of persons walking in elliptical paths in front of the camera(s). Each person walked multiple (>= 5) circuits around an ellipse, out of which the last circuit forms the data set. The protocol for the data collection, including specifications of the used imaging equipment, is not described as well. For each person, it have up to 5 covariates:

- 2 different shoe types (A, and B),
- 2 different carrying conditions (with or without a briefcase),
- on 2 different surface types (grass and concrete),

- from 2 different viewpoints (Left or Right) and
- some at 2 different time instants

Thus, there are 32 possible conditions under which persons gait could have been imaged. However, not all subjects were imaged in all conditions. The full data set can be partitioned as depicted in the following grid. File naming convention in USF dataset shown in Figure 2-1: Naming convention in USF Gait Dataset.

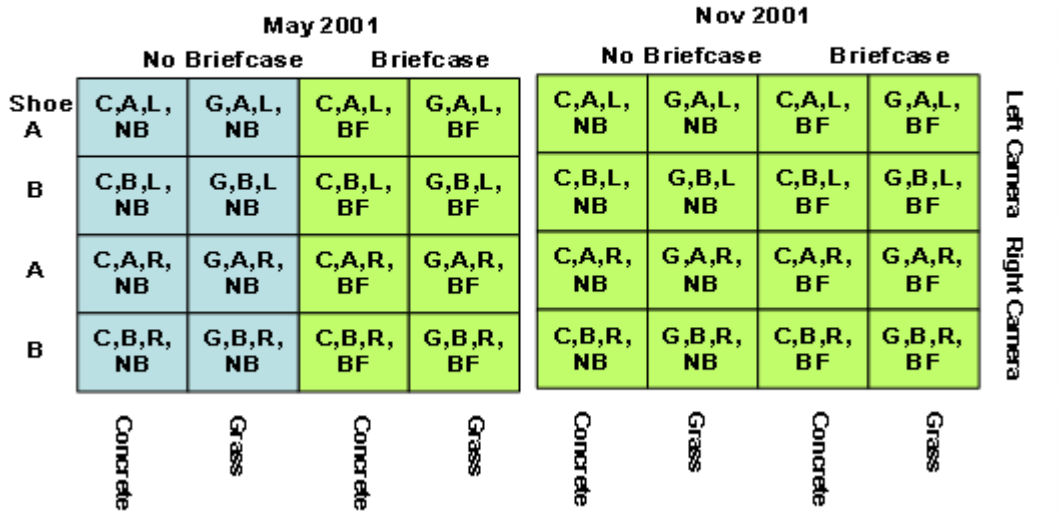


Figure 2-1: Naming convention in USF Gait Dataset

The cells with blue background as the May-2001-No-Briefcase data (used in ICPR-02 and FGR-02 papers). The dataset distributed in only the full version of the dataset which consists of 1870 sequences from 122 subjects. The total size of the data is around 1.2 Terabytes, uncompressed, and is distributed in compressed form using: one external 750GB (or more) drive.

An example frame from the USF gait dataset is shown in Figure 2-2: An Example frame from the USF Gait Dataset.

Figure 2-2: An Example frame from the USF Gait Dataset

2.1.2 University of Southampton (SOTON) Gait Dataset

The Southampton Human ID at a distance gait database consists of two major segments - a large population (~100), but basic, database and a small population, but more detailed, database. The large database is intended to address two questions: whether gait is individual across a significant number of people in normal conditions, and to what extent research effort needs to be directed towards biometric algorithms or towards computer vision algorithms for accurate extraction of subjects. The small database is intended to investigate the robustness of biometric techniques to imagery of the same subject in various common conditions (carrying items, wearing different clothing or footwear).

2.1.3 Chinese Academy of Sciences (CASIA) Gait Dataset

In the CASIA Gait Database there are three datasets: Dataset A, Dataset B (multi view dataset) and Dataset C (infrared dataset).

Dataset A (former NLPR Gait Database) was created on Dec. 10, 2001, including 20 persons. Each person has 12 image sequences, 4 sequences for each of the three directions, i.e. parallel, 45 degrees and 90 degrees to the image plane. The length of each sequence is not identical for the variation of the walker's speed, but it must ranges from 37 to 127. The size of Dataset A is about 2.2GB and the database includes 19139 images.

Figure 2-3 shows the example frame from the class-A CASIA Gait Dataset. The format of the image filename in Dataset A is 'xxx-mm_n-ttt.png', where:

- xxx: subject id,
- mm: direction,
- n: sequence number,
- ttt: frame number in a sequence

Figure 2-3: An Example frame from the Class-A CASIA Gait Dataset

Dataset B is a large multiview gait database, which is created in January 2005. There are 124 subjects, and the gait data was captured from 11 views. Three variations, namely view angle, clothing and carrying condition changes, are separately considered. Besides the video files, we still provide human silhouettes extracted from video files. The detailed information about Dataset B and an evaluation framework can be found in this paper.

Figure 2-4 shows the example frame from the class-A CASIA Gait Dataset. The format of the video filename in Dataset B is 'xxx-mm-nn-ttt.avi', where:

- xxx: subject id, from 001 to 124.
- mm: walking status, can be 'nm' (normal), 'cl' (in a coat) or 'bg' (with a bag).
- nn: sequence number.
- ttt: view angle, can be '000', '018', ..., '180'.

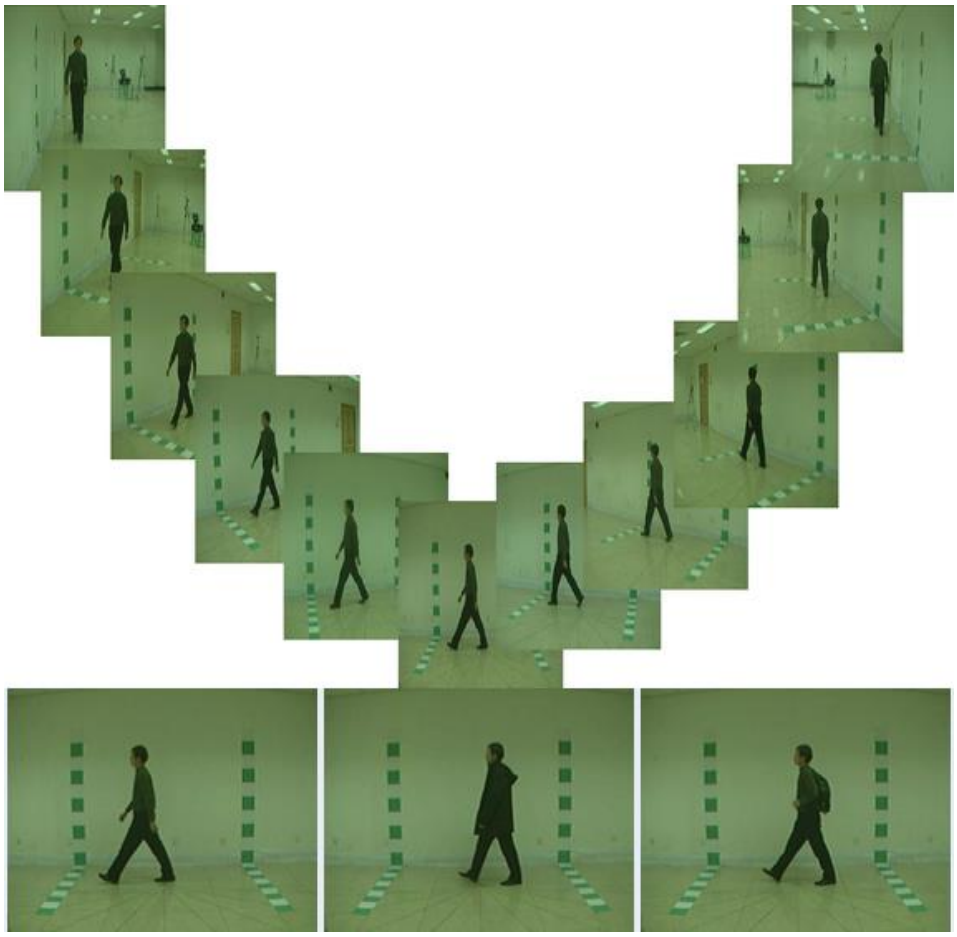


Figure 2-4: An Example frame from the Class-B CASIA Gait Dataset

Dataset C was collected by an infrared (thermal) camera in Jul.-Aug. 2005. It contains 153 subjects and takes into account four walking conditions: normal walking, slow walking, fast walking, and normal walking with a bag. The videos were all captured at night.

Figure 2-5 shows an example of Class-C CASIA Gait Database. The format of the video filename in Dataset C is '01xxxmmnn.avi', where

- xxx: subject id, from 001 to 153.
- mm: walking status, can be 'fn' (normal), 'fq' (fast walk), 'fs' (slow walk) or 'fb' (with a bag).
- nn: sequence number.

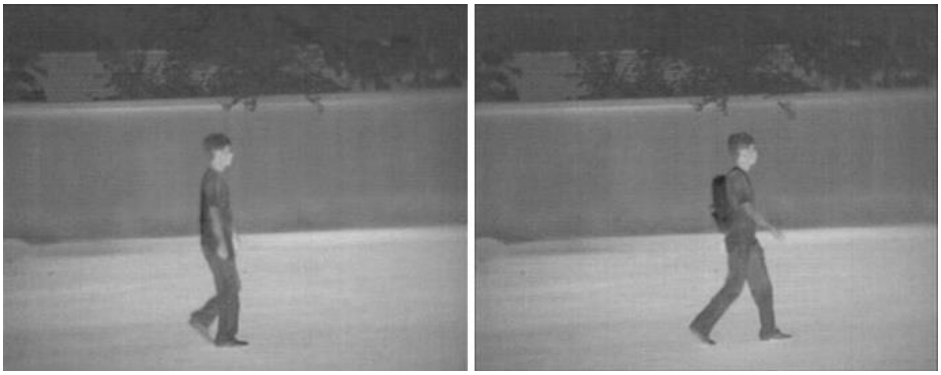


Figure 2-5: An Example frame from the Class-C CASIA Gait Dataset

Dataset D was collected synchronously by camera and Rscan Footscan in Jul.-Aug. 2009. It contains 88 subjects and takes into account real surveillance scenes and wide age distribution. This Dataset can be considered as the attempts in exploiting the relations between behavior biometrics and its corresponding prints. The videos and images are collected indoor, while all the subjects are Chinese. Figure 2-6 shows an example frame from the Class-D CASIA Gait Dataset.

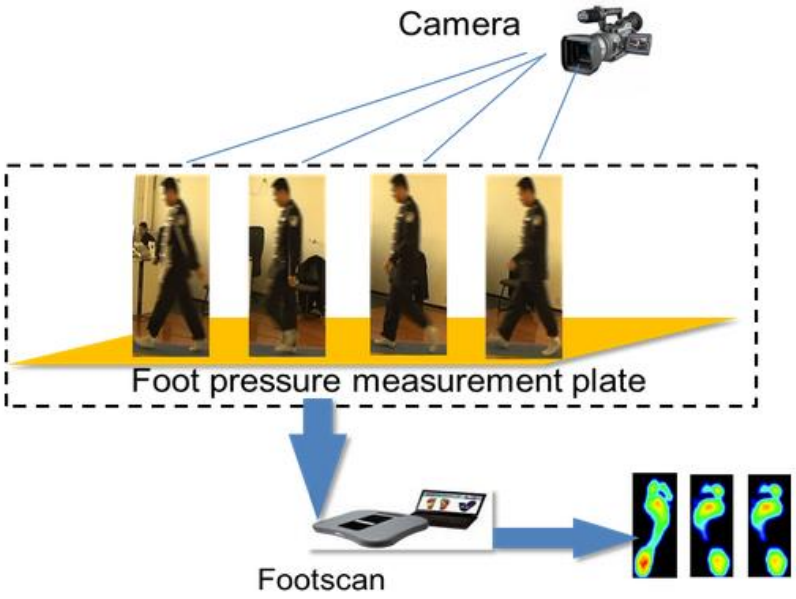


Figure 2-6: An Example frame from the Class-D CASIA Gait Dataset

Below is the table of CASIA gait video specification:

No.	Description	Detail
1.	Video Format	AVI Video

2.	Length	Around 3 – 5 seconds
3.	Size	Around 1MB
4.	Resolution	320 x 240
5.	Frame Rate	25 fps
6.	Bitrate	200 kbps

Table 2-1: CASIA Gait Video Specification

2.2 2D Discrete Wavelet Transform (2D DWT)

Discrete wavelet transform (DWT) represents an image as a subset of wavelet functions using different locations and scales [44]. It makes some decomposition images. Any decomposition of an image into wavelet involves a pair of waveforms: the high frequencies corresponding to the detailed parts of an image and the low frequencies corresponding to the smooth parts of an image. DWT for an image as a 2-D signal can be derived from a 1-D DWT. According to the characteristic of the DW decomposition, an image can be decomposed to four sub-band images through a 1-level 2-D DWT, as shown in Figure 2-7. These four sub-band images in Fig. 4 can be mapped to four sub-band elements representing LL (Approximation), HL (Vertical), LH (Horizontal), and HH (Diagonal) respectively.

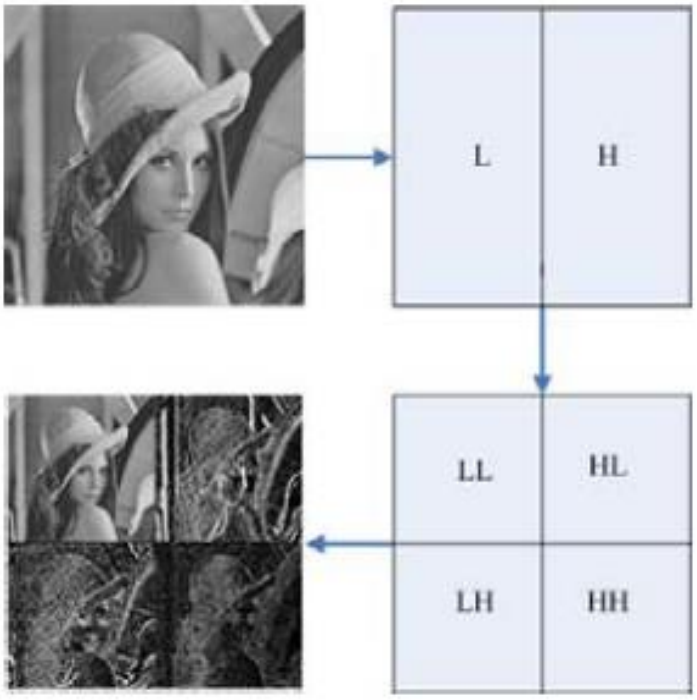


Figure 2-7: 1-Level Decomposition 2D DWT

The discrete Wavelet Transform will decompose a given signal into other signal known as the approximation and detail coefficients. A given function f(t) can be expressed through the following representation:

$$f(t) = \sum_{j=1}^{L} \sum_{K=-\infty}^{\infty} d(j,K)\varphi\left(2^{-j}t - K\right) + \sum_{K=-\infty}^{\infty} a(L,K)\theta(2^{-L}t - K) \quad \text{.....[1]}$$

Where: $\varphi(t)$ is the mother wavelet and $\theta(t)$ is the scaling function. *a*(L,K) is called the approximation coefficient at scale L and d(j,K) is called the detail coefficients at scale j. The approximation and detail coefficients can be expressed as:

$$a(L,K) = \frac{1}{\sqrt{2^L}} \int_{-\infty}^{\infty} f(t)\theta(2^{-L}t - K)dt \quad \text{.....[2]}$$

$$d(j,K) = \frac{1}{\sqrt{2^j}} \int_{-\infty}^{\infty} f(t)\varphi\left(2^{-j}t - K\right)dt \quad \text{.....[3]}$$

Based on the choice of the mother wavelet $\varphi(t)$ and scaling function $\theta(t)$, different families of wavelets can be constructed [45].

To draw easier the characteristics of the Wavelet coefficients, one can use the energy of each coefficient, then creates the 2D scatter graph for every combination of the coefficients. These are the formula of the energy for every:

$$E_{a(L,K)} = \sum_{i=1}^{video\ 2} \sqrt{\sum_{j=1}^{frame} |a(L,K)|^2} \quad \text{.....[4]}$$

$$E_{d(j,K)} = \sum_{i=1}^{video\ 2} \sqrt{\sum_{j=1}^{frame} |d(j,K)|^2} \quad \text{.....[5]}$$

Then normalize energy using the formula below:

$$E_{Total} = E_{a(L,K)} + E_{d(j,K)} \quad \text{.....[6]}$$

$$\%E_{a(L,K)} = \frac{100 \text{ x } E_{a(L,K)}}{E_{Total}} \quad \text{.....[7]}$$

$$\%E_{d(j,K)} = \frac{100 \text{ x } E_{d(j,K)}}{E_{Total}} \ \% \quad \text{.....[8]}$$

Which percentage of total energy:

$$100\%\ \text{Energy} = \left(\%E_{a(L,K)} + \%E_{d(j,K)}\right) \quad [9]$$

2.3 Silhouette Creation

The first thing to do in dataset preprocessing image is to create silhouettes. Silhouette image have to be process to every image sequences. After converting video to image sequences, the image can be proceed to create the silhouette sequences. There are two step to be considered two create silhouette, first one is object detection and the second one is noise removal. For object detection, this research will use background subtraction method. The biggest problem in an object detection is the appearance of hole in object detected. Some morphological operation used to solve the problem.

2.3.1 Background Subtraction

Background subtraction is a technique in the fields of image processing and computer vision wherein an image's foreground is extracted for further processing (object recognition etc.). Generally an image's regions of interest are objects (humans, cars, or text) in its foreground. After the stage of image preprocessing (which may include image de-noising etc.) object localization is required which may make use of this technique. Background subtraction is a widely used approach for detecting moving objects in videos from static cameras. The rationale in the approach is that of detecting the moving objects from the difference between the current frame and a reference frame, often called "background image", or "background model" [10]. Background subtraction is mostly done if the image in question is a part of a video stream.

A robust background subtraction algorithm should be able to handle lighting changes, repetitive motions from clutter and long-term scene changes. The following analyses make use of the function of $V(x,y,t)$ as a video sequence where t is the time dimension, x and y are the pixel location variables. E.g. $V(1,2,3)$ is the pixel intensity at (1,2) pixel location of the image at $t = 3$ in the video sequence.

Frame difference (absolute) at time t + 1 is

$$D(t+1) = |V(x,y,t+1) - V(x,y,t)| \quad \text{.....[10]}$$

The background is assumed to be the frame at time t. This difference image would only show some intensity for the pixel locations which have changed in the two frames. Though we have seemingly removed the background, this approach will only work for cases where all foreground pixels are moving and all background pixels are static. A threshold "*Th*" is put on this difference image to improve the subtraction (see Image thresholding).

$$|V(x,y,t) - V(x,y,t+1)| > Th \quad \text{.....[11]}$$

The accuracy of this approach is dependent on speed of movement in the scene. Faster movements may require higher thresholds.

2.3.2 Otsu's method

This proposed research using combination of threshold in [11] and histogram shape-based image thresholding (otsu's thresholding). The algorithm assumes that the image to be threshold contains two classes of pixels or bi-modal histogram (e.g. foreground and background) then calculates the optimum threshold separating those two classes so that their combined spread (intra-class variance) is minimal. Otsu's method is named after Nobuyuki Otsu.

Otsu's method exhaustively search for the threshold that minimizes the intra-class variance (the variance within the class), defined as a weighted sum of variances of the two classes:

$$\sigma_w^2(t) = \omega_1(t)\sigma_1^2(t) + \omega_2(t)\sigma_2^2(t) \quad \text{.....[12]}$$

Weights ω_1 are the probabilities of the two classes separated by a threshold t and σ_i^2 variances of these classes.

Otsu shows that minimizing the intra-class variance is the same as maximizing inter-class variance:

$$\sigma_b^2(t) = \sigma^2 - \sigma_w^2(t) = \omega_1(t)\omega_2(t)[\mu_1(t) - \mu_2(t)]^2 \quad \text{.....[13]}$$

Which is expressed in terms of class probabilities ω_1 and class means μ_i. The class probability $\omega_1(t)$ is computed from the histogram as t:

$$\omega_1(t) = \sum_0^t p\,(i) \quad \text{.....[14]}$$

While the class mean $\mu_1(t)$ is:

$$\mu_1(t) = [\sum_0^t p(i)x(i)]/\omega_1 \quad \text{.....[15]}$$

Where *x(i)* is the value at the center of the i_{th} histogram bin. Similarly, one can compute $\omega_2(t)$ and μ_t on the right-hand side of the histogram for bins greater than t.

The class probabilities and class means can be computed iteratively. This idea yields an effective algorithm.

Algorithm for Otsu's threshold	
1.	Compute histogram and probabilities of each intensity level
2.	Set up initial $\omega_i(0)$ and $\mu_i(0)$
3.	Step through all possible thresholds t = 1 … maximum intensity 1. Update ω_i and μ_i 2. Compute $\sigma_b^2(t)$
4.	Desired threshold corresponds to the maximum $\sigma_b^2(t)$
5.	One can compute two maxima (and two corresponding thresholds). $\sigma_{b1}^2(t)$is the greater max and $\sigma_{b2}^2(t)$ is the greater or equal maximum
6.	$Desired\ threshold = \frac{threshold_1 + threshold_2}{2}$

2.3.3 Sum of absolute Difference (SAD)

Sum of absolute differences (SAD) is an algorithm for measuring the similarity between image blocks. It works by taking the absolute difference between each pixel in the original block and the corresponding pixel in the block being used for comparison. These differences are summed to create a simple metric of block similarity, the L1 norm of the difference image or Manhattan distance between two image blocks.

The sum of absolute differences may be used for a variety of purposes, such as object recognition, the generation of disparity maps for stereo images, and motion estimation for video compression.

$$I = \sum_{(i,j)\in W} |I_1(i,j) - I_2(x+i, y+j)| \quad \text{.....[16]}$$

The sum of absolute differences provides a simple way to automate the searching for objects inside an image, but may be unreliable due to the effects of contextual factors such as changes in lighting, color, viewing direction, size, or shape. The SAD may be used in conjunction with other object recognition methods, such as edge detection, to improve the reliability of results.

2.4 Morphological operation

Morphological operation used in this research is a Binary morphology. In binary morphology, an image is viewed as a subset of a Euclidean space $\mathbb{R}^d$ or the integer grid $\mathbb{Z}^d$ for some dimension d.

The basic idea in binary morphology is to probe an image with a simple. Pre-defined shape, drawing conclusions on how this shape fits or misses the shaped in the image. This simple "probe" is called structuring element, and is itself a binary image (i.e. a subset of the space or grid).

Here are some examples of widely used structuring elements (denoted by B):

- Let$E = \mathbb{R}^2$; B is an open disk of radius r, centered at the origin.
- Let $E = \mathbb{Z}^2$; B is a 3x3 square, that is, B={(-1,-1), (-1,0), (-1,1), (0,-1), (0,0), (0,1), (1,-1), (1,0), (1,1)}.
- Let $E = \mathbb{Z}^2$; B is the "cross" given by: B={(-1,0), (0,-1), (0,0), (0,1), (1,0)}.

The basic operator are shift invariant operators strongly related to Minkowski addition.

Let E be a Euclidean space or an integer grid, and A a binary image in E.

2.4.1 Erosion

The erosion of the binary image A by the structuring element B is defined by:

$$A \ominus B = \{z \in E | B_{\approx} \subseteq A\} \quad [17]$$

Where B_z is the translation of B by the vector z, i.e.

$$B_{\approx} = \{b + z | b \in B\}, \forall z \in E \quad [18]$$

When the structuring element *B* has a center (e.g., *B* is a disk or a square), and this center is located on the origin of *E*, then the erosion of *A* by *B* can be understood as the locus of points reached by the center of *B* when *B* moves inside *A*. For example, the erosion of a square of side 10, centered at the origin, by a disc of radius 2, also centered at the origin, is a square of side 6 centered at the origin.

The erosion of A by B is also given by the expression:

$$A \ominus B = \bigcap_{b \in B} A_{-b} \quad [19]$$

Example application: Assume one have received a fax of a dark photocopy. Everything looks like it was written with a pen that is bleeding. Erosion process will allow thicker lines to get skinny and detect the hole inside the letter "o".

2.4.2 Dilation

The dilation of A by the structuring element B is defined by:

$$A \oplus B = \bigcup_{b \in B} A_b \quad [20]$$

The dilation is commutative, also given by:

$$A \oplus B = B \oplus A = \bigcup_{a \in A} B_a \quad [21]$$

If *B* has a center on the origin, as before, then the dilation of *A* by *B* can be understood as the locus of the points covered by *B* when the center of *B*moves inside *A*. In the above example, the dilation of the square of side 10 by the disk of radius 2 is a square of side 14, with rounded corners, centered at the origin. The radius of the rounded corners is 2.

The dilation can also be obtained by:

$$A \oplus B = \{z \in E | (B^s)_{\approx} \cap A \neq \emptyset\} \quad \text{.....[22]}$$

Where B^s denotes the symmetric of B, that is

$$B^s = \{x \in E | -x \in B\} \quad \text{.....[23]}$$

Example application: Dilation is the dual operation of the erosion. Easiest way to describe it is to imagine the same fax/text is written with a thicker pen.

2.4.3 Thin

Thin is a skeleton (or medial axis) representation of a shape or binary image, computed by means of morphological operators.

Below is the algorithm for thinning operation [46]:

Algorithm for Thinning Operation
1. Divide the image into two distinct subfields in a checkerboard pattern.
2. In the first sub iteration, delete pixel *p* from the first subfield if and only if the conditions G_1, G_2, and G_3 are all satisfied.
3. In the second sub iteration, delete pixel p from the second subfield if and only if the conditions G_1, G_2, and G_3' are all satisfied.

Condition G_1: $X_H(p) = 1$

where: $X_H(p) = \sum_{i=1}^{4} b_i$

$$b_i = \begin{cases} 1 \text{ if } x_{2i-1}=0 \text{ and } (x_{2i} = 1 \text{ or } x_{2i+1} = 1) \\ 0 \text{ otherwise} \end{cases}$$

$x_1, x_2, \ldots, x_8$ are the values of the eight neighbors of *p*, starting with the east neighbor and numbered in counter-clockwise order.

Condition G_2: $2 \leq \min\{n_1(p), n_2(p)\} \leq 3$

where:

$$n_1(p) = \sum_{k=1}^{4} x_{2k-1} \vee x_{2k}$$

$$n_2(p) = \sum_{k=1}^{4} x_{2k} \vee x_{2k+1}$$

Condition G3: $(x_2 \vee x_3 \vee \overline{x_8}) \wedge x_1 = 0$

Condition G3': $(x_6 \vee x_7 \vee \overline{x_4}) \wedge x_5 = 0$

The two sub iterations together make up one iteration of the thinning algorithm. When the user specifies an infinite number of iterations (n=Inf), the iterations are repeated until the image stops changing.

2.5 Kinect Depth Sensor Camera

Kinect is a motion sensing input device by Microsoft for the Xbox 360 video game console and Windows PCs. Based around a webcam-style add-on peripheral for the Xbox 360 console, it enables users to control and interact with the Xbox 360 without the need to touch a game controller, through a natural user interface using gestures and spoken commands. The project is aimed at broadening the Xbox 360's audience beyond its typical gamer base. Kinect competes with the Wii Remote Plus and PlayStation Move with PlayStation Eye motion controllers for the Wii and PlayStation 3 home consoles, respectively. A version for Windows was released on February 1, 2012.

Figure 2-8: Kinect Sensor Device

Kinect was launched in North America on November 4, 2010, in Europe on November 10, 2010, in Australia, New Zealand and Singapore on November 18, 2010, and in Japan on November 20, 2010. Purchase options for the sensor peripheral include a bundle with the game Kinect Adventures and console bundles with either a 4 GB or 250 GB Xbox 360 console and Kinect Adventures.

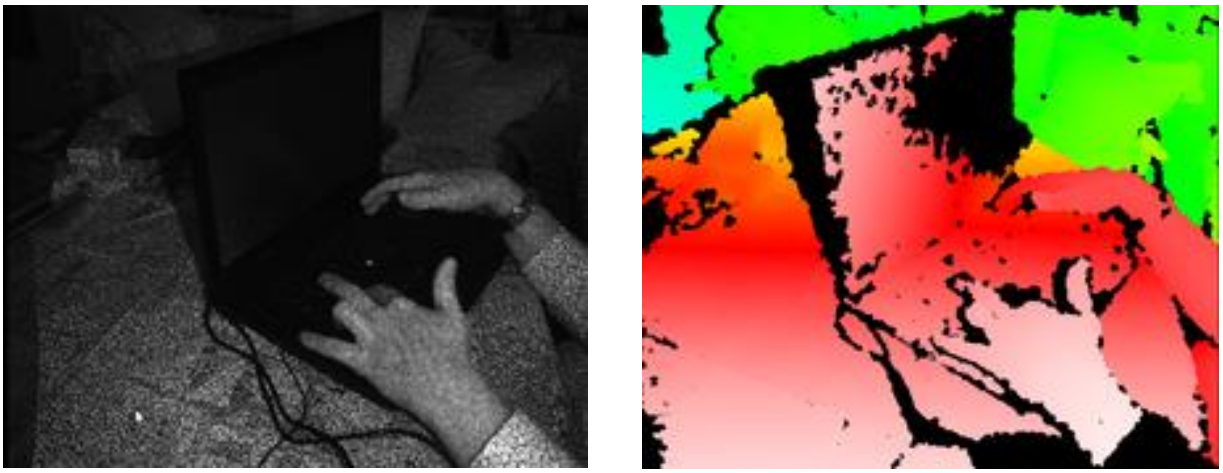
The Kinect claimed the Guinness World Record of being the "fastest selling consumer electronics device" after selling a total of 8 million units in its first 60 days. 24 million units of the Kinect sensor had been shipped as of January 2012.

Microsoft released Kinect software development kit for Windows 7 on June 16, 2011. This SDK was meant to allow developers to write Kinecting apps in C++/CLI, C#, or Visual Basic .NET.

The Kinect sensor is a horizontal bar connected to a small base with a motorized pivot and is designed to be positioned lengthwise above or below the video display. The device features an "RGB camera, depth sensor and multi-array microphone running proprietary software", which provide full-body 3D motion capture, facial recognition and voice recognition capabilities. At launch, voice recognition was only made available in Japan, the United Kingdom, Canada and the United States. Mainland Europe received the feature later in spring 2011. Currently voice recognition is supported in Australia, Canada, France, Germany, Ireland, Italy, Japan, Mexico, New Zealand, United Kingdom and United States. The Kinect sensor's microphone array enables the Xbox 360 to conduct acoustic source localization and ambient noise suppression, allowing for things such as headset-free party chat over Xbox Live.

The depth sensor consists of an infrared laser projector combined with a monochrome CMOS sensor, which captures video data in 3D under any ambient light conditions. The sensing range of the depth sensor is adjustable, and the Kinect software is capable of automatically calibrating the sensor based on gameplay and the player's physical environment, accommodating for the presence of furniture or other obstacles.

Described by Microsoft personnel as the primary innovation of Kinect, the software technology enables advanced gesture recognition, facial recognition and voice recognition. According to information supplied to retailers, Kinect is capable of simultaneously tracking up to six people, including two active players for motion analysis with a feature extraction of 20 joints per player. However, PrimeSense has stated that the number of people the device can "see" (but not process as players) is only limited by how many will fit in the field-of-view of the camera.



(a) (b)

Figure 2-9: (a) Infrared image shows the laser grid Kinect uses to calculate depth; (b) Visualization of depth map using color gradients from white (near) to blue (far)

2.6 Principal Component Analysis (PCA)

Principal component analysis (PCA) is a mathematical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linearly uncorrelated variables called principal components. The number of principal components is less than or equal to the number of original variables. This transformation is defined in such a way that the first principal component has the largest possible variance (that is, accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it be orthogonal to (i.e., uncorrelated with) the preceding components. Principal components are guaranteed to be independent only if the data set is jointly normally distributed. PCA is sensitive to the relative scaling of the original variables.

PCA was invented in 1901 by Karl Pearson. It is mostly used as a tool in exploratory data analysis and for making predictive models. PCA can be done by eigenvalue decomposition of a data covariance (or correlation) matrix or singular value decomposition of a data matrix, usually after mean centering (and normalizing or using Z-scores) the data matrix for each attribute. The results of a PCA are usually discussed in terms of component scores, sometimes called factor scores (the transformed variable values corresponding to a particular data point), and loadings

(the weight by which each standardized original variable should be multiplied to get the component score).

PCA is the simplest of the true eigenvector-based multivariate analyses. Often, its operation can be thought of as revealing the internal structure of the data in a way that best explains the variance in the data. If a multivariate dataset is visualised as a set of coordinates in a high-dimensional data space (1 axis per variable), PCA can supply the user with a lower-dimensional picture, a "shadow" of this object when viewed from its (in some sense; see below) most informative viewpoint. This is done by using only the first few principal components so that the dimensionality of the transformed data is reduced.

PCA is mathematically defined as an orthogonal linear transformation that transforms the data to a new coordinate system such that the greatest variance by any projection of the data comes to lie on the first coordinate (called the first principal component), the second greatest variance on the second coordinate, and so on.

Define a data matrix, X^T, with zero empirical mean (the empirical (sample) mean of the distribution has been subtracted from the data set), where each of the *n* rows represents a different repetition of the experiment, and each of the *m* columns gives a particular kind of datum (say, the results from a particular probe). The singular value decomposition of X is $X = W\Sigma V^T$, where the *m x m* matrix **W** is the matrix of eigenvectors of the covariance matrix XX^T, the matrix $\sum$ is an *m x n* rectangular diagonal matrix with nonnegative real numbers on the diagonal, and the *n x n* matrix **V** is the matrix of eigenvectors of X^TX. The PCA transformation that preserves dimensionality (that is, gives the same number of principal components as original variables) **Y** is then given by:

$$\begin{aligned} Y^T &= X^T W \\ &= (W\Sigma V^T)^T W \\ &= V\Sigma^T W^T W \\ &= V\Sigma^T \end{aligned}$$

V is not uniquely define in the usual case when $m < n - 1$, but **Y** will usually still be uniquely defined. Since **W** (by definition of the SVD of a real matrix) is an orthogonal matrix, each row of $\mathbf{Y^T}$ is simply a linear transformation of the corresponding row of $\mathbf{X^T}$. The first column of $\mathbf{Y^T}$ is

made up of the “scores” of the cases with respect to the “principal” component, the next column has the scores with respect to the “second principal” component, and so on.

Given a set of points in Euclidean space, the first principal component corresponds to a line that passes through the multidimensional mean and minimizes the sum of squares of the distances of the points from the line. The second principal component corresponds to the same concept after all correlation with the first principal component has been subtracted from the points. The singular values (in Σ) are the square roots of the eigenvalues of the matrix XXT. Each eigenvalue is proportional to the portion of the "variance" (more correctly of the sum of the squared distances of the points from their multidimensional mean) that is correlated with each eigenvector. The sum of all the eigenvalues is equal to the sum of the squared distances of the points from their multidimensional mean. PCA essentially rotates the set of points around their mean in order to align with the principal components. This moves as much of the variance as possible (using an orthogonal transformation) into the first few dimensions. The values in the remaining dimensions, therefore, tend to be small and may be dropped with minimal loss of information (see below). PCA is often used in this manner for dimensionality reduction. PCA has the distinction of being the optimal orthogonal transformation for keeping the subspace that has largest "variance" (as defined above).

2.7 Statistical Classifier

Statistical Classifier used to generate a decision tree developed by Ross Quinland. This statistical classifier is called C4.5 algorithm.

C4.5 builds decision trees from a set of training data using the concept of information entropy. The training data is a set $S = s_1, s_2, \ldots$ of already classified samples. Each sample s_i consists of a p-dimensional vector $(x_{1,i}, x_{2,i}, \ldots, x_{p,i})$, where the x_j represent attributes or features of the sample, as well as the class in which s_i falls.

At each node of the tree, C4.5 chooses the attribute of the data that most effectively splits its set of samples into subsets enriched in one class or the other. The splitting criterion is the normalized information gain (difference in entropy). The attribute with the highest normalized

information gain is chosen to make the decision. The C4.5 algorithm then recurses on the smaller sub lists.

This algorithm has a few base cases.

- All the samples in the list belong to the same class. When this happens, it simply creates a leaf node for the decision tree saying to choose that class.
- None of the features provide any information gain. In this case, C4.5 creates a decision node higher up the tree using the expected value of the class.
- Instance of previously-unseen class encountered. Again, C4.5 creates a decision node higher up the tree using the expected value.

In pseudocode, the general algorithm for building decision trees is:

1. Check for base cases
2. For each attribute *a*
    1. Find the normalized information gain from splitting on *a*
3. Let *a_best* be the attribute with the highest normalized information gain
4. Create a decision node that splits on *a_best*
5. Recurse on the sub lists obtained by splitting on *a_best*, and add those nodes as children of node

2.8 Support Vector Machine (SVM)

In machine learning, support vector machines (SVMs, also support vector networks) are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis. The basic SVM takes a set of input data and predicts, for each given input, which of two possible classes forms the output, making it a non-probabilistic binary linear classifier. Given a set of training examples, each marked as belonging to one of two categories, a SVM training algorithm builds a model that assigns new examples into one category or the other. A SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap

that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.

3. Methodology

All the research proposed in this thesis using 2D Model, 3D Model, and free-model Based for the feature extracted. Model-based features employ static and dynamic body parameters and are generally view and scale invariant [11][12][13]. On the other hand, model-free features usually only use binary silhouettes and do not need construction of a model for walking persons [14][15]. This research also considered using combination of model and free model based [16].

3.1 2D Model based

Model-based approaches obtain a series of static or dynamic body parameters via modeling or tracking body components such as limbs, legs, arms, and thighs. Gait signatures derived from these model parameters are employed for identification and recognition of an individual. It is evident that model model-based approaches are view-invariant and scale-independent. These advantages are significant for practical applications, because it is unlikely that reference sequences and test sequences are taken from the same viewpoint [7]. However, model-based approaches are sensitive to the quality of gait sequences. Thus, gait image sequences of high quality are required to achieve a high accuracy. Another disadvantage of the model-based approach is its large computation and relatively high time costs due to parameters calculation.

Primary model-based approaches employ static structure parameters of body as recognition features. BenAbdelkader et al. [39] present structural stride parameters consisting of stride and cadence. Figure 3-1 shows the overview method. The cadence is estimated via the walking periodicity, and the stride length is calculated by the ration of travelled distance and walking steps.

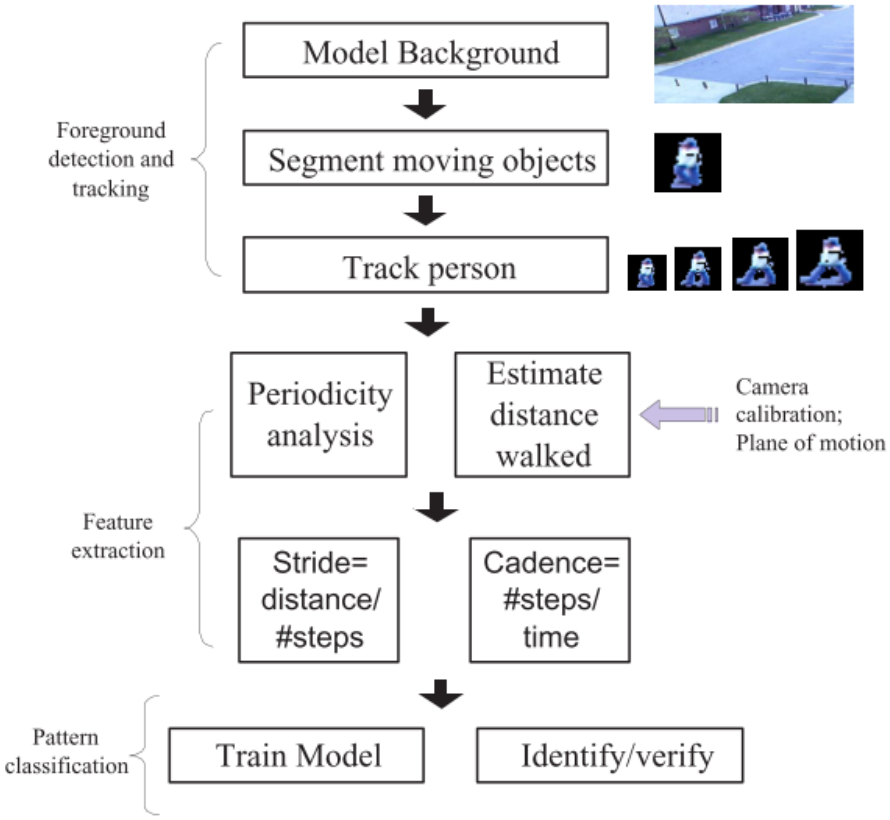


Figure 3-1: Overview of [39] Method

Bobick and Johnson [28] calculate four distances of human bodies, namely the distance between the head and foot, the distance between the head and pelvis, the distance between the foot and pelvis, and the distance between the left foot and right foot, as shown in Figure 3-2. They use these four distance to form two groups of static body parameters and reveal that the second set of parameters are more view-invariant comparing to the first set of body parameters. Their motion-capture system uses magnetic sensors to capture the three-dimensional position and orientation of the limbs of the subjects as they walk along a platform. Sixteen sensors in all are used for the head, torso, pelvis, hands, forearms, upper-arms, thighs, calves, and feet. Tanawongsuwan and Bobick [38] focus on the trajectories of joint angle from motion capture magnetic sensor data. The joint angles trajectories are computed by estimating the offsets between the 3D marker and joints as shown in Figure 3-3. Yam et al. [25] construct a structure and motion model of legs to analyze walking as well as running using biomechanics of human and pendulum motion as shown in Figure 3-4. A comparative higher recognition currency of running demonstrates that running may be more reliable for human identification due to more different gait pattern. Additionally, based on comprehensively analyzing the characteristics and

description of human gait, Cunado et al. [13] implemented Velocity Hough Transform (VHT) to extract the structure model of the thighs and the motion model of the thighs as shown in Figure 3-5. It is reported that the VHT achieved good performance of median noise immunity. Some other methods model human body parts separately. In Wang et al. [29]'s work, human body is modeled as fourteen rigid parts connected to one another at the joints as shown in Figure 3-6. The whole model has forty-eight degrees of freedom (DOFs). The tracking results, namely joint-angle trajectories signals, are considered as gait dynamics for identification and verification. They also obtain static information of body based on Procrustes shape analysis of the change of moving silhouettes, which can be independently or combinative applied to improve the recognition. More recently, Boulgouris and Chi [26] separate human body into different components and combine the result obtained from different body parts to form a common distance metric as shown in Figure 3-7. Based on the study of each part's contribution to the recognition performance, the recognition rate is improved by using the most contributing parts.

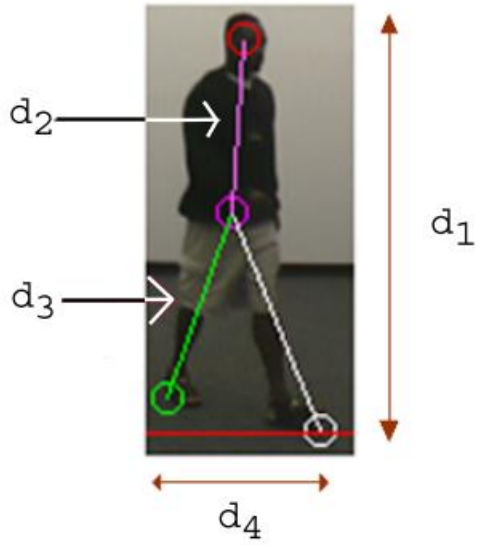


Figure 3-2: Static body parameters of [28]

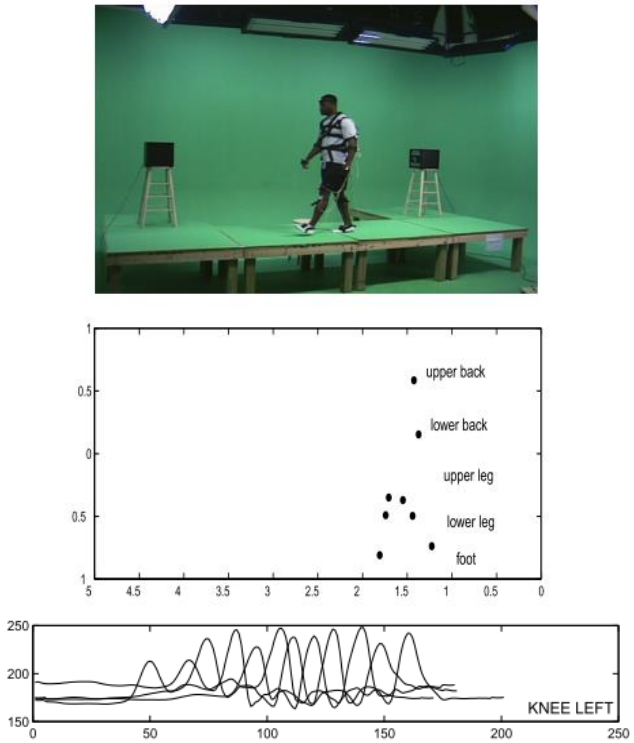


Figure 3-3: Motion capture system, markers views in the walking plane, and joint-angle trajectories of [38]

(a) Running (b) Walking

Figure 3-4: Leg motion extraction results of running and walking by temporal template matching in [25]

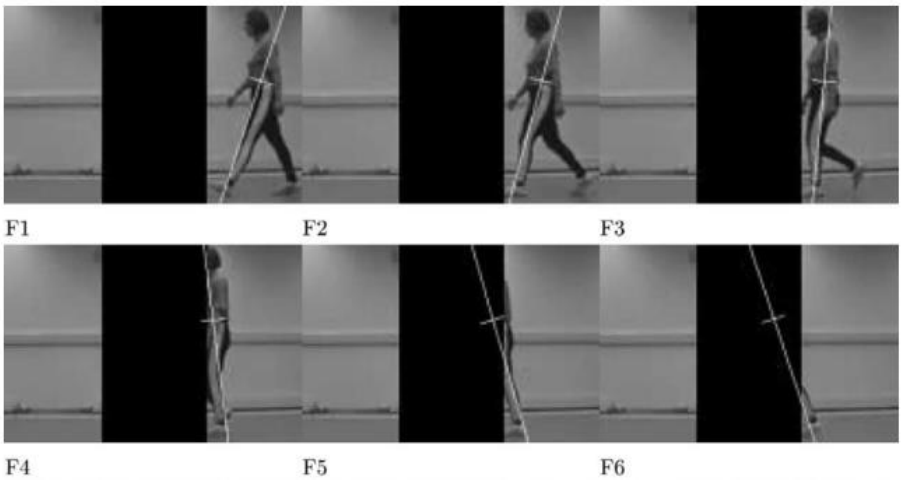


Figure 3-5: Extracted thigh model from an occluded sequence using VHT in [13]

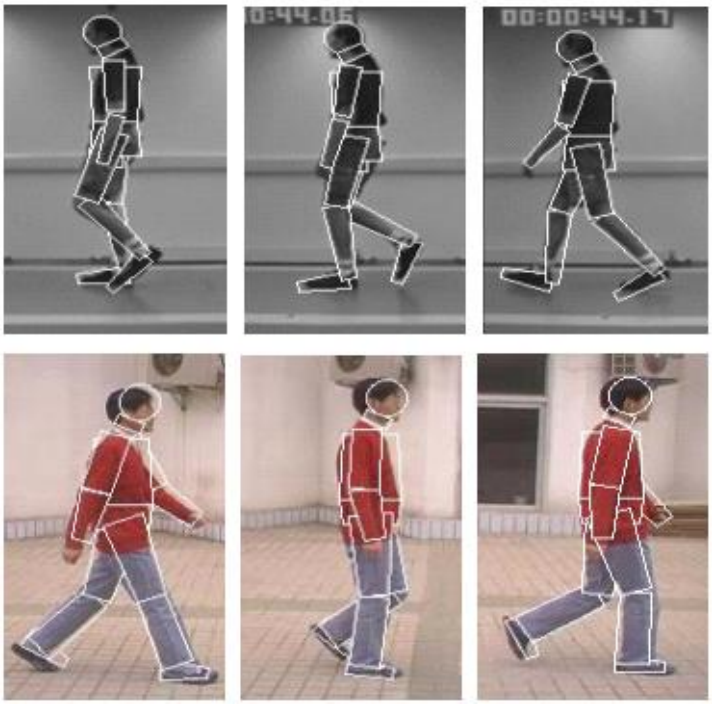


Figure 3-6: Parts of the tracking results in [29]

Figure 3-7: Sample of manually labelled silhouettes in [26]

2D Model based created in this research is using Nixon model [47]. Nixon models is based on skeleton and currently is the best 2D model for gait classification. The advantage of Nixon model is the ability to extract kinematics parameters as a feature. However, Nixon model is used

only in SOTON gait dataset and to be able to use in other dataset, some modification need to be done. Skeleton model in this research generated from skeleton morphological operation. Figure 3-8 below is the block diagram of the skeleton created to be able to handle CASIA gait dataset.

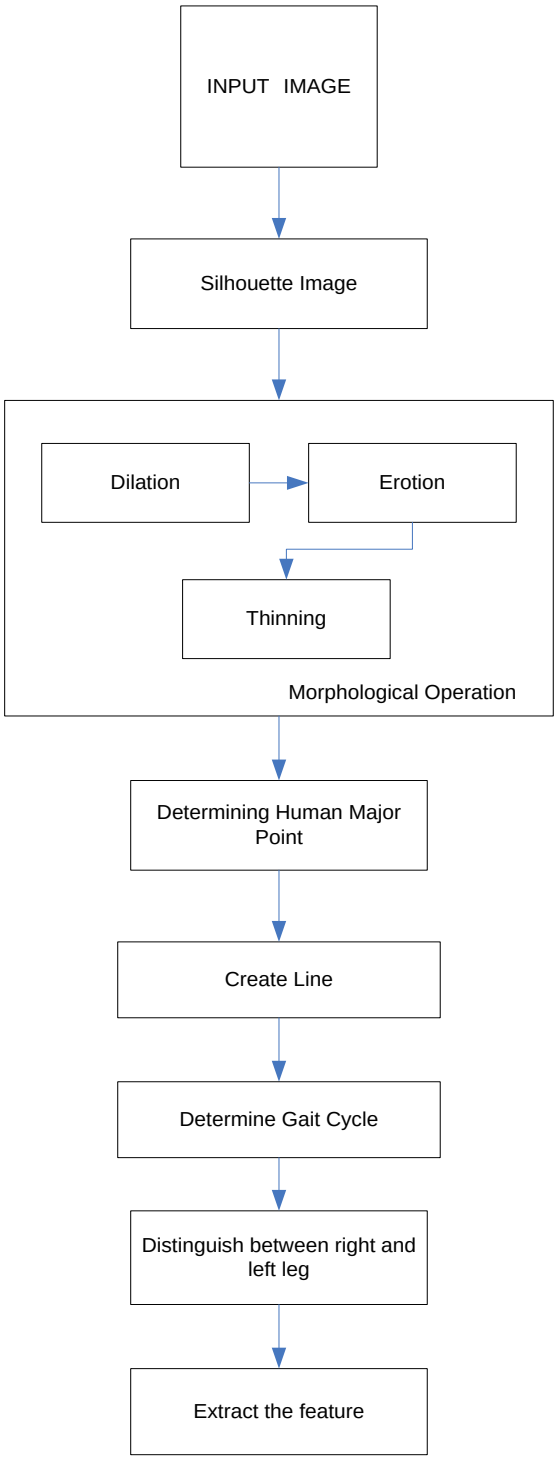


Figure 3-8. Block diagram of skeleton created for CASIA Gait dataset.

First step is do the object detection from input image using silhouette method as explained in 2.3. Silhouettes method used is also same method with silhouette in free model based shown in Figure 3-24. Next step is create skeleton using three binary morphological operations. Those are dilation, erosion, and thinning. For the dilation, use three times of *ones* structuring element while for the erosion, use six times of *ones* structuring element. Thinning algorithm is explained in 2.4.3. The skeleton model gave good result to develop human body model. However, the hand and the foot sometimes removed during the morphological operations. One also cannot use any information from arm and foot, thus they should be removed using some methods. This results is far beyond to be able to give such a features like joints angle, thus one have to fix those. This model also cannot distinguish between left and right legs. However, these results gives valuable information, for instance, the height of human body. Figure 3-9 below shows the skeleton model generated from the proposed method.

Next step is determining joint point and use them as the base for reconstructed skeleton. Wilfrid Taylor et al. measured the proportional length of human physical body. Figure 3-10 shows an average of human body parts length as compared to the body height [48]. As shown in Figure 3-10 hip is in the middle of the human body. The length of the hip – knee about 23.99% length of total human body height. The length of the knee – ankle about 16% length of total human body height. The length of hip – neck is 27.98% length of total human body height. The length of the neck – top of the head about 12.6% length of total human body height. Since hand and foot segment had removed, one will not consider using reference [48] for hand and foot measurement.

Figure 3-9: Example of skeleton image created

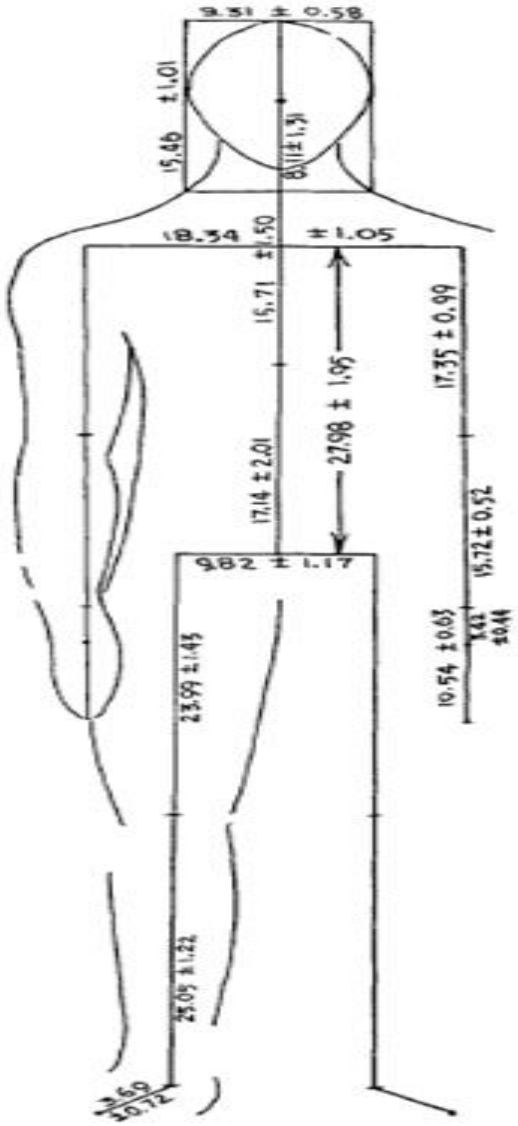


Figure 3-10: Human body measurement based on [48]

Using all the joint point determined, one can reconstruct the new smooth skeleton model. Figure 3-11 shows the proposed skeleton model reconstruction method. For human gait model, start with determine seven major points, top of the head, neck, hip, left knee, right knee, left ankle, and right ankle points. These points which will be used to reconstruct the model. Before determining the points, skeleton image have to be fixed. First step is removing hand and foot image. Using some nested loop and branch programming method one can remove them in all frames. The next step is measuring the height of the body. By measuring the average height for all image frames, body height can be measured accurately. After that, all the points can be determined based on the proportional length. Next after the major points determined, one can extract the lines. These lines will helps better analyze that the model is already good enough to be extracted the feature or not.

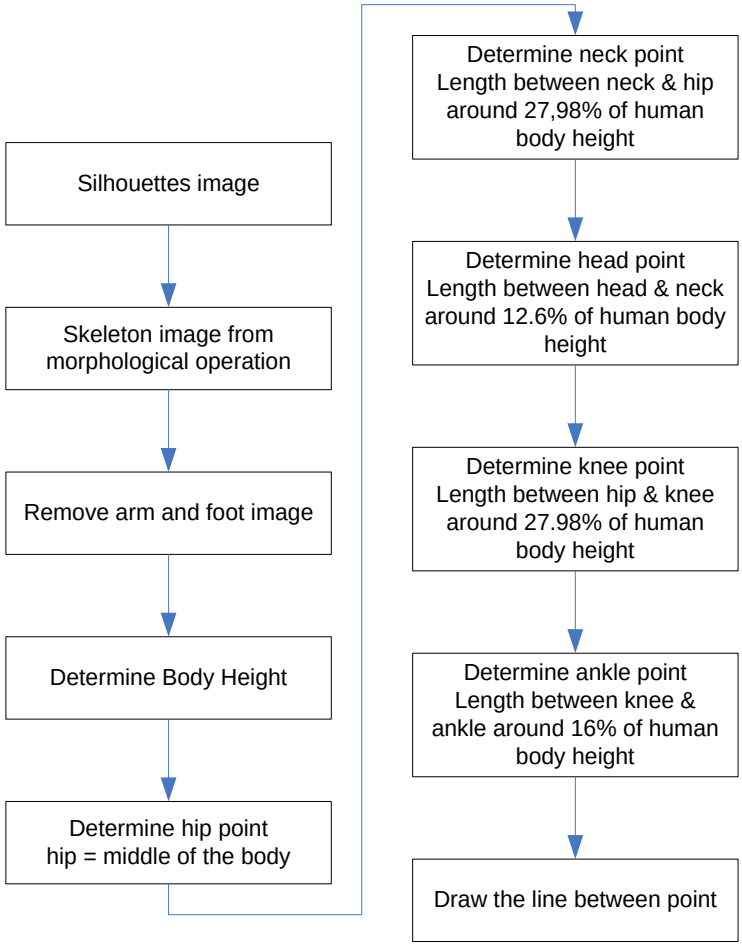


Figure 3-11: Skeleton Model Reconstruction Method

Figure 3-12 shows the results of skeleton model reconstructed.

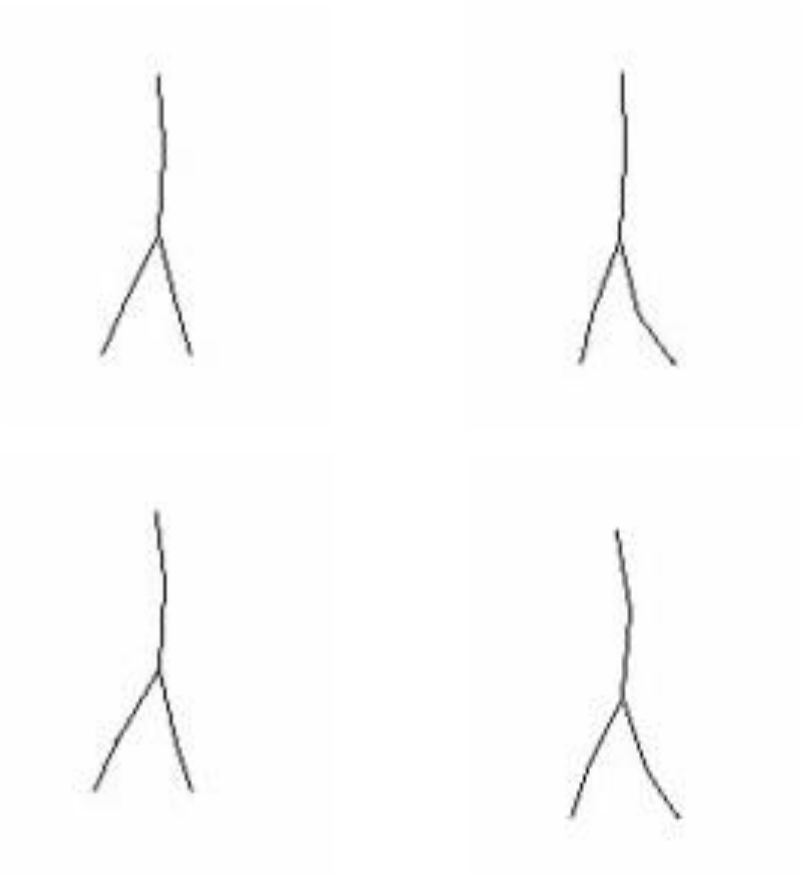


Figure 3-12: Results of skeleton model reconstructed

One thing to be considered in this 2D model is the ability to classify using only one single gait cycle. The gait cycle is used to describe the complex activity of walking. The gait cycle is described as the motion from initial placement of the supporting heel on the ground to when the same heel contacts the ground for a second time.

The gait cycle is divided into two phases, stance and swing. Stance is defined as the interval in which the foot is on the ground (about 60% of the gait cycle). Swing, on the other hands, is defined as the interval in which the foot is not in contact on the ground (about 40% of the gait cycle). Figure 3-13 shows the typical illustration of one gait cycle.

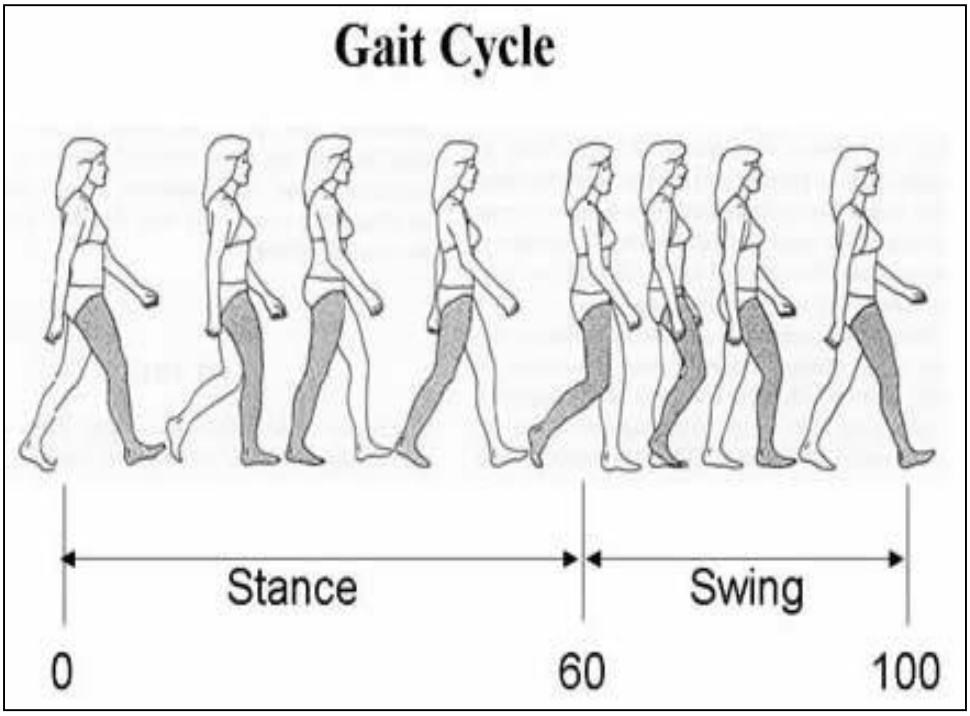


Figure 3-13: One gait cycle scenery

To get gait cycle in one image sequence, the proposed gait definition is implemented in the skeleton model. As shown in Figure 8, the gait started with stance status when the heel of stance foot reaching to the ground and is finished with heel reaching on the ground again in the next cycle. The distance between two heels are furthest in start and finish stance positions. This procedure will collect 100% of gait cycle, and remove the rest.

Figure 3-14 shows a storing the gait cycle procedure. For the CASIA gait database, in one video usually consists of around three gait cycles (two left and one right leg gait cycles or two right and one left leg gait cycles). To determine the start and the finish status of one gait cycle, one can use ankle points. If the positions of both ankle are furthest, then it will pretend as the start and the finish points. This 2D model also will distinguish between left and right leg gait because injured and diseased of human gait can just happen in one leg. Procedure for determine of one gait cycle:

1) Calculate the distance between ankle points f(x)
2) Normalize the distance: $y(x) = \left(\frac{f(x)-\min(f(x))}{\max(f(x))-\min(f(x))}\right)$
3) Calculate the first derivative

If the first derivative is bigger than threshold, then the distance is the start point of gait A, and the next point is the finish point of gait A (also the start point of gait B).

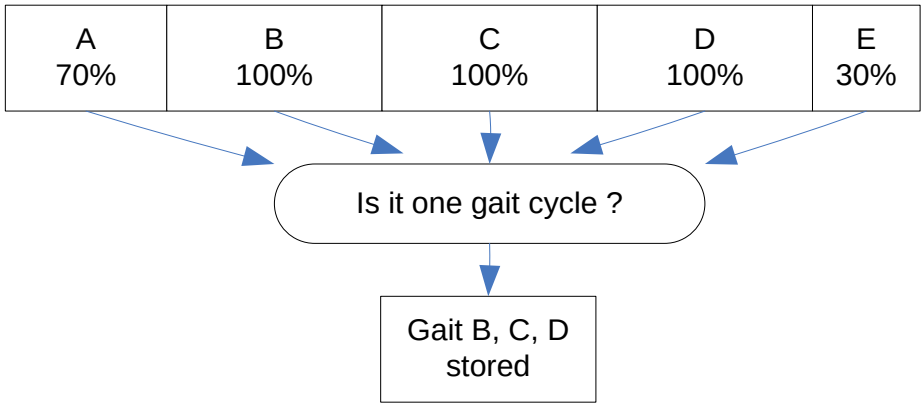


Figure 3-14: Store and remove procedure of gait cycle

. The next big problem is distinguished between left and right legs. Using one camera, the difference between both legs can be acquired. Figure 3-15 shows the procedure for distinguish two legs.

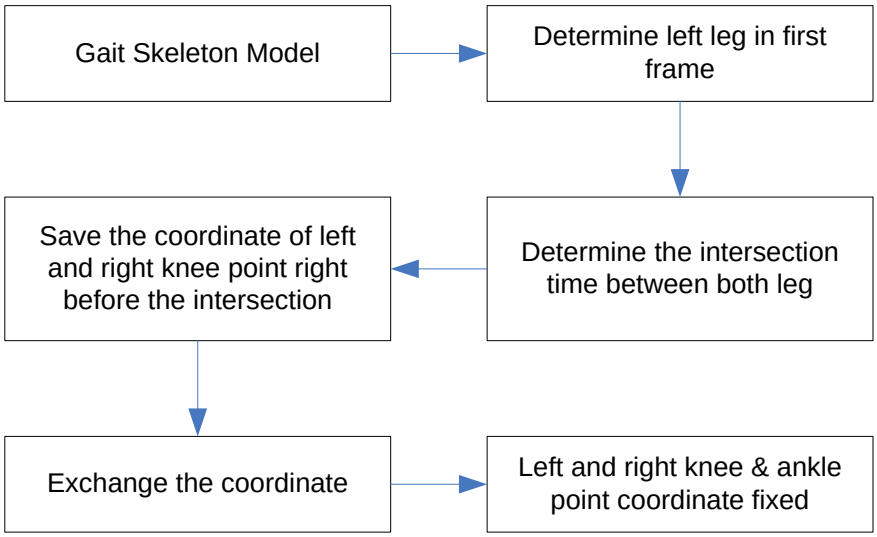


Figure 3-15: Procedure for distinguishing two legs

More difficult thing for distinguish is in the intersection time when both legs are in one shape state. First, we have to determine the left and the right legs in the first frame. The left leg mostly has bigger size than the right leg because left leg is closer to the camera. We will let the status of left and right legs same with the first frame until the intersection time. When the

intersection time comes, the left knee points have to become the right knee points and vice versa. The result of the skeleton is shown in figure 11.

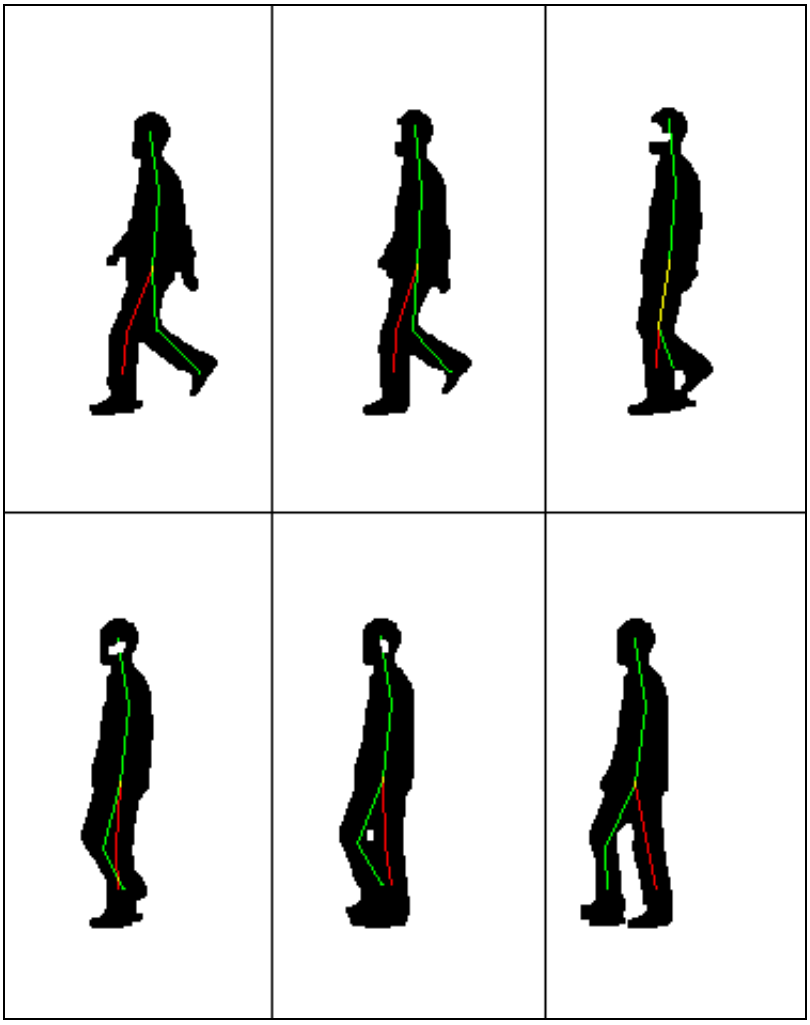


Figure 3-16: Gait skeleton model results. Red line shows the left leg and green line shows line from head until right leg

3.2 3D Model based

Around 25% of the CASIA gait dataset cannot be reconstructed using 2D model. This problem usually happen in the video with some light clothes, clothes which is similar color with the background, or any other color problems. This color problem giving inaccuracy effect to the segmentation or object detection system, thus giving low accuracy effect also in the skeleton reconstruction.

To solve the problem, the proposed research also try to do some approach in 3 dimensional point of view. The reconstructed skeleton is the 3D space skeleton based to improve coordinates accuracy and solve the problems in 2D skeleton model thus becoming more robust.

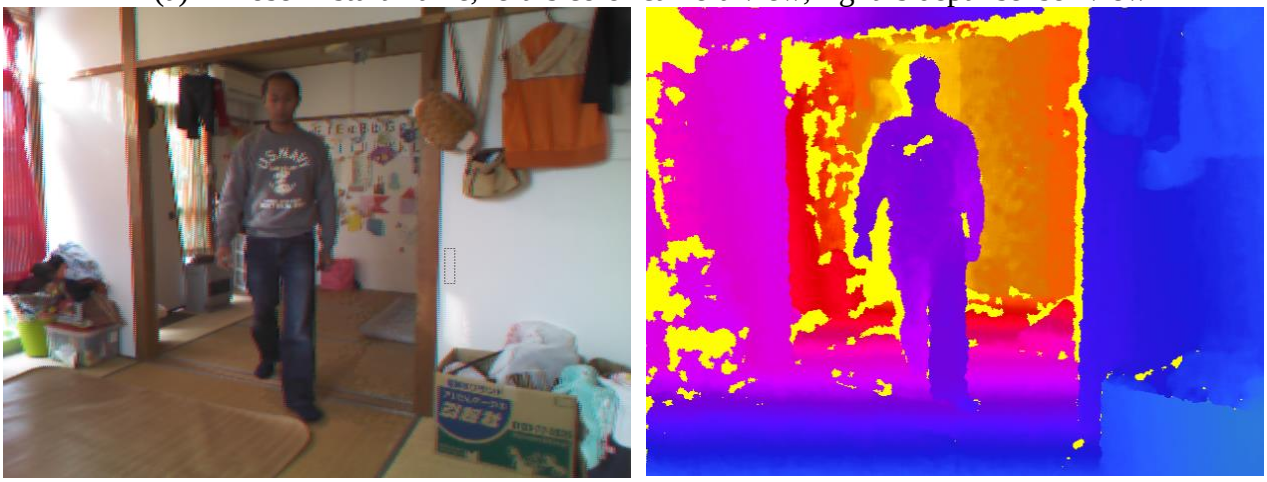
There are no such a 3D model dataset before, so that we have to create our own dataset. Our own created dataset also did not use gender as an object anymore rather using disable gait object due to some consideration:

- Gender gait need more acquisition time for collecting the subject which will be used as a dataset.
- Disable gait adding some class to create more classification problem complexity. In gender gait the class is only two, male and female. While in disable gait classification, we will decide to classify the disable quality to 5 class. This class addition aims to solve the human recognition using gait.

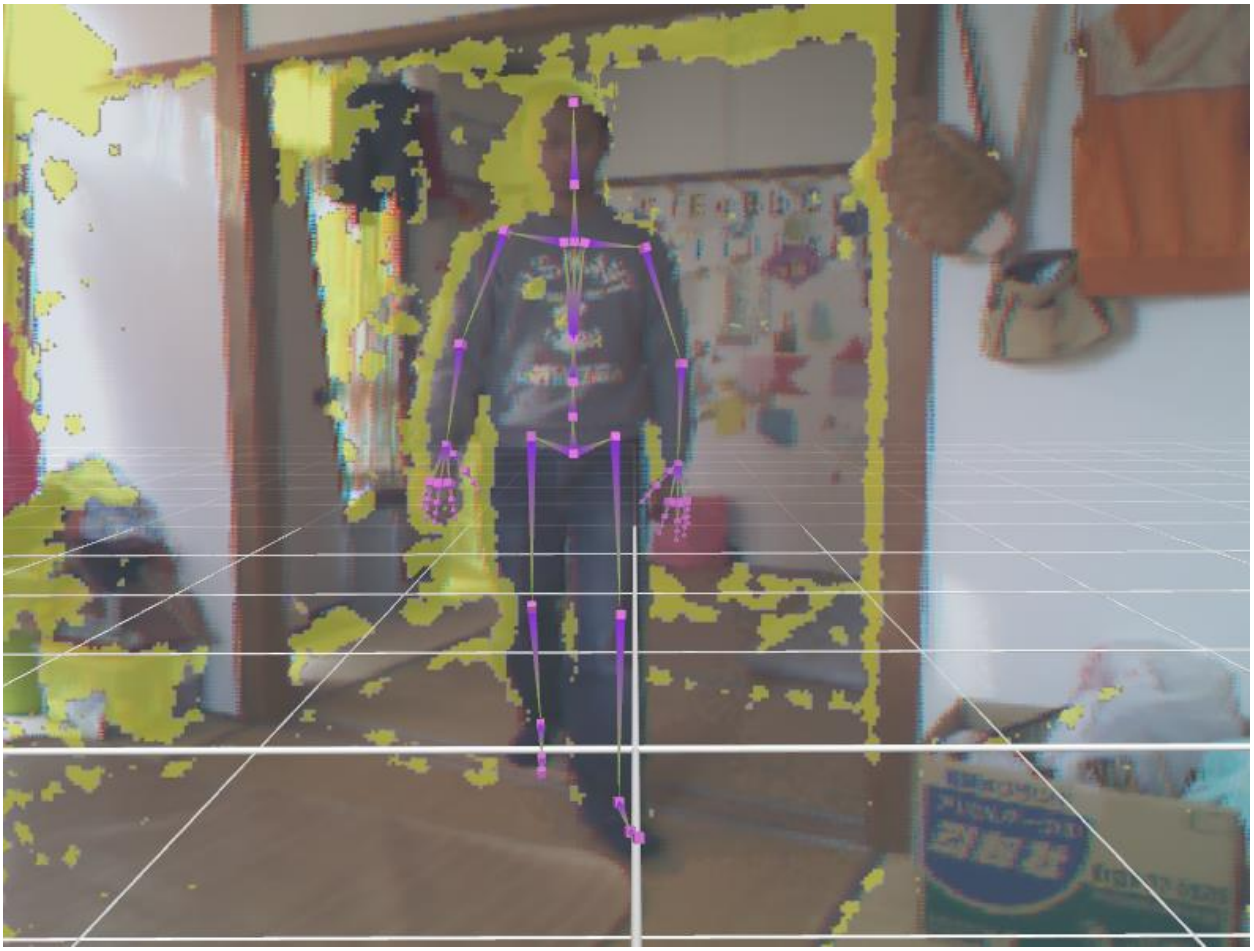
3D model creation in this research using Kinect Depth Sensor camera and IpiSoft Mocap Studio 2 application for motion object motion tracking and skeleton development. Below is the steps to get the 3D skeleton model:

1. Record subject video using 1 Kinect camera.
2. Determine the Region of Interest (ROI) from the recorded video.
3. Determine skeleton position in first frame when the subject did the T-Pose position.
4. Track skeleton position from second frame until last frame of ROI based on skeleton position in the previous frame.
5. Do the jittery removal in each frame of skeleton result from steps 4
6. Import skeleton to BVH file format (Biovision Hierarchy). BVH is animation character standard file format which is used widely by animation, model, and 3D imaging software developer.
7. Read BVH file format and do the kinematics feature extraction.
8. Do the classification using Support Vector Machine (SVM).

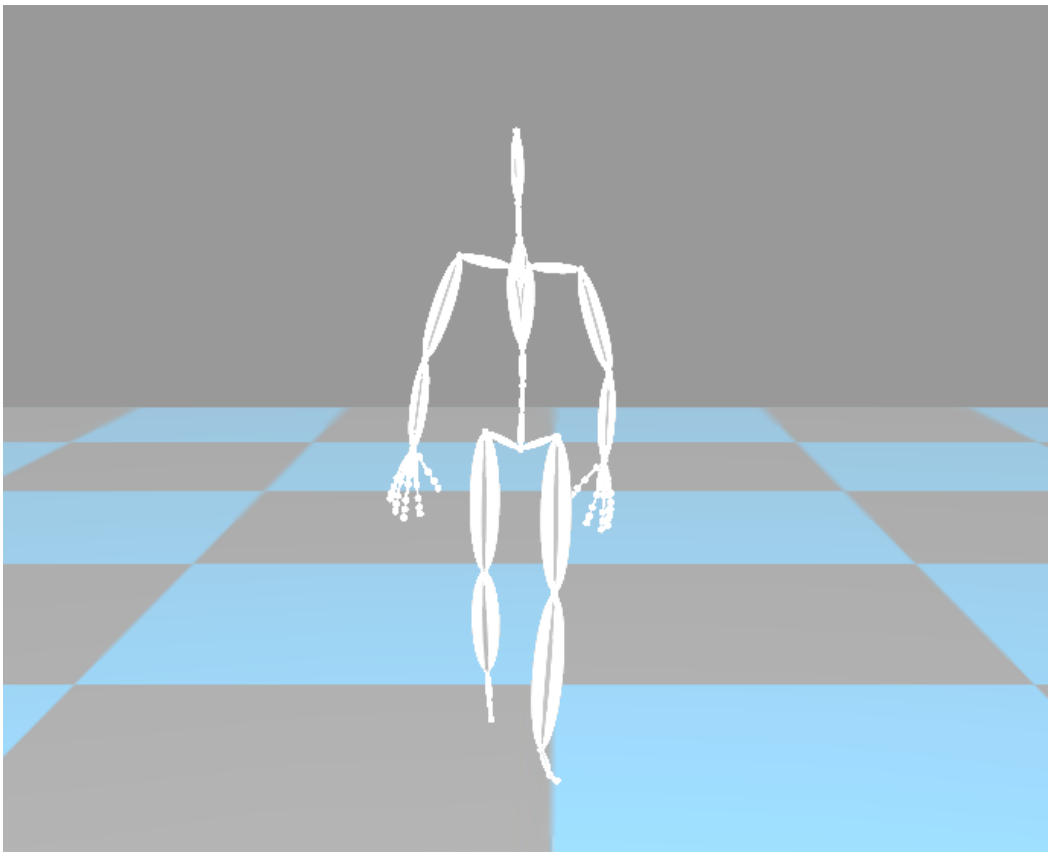
(a) T-Pose in start frame, left is color camera view, right is depth sensor view



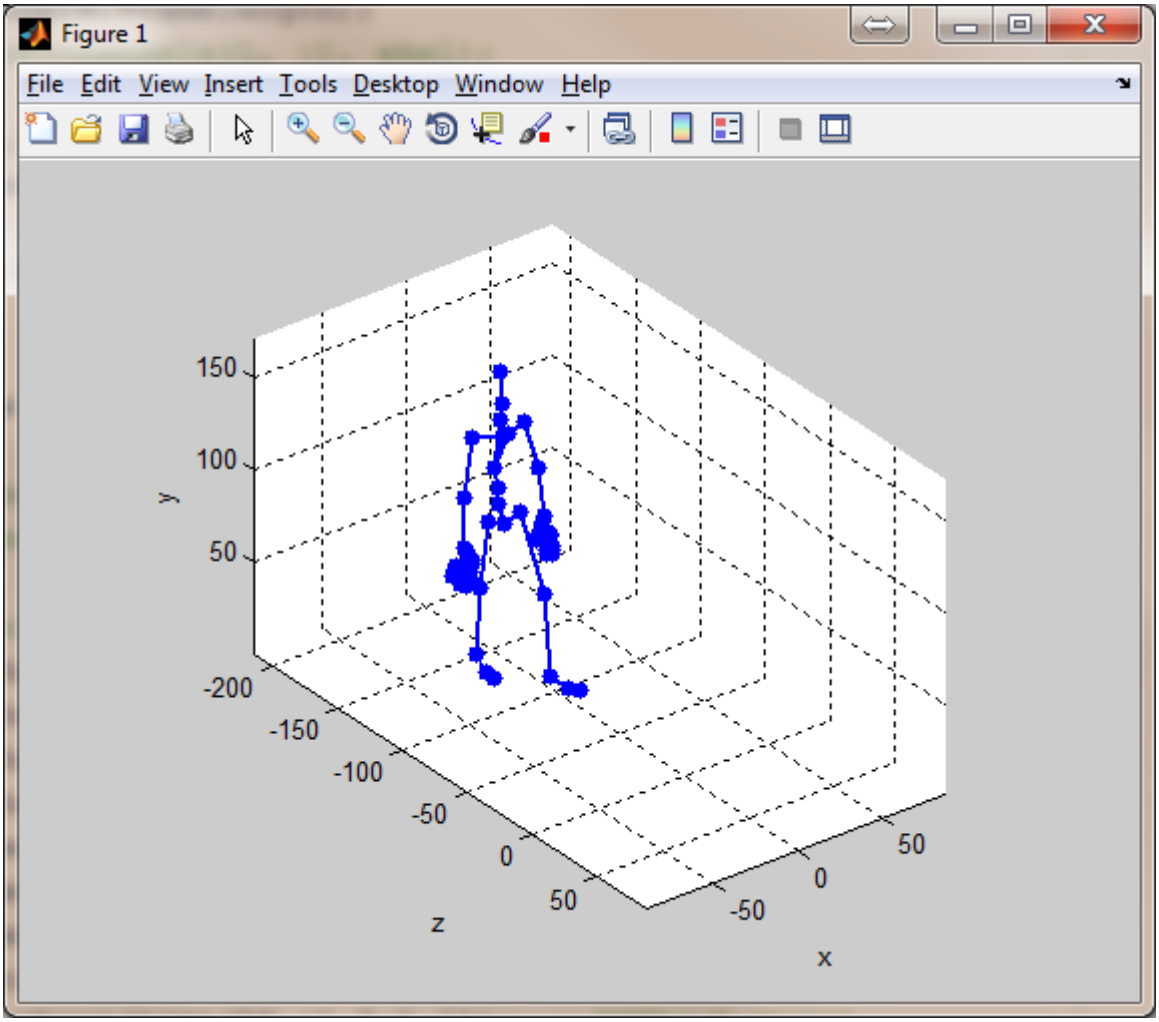
(b) Gait frame scene, left is color camera view, right is depth sensor view



(c) Skeleton tracking scene



(d) BVH file scene



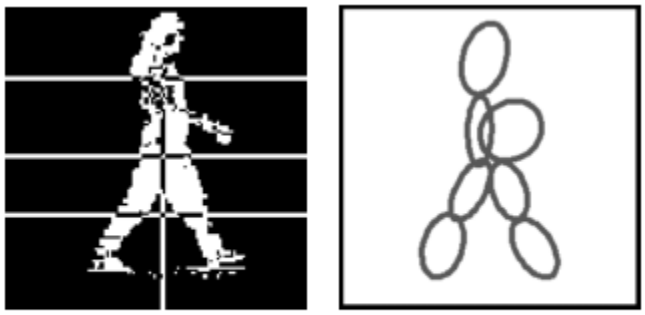
(e) BVH file read in MATLAB

Figure 3-17: 3D model proposed

3.3 Free-Model based

Model-free approaches focus on either shapes of silhouettes or the whole motion of human bodies, rather than modeling the whole human body or any parts of body. Model-free approaches are insensitive to the quality of silhouettes and have the advantage of low computational costs comparing to model-based approaches. However, they are usually not robust to viewpoints and scale. Lee and Grimson [11] divide gait silhouette into seven regions and use ellipses to fit each region as shown in . They apply Fourier Transform on the temporal signals from these ellipses and extract the magnitude components and phase components for classification. Huang and Wang [30] propose improvisation of [11] by combining the feature vector in multi-view gait

sequences as shown in Figure 3-19 and Figure 3-20. Those multi-view using 0 degree, 90 degree, and 180 degree. They divide the silhouettes into 7 (for 90 degree, i.e. front parallel view) or 5 (for 0 and 180 degree, i.e. front view and back view) parts, and then fit ellipses to each of the regions. Next, the features are extracted from each sequence by computing the ellipse parameters. For each view angle, every subject's features are normalized and combined as a feature vector. Sum rule and SVM are applied to fuse the similarity measures from 0°, 90°, and 180°. Yu et al. [49] use silhouette image contour points and extract the feature using Fourier descriptors and key Fourier descriptors and compare them in classification and achieve the conclusion that the key Fourier descriptors of human contours surpass the Fourier descriptors as shown in Figure 3-21 and Figure 3-22. They use 512, 256, 128, 64, 32, 16, 8, and 4 points respectively to express a silhouette to perform associated classification. Kale et al. [50] use the width of the outer contour of silhouettes to encode the information of silhouettes. The width is defined as the horizontal distance between the leftmost pixel and the rightmost pixel of the contour. The width of the outer contour may be unreliable due to the poor quality of silhouettes. However, the silhouette itself as features may be more suitable for low quality and low resolution data. Later, Kale et al. [51] combine the entire silhouettes and the width of outer contour silhouette as gait features.



(a) (b)

Figure 3-18: The silhouette of a foreground walking person is divided into 7 regions, and ellipses are fitted to each region in [11]

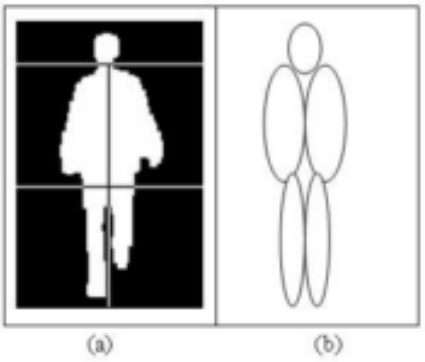


Figure 3-19: Example of the 0 and 180 degree silhouette which is divided into 5 regions, and five ellipses are fitted to these regions in [30]

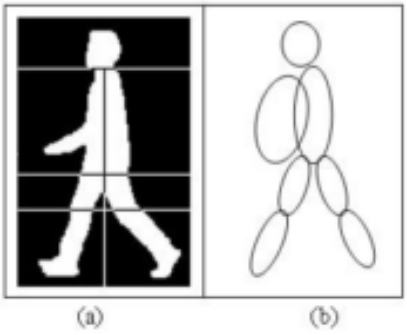
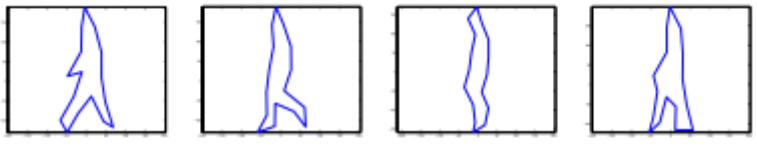


Figure 3-20: Example of the 90 degree silhouettes which is divided into 7 regions, and seven ellipses are fitted to these regions in [30]

(a) A human silhouette (b) A contour in the complex plane

Figure 3-21: contour image representation in [49]



(a) Frame 1 (b) Frame 5 (c) Frame 9 (d) Frame 13

Figure 3-22: silhouettes contour extracted using 16 points in [49]

Free-model based have advantage compare to model based. The advantages are: easy to develop and low on computation. Figure 3-23 below is the basic diagram to create free-model based.

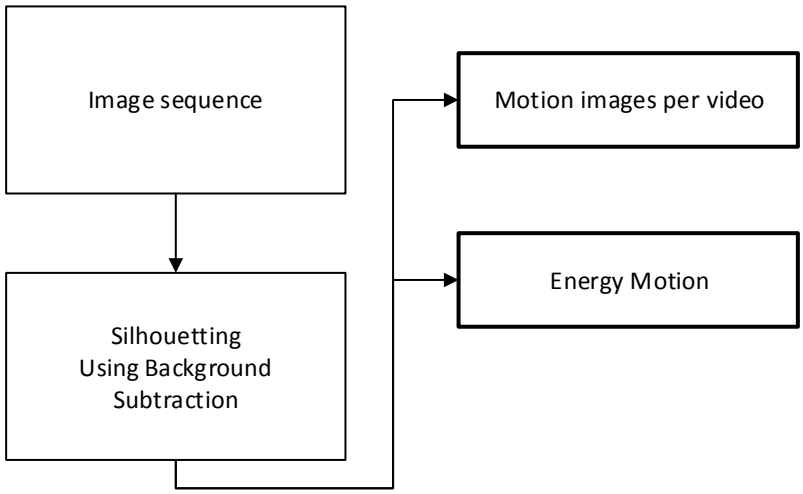


Figure 3-23: Block diagram of Free-model based created

As seen in the Figure 3-23, silhouettes is taking apart as an object detection. Silhouettes references is shown in 2.3. However, background subtraction method is very sensitive to external noise such as lighting environment, object clothes, and shadow. To overcome the problem, this research also proposed silhouette method fit with CASIA gait dataset needs. The silhouettes proposed is working in an RGB image space, and not in grayscale space. Figure 3-24 below is the block diagram of silhouette method proposed. Otsu's threshold very useful to create optimum object detected. Unfortunately, the shadow also detected. To remove the shadow, create the SAD (Sum of Absolute Difference) between RGB color space and manual threshold. The results from Otsu's and manual threshold then multiplied them. Last step is convert the color silhouettes to binary silhouettes.

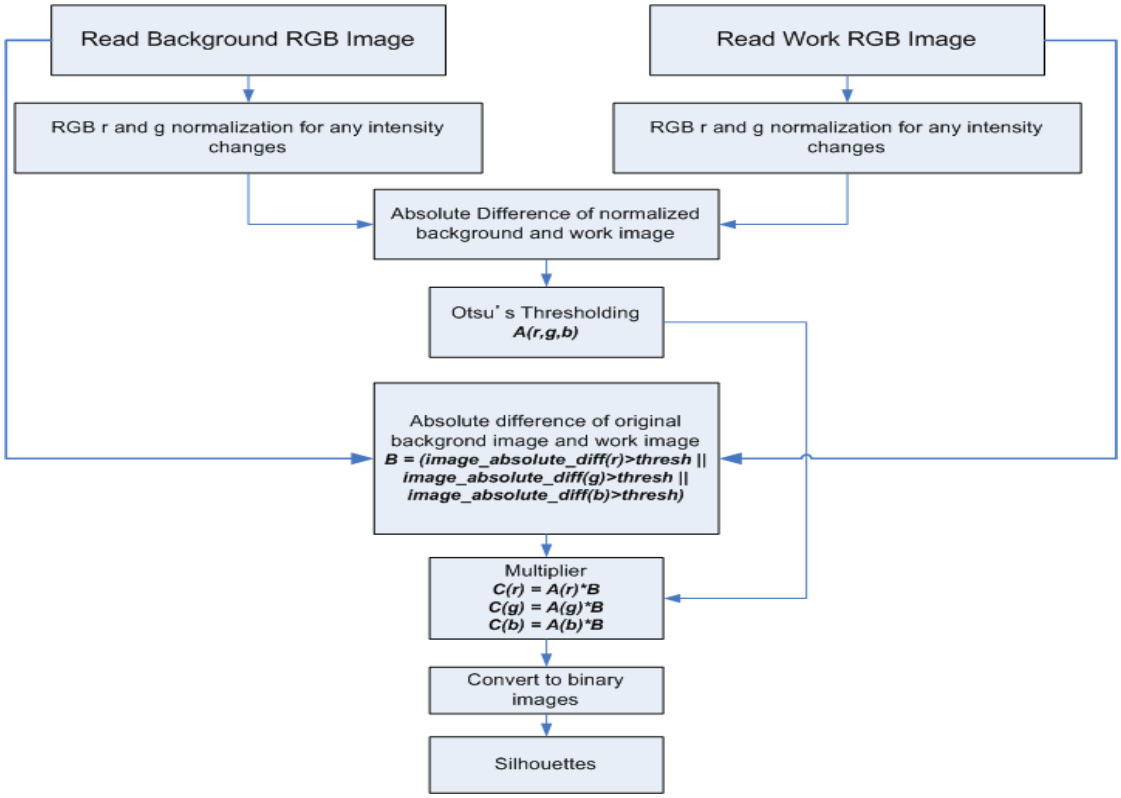


Figure 3-24: Block diagram of silhouettes method

Figure 3-25: Original image of CASIA Gait Dataset

Figure 3-26: Silhouette result from the proposed method

The first free-model based used is motion images per video. To extract motion only to the images, SAD is used between current frame and the previous frame. Below is the result of the motion feature per frame. To reduce the dimension of the feature, the proposed method combined all the motion frames in one frame. Below is the motion image sequence used in extracted feature.

Figure 3-27: Silhouettes created

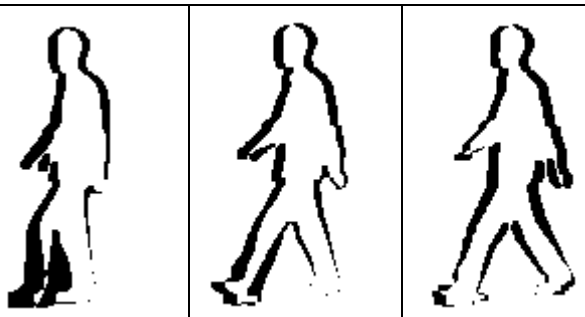


Figure 3-28: image motion created

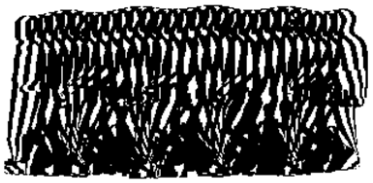


Figure 3-29 image motion sequence created

The second free-model based proposed called Gait Energy Motion (GEM). GEM represents all the silhouettes in one video file, into one new energy image. The dimensionality of the data is reduced using the energy. Using GEM will also remove the noise smoothly.

Figure 3-30 below is the example of Gait Energy Motion (GEM) for male and female.

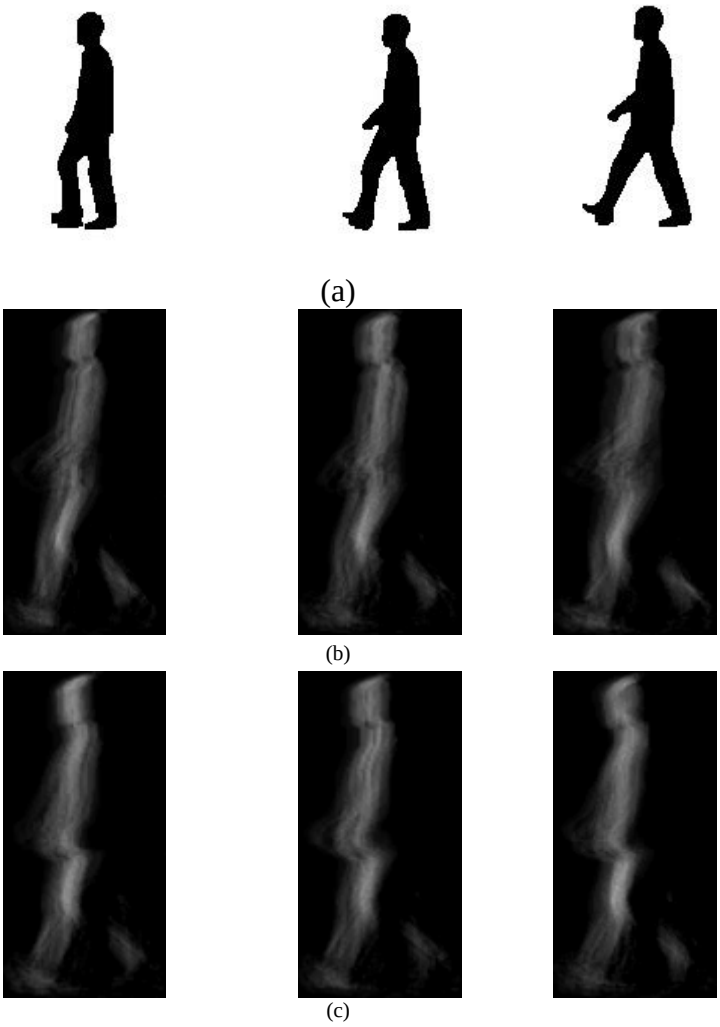


Figure 3-30: (a) Silhouettes examples, (b) GEM examples of male, (c) GEM examples of female

4. Experiments & Results

4.1 Kinematics, and Statistical Feature in Skeleton 2D Model

This chapter will explain about feature extraction in 2D skeleton model. Detail explanation about 2D skeleton model creation can be seen in 3.1. We take 22 kinematics and statistical feature from skeleton model created. Table 4-1 below shows all the feature extracted.

Type	Features
Kinematics Parameters	Total_frame, hip_left_1, hip_left_2, hip_left_3, hip_right_1, hip_right_2, hip_right_3, $neck_1$, $neck_2$, $neck_3$
Statistical Parameters	$mean(\lvert knee\_angle \rvert)$, $\sigma(\lvert knee\_angle \rvert)$, $mean(\lvert neck\_angle \rvert)$, $\sigma(\lvert neck\_angle \rvert)$, $mean(\lvert hip_{left}\_angle \rvert)$, $\sigma(\lvert hip_{left}\_angle \rvert)$, $mean(\lvert hip_{right}\_angle \rvert)$, $\sigma(\lvert hip_{right}\_angle \rvert)$, $mean(\lvert stance\_max \rvert)$, $mean(\lvert stance\_min \rvert)$, $mean(\lvert swing\_max \rvert)$, $mean(\lvert swing\_min \rvert)$

Table 4-1: 22 featured extracted from 2D Skeleton Model

Total_frame feature is taken from one video file of skeleton model. Mostly all of the feature extracted using angle feature. To calculate the angle, we can use simple trigonometry formula as shown in Figure 4-1 below and the equation [24], [25], [26], and [27] below.

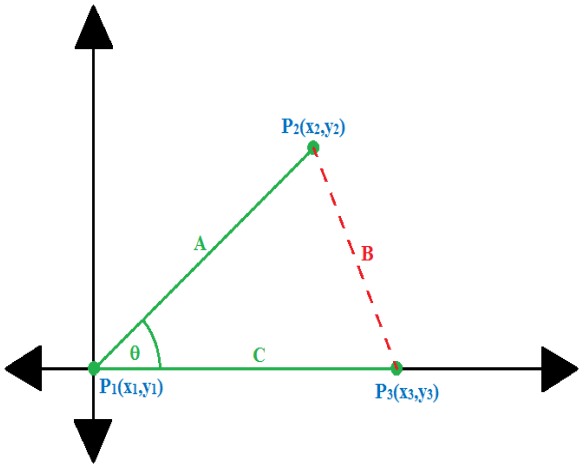


Figure 4-1: Illustration to calculate angle between two lines using points coordinates

$$A = P_2 - P_1 = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \quad \text{.....[24]}$$

$$B = P_3 - P_2 = \sqrt{(x_3 - x_2)^2 + (y_3 - y_2)^2} \quad \text{.....[25]}$$

$$C = P_3 - P_1 = \sqrt{(x_3 - x_1)^2 + (y_3 - y_1)^2} \quad \text{.....[26]}$$

Thus the angle of θ is represented as follows:

$$\theta = \cos^{-1}\left(\frac{B^2 - A^2 - C^2}{2AC}\right) \quad \text{.....[27]}$$

Knee angles for every gait state are shown in Figure 4-2. The estimated knee angles are shown in Figure 4-3 and Figure 4-4. All of those estimated results are already smoothed through local regression using weighted linear least squares method with the first order of the polynomial model. Angle estimation accuracy is defined as deviation of the estimated angle from the corresponding smoothed curved. These deviations are shown in Figure 14, and 15. It is approximately 2.5 degree. This angle estimation accuracy would be good enough for gender classification.

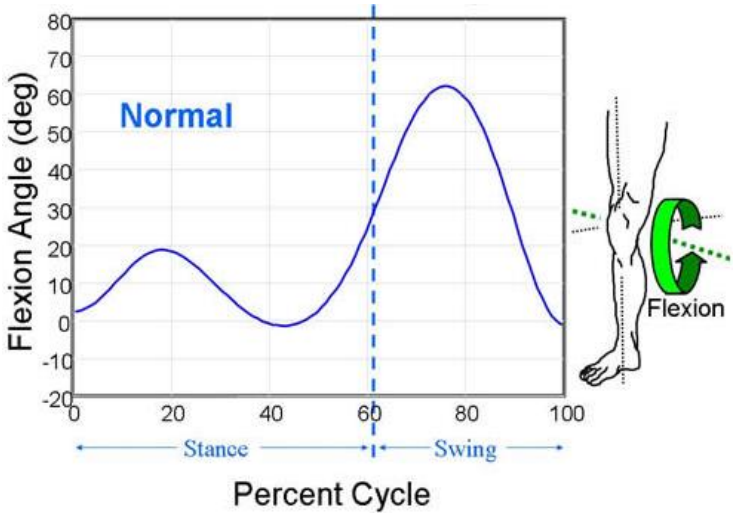


Figure 4-2: Gait cycle of normal person[52]

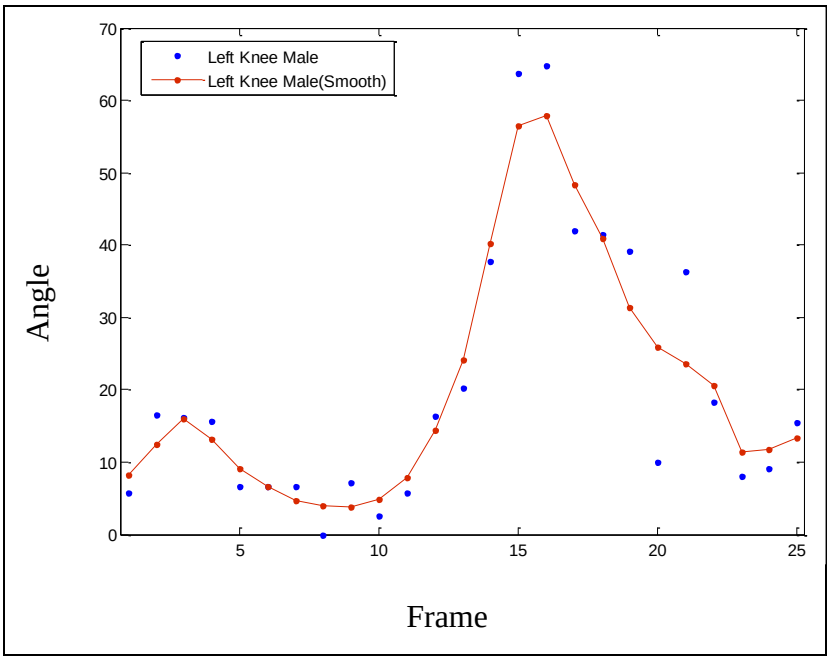


Figure 4-3: Male knee angle results from the skeleton model proposed

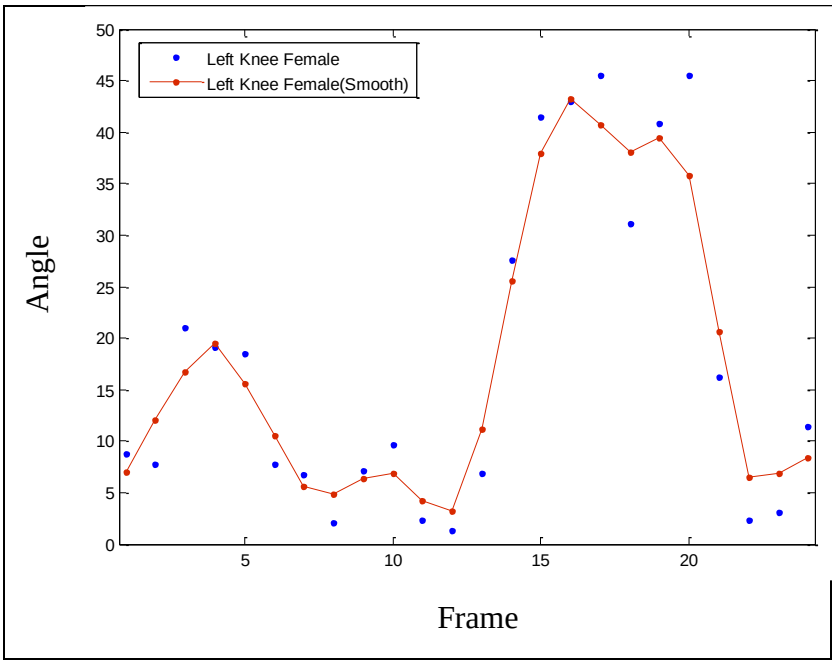


Figure 4-4: Female knee angle results from the skeleton model proposed

All of those estimated results are already smoothed through local regression using weighted linear least squares method with the first order of the polynomial model. Figure 4-5 and Figure 4-6 shows all the estimated results derived from one single step; head-neck-hip angle, neck-hip-left foot angle, neck-hip-right foot angle, and hip-active foot-ankle angles. The above mention features are selected for classification. For the knee angle, we decided to use six features for every single gait steps; maximum angle of stances, minimum angle of stances, maximum angle of swings, minimum angle of swings, knee angle average, and knee angle standard deviation.

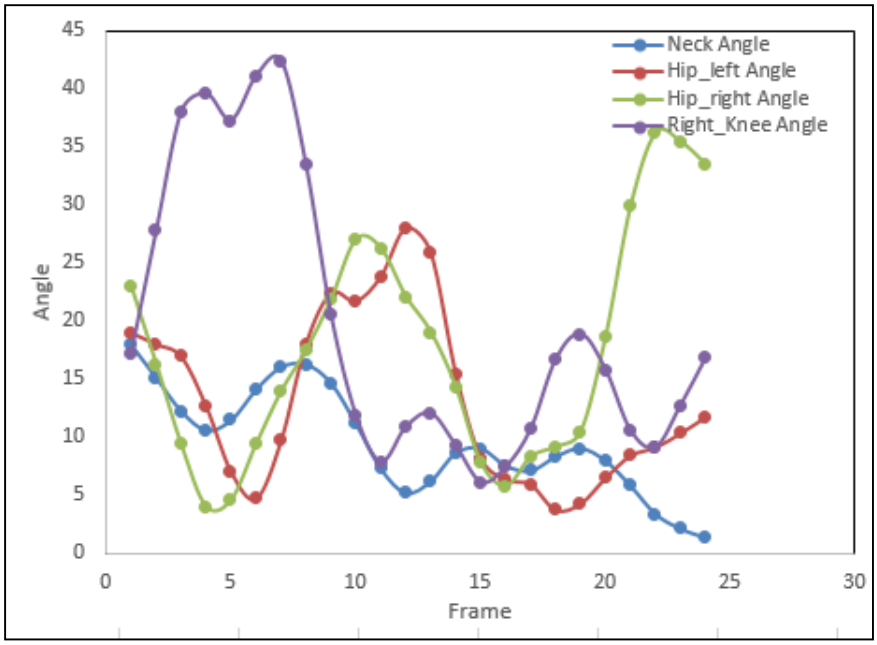


Figure 4-5: Male neck, hip-passive knee, hip-active knee, passive knee angle results from the skeleton model proposed

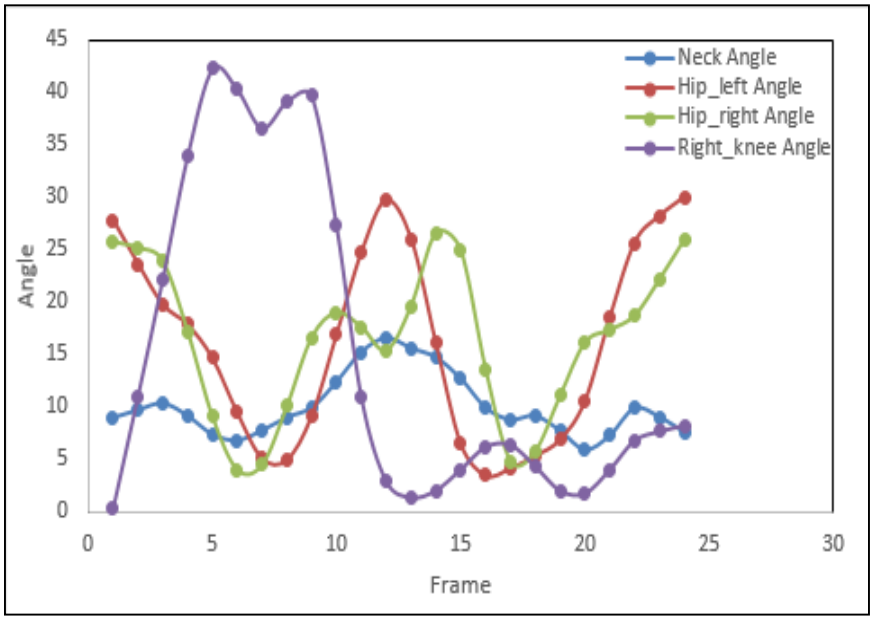


Figure 4-6: Female neck, hip-passive knee, hip-active knee, passive knee angle results from the skeleton model proposed

Figure 4-7 shows the feature from knee angles. We divided this estimated results into four segments for the neck angle, hip-left foot angle, and hip-right foot angle.

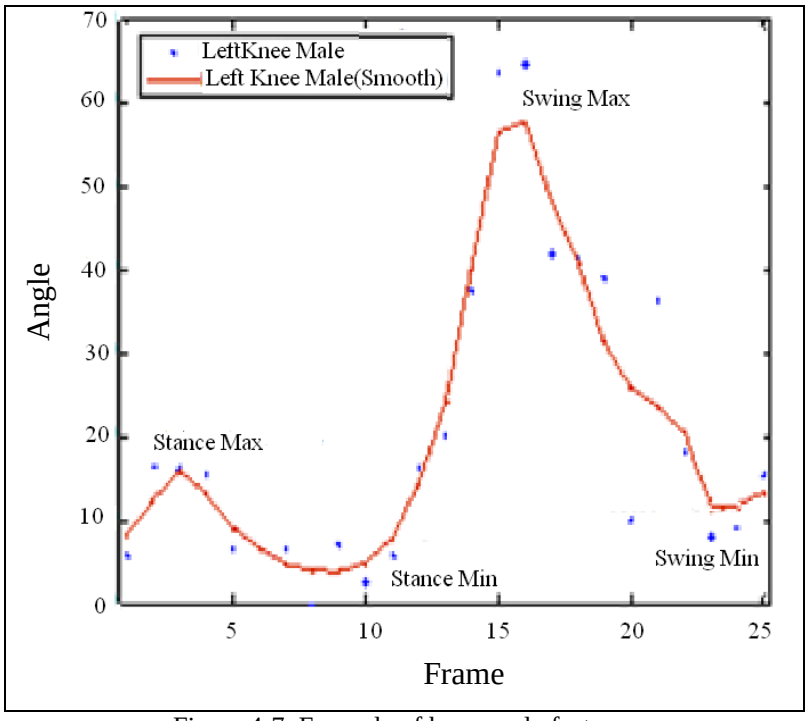


Figure 4-7: Example of knee angle features

The classification accuracy of gender classification based on C4.5 decision tree [53] with 22 of extracted features is around 70%. The classification also suggested that better result can be obtained by using only five features; mean($|hip\_left|$) , mean($|hip\_right|$) , mean($|stance\_max|$), Total_frame, and mean($|knee|$). Therefore, these five features are used for gender classification based on well-known Support Vector Machine (SVM). The training performance test using 10-fold cross validation index and the classification accuracy is 81.67%.

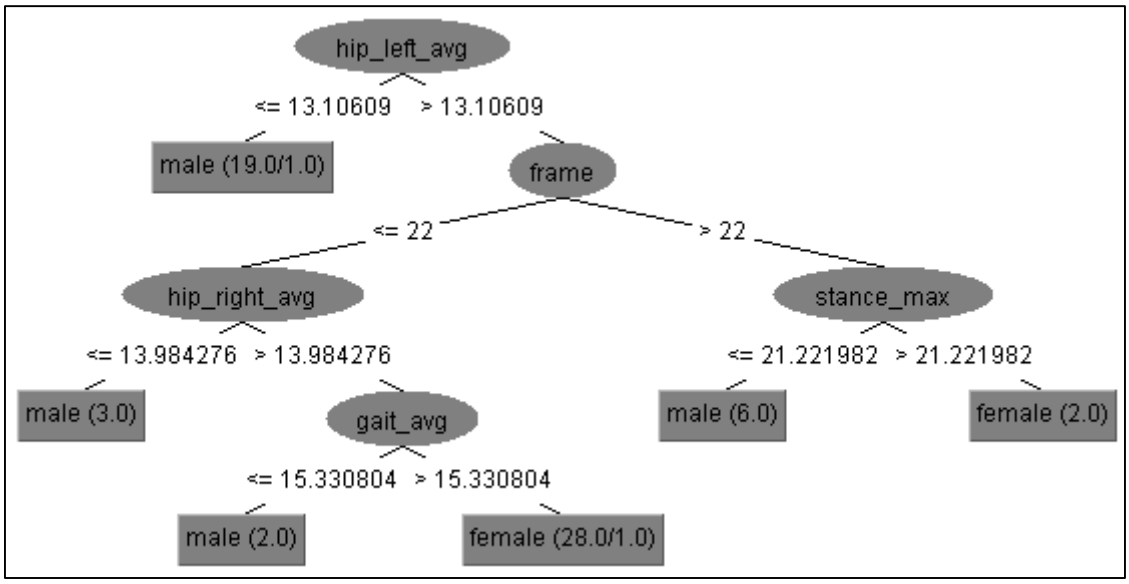


Figure 4-8: Decision Tree generated from C4.5

The previously proposed method [47] implement in personal identification (it is referred to conventional method hereafter) uses 10 features such as *amplitude, body height, cycle time, stride length, speed, mean*($|\theta_{neck}|$), $\sigma(|\theta_{back}|)$, $\sigma(|\theta_{knee}|)$, γ_{knee}, and μ_x as a comparison to the classification performance of the proposed method. γ_{knee} is the correlation between θ_{knee1} and θ_{knee2}. We implement these features to gender classification based on the same classification method (SVM). The classification accuracy is 65%. Therefore, the proposed method is superior to the conventional method by 16.7%.

F-statistics used to analyze the gait difference between classes. F-statistic is a measure to evaluate different feature discriminative capability. The greater the F-statistic values will give better discriminative capability. The F-statistic is calculated as follows:

$$F = \frac{\frac{1}{c-1}\sum_{i=1}^{c} n_i(\bar{x}_i-\bar{x})^2}{\frac{1}{n-c}\sum_{i=1}^{c}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_i)^2} \quad [28]$$

Where $x_{ij}$is the j^{th} sample of class *i*, *c* is the number of classes, n_i is the sample number of class *i*, $n = \sum_{i=1}^{c} n_i$, $\bar{x}_i$ is the mean of samples in class *i*, and $\bar{x}$ is the mean of $\bar{x}_i$.

Results from F-statistics recommend 5 best discriminative feature which are: mean($|stance\_max|$), σ($|knee|$), mean($|hip\_left|$), mean($|hip\_right|$), and σ($|hip\_right|$), since C4.5 also only suggesting 5 features. The calculated F value for each feature is shown in .

Type	No	Features	F-value
Kinematics Parameters	1	Total frame	0.0939
	2	hip_left_1	0.0666
	3	hip_left_2	0.0761
	4	hip_left_3	0.0897
	5	hip_right_1	0.0971
	6	hip_right_2	0.0387
	7	hip_right_3	0.118
	8	$neck_1$	0.132
	9	$neck_2$	0.104
	10	$neck_3$	0.105

Statistical Parameters	11	mean(|$knee$|)	0.104
	12	σ(|$knee$|)	**0.216**
	13	mean(|$neck$|)	0.149
	14	σ(|$neck$|)	0.0488
	15	mean$\left(\left\|hip_{left}\right\|\right)$	**0.614**
	16	σ$\left(\left\|hip_{left}\right\|\right)$	0.128
	17	mean$\left(\left\|hip_{right}\right\|\right)$	**0.426**
	18	σ$\left(\left\|hip_{right}\right\|\right)$	**0.151**
	19	mean(|$stance\_max$|)	**0.152**
	20	mean(|$stance\_min$|)	0.0504
	21	mean(|$swing\_max$|)	0.00367
	22	mean(|$swing\_min$|)	0.0446

Table 4-2: The calculated F value for each feature

The results from f-statistics feature using SVM is shown in Table 4-4. Table 4-4 also shows the classification performances of the conventional and the proposed methods. We use WEKA [54] as a tool for filtering and classifying the features. Table 4-3 below shows the confusion matrix from SVM classifier of all the method analyzed.

	Predicted Class							
	Previous Method		Proposed 22 Feature		F-stats Proposed 5 Feature		C4.5 Proposed 5 Feature	
	Male	Female	Male	Female	Male	Female	Male	Female
Male	18	12	14	16	18	12	21	9
Female	9	21	2	28	1	29	2	28

Table 4-3: Confusion Matrix from SVM Classifier

To calculate the precision and recall, for example C4.5 5 feature in male perspective. Define the true positive (TP), false positive (FP), and true negative (TN). From the table above, the value for TP is 21, FP is 2, and TN is 9. From those value, the precision and recall value is:

Precision = TP / (TP+FP) = 21/(21+2) = 91.3 %

Recall = TP / (TP+TN) = 21/(21+9) = 70 %

Table 4-4 shows the precision and recall for each class of the conventional and the proposed methods, respectively.

Method	Precision		Recall	
	Male	Female	Male	Female
Previous Method	66.7%	63.6%	60%	70%
Previous Method (5 features using F-Statistic)	59.4%	60.7%	63.3%	56.7%
Proposed Method (22 Features)	87.5%	63.6%	46.7%	93.3%
Proposed Method (5Feature using F-Statistic)	94.7%	70.7%	60%	96.7%
Proposed Method (5Features using C4.5)	91.3%	75.7%	70%	93.3%

Table 4-4: Precision and Recall for gender classification of the previous and proposed method

CCR is achieved from recall value average. CCR from C4.5 is (70%+93.3%)/2 = 81.65

Mean of the processing time required for each process is as follows,

1. Silhouette detection: 418.1 ms
2. Estimation of major joint angle of human body: 163.4 ms
3. Approximation with lines: 37.9 ms
4. Determination of gait cycle: 0 ms
5. Discrimination between right and left legs: 401.1 ms
6. Joint angle calculation: 34 ms
7. Classification time for all data: 70 ms

By adding all the processing time mentioned, we got total processing time mean about 1054.5 ms.

One of the highest feature discriminant is hip_left_avg. By discretize filtering the data we can see that male have smaller hip angle average compare to female hip angle average as shown in Figure 4-9 below. Male is represented with blue color and female is represented with red one.

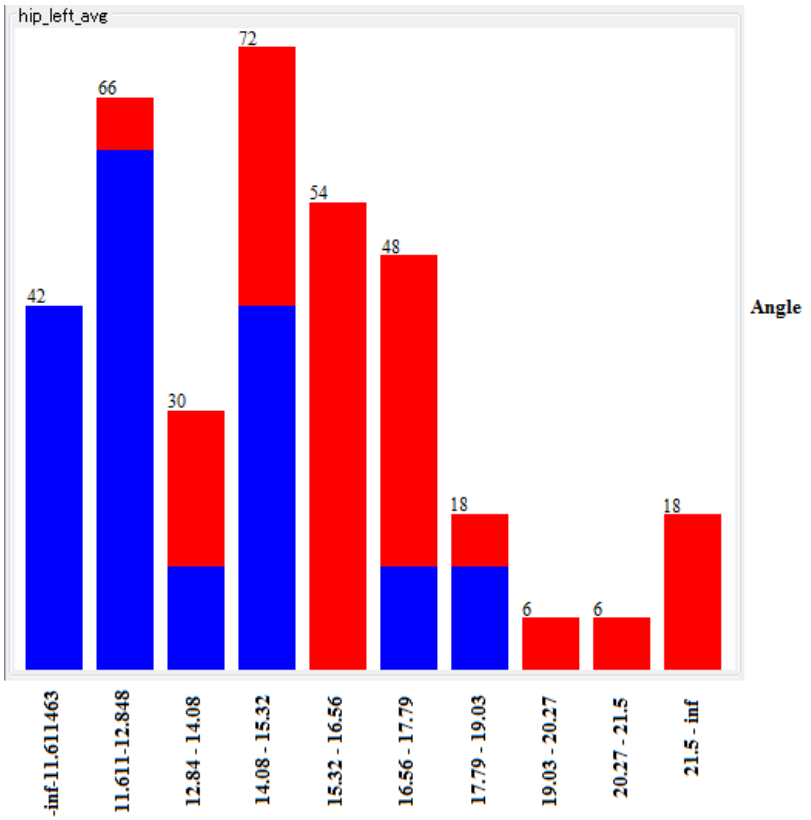


Figure 4-9: hip left average difference

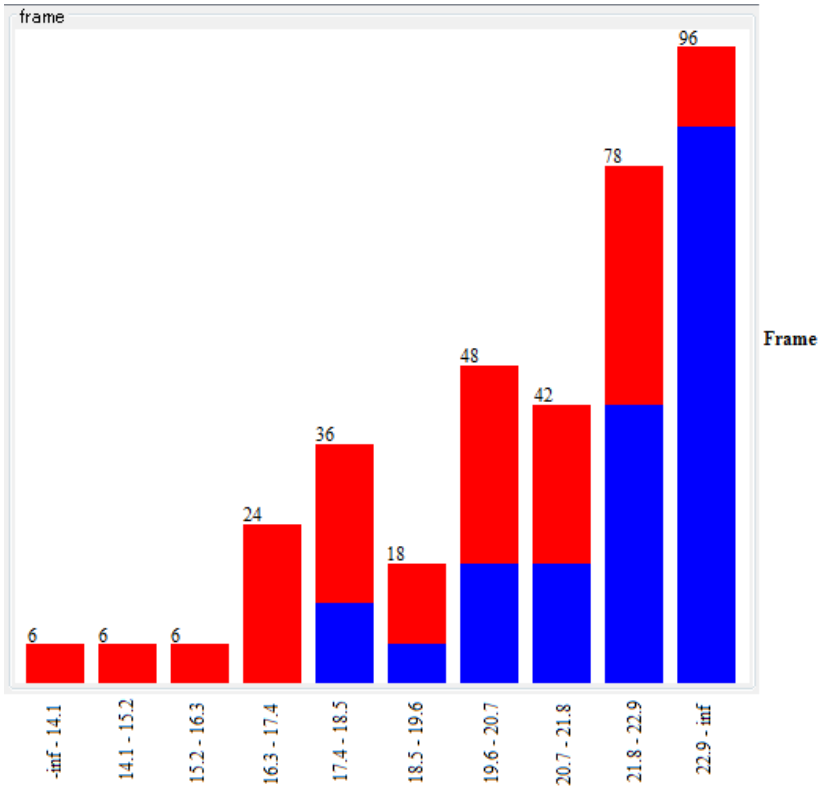


Figure 4-10: Frame difference

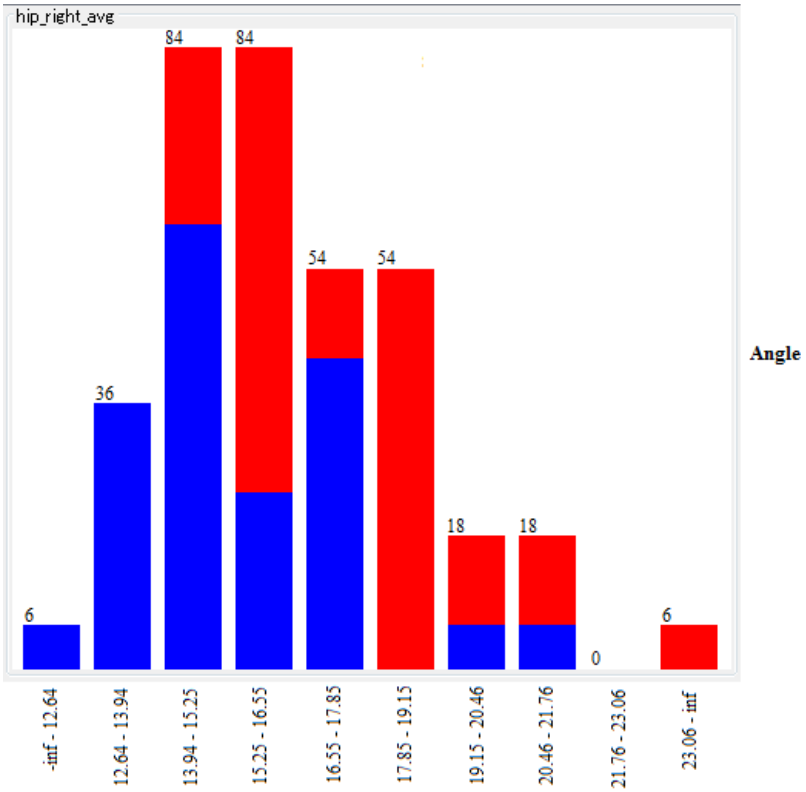


Figure 4-11: hip right average difference

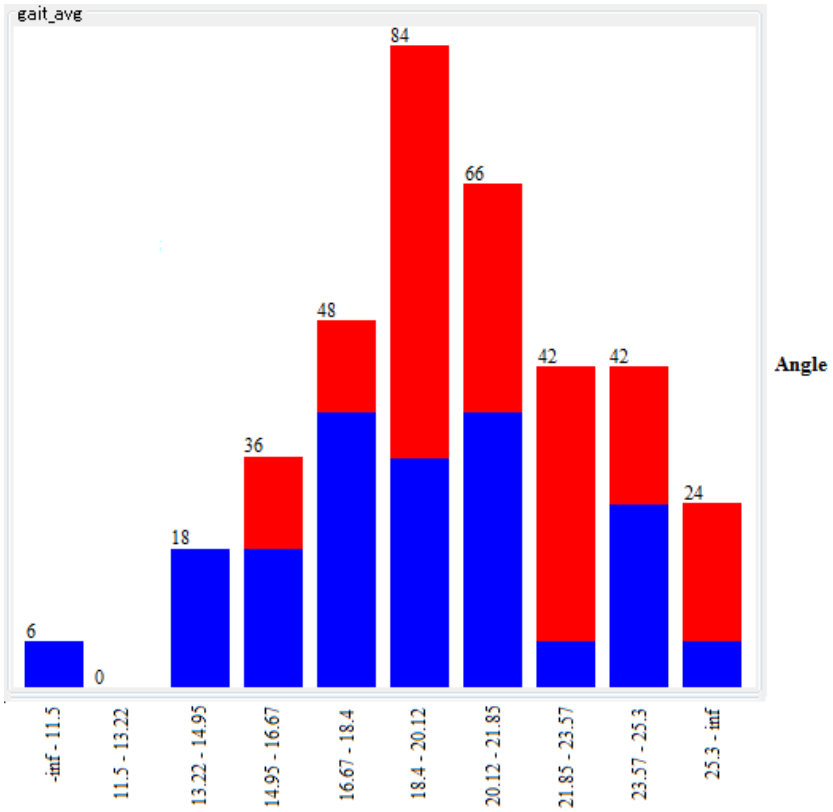


Figure 4-12: knee gait average difference

In order to confirm that all those 5 best feature is better compare to other feature combination, some experiments have done. The table below show the experiments using some random feature combination.

No.	Feature amount	Feature name	CCR
1	3	stance_max, stance_min, swing_min	55 %
2	5	swing_min, gait_std, neck_std, hip_right_avg, hip_left1	66.67 %
3	7	stance_min, swing_max, gait_std, neck_std, hip_left_std, hip_right_avg, hip_left1	58.33 %
4	9	swing_min, neck_avg, hip_left_std, hip_right_std, hip_left1, hip_left3, hip_right1, hip_right3, neck2	68.33 %
5	11	stance_max, swing_min, gait_avg, neck_avg, hip_left_avg, hip_right_avg, frame, hip_left2, hip_right1, hip_right3, neck2	71.67 %
6	13	stance_max, stance_min, swing_max, gait_std, neck_avg, hip_left_avg, hip_right_avg, hip_right_std, frame, hip_left1, hip_left2, hip_left3, hip_right3	70 %
7	15	stance_max, stance_min, swing_min, swing_max, gait_std, neck_avg, neck_std, hip_right_avg, frame, hip_left1, hip_left2, hip_right1, hip_right3, neck1, neck2	65 %
8	17	stance_max, stance_min, swing_min, gait_avg, gait_std, neck_std, hip_left_avg, hip_right_avg, hip_right_std, hip_left1, hip_left2, hip_right1, hip_right2, hip_right3, neck1, neck2, neck3, class	71.67 %
9	19	stance_max, stance_min, swing_min, gait_avg, gait_std, neck_std, hip_left_avg, hip_left_std, hip_right_std, frame,	70 %

		hip_left1, hip_left2, hip_left3, hip_right1, hip_right2, hip_right3, neck1, neck2, neck3, class	

Table 4-5: Experiments using some random feature combination

4.2 Kinematics and Statistical Feature in Skeleton 3D Model

Skeleton 3D Model methodology already explained in 3.2. This model will use our own dataset and will implement on gait disable. We use disable to add the class to 5 instead of gender classification which is only using 2 class. Those 5 class represents the quality of the disable gait. We also use fake disable gait despite of the difficulty finding the object. There will be 18 data for each class, so the total data would be 90. Currently we extracted feature from knee angle only. Below is the diagram from the knee angle extracted. Instead of using some knee parameters like stance max angle etc., we use all the angle from one gait cycle.

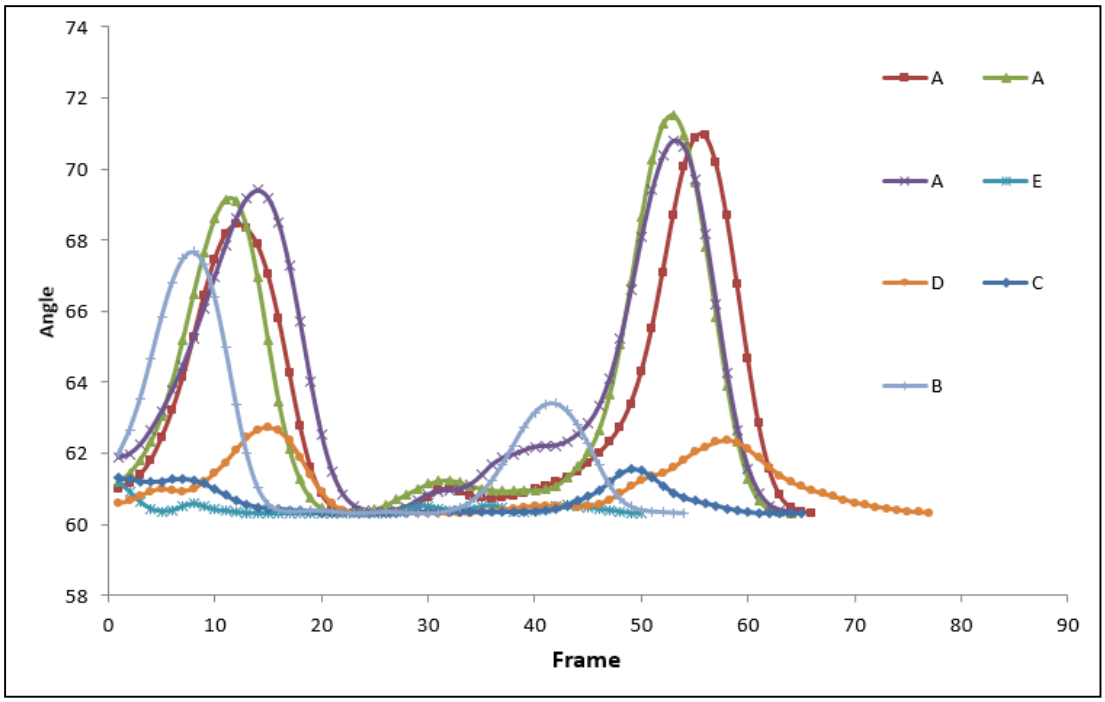


Figure 4-13: Knee angle feature extracted from 3D skeleton model

Instead of using some knee parameters like stance max angle etc., we use all the angle from one gait cycle and frame amount. To handle the speed difference between data, we can normalized all the data using the fastest speed data as the smallest division for all data. Figure 4-14 below is the normalized data of Figure 4-13.

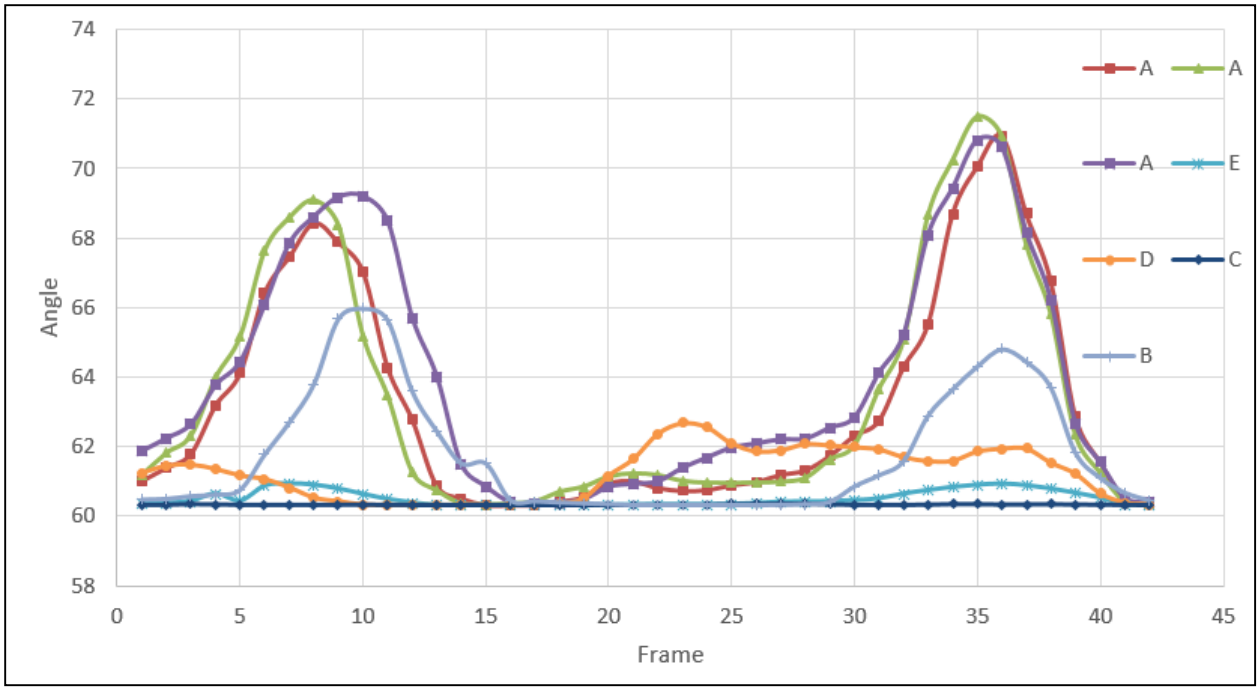


Figure 4-14: Normalized knee angle feature extracted from 3D skeleton model

Using SVM, the correct classification result is 86.36 %. Table 2-1 below is precision and recall table from SVM classification result. Table 4-7 below show the confusion matrix from SVM classification of 3D skeleton model.

Class	Precision	Recall
A	0.933	0.875
B	0.833	0.833
C	0.895	0.944
D	0.867	0.722
E	0.81	0.944

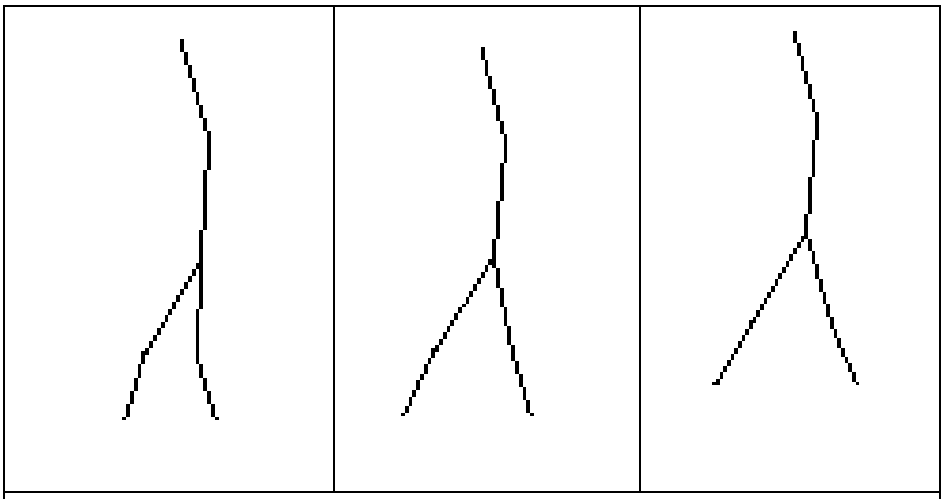
Table 4-6: Precision and Recall for disable gait classification from 3D skeleton model

	A	B	C	D	E
A	14	0	0	0	2
B	0	15	0	1	2
C	0	0	17	1	0
D	0	3	2	13	0
E	1	0	0	0	17

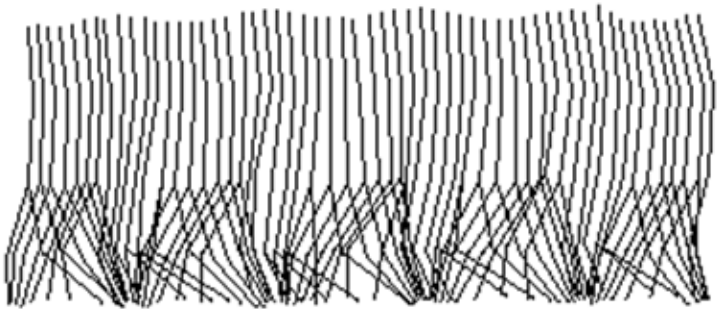
Table 4-7: Confusion matrix of disable gait from 3d skeleton model

4.3 2D Discrete Wavelet Transform (DWT) as a Feature

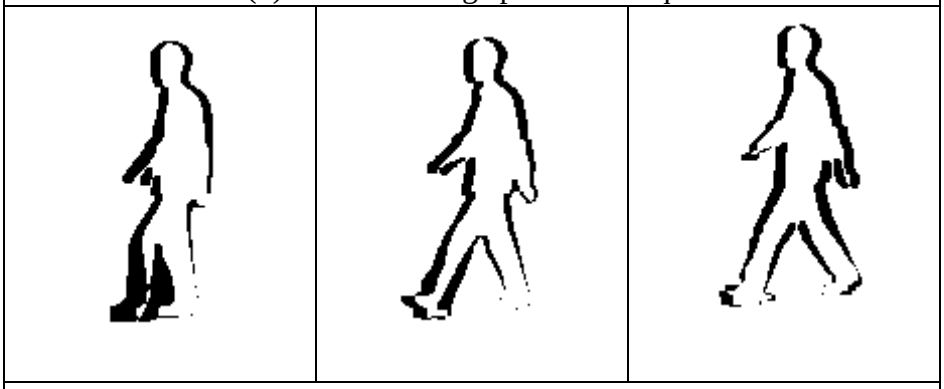
2D Discrete Wavelet Transform is a data transformation by doing composition. Detail explanation about 2D DWT can be seen in 2.2. 2D DWT very effective for image compression because its ability to decomposed the data. JPEG2000 is one of the image compression using 2D DWT method. Some references use 2D DWT to extract the feature and classify the data [35][55]. Using wavelet will reduce feature size and will increase performance if implemented in high size database. This part also will analyze CASIA gait dataset and implemented for gender classification using 2D DWT. Research method will use free-model and 2D model and see the classification results differences between those models. To analyze the ability of wavelet feature in distinguishing existed class thus we only use 4 id. Below is the free-model and 2D model images which is used to do the analysis.



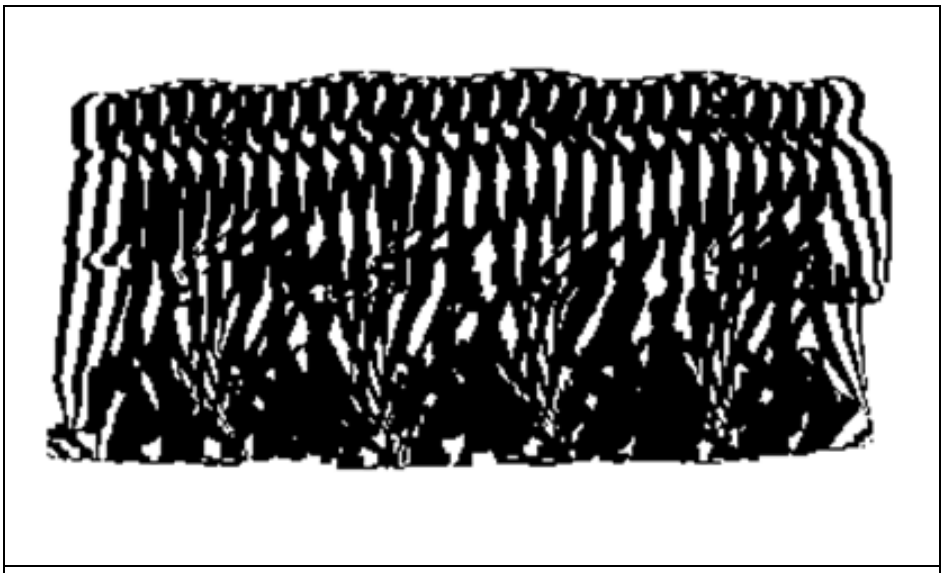
(a) Skeleton image per frame



(b) Skeleton image per video sequence



(c) Motion image per frame



(d) Motion image per video sequence

Figure 4-15: Free-model and 2d model used for analyzing database and 2D DWT

By using all those formula in equation [1] to [9], we work on some different procedures for preprocessing data:

1. Apply the wavelet transform to the single skeleton frame, averaging all the energy from one video (using the Figure 4-15(a) preprocessing data). It is a high-cost computation.
2. Apply the wavelet transform to the skeleton frame sequence, (using the Figure 4-15(b) preprocessing data). It is a high-cost computation.
3. Apply the wavelet transform to the single motion frame, averaging all the energy from one video (using the Figure 4-15(c) preprocessing data). It is a low-cost computation.
4. Apply the wavelet transform to the motion frame sequence, averaging all the energy from one video (using the Figure 4-15(d) preprocessing data). It is a low-cost computation.

We show the experiment results in the following figure. They show the state for every combination and every procedure did. One point in every figure represents data of one person in one video. One color consists of 6 points in one scatter, which means that one person have 6 video dataset. All the procedures are using Haar Wavelet at 1-level decomposition. All the possible combination is following:

1. eAeH (energy from **Approximation** and **Horizontal** Detail),
2. eAeV (energy from **Approximation** and **Vertical** Detail),
3. eAeD (energy from **Approximation** and **Diagonal** Detail),
4. eHeV (energy from **Horizontal** and **Vertical** Detail),
5. eHeD (energy from **Horizontal** and **Diagonal** Detail),
6. And eVeD (energy from **Vertical** and **Diagonal** Detail).

Figure 4-16 shows all possible energy combination per frame with 2 persons. The energy combination eHeV shows the better cluster condition. It can separate data. In the other energy combination, there are many intersection data. It will be quite difficult to classify, even if we just using 2 persons data.

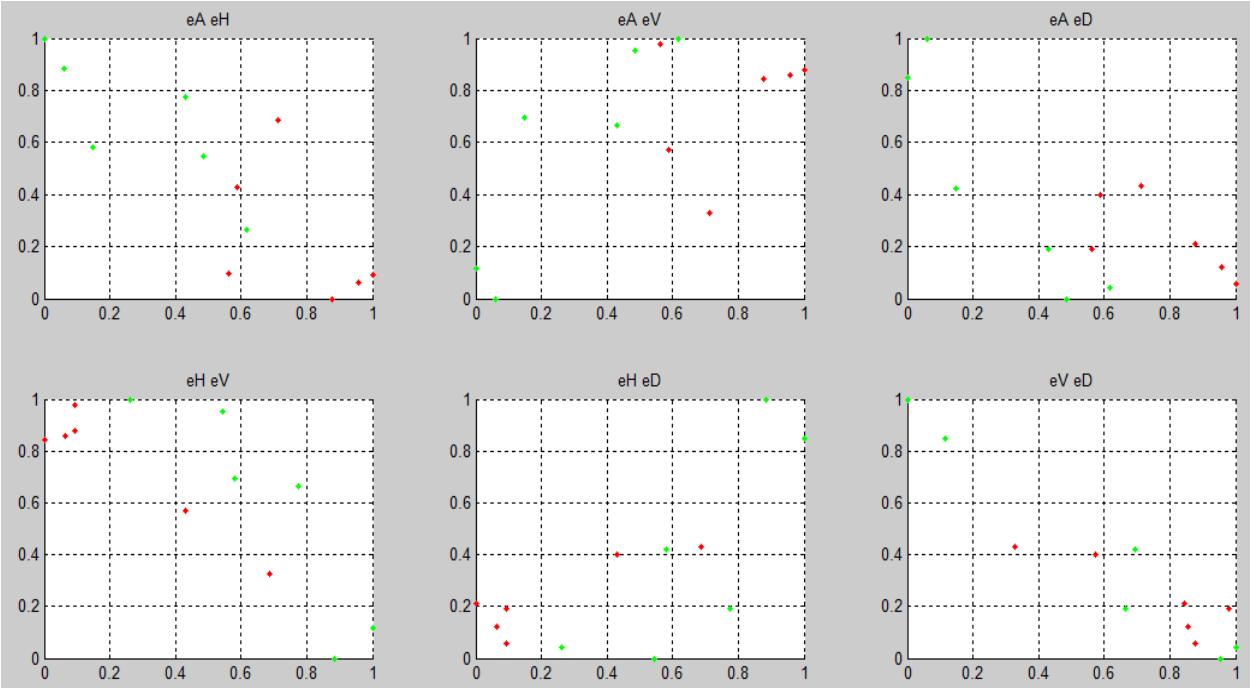


Figure 4-16: Result from single skeleton frame of 2 persons

Figure 4-17 shows all possible energy combination per sequence with 2 persons. It will be easier to do the classification, except on the eHeD combination energy. It will be difficult if we add some data as shown in Figure 4-18. In Figure 4-18, we add the data 2 more persons, so the total data is 4 persons. This figure shows that there is no possible energy combination used for classification.

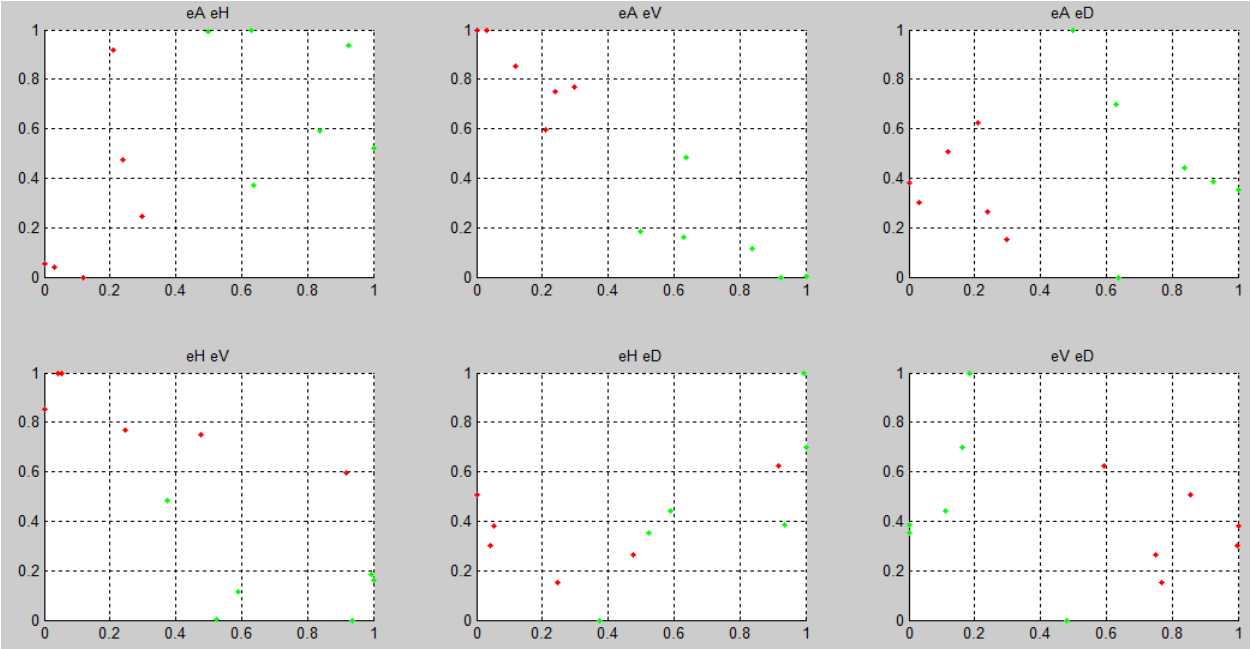


Figure 4-17: Result from skeleton frame sequence of 2 persons

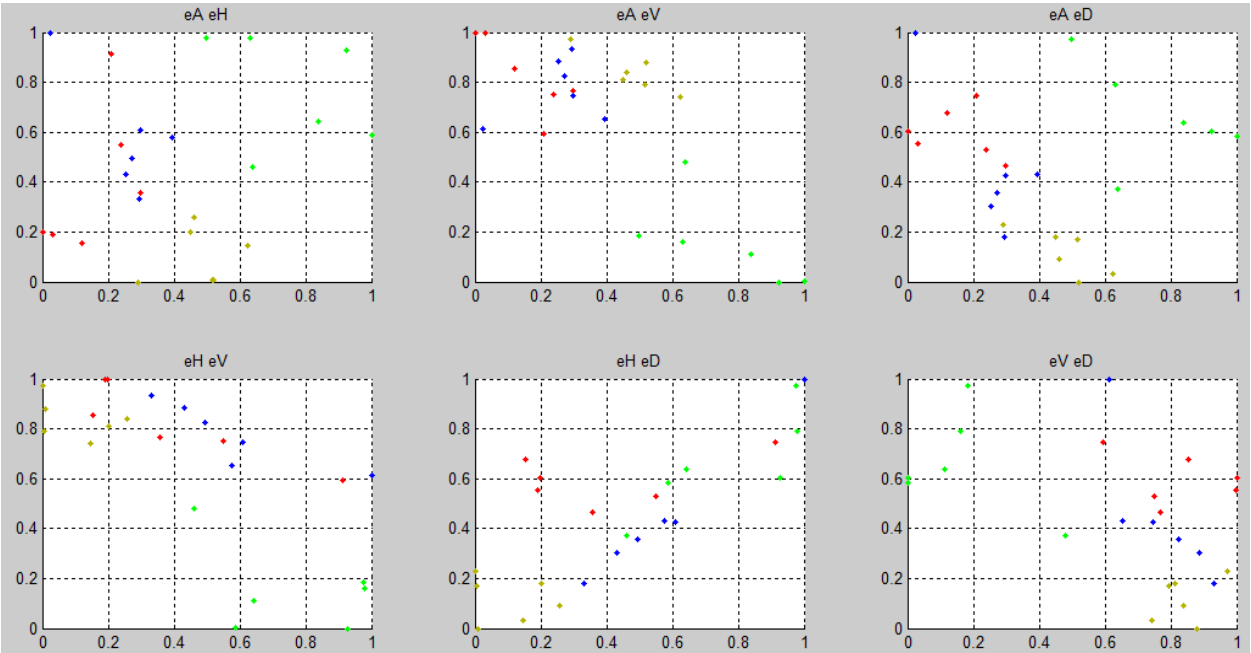


Figure 4-18: Result from skeleton frame sequence of 4 persons

Figure 4-19 shows that the result is better than the result in Figure 4-17 for 2 persons. All the combination energy looks possible to make the classification, but the best for all is the energy combination of eHeD. Figure 4-20, by using 4 person data, the best energy combination is eHeD.

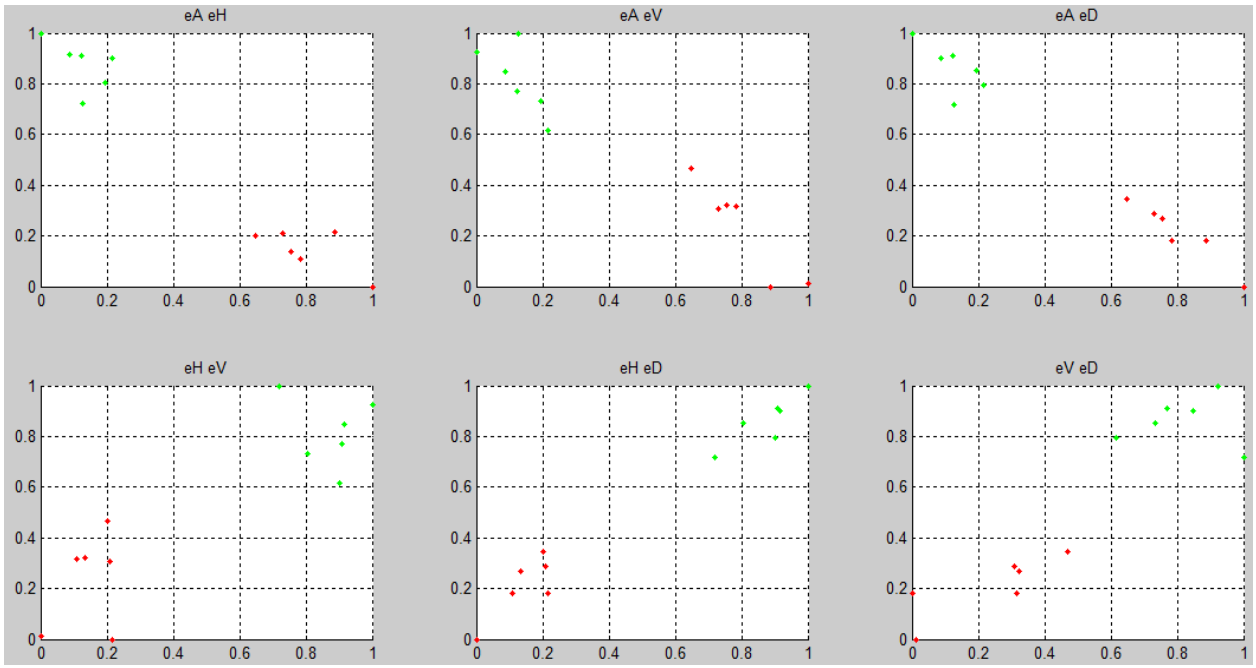


Figure 4-19: Result from single motion frame of 2 persons

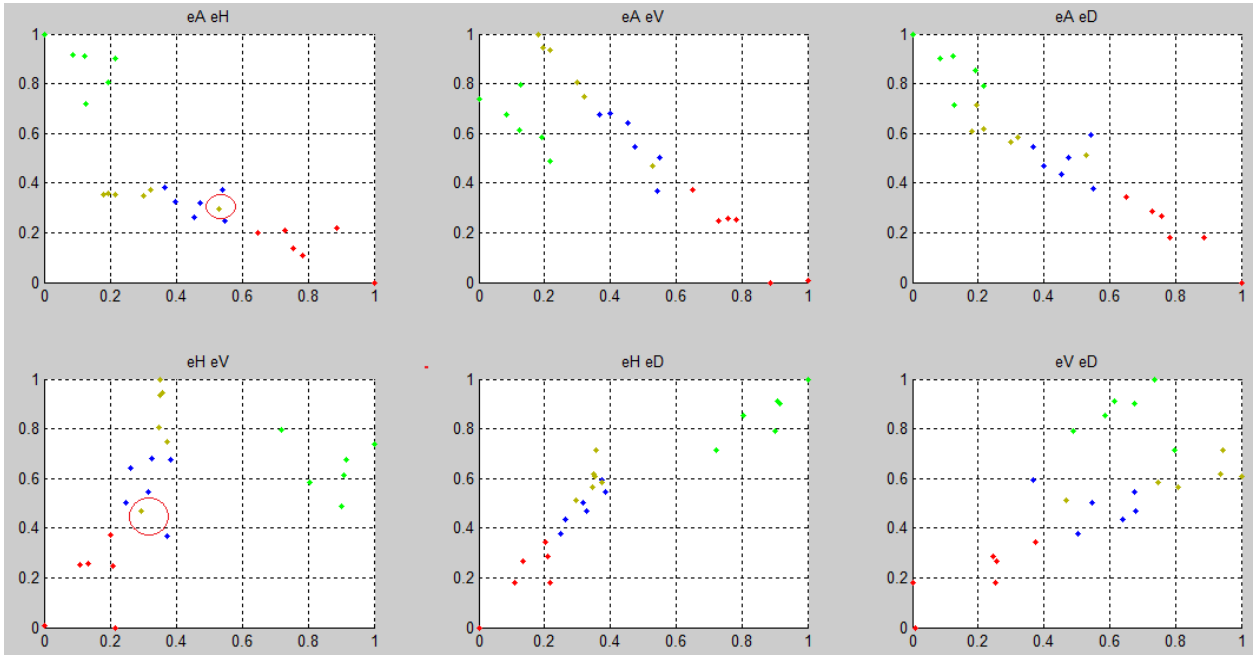


Figure 4-20: Result from single motion frame of 4 persons

Figure 4-21, by using 4 person data, we can see that the best energy combination is eHeV. The result of eHeV (in figure 10) is even better than the eHeD (in Figure 4-20).

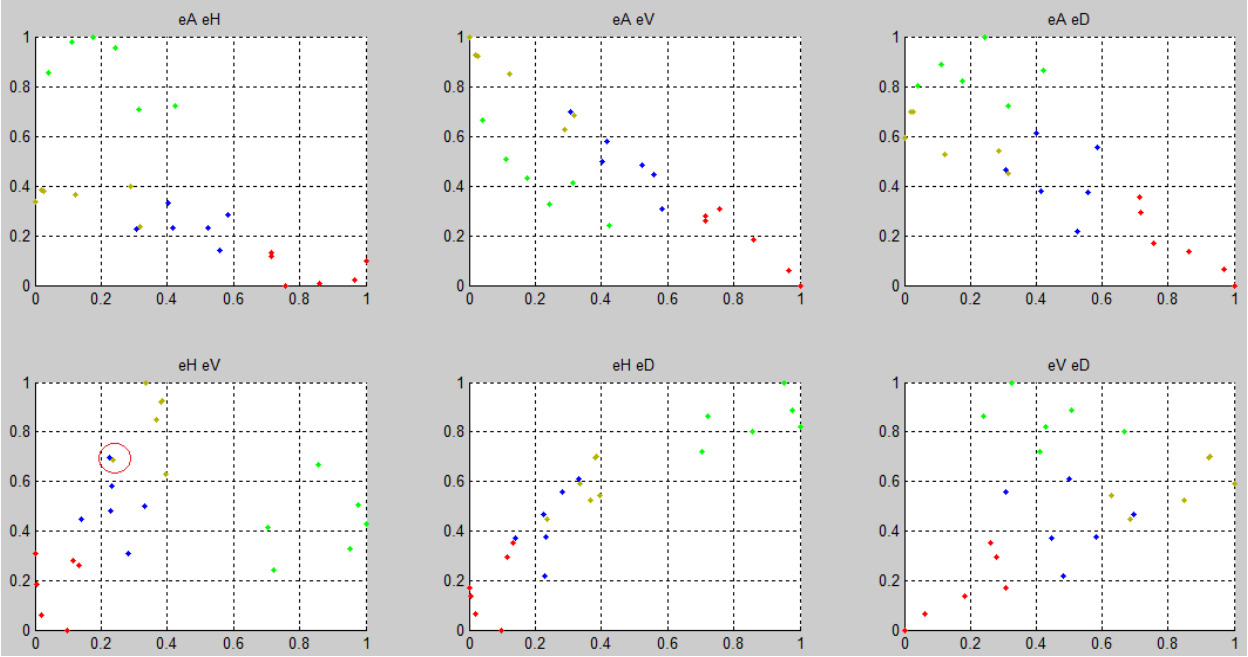


Figure 4-21: Result from motion frame sequence of 4 persons

Using the experiment results, using 2D DWT, Haar Wavelet and 1 level decomposition, we conclude as following:

(1) The best preprocessing data is a motion in frame sequence.

(2) Best combination parameters for classification are Horizontal Detail and Vertical Detail.

However, experimental results above aims only to analyzed wheter 2D DWT energy can be used for classification and if it can, which detail will be more accurate compare to other detail. Those experiments also only using two feature in every image to visualize easily when we do the classification manually. Two feature of course will be difficult to use as a gait image classification.

Next experiments is using energy for every column from image motion and analyzed classification accuracy in level 1 decomposition until its maximum level. Figure 4-22 and Figure 4-23 below showing experimental feature extraction did.

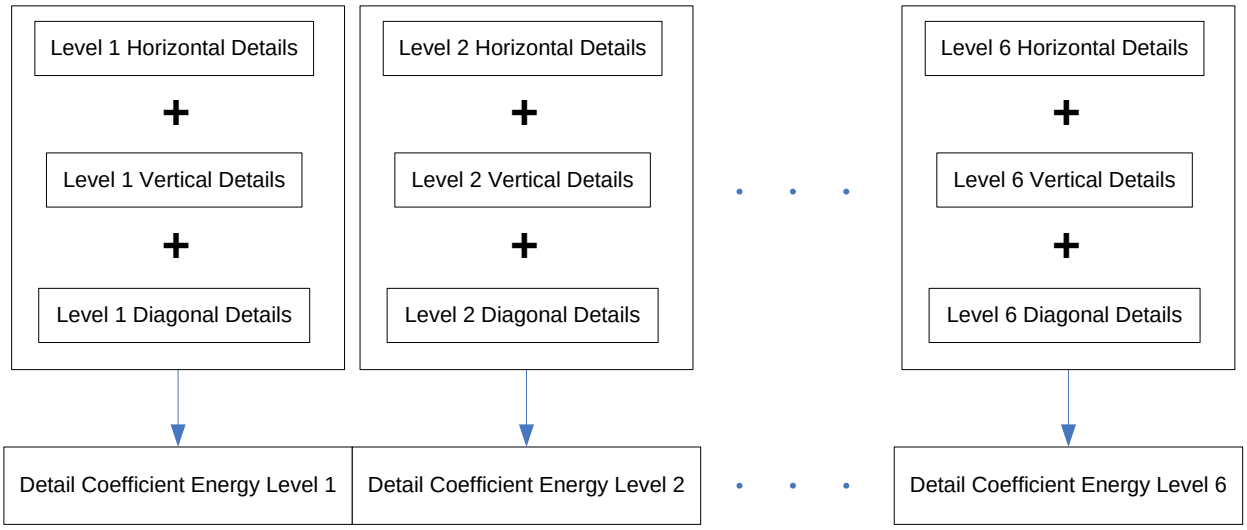


Figure 4-22: Energy Combination of Detail Coefficients

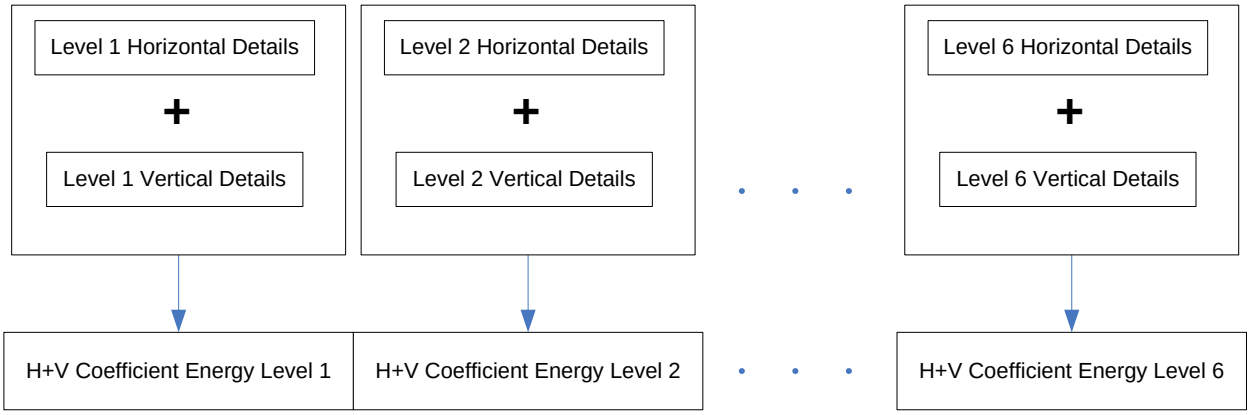


Figure 4-23: Energy Combination of H & V Coefficients

Next thing to do is reducing noise and get the necessary pattern. Most of the pattern recognition usually reduces the noise before make such a classification [37]. This is well known as the under-sample problem. Thus, a noise reduction algorithm is required to obtain useful and informative features for classification. Principal component analysis (PCA) and linear discriminate analysis (LDA) are traditional but widely used feature reduction methods. The proposed method will use Wavelet based multi-scale PCA to reduce the feature data.

Figure 7.a shows the feature data for the female gender using the proposed methods. Figure 7.b is the feature data chart for the male gender. Figure 7.c is the chart data after made such

feature reduction in the female gender. Figure 7.d is the chart data after made such feature reduction in the male gender.

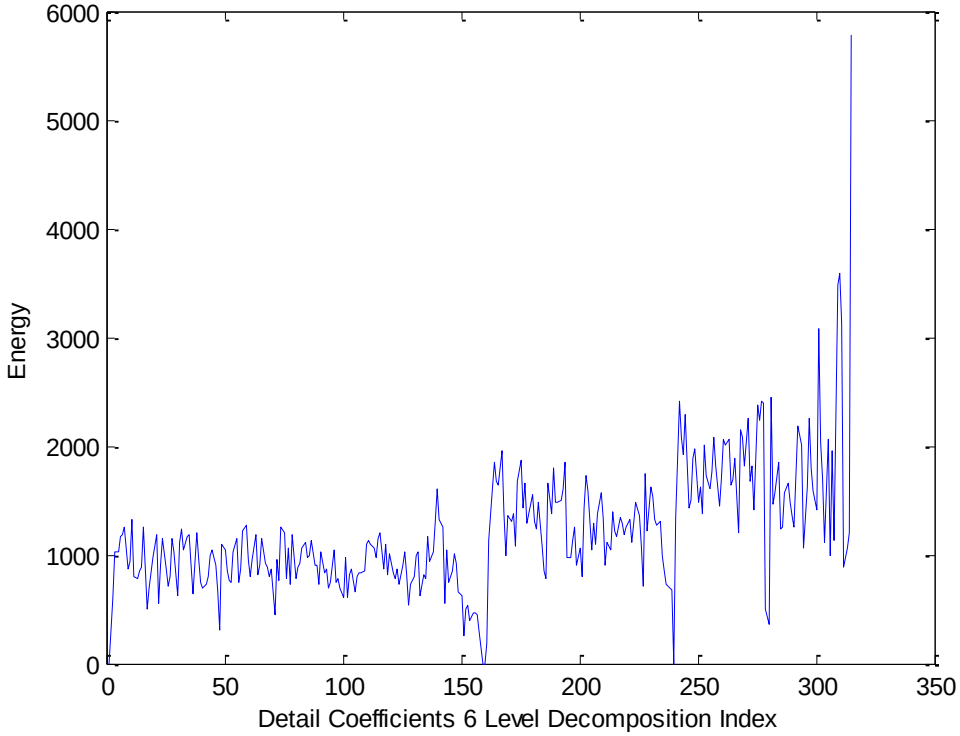


Figure 4-24: Example of 6 Level Detail Coefficients Decomposition Index in Male data

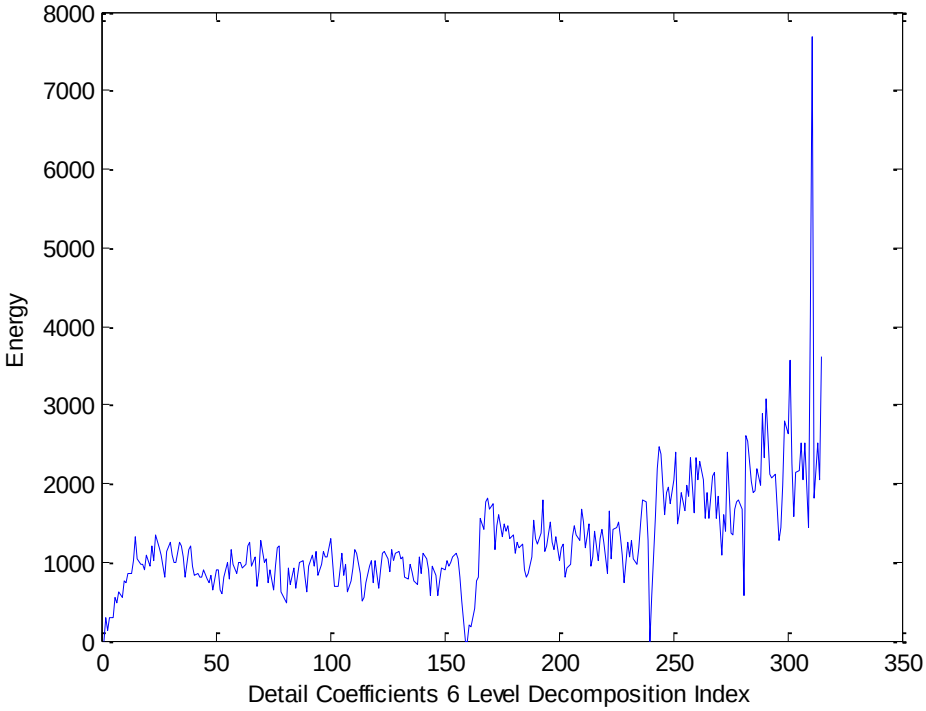


Figure 4-25: Example of 6 Level Detail Coefficients Decomposition Index in Female data

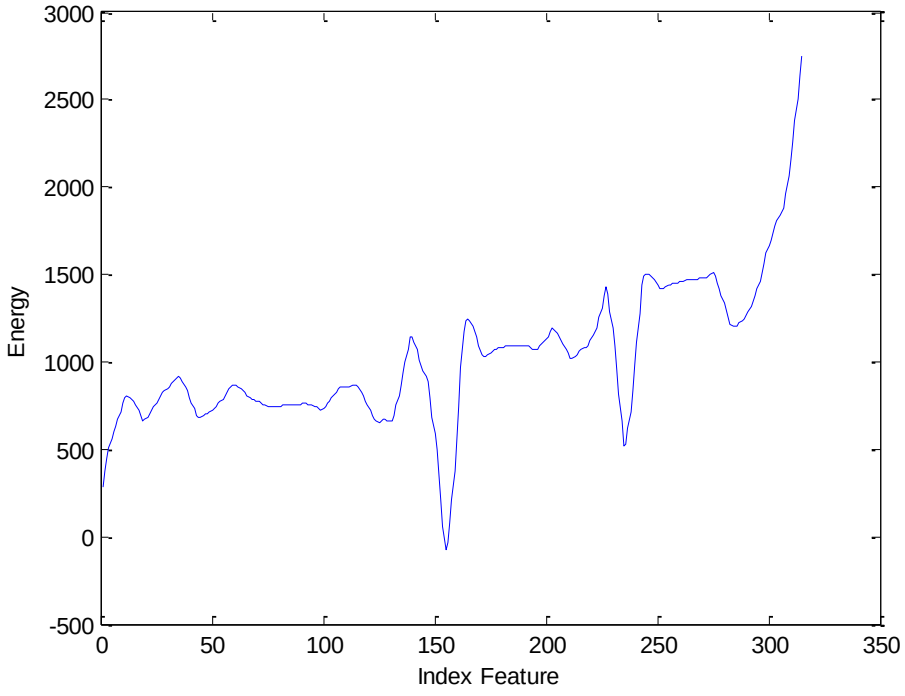


Figure 4-26: Image in Figure 4-24 after doing the noise reduction

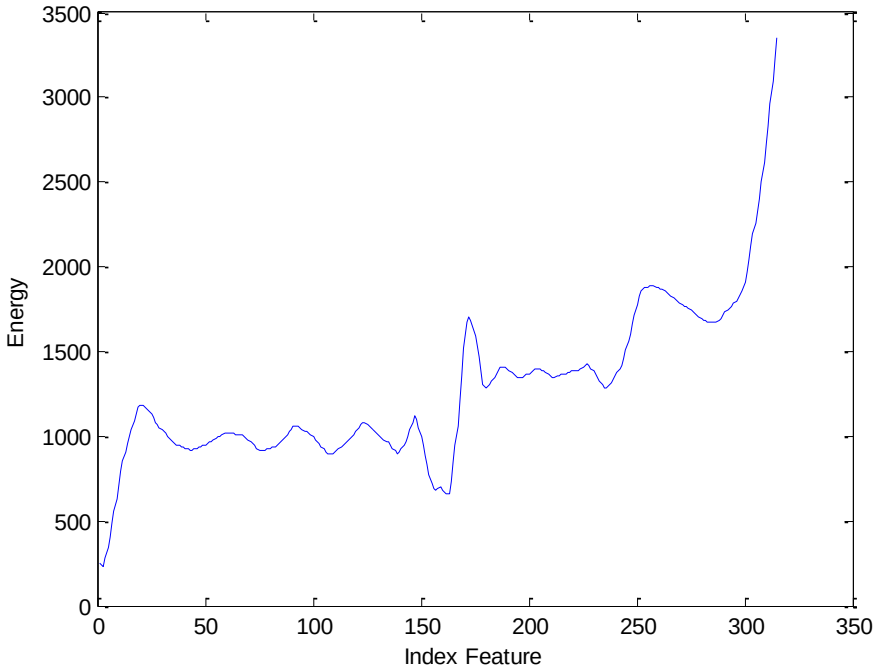


Figure 4-27: Image in Figure 4-25 after doing the noise reduction

The experiment will implement to the CASIA (Chinese Academy of Sciences) Gait B class dataset in 90 degrees point of view.

Table 4-8 below is the training specification from the experiments. Using the combination of Detail coefficients, the result for the classification task is shown in Table 4-9.

Gender	Training (Video)			Testing (Video)
	Train	Validation	Test	
Male	162	9	9	42
Female	162	9	9	42

Table 4-8: Training Specification

	Actual Classification	
Predicted Classification	True Positive = 50 %	False Positive = 9.5 %
	False Negative = 0.0 %	True Negative = 40.5 %

Table 4-9: Classification Results of the Method for Combination of Delta Coefficients

The precision, recall, and accuracy results are:

- Precision = 50 / (50+9.5) % = 84.033 %
- Recall = 50 / (50+0) = 100 %
- Accuracy = (50+40.5) / (50+9.5+0+40.5) = 90.5 %

The accuracy of classification reached 90.5% using back-propagation neural network.

Using the combination of Horizontal and Vertical coefficients, the result for the classification task shown in the Table 4-10 below:

	Actual Classification	
Predicted Classification	True Positive = 50 %	False Positive = 7.1 %
	False Negative = 0.0 %	True Negative = 42.9 %

Table 4-10: Classification Results of the Method for Combination of Horizontal and Vertical Coefficients

The precision, recall, and accuracy results are:

- Precision = 50 / (50+7.1) % = 87.565 %
- Recall = 50 / (50+0) = 100 %
- Accuracy = (50+42.9) / (50+7.1+0+42.9) = 92.9 %

The accurate of classification reached 92.9% using back-propagation neural network.

We also try to do the experiment in different level and different type of 2D DWT. The result of the experiment is shown in the Table 4-11 below:

Type	Lv1	Σ(lv1,lv2)	Σ(lv1..lv3)	Σ(lv1..lv4)	Σ(lv1..lv5)	Σ(lv1..lv6)
Haar	83.3%	84.4%	87.8%	86.8%	90.6%	92.9%
DB2	52.8%	55.2%	58.7%	58.8%	61.76%	63.3%

Table 4-11: CCR with different level and 2D DWT type

4.4 Gait Energy Motion as a Feature

We have shown that by using 2D DWT energy we could extract feature and give a promising result. Need to be consider that by seeing previous research results, accuracy achieved from an experiment in 4.3 as shown in Table 4-14. Based on the results achieved, we try to fix and improve the image representation created. Some reason that we use as a consideration from the experiments conducted are:

- Image representation size is still too big. Using 320 x 240 image size, will make time process and transformation to 2D DWT longer. Thus, we will reduce the image representation size smaller.

- Pixel value from image representation have a small range value as seen in in Figure 4-15 (d). If the motion in frame x-1 have a pixel value of 1and add with motion in frame x which have the same location, then the pixel value still 1. We need to change the pixel value range.

Using all the considerations above, we propose one image representation which we call Gait Energy Motion (GEM). GEM is defined as:

$$F(i,j) = \frac{1}{T}\sum_{t=2}^{T}|I(i,j,t) - I(i,j,t-1)| \quad \text{.....[29]}$$

Where *I(i,j,t)* is a binary silhouette image at current frame *t*, and *I(i,j,t-1)* is a binary silhouette image at previous frame *t*.

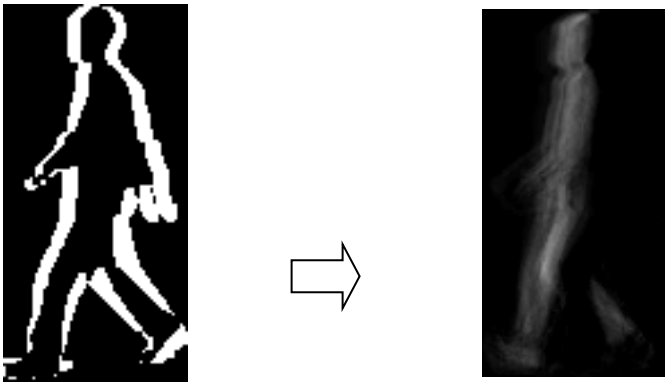


Figure 4-28: Image of GEM

Below shows the flowchart for generating GEM.

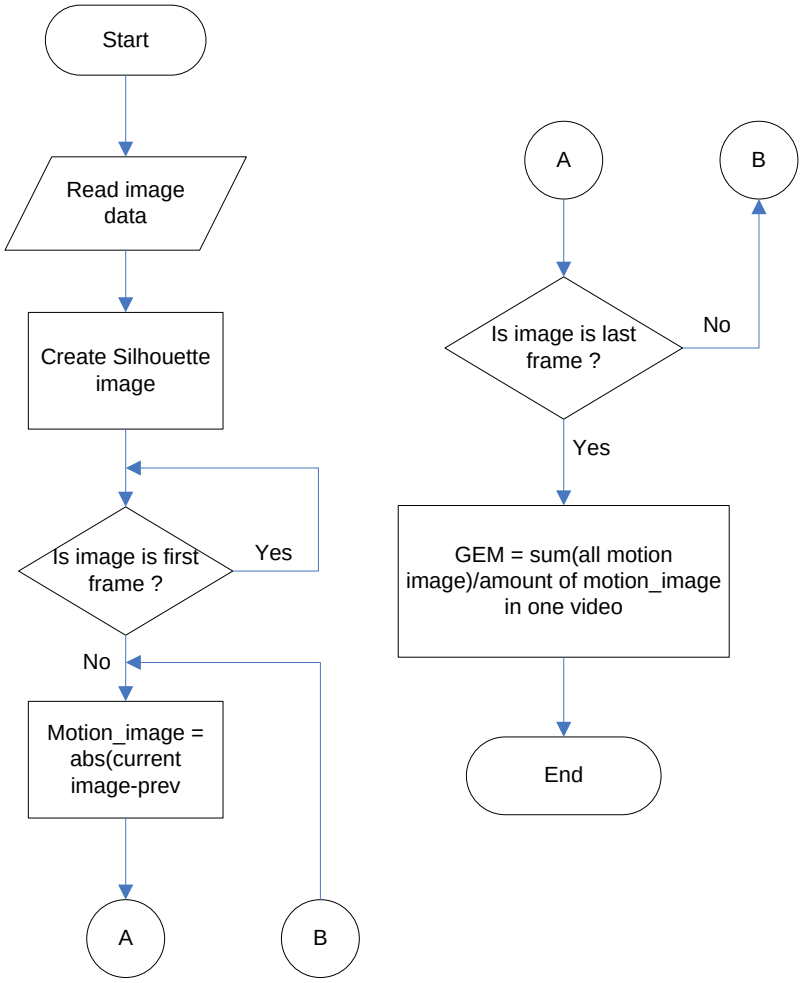


Figure 4-29: Flowchart for generating GEM

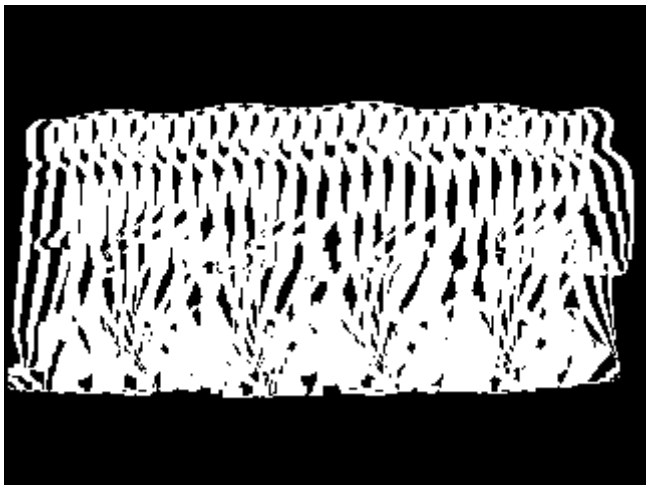
There is 124 unique human gait in B class dataset. From the total of 124 dataset, only 31 data of female gait existed. To make the data balance, we will also use 31 male gait dataset. Every human has 6 video of perpendicular view, so we pretend to use every single video as one data. The total of the data is 372, consists of 186 video of male gait and 186 video of female gait.

We use SVM as a classification method and using 10 cross fold validation as a training method [32]. We also try to 2D DWT to analyze and compare the result. The result is in the Table 4-12 below:

Method	Classification Time	CCR
Gait Energy Motion (GEM)	722.2 ms	97.47 %
GEM with Velocity	**744.5 ms**	**97.63 %**
2D DWT GEM Lv 1.	566.1 ms	97.22 %
2D DWT GEM Lv 1. with Velocity	630 ms	97.32 %

Table 4-12: CCR table for some methods analyzed

Very interesting to analyze is the classification time for every method. Every pixel is pretend to be the feature want to classify. The classification cost will be higher if the image size is bigger. GEM is smaller image compare to original 2D free-model, because GEM only took the motion part, not all part of the image. Using Approximation detail in Discrete Wavelet Transform will reduce the classification cost because the image size is rescale 4 times smaller in level 1.



(a) 320 x 240

(b) 81x 163

Table 4-13: Image size comparation between (a) motion sequence and (b) GEM

Since we use Gait Energy Motion (GEM) which is taking the motion parameter in the image, the difference between classes should be the motion. Which part of the human body motion giving the significance difference between classes is also another basic question.

In [32], Shiqi Yu et al. used analysis of variance (ANOVA) F-statistics to analyze the gait difference between class. We will use the same method to analyze GEM difference. The ANOVA F-statistic is a measure to evaluate different features discriminative capability. The greater the F-statistic values will give better discriminative capability. The F-statistic is calculated as shown in equation [28].

If we implement in our research, the formula above will become:

$$F = \frac{180((\text{mean}_{\text{gem\_male}} - \text{mean}_{\text{gem\_all}})^2 + (\text{mean}_{\text{gem\_female}} - \text{mean}_{\text{gem\_all}}))^2}{\frac{1}{370}(\sum(\text{gem}_{\text{male(i,j)}} - \text{mean}_{\text{gem\_male}})^2 + \sum(\text{gem}_{\text{female(i,j)}} - \text{mean}_{\text{gem\_female}})^2)} \quad \text{.....[30]}$$

The calculated F-statistic values are shown in Figure 4-30: F-Statistics image. Whiter color means better discriminative capability. The highest discriminative value is seen in the left foot motion. Right foot motion also seen some discriminative value, but because of the longer distance than the left foot, the value is not too high. Other area that have higher discriminative value than others is the hand motion.

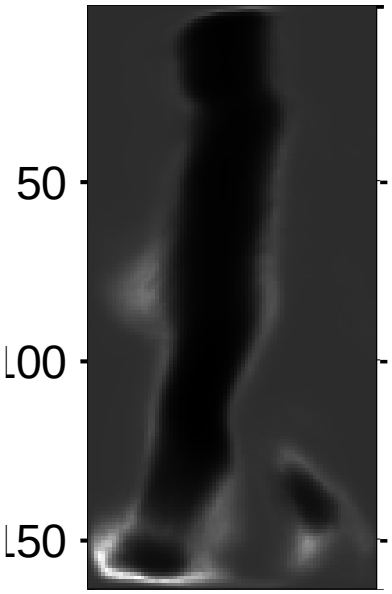


Figure 4-30: F-Statistics image

Below is the table comparison of CCR results with another results.

Method	Dataset	CCR
Lee and Grimson [11]	25 males & 25 females	85.0%
Huang and Wang [28]	25 males & 25 females	85.0%
2D DWT Energy proposed [19]	31 males & 31 females	92.9%
Li et al. [2]	31 males & 31 females	93.28%
GEM Proposed	31 males & 31 females	97.63%

Table 4-14: CCR for some published methods

5. Concluding Remarks

5.1 Conclusions

From some feature we extracted and also from the experimental results, we can make some conclusion:

1. Using silhouettes image from CASIA gait dataset:

   a. By using 2D skeleton model and extracting kinematics and statistical feature, we got 81.65% CCR.

   b. By creating motion image sequence feature extraction and using 2D DWT Horizontal dan Vertical Coefficients, we got 92.9% CCR

   c. By creating Gait Energy Motion (GEM) and extracting the feature, we got 97.63% CCR.

2. Using Depth image from Kinect camera, then extracting the kinematics feature from 3D skeleton model created, we got 86.36% CCR.

3. From the results in 1.a and table 4.2 we can see that the feature with biggest discriminant value is hip angle, thus we can conclude that to distinguish between male and female through their gait and using perpendicular view single camera, we should consider using hip angle as the most important feature. This feature never been extracted in the previous research.

By compare 2D skeleton model based and GEM free-model based, we can conclude the advantages between one another:

- Skeleton model giving more feature amount. From the experiments, we can get 22 features. Previous Nixon model [47] has 20 features mostly different with our proposed method. GEM as a free-model is an image feature extraction. Using some transformations and statistical feature we can only get fewer features. More feature amount will give promising

accuracy result if the complexity of data or sum of class become higher for example used in human gait recognition.

- GEM compare to skeleton is easier to develop and a lower computation time. Skeleton model need around 1054.5 millisecond computation time, while GEM free-model need around 300 millisecond computation time.

5.2 Future Works

The following task is suggested for future works for gait gender classification.

1. Using higher camera resolution in motion image feature extraction will be challenging because of higher image resolution. This also include implementation in Digital Pocket Camera, Digital Single-lens Reflex camera (DSLR), Smartphone camera, or even wearable devices like Google Glass.

2. Gender classification will be helpful for human recognition as one of the feature along with age classification and any feature extracted in human recognition.

3. Combination of skeleton model and gait energy motion also promising to extract feature in human recognition.

4. Preparing and creating 3D skeleton gait dataset should be considered.

References

[1] X. Qinhan, “Technology review - Biometrics-Technology, Application, Challenge, and Computational Intelligence Solutions,” *IEEE Computational Intelligence Magazine*, vol. 2, pp. 5–25, 2007.

[2] X. Li, S. Maybank, and S. Yan, “Gait components and their application to gender recognition,” *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 38, no. 2, pp. 145–155, 2008.

[3] Z. Yang, M. Li, and H. Ai, “An experimental study on automatic face gender classification,” *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, vol. 3, pp. 1099 – 1102, 2006.

[4] H.-C. Lian and B.-L. Lu, “Multi-view gender classification using multi-resolution local binary patterns and support vector machines.,” *International journal of neural systems*, vol. 17, no. 6, pp. 479–87, Dec. 2007.

[5] B. Moghaddam, “Learning gender with support faces,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 5, pp. 707–711, May 2002.

[6] M. Li, K. J. Han, and S. Narayanan, “Automatic speaker age and gender recognition using acoustic and prosodic level information fusion,” *Computer Speech & Language*, vol. 27, no. 1, pp. 151–167, Jan. 2013.

[7] N. V. Boulgouris, D. Hatzinakos, and K. N. Plataniotis, “Gait recognition: A Challenging Signal Processing Technology For Biometric Identification,” *IEEE Signal Processing Magazine*, vol. 22, no. 6, pp. 78–90, 2005.

[8] M. S. Nixon and J. N. Carter, “Automatic Recognition by Gait,” *Proceedings of the IEEE*, vol. 94, pp. 2013–2024, 2006.

[9] D. Kim and J. Paik, “Gait Recognition Using Active Shape Model and Motion Prediction,” *IET Computer Vision*, vol. 4, pp. 25–36, 2010.

[10] M. Piccardi, “Background subtraction techniques: a Review,” *IEEE International Conference on Systems*, vol. 4, pp. 3099–3104, 2004.

[11] L. Lee and W. Grimson, “Gait Analysis for Recognition and Classification,” in *Proceedings of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, 2002, pp. 148–155.

[12] D. K. Wagg and M. S. Nixon, "On Automated Model-Based Extraction and Analysis of Gait," in *Proceedings of the Sixth IEEE International Conference on Automatic Face and Gesture Recognition*, 2004, pp. 11–16.

[13] D. Cunado, M. S. Nixon, and J. N. Carter, "Automatic Extraction and Description of Human Gait Models for Recognition Purposes," *Computer Vision and Image Understanding*, vol. 90, no. 1, pp. 1–41, 2003.

[14] L. Wang, T. Tan, S. Member, H. Ning, W. Hu, W. Liang, and T. Tieniu, "Silhouettes Analysis-based Gait Recognition for Human Identification," *IEEE Transactions on Pattern Analysis and Machine Intelligence2*, vol. 25, no. 12, pp. 1505–1518, 2003.

[15] S. Sarkar, P. J. Phillips, Z. Liu, I. R. Vega, P. Grother, and K. W. Bowyer, "The humanID gait challenge problem: datasets, performance, and analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 162–177, 2005.

[16] K. Arai and R. Asmara, "Human Gait Gender Classification in Spatial and Temporal Reasoning," *IJARAI International Journal of Advanced Research in Artificial Intelligence*, vol. 1, no. 6, pp. 1–6, 2012.

[17] C. BenAbdelKader, R. Cutler, and L. Davis, "Motion-based Recognition of People in EigenGait Space," in *Proceeding of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, 2002, pp. 267–272.

[18] R. Asmara, A. Basuki, and K. Arai, "A Review of Chinese Academy of Sciences (CASIA) Gait Database As a Human Gait Recognition Dataset," in *Industrial Electronics Seminar*, 2011, pp. 267–271.

[19] K. Arai and R. Asmara, "Human Gait Gender Classification using 2D Discrete Wavelet Transforms Energy," *IJCSNS International Journal of Computer Science and Network Security*, pp. 62–68, 2011.

[20] S. Aravind, R. Amit, and C. Rama, "A Hidden Markov Model Based Framework for Recognition of Humans from Gait Sequences," in *Proceeding of the 2003 IEEE International Conference on Image Processing*, 2003, pp. II–93–6.

[21] M. H. Cheng, M. F. Ho, and C. L. Huang, "Gait Analysis for Human Identification through manifold learning and HMM," *Pattern Recognition*, vol. 41, pp. 2541–2553, 2008.

[22] C. Changhong, L. Jimin, Z. Heng, H. Haihong, and T. Jie, "Factorial HMM and Parallel HMM for Gait Recognition," *IEEE Transactions on System, Man, and Cybernetics, Part C: Application and Reviews*, vol. 39, pp. 114–123, 2009.

[23] L. Zongyi and S. Sarkar, "Improved gait recognition by gait dynamics normalization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, pp. 863–876, 2006.

[24] C. Chen, J. Liang, H. Zhao, H. Hu, and J. Tian, "Frame Difference energy image for gait recognition with incomplete silhouettes," *Pattern Recognition Letter*, vol. 30, pp. 977–984, 2009.

[25] C. Yam, M. S. Nixon, and J. N. Carter, "Automated person recognition by walking and running via model-based approaches," *Pattern Recognition*, vol. 37, no. 5, pp. 1057–1072, May 2004.

[26] N. V. Boulgouris and Z. X. Chi, "Human gait recognition based on matching of body components," *Pattern Recognition*, vol. 40, no. 6, pp. 1763–1770, Jun. 2007.

[27] Y. Wang, S. Yu, and T. Tan, "Gait recognition based on fusion of multi-view gait sequences," *Advances in Biometrics*, vol. 3832, pp. 605–611, 2005.

[28] A. Bobick and A. Johnson, "Gait recognition using static, activity-specific parameters," in *Computer Vision and Pattern Recognition, Proceedings of the 2001 IEEE Computer Society Conference on*, 2001, vol. 1, pp. I–423 – I–430.

[29] L. Wang, H. Ning, T. Tan, and W. Hu, "Fusion of static and dynamic body biometrics for gait recognition," *Proceedings Ninth IEEE International Conference on Computer Vision*, vol. 2, no. Iccv, pp. 1449–1454, 2003.

[30] G. Huang and Y. Wang, "Gender classification based on fusion of multi-view gait sequences," *Computer Vision–ACCV 2007*, vol. 4843, pp. 462–471, 2007.

[31] S. Rahati, R. Moravejian, and F. M. Kazemi, "Gait Recognition Using Wavelet Transform," in *Fifth International Conference on Information Technology: New Generations (ITNG 2008)*, 2008, pp. 932–936.

[32] S. Yu, T. Tan, K. Huang, K. Jia, and X. Wu, "A study on gait-based gender classification.," *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 18, no. 8, pp. 1905–10, Aug. 2009.

[33] W. Kusakunniran, Q. Wu, J. Zhang, and H. Li, "Multi-view Gait Recognition Based on Motion Regression Using Multilayer Perceptron," *2010 20th International Conference on Pattern Recognition*, pp. 2186–2189, Aug. 2010.

[34] S. Zheng, J. Zhang, and K. Huang, "Robust view transformation model for gait recognition," in *Image Processing (ICIP), 2011 18th IEEE International Conference on*, 2011, pp. 2073 – 2076.

[35] K. Arai and R. Asmara, "Gait Recognition Method Based on Wavelet Transformation and Its Evaluation With Chinese Academy of Sciences (CASIA) Gait Database as A Human Gait Recognition Dataset," in *Ninth International Conference on Information Technology - New Generation*, 2012, pp. 656–661.

[36] C. BenAbdelkader, R. Cutler, and L. Davis, "View-invariant estimation of height and stride for gait recognition," *Biometric Authentication*, vol. 2359, pp. 155 – 167, 2006.

[37] D. Xu, S. Yan, D. Tao, S. Lin, and H.-J. Zhang, "Marginal Fisher analysis and its variants for human gait recognition and content- based image retrieval.," *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 16, no. 11, pp. 2811–21, Nov. 2007.

[38] R. Tanawongsuwan and A. Bobick, "Gait recognition from time-normalized joint-angle trajectories in the walking plane," *Computer Vision and Pattern Recognition, Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 2, pp. II–726 – II–731, 2001.

[39] C. BenAbdelkader, "Stride and cadence as a biometric in automatic person identification and verification," in *Automatic Face and Gesture Recognition, 2002. Proceeding. Fifth IEEE International Conference on*, 2002, pp. 372–377.

[40] J. Yoo, D. Hwang, and M. Nixon, "Gender classification in human gait using support vector machine," *Advanced concepts for intelligent vision systems, Lecture Notes in Computer Science*, vol. 3708, pp. 138–145, 2005.

[41] K. Arai and R. Asmara, "Human Gait Gender Classification in Natural Condition Utilizing NIR Camera," in *IIEEJ*, 2012, pp. 203–210.

[42] K. Arai and R. Asmara, "Human Gait Estimation for Handicapped Person," in *IIEEJ*, 2012, pp. 115–122.

[43] J. Wang, M. She, S. Nahavandi, and A. Kouzani, "A Review of Vision-Based Gait Recognition Methods for Human Identification," *2010 International Conference on Digital Image Computing: Techniques and Applications*, pp. 320–327, Dec. 2010.

[44] I. Daubechies, *Ten Lectures on Wavelets*, vol. 61, no. CBMS-NSF. SIAM, 1992, p. 357.

[45] G. Strang and T. Nguyen, *Wavelets and Filter Banks*, vol. c. Wellesley-Cambridge Press, 1996, p. 520.

[46] L. Lam, S. Lee, and C. Suen, "Thinning methodologies-a comprehensive survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, no. 9, pp. 869– 885, 1992.

[47] J. Yoo and M. Nixon, "Feature extraction and selection for recognizing humans by their gait," *Advances in Visual Computing*, vol. 4292, pp. 156–165, 2006.

[48] W. Dempster and G. Gaughran, “Properties of body segments based on size and weight,” *American Journal of Anatomy*, vol. 18, no. 7414, pp. 33–54, 1967.

[49] S. Yu, L. Wang, W. Hu, and T. Tan, “Gait Analysis for Human Identification in Frequency Domain,” *Third International Conference on Image and Graphics (ICIG’04)*, pp. 282–285, 2004.

[50] A. Kale, A. N. Rajagopalan, N. Cuntoor, V. Kr, and C. Park, “Gait-based Recognition of Humans Using Continuous HMMs,” in *the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, 2002, pp. 1–6.

[51] A. Kale, A. Sundaresan, a N. Rajagopalan, N. P. Cuntoor, A. K. Roy-Chowdhury, V. Krüger, and R. Chellappa, “Identification of humans using gait.,” *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 13, no. 9, pp. 1163–73, Sep. 2004.

[52] “Gait Cycle of normal person.” [Online]. Available: http://me.queensu.ca/People/Deluzio/DataAnalysis.html.

[53] J. . Quinlan, *C4.5: Programs for Machine Learning*. Morgan Kaufman Publisher, 1993.

[54] G. Holmes, A. Donkin, and I. H. Witten, “WEKA : A Machine Learning Workbench.”

[55] E. Yih, G. Sainarayanan, and A. Chekima, “Palmprint Based Biometric System: A Comparative Study on Discrete Cosine Transform Energy, Wavelet Transform Energy and SobelCode Methods,” *Biomedical Soft Computing and Human Sciences*, vol. 14, no. 1, pp. 11–19, 2009.

Appendix

C4.5 Algorithm

C4.5 is an algorithm used to generate a decision tree developed by Ross Quinlan. C4.5 is an extension of Quinlan's earlier ID3 algorithm. The decision trees generated by C4.5 can be used for classification, and for this reason, C4.5 is often referred to as a statistical classifier.

C4.5 builds decision trees from a set of training data in the same way as ID3, using the concept of information entropy. The training data is set $S = s_1, s_2 \ldots$ of already classified samples. Each sample s_i consists of a p-dimensional vector $(x_{1,i}, x_{2,i}, \ldots, x_{p,i})$, where the x_j represent attributes or features of the sample, as well as the class in which s_i falls.

At each node of the three, C4.5 chooses the attribute of the data that most effectively splits its set of samples into subsets enriched in one class or the other. The splitting criterion the normalized information gain (difference in entropy). The attribute with the highest normalized information gain is chosen to make the decision. The C4.5 algorithm then recurses on the smaller sublists.

This algorithm has a few base cases.

- All the samples in the list belong to the same class. When this happens, it simply creates a leaf node for the decision tree saying to choose that class.
- None of the features provide any information gain. In this case, C4.5 creates a decision node higher up the tree using the expected value of the class.
- Instance of previously-unseen class encountered. Again, C4.5 creates a decision node higher up the tree using the expected values.

Pseudo code

In pseudo code, the general algorithm for building decision trees is:

1. Check for base cases
2. For each attribute a
    1. Find the normalized information gain from splitting on a

3. Let a_base be the attribute with the highest normalized information gain
4. Create a decision node that splits on a_best
5. Recurse on the sublists obtained by splitting on a_best, and add those nodes as children of node

Support Vector Machine

In machine learning, support vector machines (SVMs, also support vector networks) are supervised learning models with associated learning algorithms that analyze data and recognize patterns, used for classification and regression analysis. The basic SVM takes a set of input data and predicts, for each given input, which of two possible classes forms the output, making it a non-probabilistic binary linear classifier. Given a set of training examples, each marked as belonging to one of two categories, an SVM training algorithm builds a model that assigns new examples into one category or the other. An SVM model is a representation of the examples as points in space, mapped so that the examples of the separate categories are divided by a clear gap that is as wide as possible. New examples are then mapped into that same space and predicted to belong to a category based on which side of the gap they fall on.

In addition to performing linear classification, SVMs can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces.

Formal Definition

More formally, a support vector machine constructs a hyperplane or set of hyperplanes in a high- or infinite-dimensional space, which can be used for classification, regression, or other tasks. Intuitively, a good separation is achieved by the hyperplane that has the largest distance to the nearest training data point of any class (so-called functional margin), since in general the larger the margin the lower the generalization error of the classifier.

Whereas the original problem may be stated in a finite dimensional space, it often happens that the sets to discriminate are not linearly separable in that space. For this reason, it was proposed that the original finite-dimensional space be mapped into a much higher-dimensional space, presumably making the separation easier in that space. To keep the computational load

reasonable, the mappings used by SVM schemes are designed to ensure that dot products may be computed easily in terms of the variables in the original space, by defining them in terms of a kernel function *K(x,y)* selected to suit the problem. The hyper planes in the higher-dimensional space are defined as the set of points whose dot product with a vector in that space is constant. The vectors defining the hyper planes can be chosen to be linear combinations with parameters α_1 of images of feature vectors that occur in the database. With this choice of a hyper plane, the points x in the feature space that are mapped into the hyper plane are define by the relation:

$\sum_i \alpha_i K(x_i, x) = constant$, note that if *K(x,y)* becomes small as *y* grows further away from *x*, each term in the sum measures the degree of closeness of the test point *x* to the corresponding data base point x_i. In this way, the sum of kernels above can be used to measure the relative nearness of each test point to the data points originating in one or the other of the sets to be discriminated. Note the fact that the set of points *x* mapped into any hyper plane can be quite convoluted as a result, allowing much more complex discrimination between sets which are not convex at all in the original space.

Linear SVM

Given some training data D, a set of n points of the form

$$\mathfrak{D} = \{(x_i, y_i) | x_i \in \mathbb{R}^p, y_i \in \{-1,1\}\}_{i=1}^{n}$$

Where the y_i is either 1 or -1, indicating the class to which the point X_i belongs. Each X_i is a *p*-dimensional real vector. We want to find the maximum-margin hyper plane that divides the points having $y_i = 1$ from those having $y_i = -1$. Any hyper plane can be written as the sets of points x satisfying

$$w \cdot x - b = 0,$$

where . denotes the dot product and w the normal vector to the hyper plane. The parameter $\frac{b}{\|w\|}$ determines the offset of the hyper plane from the origin along the normal vector w.

If the training data are linearly separable, we can select two hyper planes in a way that they separate the data and there are no points between them, and then try to maximize their distance.

The region bounded by them is called “the margin”. These hyper planes can be described by the equations

$\mathrm{w} \,.\, \mathrm{x} - \mathrm{b} = 1$

and

$\mathrm{w} \,.\, \mathrm{x} - \mathrm{b} = -1$

By using geometry, we find the distance between these two hyper planes is $\frac{2}{\|w\|}$, so we want to minimize $\|w\|$. As we also have to prevent data points from falling into the margin, we add the following constraint: for each *i* either

$\mathrm{w} \,.\, x_i - b \geq 1$ for $\mathrm{x_i}$ of the first class

or

$\mathrm{w} \,.\, x_i - b \leq 1$ for $\mathrm{x_i}$ of the second.

This can be rewritten as:

$y_i(w \,.\, x_i - b) \geq 1,$ for all $1 \leq \mathrm{i} \leq \mathrm{n}$.

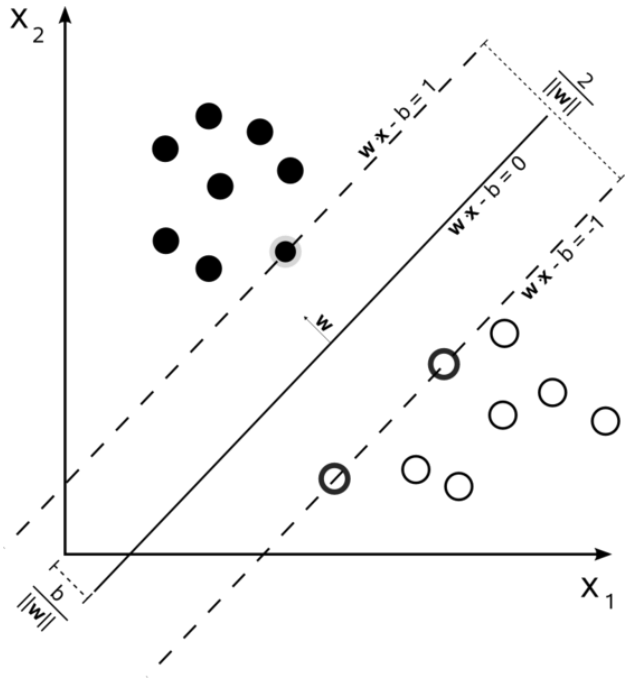
We can put this together to get the optimization problem:

Minimize (in w,b)

$\|w\|$

Subject to (for any i = 1, …, n)

$y_i(w \,.\, x_i - b) \geq 1$



Maximum-margin hyper plane and margins for an SVM trained with samples from two classes. Samples on the margin are called the support vectors.

F-test Statistics

An F-test is any statistical test in which the test statistic has an F-distribution under the null hypothesis. It is most often used when comparing statistical models that have been fitted to a data set, in order to identify the model that best fits the population from which the data were sampled. Exacts F-test mainly arise when the models have been fitted to the data using least squares.

Common examples of F-tests

Examples of F-tests include:

- The hypothesis that the means of several normally distributed populations, all having the same standard deviation, are equal. This is perhaps the best-known F-test, and plays an important role in the analysis of variance (ANOVA).
- The hypothesis that a proposed regression model fits the data well.

- The hypothesis that a data set in a regression analysis follows the simpler of two proposed linear models that are nested within each other.
- Scheffe's method for multiple comparisons adjustment in linear models.

Formula and calculation

Most F-tests arise by considering a decomposition of the variability in a collection of data in terms of sums of squares. The test statistic in an F-test is the ratio of two scaled sums of squares reflecting different sources of variability. These sums of squares are constructed so that the statistic tends to be greater when the null hypothesis is not true. In order for the statistic to follow the F-distribution under the null hypothesis, the sums of squares should be statistically independent, and each should follow a scaled chi-squared distribution. The latter condition is guaranteed if the data values are independent and normally distributed with a common variance.

Multiple-comparison ANOVA problems

The F-test in one-way analysis of variance is used to assess whether the expected values of a quantitative variable within several pre-defined groups differ from each other. For example, suppose that a medical trial compares four treatments. The ANOVA F-test can be used to assess whether any of the treatments is on average superior, or inferior, to the others versus the null hypothesis that all four treatments yield the same mean response. This is an example of an "omnibus" test, meaning that a single test is performed to detect any of several possible differences. Alternatively, we could carry out pairwise tests among the treatments (for instance, in the medical trial example with four treatments we could carry out six tests among pairs of treatments). The advantage of the ANOVA F-test is that we do not need to pre-specify which treatments are to be compared, and we do not need to adjust for making multiple comparisons. The disadvantage of the ANOVA F-test is that if we reject the null hypothesis, we do not know which treatments can be said to be significantly different from the others — if the F-test is performed at level α we cannot state that the treatment pair with the greatest mean difference is significantly different at level α.

The formula for the one-way ANOVA F-test statistic is

$$F = \frac{explained\ variance}{unexplained\ variance},$$

or

$$F = \frac{between\text{-}group\ variability}{within\text{-}group\ variability}.$$

The "explained variance", or "between-group variability" is

$$\sum_{ij} n_i(Y_{i\cdot} - \bar{Y})^2 \,/\, (K-1)$$

Where $\overline{Y_{i\cdot}}$ denotes the sample mean in the i^{th} group, n_i is the number of observations in the i^{th} group, $\bar{Y}$ denotes the overall mean of the data, and K denotes the number of groups. The "unexplained variance", or "within-group variability" is

$$\sum_{ij} \left(Y_{ij} - \overline{Y_i}\right)^2 \,/\, (N-K),$$

Where Y_{ij} is the j^{th} out of K groups and N is the overall sample size. This F-statistic follows the F-distribution with K-1, N-K degrees of freedom under the null hypothesis. The statistic will be large if the between-group variability is large relative to the within-group variability, which is unlikely to happen if the population means of the groups all have the same value. Note that when there are only two groups for the one-way ANOVA F-test, $F = t^2$ where t is the Student's t statistic.

Regression problems

Consider two models, 1 and 2, where model 1 is 'nested' within model 2. Model 1 is the restricted model, and Model 2 is the unrestricted one. That is, model 1 has p1 parameters, and model 2 has p2 parameters, where p2 > p1, and for any choice of parameters in model 1, the same regression curve can be achieved by some choice of the parameters of model 2. (We use the convention that any constant parameter in a model is included when counting the parameters. For instance, the simple linear model y = mx + b has p = 2 under this convention.) The model with more parameters will always be able to fit the data at least as well as the model with fewer parameters. Thus typically model 2 will give a better (i.e. lower error) fit to the data than model 1. But one often wants to determine whether model 2 gives a significantly better fit to the data. One approach to this problem is to use an F test.

If there are n data points to estimate parameters of both models from, then one can calculate the F statistic, given by

$$F = \frac{\left(\frac{RSS_1 - RSS_2}{p2 - p1}\right)}{\left(\frac{RSS_2}{n - p2}\right)}$$

Where RSS_1 is the residual sum of squares of model i. If your regression model has been calculated with weights, then replace RSS_i with $\chi 2$, the weighted sum of squared residuals. Under the null hypothesis that model 2 does not provide a significantly better fit than model 1, F will have an F distribution, with (p2 − p1, n − p2) degrees of freedom. The null hypothesis is rejected if the F calculated from the data is greater than the critical value of the F-distribution for some desired false-rejection probability (e.g. 0.05).