Robust gait recognition using adaptive random depth subspace from depth information

By QU YUAN

A thesis submitted to
School of Information Science,
Japan Advanced Institute of Science and Technology,
in partial fulfillment of the requirements
for the degree of
Master of Information Science
Graduate Program in Information Science

Written under the direction of
Associate Professor Atsuo Yoshitaka

September, 2014

Robust gait recognition using adaptive random depth subspace from depth information

By QU YUAN (1210213)

A thesis submitted to
School of Information Science,
Japan Advanced Institute of Science and Technology,
in partial fulfillment of the requirements
for the degree of
Master of Information Science
Graduate Program in Information Science

Written under the direction of
Associate Professor Atsuo Yoshitaka

and approved by
Associate Professor Atsuo Yoshitaka
Professor Jianwu Dang
Associate Professor Kazunori Kotani

August, 2014 (Submitted)

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Research background

Identifying individual by biometrics is a challenging task and has been studied extensively. Biometric modalities are the distinctive and measurable characteristics for labeling and describing individuals, and these modalities are often categorized as physiological and behavioral characteristics. The former characteristics are related to the shape of a single body part. Examples include, but are not limited to fingerprint, veins, face, DNA, palm print, hand geometry and iris. The latter characteristics are related to the pattern of behavior of a person, such as signature and voice.

Human walking is a complex locomotion which involves synchronized movements of body parts, joints and the interaction among them. Every individual has his/her own walking pattern which is termed as 'gait'.

As a behavioral characteristic, human gait is advantageous for individual recognition over other traditional biometric modalities such as iris, DNA and fingerprint. Since gait features can neither be imitated nor hidden easily. Furthermore, walking data acquisition doesn't require a close distance and high quality sensors. In visual surveillance the distance between a camera and a people under surveillance is often large. The camera may be set on the top of a building, in the ceiling of a corridor, or wherever a panorama of the surveyed area could be obtained. In addition, unobtrusiveness without cooperation or contact of the subject also makes gait data acquisition secluded. For example, the subject may need to put his or her hand in a scanner to obtain the palm prints or finger prints. The requirements are even more strict for iris recognition, since the subjects have to look into an eyepiece. Above-mentioned advantages make gait recognition attractive in many applications such as intelligent video surveillance, access control, biomedical research. Recently, gait recognition has been used as a automated surveillance tool for tracking and detecting the criminal when other biometric modalities are obscured from a crime scene.

1.2 Overview of Gait recognition

1.2.1 Gait recognition using Binary image

Over the last twenty years, researchers presented numerous successful gait recognition techniques with binary images. These techniques were classified into two major categories called model-based and appearance-based approach [1].

Model-based approaches generally represent the dynamic and static components of gait by modeling the intrinsic kinematics of human motion and body structure. The models are also constructed by some locomotion like speed, the length of the rhythm, pitch, and so on. Then, the extracted gait features is matched with these model components directly. Thus, these approaches are less affected by covariates related either to the subject itself (e.g., different clothing) or to the environment (e.g., walking on the different ground). For instance, Johansson et al. [2] presented point light walker method by watching video sequences of lights affixed to several key joints of a walker. However, wearable devices for gait recognition are expensive and unreliable in real-world applications. Cunado et al. [3] presented a model-based moving feature extraction approach by using a Fourier series to describe the motion of the thigh joint trajectories, and applied temporal evidence gathering techniques to extract the moving model from a sequence of images. The simulation results highlighted potential performance benefits in the presence of noise and high levels of self-occlusion. Nonetheless, this method is applicable to the case where the region of subject's thighs is visible. A robust method proposed by Ding et al. [4], that was based on modeling shape variation information extracted by Local Linear Embedding (LLE), and was recasting the recognition problem into a robust model validation form. However, the major limitation of model-based approaches is that it requires high quality gait images and high computational cost for motion model construction for each body part [5].

In contrast, most of the current approaches are appearance-based, which directly utilized the silhouettes of gait sequences for feature extraction without developing any model. As a representative appearance-based approach, Gait Energy Image (GEI) is a spatio-temporal template calculated by averaging all images extracted from video in a full gait cycle [6]. This averaging process seemingly discards the information, however, since the noise in each frame is independent, averaging removes a substantial part of the redundant and noise information. This method achieves good experimental results with less computational cost. As an improvement of Gait Energy Image method, Gait Entropy Image (GEnI) was proposed by Bashir et al. [7], which computed the entropy in each single image. The Dynamic body areas which undergo consistent relative motion during a gait cycle (e.g. legs, arms) will result to a high gait entropy value, whereas those areas that remain static (e.g. trunk) would give rise to low values. This method is capable of reducing the negative effect of covariates in individual recognition. The appearance-based approaches are suitable for practical application since they operate on binary silhouettes which could be of low quality, and no color or texture information is needed. Nevertheless, most of appearance-based approaches are view dependent and perform best when a side view

point image is applied. In addition, as the limitation of the high sensitivity for the body shape changing, most of appearance-based approaches are hard to be reliable enough in some real-life scenarios, such as variations of clothes and carrying a baggage.

Focusing on robust gait recognition of individual, several part-appearance-based approaches were proposed [8] [9] [10] [11]. These approaches treat each body part independently and obtain good performance in large intra-class variation database. However, all of them are based on a strong assumption where the clothing types appear in both probe (test) set and training stage. Based on this assumption, each body part has a fixed weight for every sequence in probe (test) set ignoring the clothing types and walking postures. Moreover, considering real-world gait recognition of individual, it is not realistic to train a model for all possible walking conditions for each subject. Table 1.1 shows the summary of studies related to gait recognition using binary images.

Table 1.1: Summary of studies using binary images

Model-based methods	**Appearance-based methods**
Johansson et al. (1975) [2]	Han et al. (2006) [6]
Niyogi et al. (1994) [1]	Bashir et al. (2009) [7]
Cunado et al. (2003) [3]	Hossain et al. (2010) [11]
Yam et al. (2004) [5]	Rokanujjaman et al. (2012) [9]
Ding et al. (2008) [4]	Rokanujjaman et al. (2013) [10]
	Islam et al. (2013) [8]

1.2.2 Gait recognition using Depth image

Recent advances in human-computer interaction (HCI) technology resulted in development of new depth sensors that enable users to interact with the computer by movement. One of the leading products in that area is the Microsoft Kinect. The depth sensor provided superiority to solve the problem of viewpoint dependency, where the depth information of physical distance from the sensor to a human body part at each pixel can be captured efficiently. Based on this advantage, the accuracy of frontal gait recognition in appearance-based method has great potential to be improved. However, compared with the number of algorithms that are applied for binary images in gait recognition, studies on depth gait recognition is relatively fewer.

Sivapalan et al. [12] improved the concept of GEI to 3D and proposed a new gait representation called Gait Energy Volumes (GEV) for depth gait recognition in frontal view point image. Their experimental results show improvement over the traditional appearance-based approaches (e.g. GEI) in frontal depth gait recognition. In case where only normal walking condition is involved in his database, there is no way to prove that their method is effective for robustness test. Meanwhile, Hofmann et al. [13] proposed an effective method for depth gait representation. Motivated by the idea of histograms of oriented gradients (HOG) [14], depth gait sequence is represented by a Depth Gradient Histogram Energy Image (DGHEI) which improves the accuracy in depth gait recognition.

Differently from above-mentioned appearance-based method using depth images, Milovanovic et al. [15] proposed a model-based method using the skeleton structure of human body based on the technique of skeleton tracking to extract the key joints. Because of the drawbacks of the recording device, only 20 joints are capable to be extracted. It caused much worse performance than the above-mentioned appearance-based method using depth image. Table 1.2 shows the summary of studies related to gait recognition using depth images.

Table 1.2: Summary of studies using depth images

Model-based methods	**Appearance-based method**
Milovanovic et al. (2013) [15]	Sivapalan et al. (2011) [12]
	Hofmann et al. (2012) [13]

1.2.3 Gait databases

Since the area of gait recognition has been studied for several decades, the research community provided several publicly available databases for comparative performance evaluation. These gait databases involve different walking conditions, which include, but are not limit to different viewing angle, clothing, shoe types, walking site, carrying condition and time. The first available gait database is the UCSD ID Database [16], which contains merely six subjects in normal walking condition. The first available gait database with the variations of walking condition is CMU Mobo Database [17]. Other databases such as SOTON Small Database [18] and CASIA Database B [19] include both carrying condition and changing clothing condition.

The most comprehensive database using RGB sensor, which provides a large set of subjects (more than 100 subjects) as well as a substantial set of walking condition variations

(more than 4 kinds of variations), is the Human ID Gait Challenge database [20]. The TUM GAIT from Audio, Image and Depth (GAIT) database [21], which is the only database available in public, using depth sensor as recording device. Gait recognition for that dataset is considered to be challenging in the field of gait recognition. This database contains total 305 subjects with multiple walking conditions, such as time, carrying condition and shoe variation. Only side view is used throughout the recording process.

1.3 Motivation and Objectives

As describe in section 1.2, researchers proposed a number of approaches for identifying individual by gait. However, there are a number of covariates that affect the performance of gait recognition of individual when appearance-based approaches are applied. The study in [20] considers three major covariates that affect gait recognition: viewing angle, changing clothes, carrying objects. The worst performance is confirmed when a frontal viewpoint is applied that is captured by RGB sensors. This is because the directions of legs' movement are invisible in the human body silhouettes while subject walks toward the sensor.

Differently from traditional RGB sensors, gait information provided by depth sensors, such as Microsoft Kinect, contains the depth information of physical distance from the sensor to a human body part at each pixel while walking, and it shows good performance in frontal gait recognition. However, the existing methods focusing on depth gait data demonstrated low accuracy by some cofactors such as clothes or carrying objects, since they treated each body part equally.

In this work, the author concentrates on exploring the applicability of depth information for robust gait recognition of individuals from frontal viewpoint, and proposes an appearance-based approach named adaptive random depth subspace (ARDS) which satisfies the following:

- The ARDS avoids the negative effect of covariates (changing clothes and carrying objects) and improves the accuracy in gait recognition of individual.
- Only normal walking condition (wearing T-shirt, regular pair of pants and natural swinging arms) is assumed in gallery set as references, which is realistic in real-world applications.
- The proposed method is capable of selecting gait features from each body part adaptively without assuming the walking condition in each test gait sequence.

Until now, the only available depth gait database in public, TUM GAIT from Audio, Image and Depth (GAIT), was offered for the investigation of gait recognition of individual using depth images. However, there are three conditions in this database that do not meet the experimental requirement in this work.

- Gait data was recorded only from side viewpoint in their database.
- 32 of 305 subjects were recorded both in winter and spring and therefore there are only two clothing types (down jacket and full T-shirt) for each of the 32 subjects.
- Only one kind of carrying condition (knapsack) was recorded in their database.

In summary, considering the lack of intra-class variation in the TUM GAIT from Audio, Image and Depth (GAIT) database, the author decided to prepare a new depth gait database with larger intra-class variation of frontal depth video frames by himself, although the GAIT is one of the largest database with respect to subject number. For gait recognition of individual using depth images, a low-cost and easy-to-install system is more suitable, that encouraged the author to choose the Microsoft Kinect XBOX 360 sensor to acquire depth images. The proposed appearance-based approach ARDS will be tested on this new depth gait database and evaluated its effectiveness by comparing with state-of-the-art methods.

In summary, the major contribution of this work is that the author develop a novel appearance-based approach ARDS for the purpose of attaining the improved accuracy in robust depth gait recognition of individual.

1.4 Structure of Thesis

Structure of this paper is as follows:

- **Chapter 1: Introduction**

  In this chapter, the motivation and objectives of this study are described. Besides, a literature survey of related work is described along with their advantages and drawbacks. Especially, a summary of publicly available gait databases is reported to evaluate the insufficient walking conditions for real scenarios.

- **Chapter 2: Preprocessing of Depth image**

  In Chapter 2, the author introduces the preprocessing of depth gait recognition. The preprocessing methods include background subtraction, gait cycles detection, noise reduction and human silhouette alignment.

- **Chapter 3: Gait Recognition with Depth Image**

  In Chapter 3, the author discusses an appropriate depth gait representation that is computed from the series of static gait images. Then, the author proposes adaptive random depth framework that is applied to select subspaces form gait representation as the local descriptors. Classification strategy is also presented to decide the subject ID in test sequences.

- **Chapter 4: Evaluation of ARDS**

  In Chapter 4, a new robust depth gait database is presented. Then, the experimental results of proposed method are shown by comparing them with state-of-the-art methods. Finally, the effectiveness evaluation of two significant parameters in the proposed method is carried out.

- **Chapter 5: Conclusion and Future Work**

  The summary of presented method will be shown in Chapter 5. Then, the author discusses the possible improvement and extension of the proposed method.

Chapter 2

Preprocessing of Depth Image

Generally, preprocessing of depth image is the first step in depth gait recognition. By capturing the gait data with the Microsoft Kinect sensor, the author obtained the gait data from two sources, which consist of a color stream and a depth stream. The color video frames were extracted frame-by-frame as jpeg images and depth gait data was extracted as raw 16 bit binary format from the depth stream before the preprocessing stage. Because of the purpose of this study, the gait data extract from color video frame was not used in this work. The author mainly introduces the preprocessing of depth image in the following section. The following flow chart shows the procedure of processing.

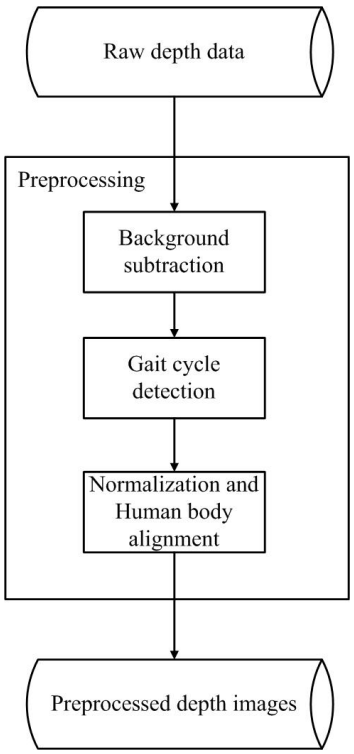


Figure 2.1: Flow chart of preprocessing

2.1 Background subtraction

Background subtraction is a technique in the fields of image processing and computer vision where foreground is extracted for further processing. Generally, an image region of interest includes human body in its foreground, where segmentation is performed to remove a floor, ceiling and other objects.

A robust background subtraction algorithm should handle the following conditions [22]:

- Lighting changes
- Repetitive motions from clutter.
- Long-term scene changes

In traditional gait recognition using binary images, the walking subject in each color frame is extracted from the background and converted into binary image by detecting the walking subject by taking the difference between the current frame and a reference background frame. Basically, the reference background frame must be the representation of the scene with no moving objects and must be kept updated regularly so as to adapt to the varying luminance conditions and geometry settings. In this case, the gait data should be captured in a controlled environment. The Figure 2.3 (a) shows the result of background subtraction.

In gait recognition using depth images, depth information is applied to segment a subject from the background reliably since the subject is less affected by typical color segmentation issues, such as shadows or highlighted regions [23] . The reference background frame is defined by depth information in the scene without subject. Since the distance between subject and background are relatively large, a large margin exists and simple thresholding of the depth difference results in good segmentation [13].

In order to extract the human body silhouette form the depth gait data, subject's silhouette segmentation is carried out by adopting a novel skeleton-tracking technique provided by Microsoft [24] in this work. The skeleton-tracking technique accepts the depth data coming from the Kinect sensor and performs a process called 'skeletonization'. It detects individual users who are within the view of the sensor and assigns a tracking ID, then tracks the positions of each body parts. However, the algorithm of tracking is not open to public.

In this work, the raw depth data captured by the Kinect sensor is encoded by 16 bit for each pixel in the depth frame. The 0-2th bits correspond to the storage of player index (tracking ID) provided by aforementioned technique, while 3-15th bits signify the storage of depth value. Hence this bit mask provides a simple and rapid way to classify the pixels into two categories: Pixels of background and Pixels within a human body. The final

result is visualized as shown in Figure 2.3 (b) by grayscale. The higher intensity value in each grayscale corresponds to longer physical distance at this position

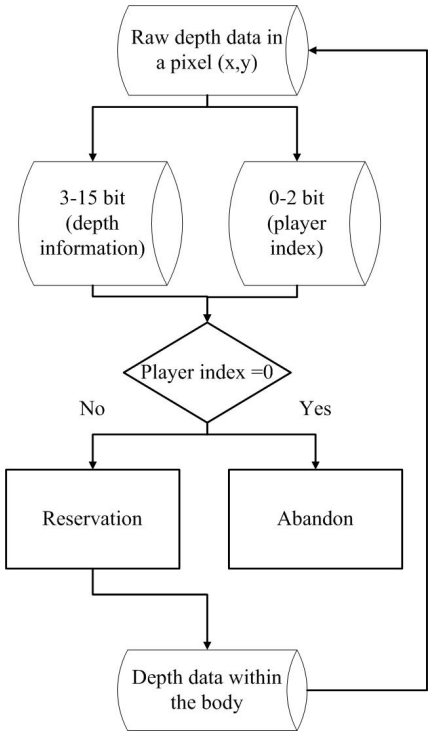
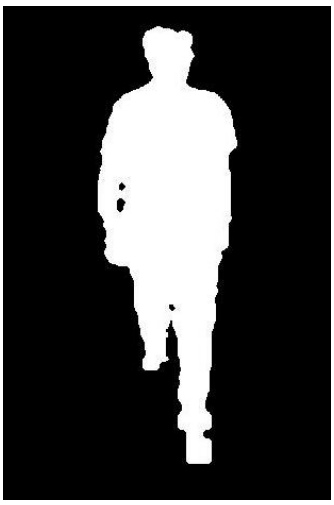


Figure 2.2: Flow chart of Background subtraction using depth information



(a) Background subtraction result of binary image

(b) Background subtraction result of grayscale image

Figure 2.3: Background subtraction result

2.2 Gait cycle detection

Gait cycle detection is generally the first step in the process of gait recognition, since the walking pattern of individual is represented by the gait features extracted through each gait cycle completely. One of the research direction in the gait recognition is the gait cycle detection.

Medical studies [23] [25] suggests that the human gait forms periodic motion and therefore, cyclic phases of the walking sequence can be extracted. In this case, the body posture during the walking process changes periodically and the upper and lower limbs move symmetrically. Especially, if a person is affected by neurological diseases or injury, asymmetry of lower limb movement will occur during the walking process. However, in this work, the author assumes the healthy adults as the subject and excludes the above-mentioned cases.

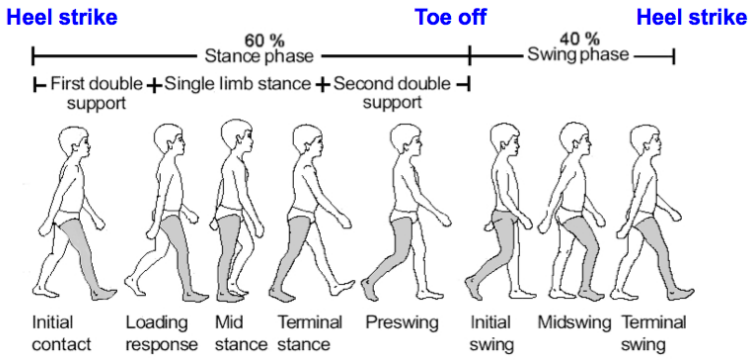


Figure 2.4: Gait cycle [26]

Figure 2.4 shows an example of gait cycle that consists of successive foot contacts of the same limbs [26]. One gait cycle can be described as the motion from the initial placement of the supporting heel on the ground until the same heel contacts the ground for the second time. The gait cycle can be divided into two phases:

- Stance: defined as the interval in which the foot is on the ground (60% of duration of the gait cycle)
- Swing: defined as the interval in which the foot is not in contact with the ground (40% of duration of the gait cycle) clutter.

In this study, author proposed a simple effective method, which is called 'symmetry quantization', to detect the gait cycle. The basic idea is to detect the difference of pixel number between lower left and lower right body of the subject in each frame. In the first step, the author calculates the centroid of human body silhouette by the following

equation:

$$x_c = \frac{1}{K}\sum_{i=1}^{K} x_i \tag{2.1}$$

$$y_c = \frac{1}{K}\sum_{i=1}^{K} y_i \tag{2.2}$$

Here K denotes the total number of the pixel within a human body silhouette. Variable x_i and y_i are the x-coordinate and y-coordinate of pixel (x_i,y_i), respectively. (x_c,y_c) is the coordinate of centroid of human body silhouette.

To quantize the symmetry of t^{th} gait frame in an input gait video, a horizontal section which is crossed the centroid is set by the author to divide the gait frame into two subimages (upper subimage and lower subimages). Then, the author discards the upper subimage and remain the lower subimage for gait cycle detection. Let I^t_{left} and I^t_{right} be the pixel number of left and right regions of t^{th} lower subimage, respectively. The difference C^t is calculated by the following equation:

$$C^t = I^t_{left} - I^t_{right} \tag{2.3}$$

The Figure 2.5 shows the differences calculated from each frame of a subject. The variability of difference is high because of the asymmetric limbs swing. The positions of zero ((b),(d)) correspond to the instantaneous gait posture where one leg moves across the other. Here, two arms fall naturally, which lead the minimal difference between two image regions. The positions of maxima ((a),(e))and minima ((c))correspond to the instantaneous gait postures of 'left-food strike' and 'right-food strike'. The Figure 2.6 shows the corresponding gait frame at position (a),(b),(c),(d),(e).

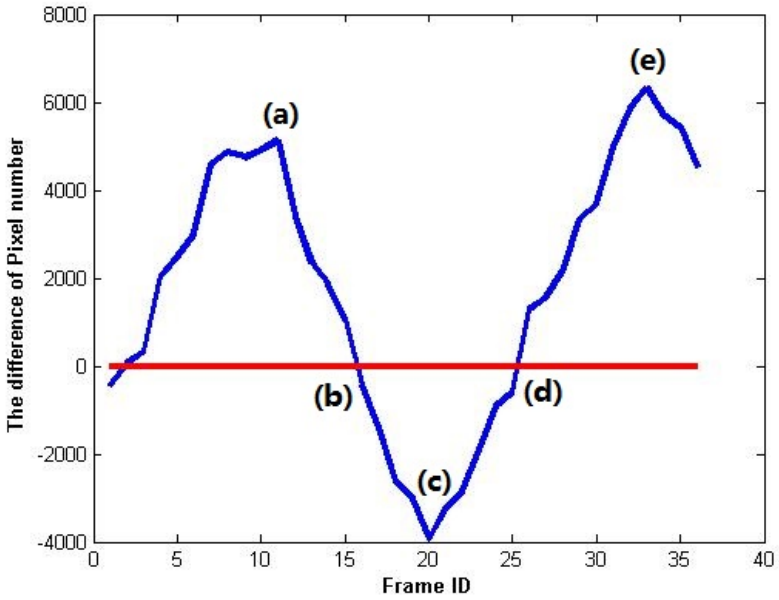


Figure 2.5: Gait cycle detection

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

Figure 2.6: The sample frames in the positions of (a),(b),(c),(d),(e)

Note that due to the detection range limitation of Microsoft Kinect, it is difficult to record multiple gait cycles continuously. The situation that two maxima and two minima of pixel number difference occur simultaneously will not be appear in one procedure of gait cycle detection. Based on the definition of gait cycle described in this section, the images whose difference value fall into the range between two maxima ((a)-(e)) or two minima are extracted and constituted a gait sequence for the following processing. However, the proposed gait cycle detection method perform best when the lower limbs of a subject aren't disturbed. In some situations that the subjects wear long coat or skirt, the performs of this method are much worse. In this case, most of the procedure of gait cycle detection are accomplished manually based on the definition of gait cycle.

2.3 Depth image normalization

In this section, the author briefly introduces the procedures of depth image normalization and human body alignment. The effect of varying depth due to the distance changes between the walking subjects and the Kinect sensor should be reduced. As described in section 2.2, the coordinates of the centroid (x_c,y_c) of the human body is computed by equation 2.1 and equation 2.2. Then, the depth value of each pixel within the human body silhouette is normalized by a translate procedure, as shown in the following equation:

$$\mathrm{R}(x,y) = D(x,y) - D(x_c,y_c) \tag{2.4}$$

Here, $\mathrm{R}(x,y)$ denotes the relative distance between the centroid of human body surface and each pixel (x,y). $D(x,y)$ is the distance of the subject point corresponding the pixel (x,y) on the depth silhouette. $D(x_c,y_c)$ is the absolute distance measured by depth sensor, which denotes the distance between each human body part and the depth sensor. Finally, a median filer is used to reduce the noises. Thus, the procedure of distance normalization is completed.

In next step, each human body silhouette is normalized by scale variation to the same

size. The size of human body silhouette stays nearly the same during a gait cycle in side viewpoint images. However, in the frontal viewpoint, the closer subject walks, the larger silhouette will be obtained. Assume that I_t is the t^{th} frame in a full gait cycle. $a(t)$ is the zoom factor of body height while $b(t)$ is the zoom factor of body width for I_t. $a(t)$ and $a(t)$ is defined as the following equation:

$$a(t) = \frac{H(n)}{H(t)} \tag{2.5}$$

$$b(t) = \frac{W(n)}{W(t)} \tag{2.6}$$

Where $H(n)$ and $W(n)$ denotes the height and width of human body silhouette in the last frame in a gait cycle. Meanwhile, $H(t)$ and $W(t)$ denotes the height and width of I_t. In this work, the author applied the bilinear interpolation for size normalization. Bilinear interpolation is an extension of linear interpolation for interpolating functions of two variables on a regular grid. The author first conducts linear interpolation in the x-direction and y-direction respectively:

$$f_a = f(x, y) + (f(x + 1, y) - f(x, y)) \times p \tag{2.7}$$

$$f_b = f(x, y + 1) + (f(x + 1, y + 1) - f(x, y + 1)) \times p \tag{2.8}$$

Where $f(x, y)$ is the original distance value in pixel (x, y). Then, the desired estimate of $G(x, y)$ is computed by the following equation:

$$G(x, y) = f_a + (f_b - f_a) \times q \tag{2.9}$$

Here, p and g are two weights, which are calculated from the distance value of neighbour pixels. Finally, each human body silhouette is aligned to the center of subject's neck. The normalized images will be used in the following process for depth gait recognition of individual.

Chapter 3

Gait Recognition with Depth Image

3.1 Depth gait representation

In this section, the author will briefly introduce the gait representation technique based on depth oriented gradient histogram [13], which will be applied in our proposed method as a gait template for selecting adaptive random depth subspace.

Each subject's gait sequence is defined as a series of static images which are extracted frame-by-frame from the gait video in one full gait cycle. The method of extracting gait features from gait sequence is always the key problem in appearance-based gait recognition. Most existing approaches extract silhouettes frame-by-frame from a gait sequence, by composing a feature set to express the whole gait process. As a matter of fact, this may result in a high dimensional feature space including a large amount of redundant information.

An excellent gait representation should satisfy the following characteristics:

- *Containing both dynamic (e.g. locomotion frequency) and static (e.g. shape of each body part) information.*
- *Paying less storage space and computation cost.*
- *Being less sensitive to the noises.*
- *Containing plentiful discrimination clues.*

As the described in Chapter 1, multiple gait representations are presented for gait recognition of individual using binary images, such as GEI, GEnI. Both of these two representations, as shown in Figure 3.1, follow the aforementioned characteristics [6] [7]. In Figure 3.1 (a), a pixel with higher intensity value in GEI means that human walking occurs more frequently at this position (i.e., with higher energy). As an improvement of GEI, the GEnI captures most of motion information and is robust to covariate condition changes

that affect the static areas of human body, as shown in Figure 3.1 (b). However, in traditional gait recognition with RGB images, each pre-processed image is converted into binary image. In contrast, the depth information within person's silhouette is represented by grayscale in our work. The grayscale contain a lot of edges and depth gradients information, which provide additional discrimination clues in the following recognition. Thus, a new depth gait representation is proposed in this work, which is motived by DGHEI.

(a) GEI (b) GEnI

Figure 3.1: Gait representation using binary images

The method of extracting gait features from gait sequence is always the key problem in appearance-based gait recognition. Most existing approaches extract silhouettes frame-by-frame from a gait sequence, by composing a feature set to express the whole gait process. As a matter of fact, this may result in a high dimensional feature space including a large amount of redundant information. Gait Energy Image provides an efficient solution that averaging all images extracted from video in a full gait cycle. Thus, in this work, the preprocessed depth gait images are averaged over a gait sequence at first:

$$\mathbf{H}(\mathbf{i},\mathbf{j},\mathbf{d}) = \frac{1}{T}\sum_{t=1}^{T}\mathbf{I^t}(\mathbf{i},\mathbf{j},\mathbf{d}) \tag{3.1}$$

Here, i and j denote the position of a pixel and d denotes the gray intensity value of this pixel. T denotes the number of gait image in a gait sequence. Then, the following procedure of computation of depth gait representation is similar with DGHEI computation.

The idea of DGHEI is motivated by the concept of 'histograms of oriented gradients (HOG)', which are feature descriptors used in computer vision and image processing for the purpose of object detection.

The basic idea of HOG is that local object appearance and shape can often be characterized rather well by the distribution of local intensity gradients or edge directions, even without precise knowledge of the corresponding gradient or edge positions [14].

The HOG descriptor maintains many advantages over other descriptors.

1.The HOG descriptor operates on localized cells, which makes the method upholds invariance to geometric and photometric transformations, except for object orientation. Such changes would only appear in larger spatial regions.

2.Coarse spatial sampling, fine orientation sampling, and strong local photometric normalization permits the individual body movement of pedestrians to be ignored as long as they maintain a roughly upright position.

3.The HOG computational cost is less than other descriptors such as SIFT, since the HOG descriptor ignores the scale-invariant and rotation-invariant condition.

Following the procedure of HOG computation, the extraction of DGHEI from gait images in a first step consists of calculating histograms of oriented gradients at each preprocessed depth gait images I^t. As evaluated by Dalal et al. [14], the most common method is simply to apply the mask, which is termed as '1-dimensional center point discrete derivative mask', in one or both of the horizontal and vertical directions. Specifically, this method requires filtering the color or intensity of the image with the following filter mask:

$$[-1, 0, 1] \tag{3.2}$$

Note that two arrangements **R-HOG** and **C-HOG** (for rectangular and circular HOG) are proposed by Dalal et al. [14]. DGHEI apply the former as block geometries. Thus, the magnitude γ and orientation θ of the gradient at a pixel (x,y) is calculated as the following equations:

$$u(x, y) = H(x - 1, y) - H(x + 1, y) \tag{3.3}$$

$$v(x, y) = H(x, y - 1) - H(x, y + 1) \tag{3.4}$$

$$\gamma(x, y) = \sqrt{u(x, y)^2 + v(x, y)^2} \tag{3.5}$$

$$\theta(x, y) = arctan(u(x, y), v(x, y)) + \pi \tag{3.6}$$

Here, $H(x, y)$ denotes the gray intensity value at (x,y). Then, gradient orientations at each pixel are quantized discretized into one of nine orientations:

$$\hat{\theta}(x, y) = \left\lfloor \frac{9 \cdot \theta(x, y)}{2\pi} \right\rfloor \tag{3.7}$$

In the next step, each pixel within a square image chunk (typically by the fixed size of 8×8 pixels based on the experiment results conducted by Dalal et al. [14]) is weighted by an orientation-based histogram channel with regard to the orientation θ^t.

Thus, the non-overlapping chunk, which is called 'cell', is represented by 9-bin histogram of oriented gradients. The sample of block and cell in R-HOG is indicated as the following Figure 3.2.

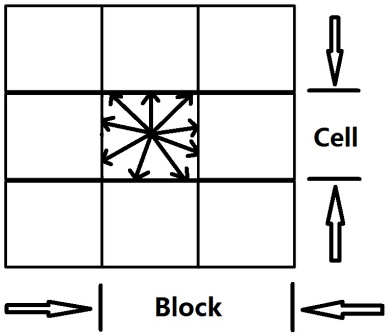


Figure 3.2: An example of block and cell (R-HOG)

Gradient intensity varies over a wide range because of local variations in illumination and foreground-background contrast [14]. Therefore effective local contrast normalization turns out to be essential for good performance. This can be done by accumulating a measure of local histogram over a larger spatial regions ('blocks') and using the results to normalize all of the cells in the block. The size of the 'block' is configured 16×16 in this study.

Here, the author applies $L2$ *norm* to normalize the reconstructed cell-based matrix. Let $\mathbf{D}$ be the unnormalized descriptor, the $L2$-*norm* normalization is described as the following equation:

$$\mathbf{D_n} = \frac{\mathbf{D}}{\sqrt{\|\mathbf{D}\|_2^2 + e^2}} \tag{3.8}$$

Here, e is the regularization parameter. Dalal et al. [14] evaluated the effectiveness of e, but the results were insensitive to e over a wide range. In this work, the e is configured as 0.001. $\mathbf{D_n}$ is the normalized vector storing the final histogram of all of the blocks one by one in a top-left to bottom-right image scan manner, the cells histogram are stored in the same manner in each block's section of $\mathbf{D_n}$.

Then, the normalized descriptor $\mathbf{D_n}$ is reconstructed into a matrix based on the location in the original image.

Figure 3.3: Visualization of cell-based gait matrix

In summary, the depth gait representation can be considered as an extension to GEI which applies depth information instead of binary human body silhouette. The final depth representation is visualized as shown in Figure 3.3. In this figure, the origin of each gradient vector denotes the position of a cell, and each cell is considered as a minimum feature unit for random depth subspace selection. The procedure of depth gait representation is illustrated in Figure 3.4.

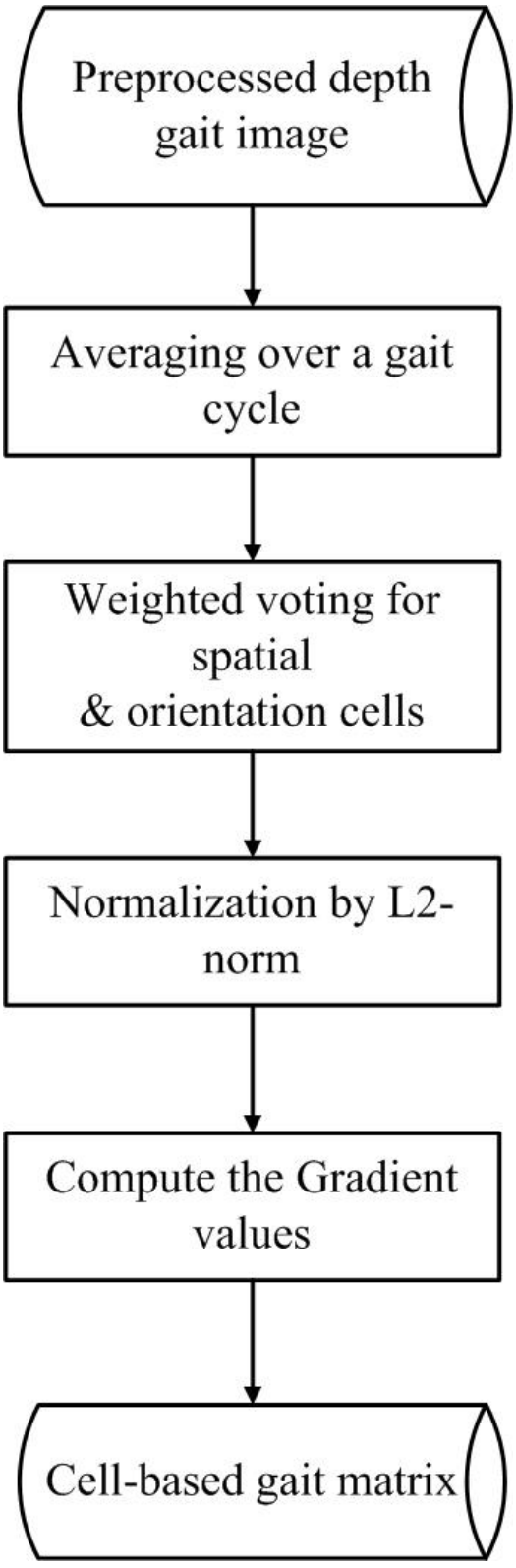


Figure 3.4: Flow chart of depth gait representation computation

3.2 Adaptive random depth subspace framework

3.2.1 Body-region segmentation

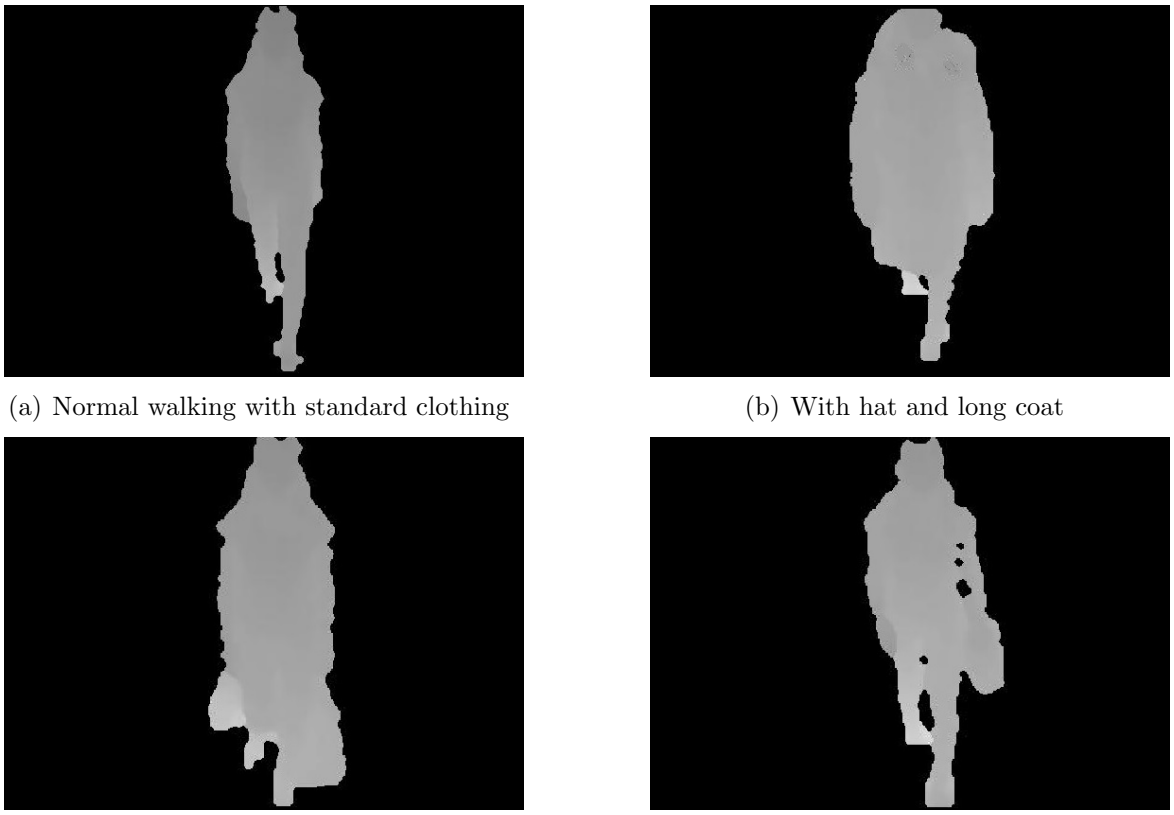
In a large intra-class variation database, each subject walks with various conditions with respect to clothing and carrying baggage. However, the notable difference of clothes types, including long coats, down jackets, hats and skirts, complicates the identification process. In this case, parts of depth gait image may be affected by partial occlusion or body shape changing.

Thus, the body shape and depth gradient may not be reliable in these body parts. In contrast, other unaffected parts maintain nearly the same. These parts contain the useful discrimination clues for individual recognition.

In conclusion, different clothing conditions and walking postures generally affect different parts of the body. The following figures show the grayscales of a walking subject in 3 different conditions. For example, consider a combination of a pair of normal pants and a half T-shirt as being the standard clothes of a subject as shown in figure 3.5 (a).

In some real-life scenarios, a subject puts a hat and wears a long coat in addition to the standard clothes, which only affects the head, trunk and thigh respectively, as shown in Figure 3.5 (b). In case where a skirt and a half T-shirt are dressed by the subject, the shape of lower body parts is changed. The oriented gradient information extraction of legs is disturbed by interference, such as the swing of skirt during walking, as illustrated in Figure 3.5 (c). However, this walking condition allows the other parts to be used for successful recognition of individual, such as head and trunk.

Similarly, which will also be the case for different walking condition is carrying objects. If the subject walks carrying a shoulder bag, the oriented gradient of one arm is quite variable compared with natural swinging arms as shown in Figure 3.5 (d). Meanwhile, the shape of human silhouette is also changed. In contrast, the oriented gradient information of legs is unobscured and undisturbed. Therefore, body-region segmentation is a mandatory for selecting the appropriate parts which are used for gait feature extraction.



(a) Normal walking with standard clothing

(b) With hat and long coat

(c) With skirt

(d) With shoulder bag

Figure 3.5: The samples of different walking conditions

To tackle robust gait recognition for individual under various walking conditions, the author segmented the human body into 4 non-overlapping parts based on known anatomical properties discussed by Dempster et al. [27]. In their work, mean values of key joints' position are presented based on an investigation of cadavers of normal appearance. In our work, the author applies the these results to assume the vertical positions of neck, pelvis and knee without detecting each body part.

When a body height is H, the human body is segmented according to the vertical position of the neck, pelvis and knee, thereby into 4 parts. The upper most section is neck (0.870H from the bottom) and the first part covers from the top of the head to the neck; then the following one is pelvis (0.480H from the bottom) and the second part covers from the neck to the pelvis; the last section is the knee (0.258H from the bottom) and that part covers from the pelvis to knee; the fourth part covers from the knee to the ground. Thus, the depth gait representation is divided into 4 parts marked 1, 2, 3 and 4, as shown in Figure 3.6. Body parts 1, 2, 3 and 4 represent separated regions which are called head, trunk, thigh and crus in anatomy, respectively.

Therefore, the proposed method considers the various contributions of different part for

robust individual recognition. Note that the discrimination efficiency of the larger part is high, while the probability of larger part being free from different clothing condition is low, and vice versa for the smaller parts [11]. The gait features should be selected and extracted independently from each body part in the following work.

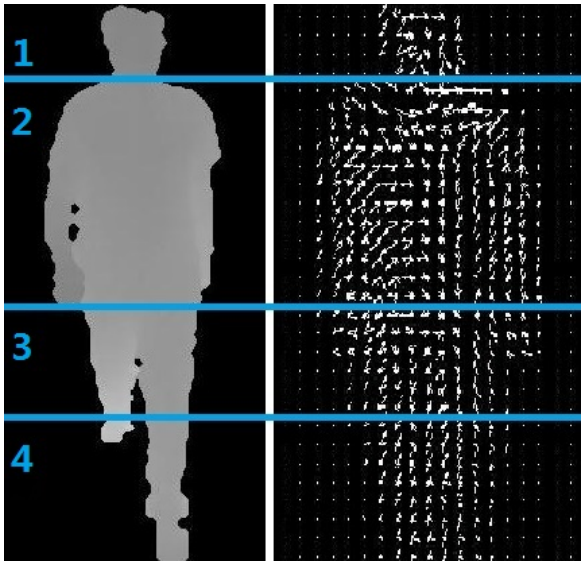


Figure 3.6: Body-region segmentation

3.2.2 Random depth subspace selection

As illustrated by Figure 3.7, in the appearance-based gait recognition, the extracted features can be divided into two classes based on the spatial locality, which are termed as global descriptor and the local descriptor.

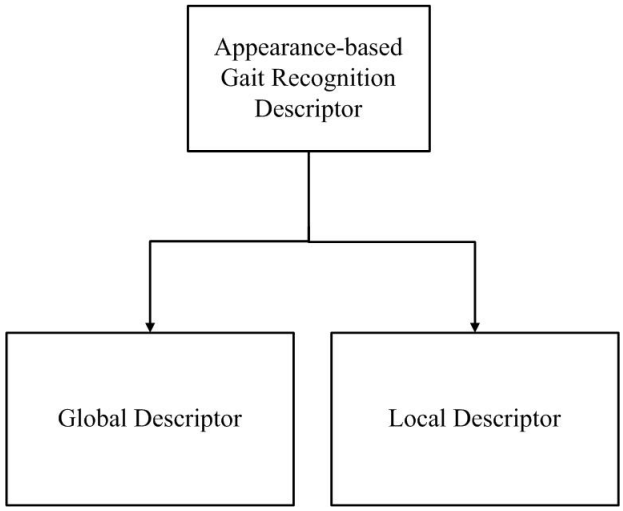


Figure 3.7: Descriptors in appearance-based gait recognition

In this section, the author will introduce the procedure of subspace selection and gait feature reconstruction by ARDS framework. The subspace selection follows the following hypothesis:

- *Local descriptor contains enough discrimination clues to identify a person, and avoiding overfitting is essential for robust gait recognition of individual.*

As a representative algorithm applying local descriptor , RSM-2DPCA method is proposed by Guan et al. [28] for gait recognition of individual. In their work, the eigenvectors of the projected feature matrix are selected randomly as the subspace for classification. However, each subspace in their work contains the whole-body gait information regardless of the subject's clothing type. The interferences by clothing variation are still retained. Meanwhile, as it was reported in [8], the subspace generated by RSM increases the computation complexity. Furthermore, this method used the same subjects in both training and testing phases to make the full hypothesis. In real applications, a testing subject is not always available for training in advance.

In this work, the idea of ARDS framework is motivated by Isam et al. [8]. However, because of random selection of image chunks (non-overlapping square regions in the binary gait image), the spatial information is lost in their work. In contrast, the author selects chunks (as named 'cell' in our work) at the same location between probe and gallery and reconstructs the gait feature by the luxuriant spatial information. Moreover, their approach is based on a strong assumption where the clothing types appear both in a probe (test) set and a training set. Following their assumption, the author assumes that each body part has a fixed weight for every sequence in probe (test) set ignoring the clothing types. To overcome this problem, the author applies an adaptive feature selection algorithm to extract features in each body part adaptively, which is realistic for assumed applications. The solution to the problem of how to select the subspace adaptively will be described in detail in section 3.2.3.

Before presenting the process of random depth subspace selection, the author will introduce the definition of subspace. Each subspace is defined as a sub matrix for gait feature reconstruction by selecting N cells from all candidate cells in p^{th} part of depth gait representation randomly. The coordinates of each cell are represented by i and j. Thus, for recognizing an individual from an input test sequence, the procedure of random depth subspace selection and construction in p^{th} part is given in Algorithm 1. In particular, all of cells which contain the same value of zero are discarded from the candidate cells, because most of these cells correspond to the background and do not contribute discrimination for individual recognition. In this algorithm, the gallery (reference) set includes all subjects and each subject has several gait sequences with normal walking condition (described in Section 1.3). The probe (test) set contains the same subjects in gallery set and several gait sequences with various walking conditions are captured for each subject.

Algorithm 1

Input:

T_p:The cell-based matrix in p^{th} part of one input gait sequence in probe set.

G_p^k: The cell-based matrix in p^{th} part of the k^{th} gait sequence in gallery set.

$Tcell_{ij}$: The cell of T_p which is located at(i,j).

$Gcell_{ij}^k$: The cell of G_p^k which is located at (i,j).

N: Total number of cells in each subspace.

K: Total number of gait sequences in the gallery set G_p^k.

m_p,n_p: $m_p \times n_p$ cells constitute each cell-based matrix in p^{th} part.

Output:

ST: Subspace of one input gait sequence in probe set.

SG^k: Subspace of the k^{th} gait sequence in the gallery set.

Procedure:

(1) $ST = \emptyset$;

(2) $SG^k = \emptyset$;

(3) **for** $n \leftarrow 1$ to N **do**

(4) select a pair values of i, j in the range randomly

$[i = 1, 2, 3 \ldots, m_p; j = 1, 2, 3 \ldots, n_p]$

(5)$ST \leftarrow [ST; Tcell_{ij}]$; ($[A; B]$ denotes that matrices A and B

integrate vertically into a new matrix)

(6) **for** $k \leftarrow 1$ to K **do**

(7) $SG^k \leftarrow \left[SG^k; Gcell_{ij}^k\right]$;

(8) **end**

(9) **end**

Assume that the author select L_p subspaces from the p^{th} part in a probe sequence, then the same number of subspaces L_p will be selected from corresponding part for each sequence belonging to gallery set. Thus the total number of subspace in each sequence is:

$$L = \sum_{p=1}^{P} L_p \tag{3.9}$$

In this equation, j is the total number of body parts and $P = 4$ in our experiment (as described in Section 3.2.1).

Then, the L subspaces are selected from a sequence of gallery set and test sequence respectively by running Algorithm 1 L times. Thus the reconstructed gait feature matrix is calculated by connecting each subspace one by one as following:

$$\mathbf{ALLSG^k} = \left((\mathbf{SG^k_1})^T, (\mathbf{SG^k_2})^T, (\mathbf{SG^k_3})^T, ..., (\mathbf{SG^k_L})^T\right) \tag{3.10}$$

$$\mathbf{ALLST} = \left((\mathbf{ST_1})^T, (\mathbf{ST_2})^T, (\mathbf{ST_3})^T, ..., (\mathbf{ST_L})^T\right) \tag{3.11}$$

Where

- $\mathbf{ALLSG^k}$ is the reconstructed gait feature matrix for k^{th} sequence in gallery set.

- **ALLST** is the reconstructed gait feature matrix for the input test sequence.

- $\mathbf{SG^k}$ is a row vector which denotes a subspace of k^{th} sequence in gallery set calculated by Algorithm 1

- **ST** is a row vector which denotes a subspace of the input test sequence.

The Figure 3.8 visualizes the procedure of subspace selection and gait feature matrix reconstruction. Each chunk in this figure denotes a 'cell' in cell-based matrix matrix. Each straight line denotes a process of cell selection from the cell-based matrix matrix. Note that the location of the selected cell in an input test sequence is consistent with all sequences in gallery set for retaining the spatial information efficiently. Assume that N cells are selected randomly from each cell-based matrix, a subspace of each gait sequence consists of N cells, as shown in this figure ($N = 4$ in this figure). Finally, after repeating the above-mentioned process L times, the $\mathbf{ALLSG^k}$ and **ALLST** is calculated as equation 3.10 and equation 3.11, respectively.

The $\mathbf{ALLSG^k}$ and **ALLST** are considered as the new gait feature matrices for classification instead of the cell-based matrix in the section 3.3.

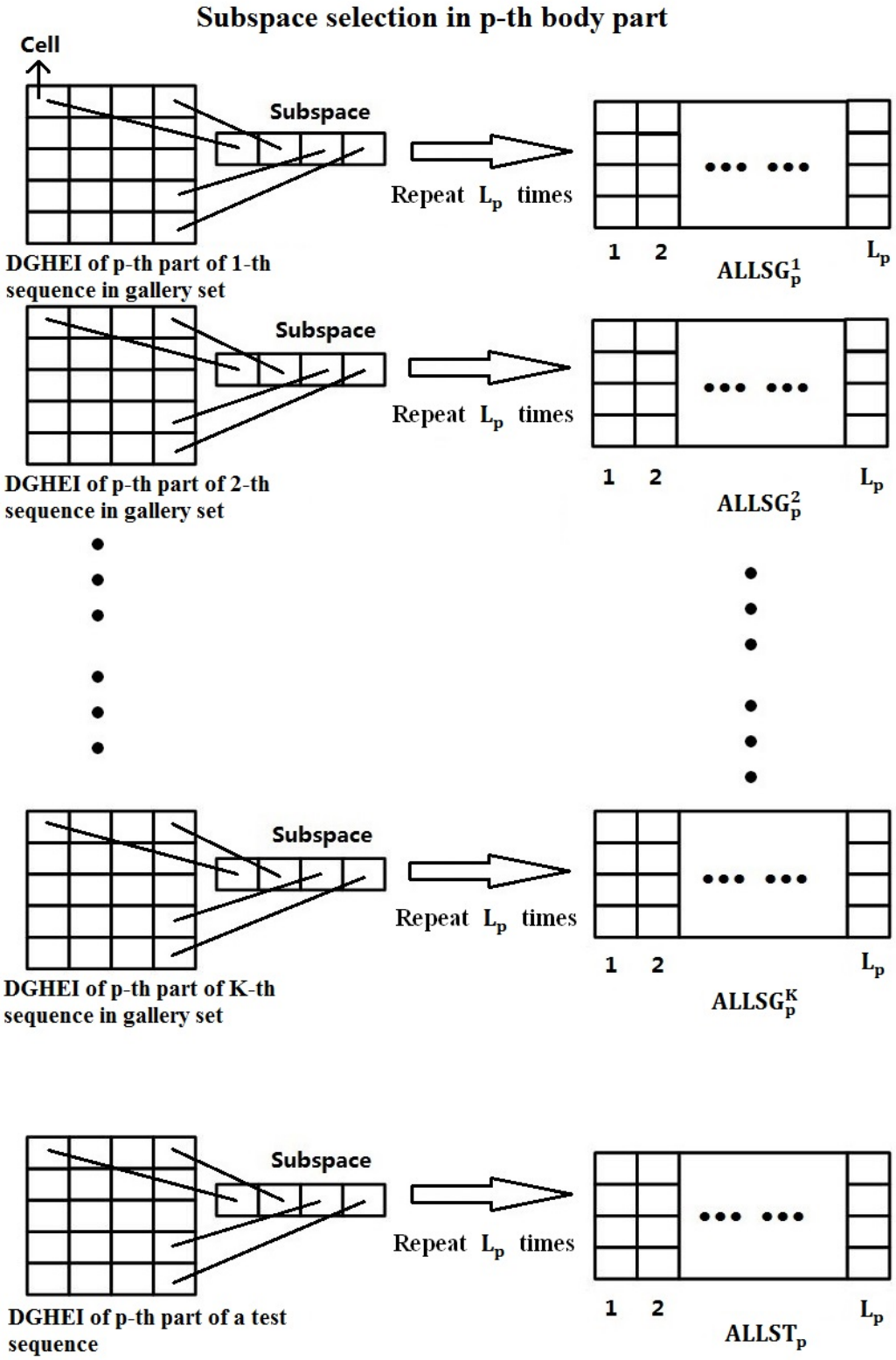


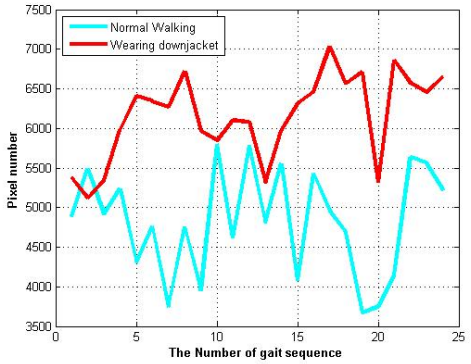
Figure 3.8: Procedure of subspace selection and gait feature reconstruction

3.2.3 Adaptive proportion assignment

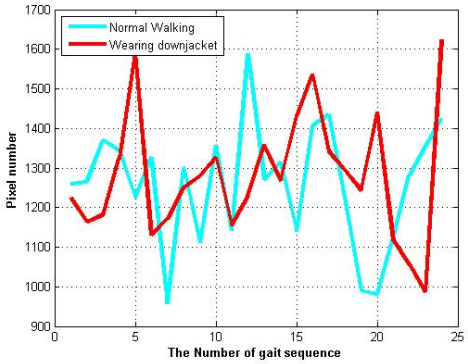
In this section, the author will apply an adaptive feature selection technique to decide the number of subspaces (L_p) for p^{th} body part. The proportion of L is variable for each input sequence in a probe set.

The most of existing methods for reducing the negative effect of clothing variation in gait recognition have a strong assumption that the clothing types appearing in the a probe set are also involved in the training stage. Hossain et al. [11] constructed a dataset with large clothing variations and proposed clothing-invariant gait recognition. They defined eight parts: four consecutive parts and four overlapping parts based on anatomical study. The clothing types in their database are categorized into similar and dissimilar cloths by measuring normalized Euclidean distance. Then they adaptively assigned more weights for the clothing types which have the less distance to the standard clothes and less weights for the clothing types which have the larger distance to the standard clothes. The weight is computed based the above mentioned assumption. However, it is not realistic for real-world applications and these assumptions, and they should be removed for practical system.

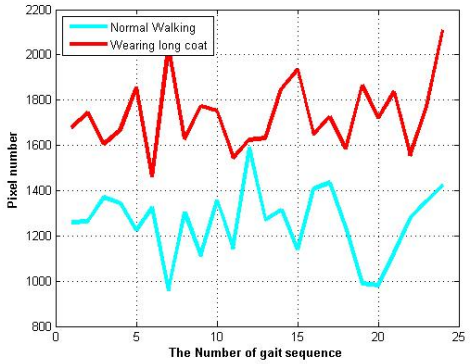
Without assuming specific clothes type and baggage carrying condition for each input test sequence, we cannot decide what part of cell-based matrix is affected by walking condition. However, Li et al. [29] tested on a RGB gait database with the variation of walking condition, and obtained the results that notable changes, especially the pixel number, appeared in subject's silhouette image under the variety of walking conditions.



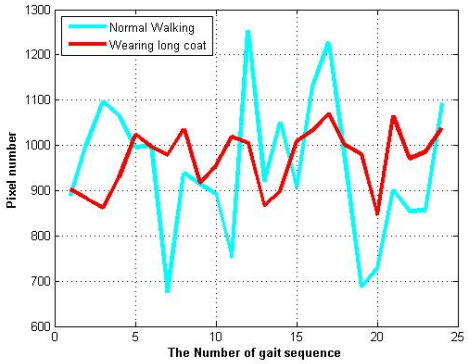
(a) 'Trunk' part Wearing down jacket and Normal walking



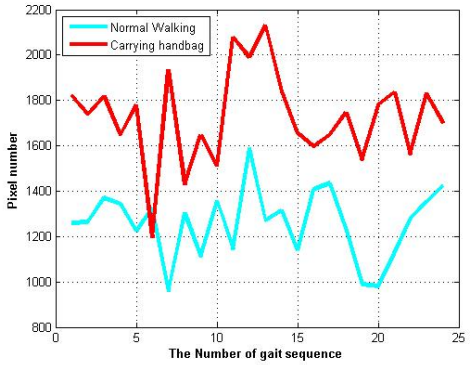
(b) 'Thigh' part Wearing down jacket and Normal walking



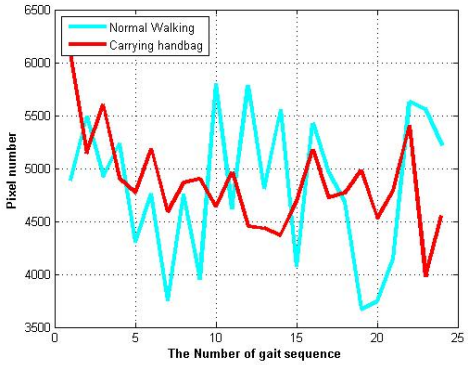
(c) 'Thigh' part Wearing Long coat and Normal walking



(d) 'Crus' part Wearing Long coat and Normal walking



(e) 'Thigh' part Carrying a handbag and Normal walking



(f) 'Trunk' part Carrying a handbag and Normal walking

Figure 3.9: Pixel number distributions

Figure 3.9 shows examples of the pixel number distribution in different parts under six different walking conditions. When a subject walks with down jacket, it is observed that the number of pixels mainly falls into two scopes in 'trunk' parts. As described in Section 3.2.2, the subspaces extracted from this part when a subject walks with down jacket is interfered features for individual recognition. However, the pixel number in 'thigh' part still remains nearly the same with the normal walking condition, which allows to identify a person by the discrimination clues in this part. Similarly, when a subject walks with long coat or handbag, it is observed that the number of pixels mainly falls into two scopes in 'thigh' part while the pixel number in 'crus' and 'trunk' part still remains nearly the same with the normal walking condition.

In this work, the author applies the fuzzy logic to decide if some body parts are affected by the variation of walking condition. In fuzzy logic, fuzzy concept is often regarded as concept is neither completely true nor completely false. Fuzzy proposition (such as 'the p^{th} human body part is affected') is presented by a value which could fall into a probable

range defined by quantitative limits or parameters, and it can be usefully described with imprecise categories (such as 'high', 'medium' or 'low') using some kinds of qualitative scale. The basic idea in this work is that a real number is assigned to a fuzzy proposition ('the p^{th} human body part is affected') as a probability within a range between 0 to 1. The value of 1 means that the statement is completely true and 0 means that the statement is completely false, while values less than 1 but greater than 0 denote that the statements are 'partly true', to a given, quantifiable extent. The real number is termed as 'fuzzy membership degree'.

The purpose of measuring fuzzy membership degree is to assess quality degradations caused by changing walking condition, and assign the number (L_p) of subspace for each body part of an input test sequence (as describe section 3.2.2). Note that, to minimize the negative effect by the variation of walking condition, the author need to impose smaller number of the selected subspace in the body part which has a lower probability of being unaffected by changing the walking condition from normal walking condition. Meanwhile, the author assigns larger number of selected subspace in the body part which has a greater probability of being unaffected by changing the walking condition from normal walking condition.

For this purpose, the important aspect of fuzzy membership degree measurement is that it allows for determining the weights of input values in an adaptive and data-dependent way [30]. As such the proposed method is able to provide an adaptive mechanism for determining the proportion of L.

A. Pixel number measurement

Assume that each subject in probe set or gallery set contains T images presented a complete walking process through one gait cycle. Let I_p^t denotes the p^{th} subimage segmented from the t^{th} image and $p \in (1, 2, 3, 4)$ as described in section 3.1. Then, a subject's mean pixel number (x_p) of p^{th} body part is computed as the following equation:

$$x_p = \frac{1}{T} \sum\nolimits_{t=1}^{T} I_p^t \tag{3.12}$$

In a gallery set, an upper and a lower bound is computed for the following processing. Assume that there are total N subjects in gallery set, the upper bound Ω_p^{Up} and lower bound Ω_p^{Lo} of i^{th} mean pixel number (x_p^i) of p^{th} body part is calculated as equation 3.13 and equation 3.14:

$$\Omega_p^{Lo} = min_{i=1}^{N} x_p^i \tag{3.13}$$

$$\Omega_p^{Up} = max_{i=1}^{N} x_p^i \tag{3.14}$$

B. Proportion assignment

Based on the distribution of pixel number, the author calculates a fuzzy membership degree by the fuzzy membership for an input test sequence as follows:

$$Mem_p = \begin{cases} 0, \Omega_p^{Up} \leq x_p \\ 1 - \frac{x_p - \Omega_p^{Lo}}{\Omega_p^{Up} - \Omega_p^{Lo}}, \Omega_p^{Lo} < x_p < \Omega_p^{Up} \\ 0, x_p \leq \Omega_p^{Lo} \end{cases} \tag{3.15}$$

In equation 3.15, Mem_p denotes the fuzzy membership degree towards p^{th} part. Higher value of the fuzzy membership degree means that human silhouette has a greater probability of being unaffected by walking conditions in p^{th} part. x_p is the average pixel number of all images in p^{th} part. Meanwhile, Ω_p^{Up} and Ω_p^{Lo} is the upper and lower bound of pixel number in a gallery set, respectively. The thresholds are decided only from the normal walking sequences in the gallery set.

From this equation, it is observed that the author will not select the subspaces from the p^{th} body part if the number of its pixels is deviated from the range of $[\Omega_p^{Lo}, \Omega_p^{Up}]$. Note that in order to decide the unaffected parts, one possible solution is to find the parts whose value falls into the range of $[\Omega_p^{Lo}, \Omega_p^{Up}]$ and to discard the rest of body parts directly. Then, the number of selected subspaces is equal to the same value in each body part. In practice, however, some of the selected body parts may be affected by the variation of walking condition and lead to recognition error, that is mainly caused by inaccurate estimation (or measurement) of pixel number. To overcome the aforementioned limitation, the proposed proportion assignment takes advantage of fuzzy membership degree.

Then the proportion (Pro_p) of L to be calculated for p^{th} body part within a probe gait sequence is obtained as follows:

$$\mathrm{Pro}_p = \frac{Mem_p}{\sum_{p=1}^{P} Mem_p} \tag{3.16}$$

Finally, the number L_p of subspaces selected from the p^{th} part in a test sequence is calculated as follows:

$$L_p = \mathrm{Pro}_p \times L \tag{3.17}$$

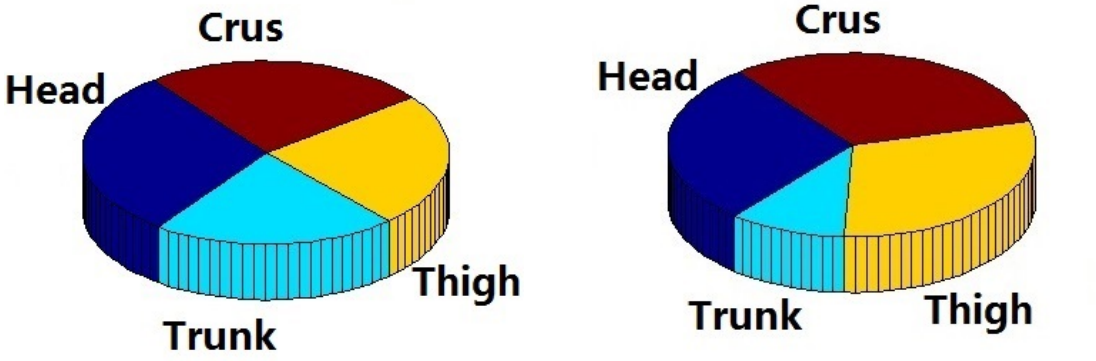
Note that the proportion obtained from equation 3.16 satisfies the constraint of depth subspace selection framework as shown in following equation:

$$0 \leq \mathrm{Pro}_p \leq 1 \tag{3.18}$$

$$\sum_{p=1}^{P} \mathrm{Pro}_p = 1 \tag{3.19}$$

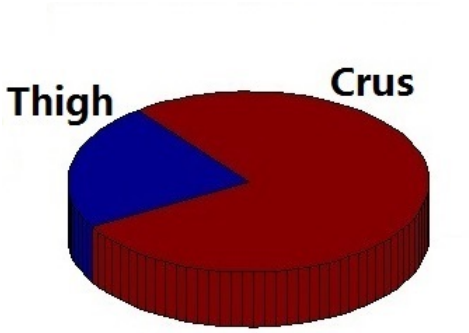
Figure 3.10 shows the results of proportion assignment when a subject walks under eight different walking conditions. The numbers of selected subspaces are similar in four kinds of body parts under the normal walking condition because each part is unobstructed and contains useful discrimination clues for individual recognition. Nevertheless, when a person carries a bag while walking, the pixel number increases and goes beyond the thresholds for two lower body parts due to the bag's volume. In this case, the author will select subspaces only from 'head' part and 'trunk' part for discarding the interfered information from the region of a bag.

Thus, for an input test sequence without knowing walking condition, the number (L_p) of selected subspace is decided by the proposed method. The value of L_p will be applied in the procedure of random subspace selection. In addition, from an application point of view, the proposed method is simple to implement because the fuzzy membership function (equation 3.15) generally requires no complex fuzzy rules during the computation of membership degree.

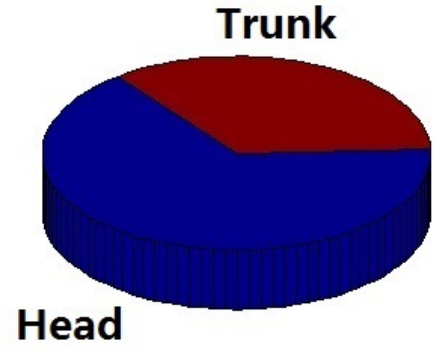


(a) Normal walking condition

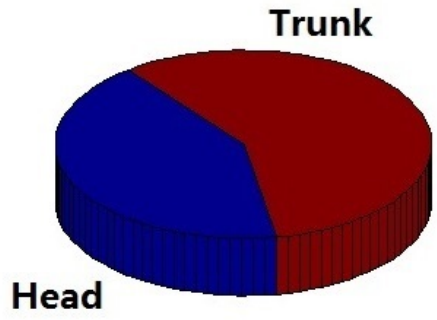
(b) Wearing a down jacket



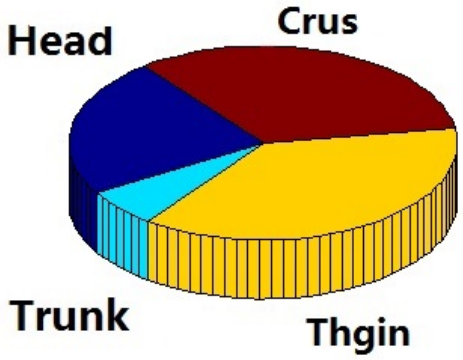
(c) Wearing a long coat



(d) Carrying a handbag



(e) Wearing a skirt



(f) Carrying a shoulder bag

Figure 3.10: Subspace number proportion under different walking conditions

3.3 Individual classification

In this section, the author will introduce the individual classification strategy in this work. As described in section 3.2.2, one reconstructed gait feature matrix is constituted by subspaces from a test sequence as well as K reconstructed gait feature matrices from all sequences in gallery set. Assume that there are total N subjects (classes) in gallery

set. Each subject contains T depth gait sequences, e.g. there are T reconstructed gait feature matrices. Thus, the value of K is calculated by the following equation:

$$K = NT \tag{3.20}$$

In the next step, the author calculates Euclidean distance to measure the similarity of two reconstructed gait feature matrix:

$$\mathbf{d^k(i)} = \left\| \mathbf{ALLST(i)} - \mathbf{ALLLSG^k(i)} \right\|; i = 1, 2, ..., L \tag{3.21}$$

Where $\mathbf{d^c(i)}$ denotes the matrix-to-matrix Euclidean distance between a test sequence and c^{th} class in a gallery set. Then, the NN (Nearest Neighbor) classifier is used to identify the best match for each $\mathbf{d^k(i)}$.

$$\mathbf{L(i)} = min_{c=1}^{N} \mathbf{d^k(i)}; i = 1, 2, ...L \tag{3.22}$$

Here, $\mathbf{L(i)}$ is a row vector with L class labels. The final decision is taken by majority voting from these L class labels.

Thus, a subject in a probe set is classified into a class by matching this subject with all subjects in a gallery set as above mentioned classification strategy. The flow chart of individual classification is illustrated in Figure 3.11.

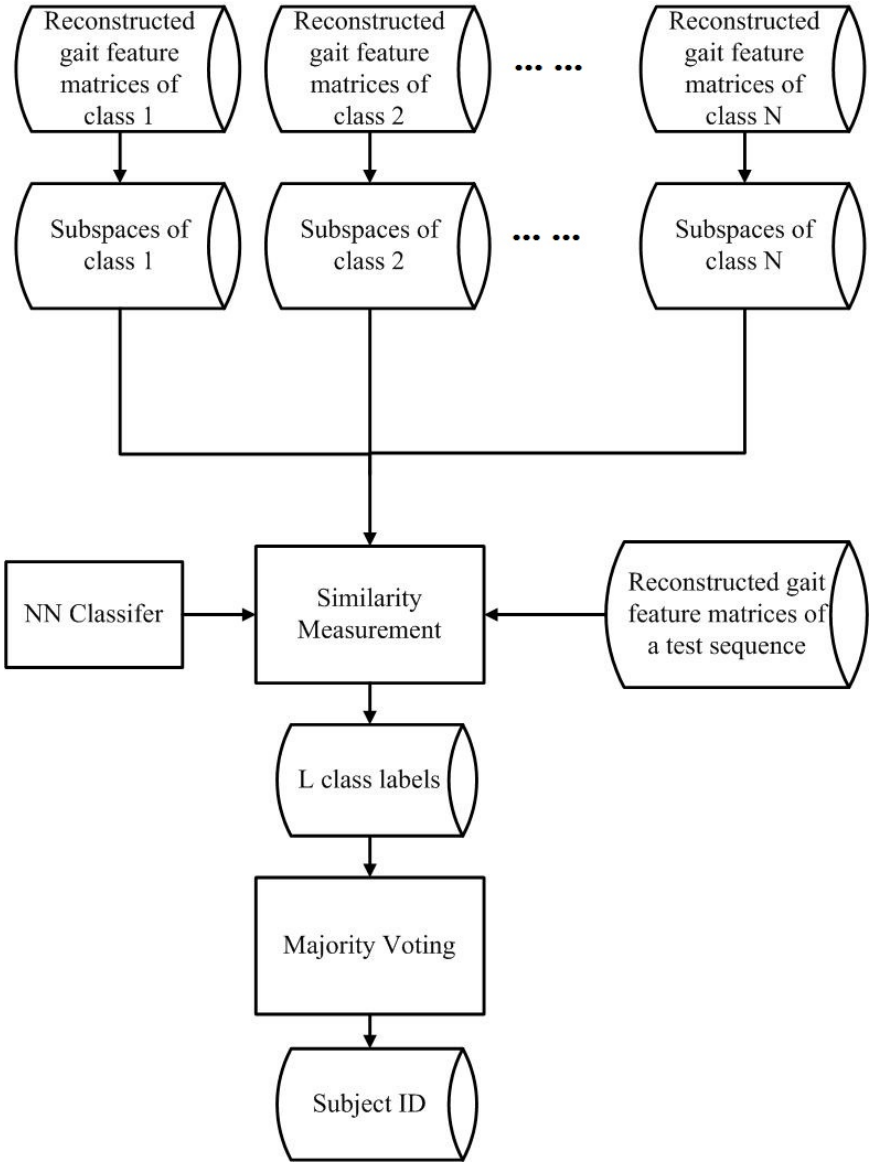


Figure 3.11: Procedure of individual classification

A flow chart of recognizing an individual from an input test sequence is shown in Figure 3.12. The gallery set consists of K normal walking sequences while the test sequence is a subject's gait sequence of walking with unknown clothes type and postures. Preprocessed images are applied for both cell-based matrix calculation and adaptive proportion assignment. Each subspace from the test set is matched with all the subjects in gallery set and best match is achieved by the classification.

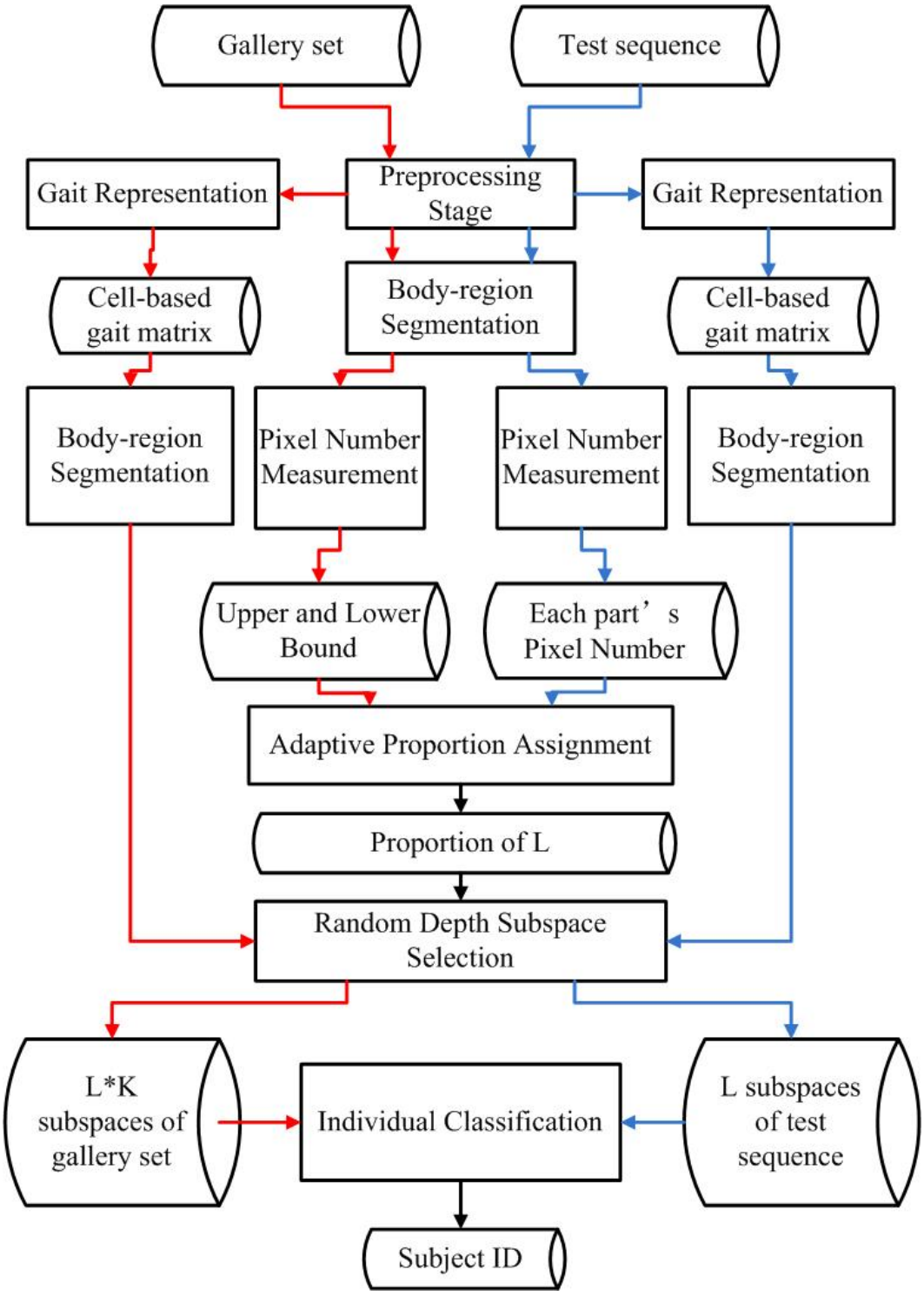


Figure 3.12: Procedure of individual recognition

Chapter 4

Evaluation of ARDS

4.1 Depth gait database

In this section, the author will introduce a new depth gait database that is applied to evaluate the effectiveness of ARDS. Since no depth gait database containing multiple walking conditions from frontal viewpoint is available in public domain, the author prepared depth gait database of frontal depth video frames by himself. The sensor configuration and the recording procedure is described in detail.

4.1.1 Sensor

Microsoft Kinect for XBOX 360 [31] was used for recording in this work as shown in Figure 4.1. A set of technologies that enable humans to interact naturally with computers, the Kinect provides the author with the foundation needed to create and deploy interactive applications. This sensor provides a RGB stream, a depth stream and four-channel audio. The gait data of each subject is recorded as both color (not used) and depth(resolution: 640×480 pixels, frame rate: approximately 30 fps).

4.1.2 Recording site

The author selected a narrow corridor in a lab as the recording site, since the infrared light from the sunlight interferes with the depth stream acquisition. Moreover, this recording site, while mainly necessary for research, still closely simulate a real-life application scenarios such as access control in a narrow corridor.

A larger distance of the subjects from the sensor is favorable for gait recognition, since multiple gait cycles will be recorded in the field of view. However, the range that the Kinect is able to cover depends on the limitations of the sensing technology. The device records the depth info from 0.8 m upto 4 m. According to the manufacturer [31], the optimal distance between the subject and the sensor should be between 1.8 m and 3.6 m. To comply with the aforementioned condition, the Kinect is positioned at appropriate height from the floor with a downward view angle of 20°, allowing the depth sensor to

capture at least one full cycle of each walking subject. Figure 4.2 shows the configuration for recording the gait data.

Figure 4.1: Microsoft Kinect for XBOX 360

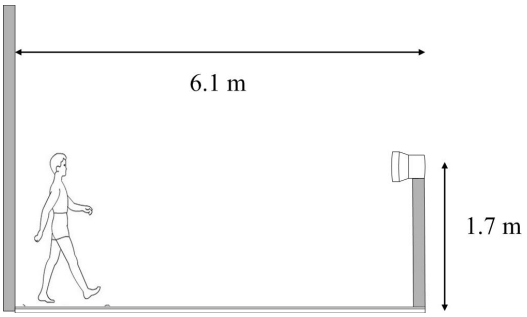


Figure 4.2: Recording setup

4.1.3 Recording procedure

For recording gait data, subjects were chosen by confirming for their willingness to participate. Each subject was instructed to start at the back maker and walk towards the sensor up to the front marker for 28 times. Placing the back markers one meter beyond the maximize detection distance as a sufficient distance for a stable walking speed. The RGB and depth gait data was recorded by the format of 'gait.xed', which is the raw format in Kinect Studio [32].

4.1.4 Database statistics

This depth gait database consists of 12 subjects with seven walking conditions for each subject, including 4 normal walking sequences (wearing T-shirt, regular pants and natural swinging arms), 8 to 12 clothes-varying sequences and 12 to 16 sequences of other walking conditions, named as Set A, Set B and Set C respectively. The contents of each divided set are described in section 4.2.

Guan et al. [28] investigated the effects of different clothing types by dividing each clothes type combination into four groups and calculated the individual recognition rate for each group. Then Guan et al. [28] generated the challenging clothing types, such as rain coat, skirt and down jacket, lead to the worst performance. The average accuracy for the group of challenging clothing types is dramatically decreased to approximately 50 %. It should be pointed out under the circumstance where subject walks covering whole body with challenging clothing type (e.g. hat + down jacket + skirt), it is unrecognizable even by human being, since all of the human gait features are obscured by clothing.

In our work, each subject is required to walk with wearing one or two challenging clothing types and a standard clothing type (e.g. down jacket + regular pants, hat + long coat + regular pants) as the probe sequence. The author chose some typical challenging dresses, such as skirt, down jacket and long coat, for simulating real circumstance, regardless of

other skintight dresses like suits or mufflers. Since the skintight dresses lead to small-scale changes in appearance of human silhouette and contribute minimal impact towards the gait recognition of individual. Meanwhile, four common walking postures are selected for robustness, including walking with a handbag ,putting his/her hands in the pockets and so on.

The Table 4.1 lists the walking conditions in depth gait database with affected body parts, while Figure 4.3 shows the sample frames from RGB video and corresponding depth video. Note that the first row of figure is considered as the normal walking condition in the experiments. The depth frames are usually noisy with black holes, especially in the subject's hair region. Since infrared ray emitted by Kinect sensor is unable to be reflected by some special Material.

Table 4.1: Walking condition in different body part

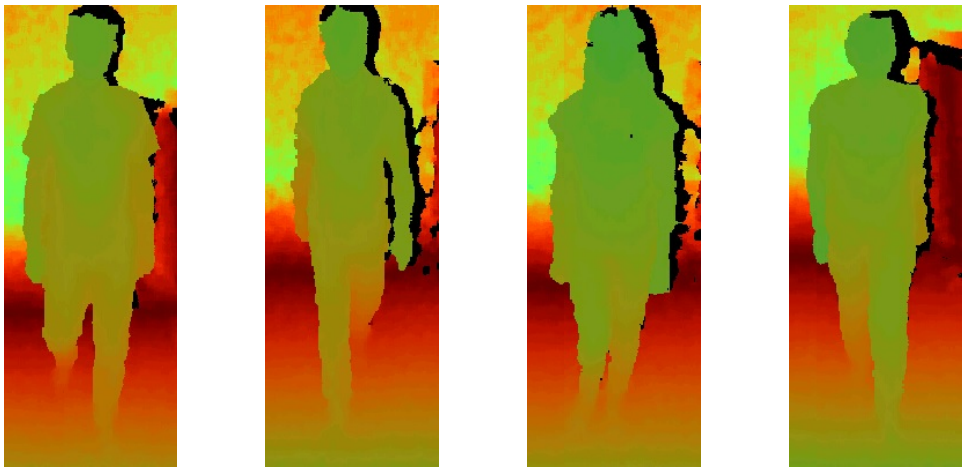
Affected body parts	**Walking conditions**
Upper body	T-shirt Hat Down jacket Long coat Shoulder bag Burying hands in the pockets.
Lower body	Regular pants Short pants Long coat Skirt Handbag

(a) Normal walking

(b) Normal walking

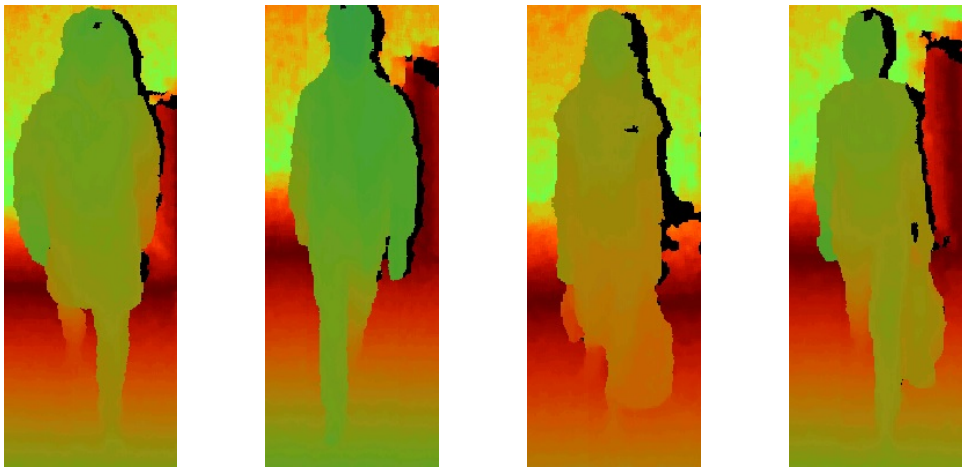
(c) Normal walking

(d) Normal walking



(e) Normal walking (f) Normal walking (g) Normal walking (h) Normal walking

(i) Hat and long coat (j) Down jacket (k) Skirt (l) handbag



(m) Hat and long coat (n) Down jacket (o) Skirt (p) handbag

Figure 4.3: Samples in robust depth gait database

4.2 Configuration of Experiment

To evaluate the robustness of proposed method, the author segments Set A into two different subsets: Set A1 consists of the 2 sequences of normal walking, taken as the gallery set; Set A2 includes the rest 2 of sequences. Note that only normal walking sequences constitute the gallery set, which is more realistic in real applications. For each subject, the probe set is composed of three subsets: Set A2, Set B and Set C, as illustrated in Table 4.2. The purpose of testing three subsets is to evaluate the effectiveness of each method toward different walking conditions.

In addition, the gallery set is also referred to for deciding the thresholds as described in section 3.2.3. Since the proposed method is a random system, only one experiment is not enough for evaluation and comparison. The author measured the mean accuracy by running ten times. The author compared the proposed method with state-of-the-art methods for robust gait recognition using depth data

Considering real-world gait recognition of individual, it is not realistic to train a model for all possible walking conditions for each subject. In this case, the methods breached the above-mentioned requirements are eliminated in comparison prior to the experiment.

Table 4.2: Walking conditions in different Set

Set A1	**Set A2**	**Set B**	**Set C**
Normal	Normal	Down jacket	Shoulder bag
		Long coat	Hand bag
		Hat	Putting one hand in the pockets.
		Skirt	Putting two hands in the pockets.

4.3 Results and Discussion

4.3.1 Performance comparison by algorithm

Table 4.3 shows the maximum and minimum of the accuracy obtained by ten times of runs for presented depth gait database. Here, two parameters, N and L, are set to 3 and 300, respectively.

- N is the number of cells in each subspace.
- L is the total number of subspaces in each gait sequence.

Table 4.3: Maximum and minimum accuracy of proposed method by running ten times

Condition: L=600, N=3	**Set A2**	**Set B**	**Set C**	**ALL**
Maximum	91.67%	70.97%	78.66%	76.60%
	(22/24)	(88/124)	(129/164)	(239/312)
Minimum	91.67%	65.32%	75.00%	72.44%
	(22/24)	(81/124)	(123/164)	(226/312)
Mean	91.67%	69.27%	77.07%	75.10%

The author implemented the GEV and DGHEI in our depth gait database, and compared experimental results with these two depth gait recognition methods for demonstrating the effectiveness of proposed method.

As shown in Table 4.4, GEV and DGHEI, the methods focusing on depth gait images, extended the concept of the GEI to 3D and showed good performance in Set A2. However, the performance becomes worse when the intra-class variation exceeds individual variation for different walking conditions such as Set B and Set C. This is because the GEV features, as well as DGHEI, are extracted from the whole body silhouettes, which is affected by the variation of walking conditions. Performance is remarkably improved in DGHEI because some body parts with significant discrimination clues, such as legs, are not obstructed. In this case, using DGHEI to extract the gradient information from these body parts becomes smooth, and it is possible to obtain better accuracy than GEV method.

Moreover, this result also revealed that the lower limb, which contain multiple dynamic locomotion changes, provides more discrimination clues than the upper body.

As seen in the Table 4.4, the accuracy (Mean accuracy by running ten times) of proposed method is superior to other methods in these data sets. Specifically, proposed method improves the accuracy higher in Set B than other subsets. Based on the local gradient information of body and adaptive proportion assignment, proposed method is capable of avoiding the interfered information under the variation of clothing effectively.

Table 4.4: Performance comparison by algorithm (depth gait recognition methods)

Methods	**Set A2**	**Set B**	**Set C**	**ALL**
GEV [12]	91.67%	55.56%	62.20%	61.86%
	(22/24)	(69/124)	(102/164)	(193/312)
DGHEI [13]	91.67%	42.74%	73.78%	62.82%
	(22/24)	(53/124)	(121/164)	(196/312)
Proposed	91.67%	69.27%	77.07%	75.10%

The author also tested the some binary gait recognition methods for the same database. As shown in Table 4.5, it is obvious that these algorithm using binary silhouettes which are extracted from the depth images achieved worse performance than these which is applied for the depth information, because the physical distance provides additional discrimination clues from the depth images.

Table 4.5: Performance comparison by algorithm (binary gait recognition methods)

Methods	**Set A2**	**Set B**	**Set C**	**ALL**
GEI [6]	91.67%	30.65%	40.85%	40.71%
	(22/24)	(38/124)	(67/164)	(119/132)
GEnI [7]	91.67%	45.97%	43.90%	48.40%
	(22/24)	(57/124)	(72/164)	(151/312)
Proposed	91.67%	69.27%	77.07%	75.10%

4.3.2 Analysis of misclassification cases

For further improvement, the author conducted an investigation of the mismatch situation by applying proposed method for the depth gait database. Figure 4.4 shows the two common wrong cases in Set B. As illustrated in Figure 4.4, most of misclassification occurred when the vast majority of lower limbs are obstructed as well as the upper body.

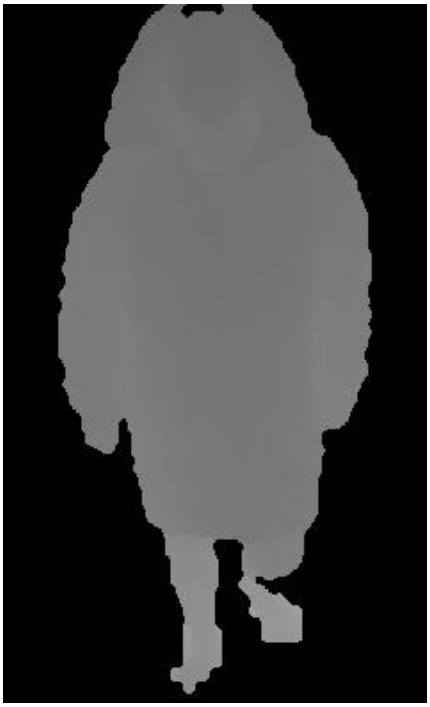


Figure 4.4: Misclassification case in Set B

Figure 4.5 shows common wrong cases in Set C. As illustrated in the figure, Most of misclassification occurred when the subject walks with carrying a handbag.

There are two reasons of misclassification. First, the pixel number is no significant change between above mentioned walking condition and normal walking condition based on proposed adaptive proportion assignment method. In this case, the number of selected subspaces is similar in each body part in above mentioned walking condition. Thus, the obvious changes of oriented gradient information in arms between the test subject (one arm maintains static to hold the bag while the other one waves naturally) and its correct class (two arms wave naturally) cause the mismatch. Second, as described in section 4.3.1, lower limbs contain significant discrimination clues. In case of carrying a handbag, most of the test subjects' lower body parts were not taken into account for subspace selection due to the larger variation of pixel number compared with the normal walking condition. Thus, luxuriant discrimination clues were discarded before the classification stage, although some useful information are remained in upper body part.

(a) Misclassification case 3 (b) Misclassification case 4

Figure 4.5: Misclassification case in Set C

4.3.3 Investigation of parameters in ARDS

In proposed method, there are two significant parameters which need to be investigated for effectiveness: (1)N; the number of cells in each subspace; (2) L; the total number of subspaces in each gait sequence.

The author obtained the mean accuracy by running ten times under each combination of N and L as shown in Figure 4.6. In the experiment, each value of N has been tested by setting the range from 2 to 8. The best mean accuracy is obtained for N=3. Though it is not proved in Isam et al. [8], the mean accuracy will be decreased as well, when the value of N decrease from the value which lead to the best accuracy (N=3 in our experiment). In case where the size of each subspace is too small, the corresponding subspaces from the same subject have more probability to be mismatched by the slight discrepancies of the same body part between probe and gallery. Meanwhile, if N is too large, the mean accuracy will be decreased due to overfitting problem.

Based on the theory of random subspace method [33], the mean accuracy is stable in a small scale for each decided value of N with the growth of L. However, as the number of subspace (L) rises, the computation cost of random depth subspace selection goes up. Thus we decided the parameter combination as N=3 and L=600, which is lead to the best performance in our experiment.

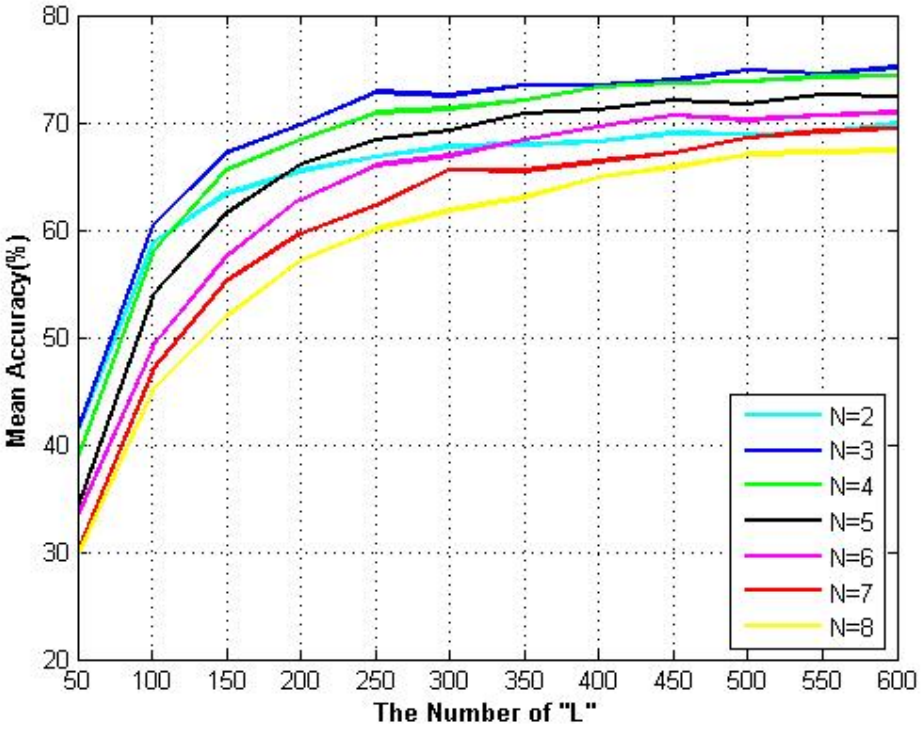


Figure 4.6: Performance comparison with respect to N

Chapter 5

Conclusion and Future Work

This paper introduced an effective method, which is called Adaptive Random Depth Subspace (ARDS) framework for robust depth gait recognition for individual. Different form the related method of depth gait recognition, we extend the whole-based feature extraction to part-based selection for adjusting the walking conditions variation. For real-world application, we combine the adaptive feature selection technique with the random selection procedure to assign the number of subspaces for each body part. Experimental results showed superior performance to the existing methods.

Moreover, the author established a new robust depth gait database with larger intra-class variation of frontal depth video frames himself, since the existing depth gait database depends on side viewpoint and lacks of walking conditions.

The proposed ARDS method has a disadvantage that the body-part segmentation is accomplished by the premise of anatomical body properties, which may cause the mismatch due to the body proportion variation of the subjects. Thus, open issues include the investigation of other techniques for dynamic body-part detection and segmentation, and testing the proposed method with more number of subjects.

Here, the author has only focused on applying the depth gait images for recognition of individual. The accuracy is expected to be improved if a combination of different data sources methods is proposed. Thus, combining the skeleton and appearance features for gait recognition will be a direction for future research.

Acknowledgement

I would like to take this opportunity to express my deep gratitude to all the people who supported and encouraged me throughout my time at Master's program. My deepest gratitude goes first and foremost to Associate Professor Atsuo Yoshitaka, my supervisor, for his constant encouragement and guidance through my study. Meanwhile, I sincerely appreciate his instructive advice and help in my research and my daily life at JAIST.

I am also greatly indebted to the Associate Professor Kazunori Kotani who have instructed and helped me a lot in the past one year. Meanwhile, I also grateful for the suggestions that Professor Jianwu Dang have given to me through my study in Tianjin university and JAIST.

My sincere gratitude is further more to sent all members in image processing lab. They helped me a lot when I got troubles in the new environment.

Last, I should finally like to express my gratitude to my beloved parents who have always been helping me out of difficulties and supporting without a word of complaint.

Bibliography

[1] Sourabh A Niyogi and Edward H Adelson. Analyzing and recognizing walking figures in xyt. In *Computer Vision and Pattern Recognition, 1994. Proceedings CVPR'94., 1994 IEEE Computer Society Conference on*, pages 469–474. IEEE, 1994.

[2] Gunnar Johansson. Visual perception of biological motion and a model for its analysis. *Perception & psychophysics*, pages 201–211, 1973.

[3] David Cunado, Mark S Nixon, and John N Carter. Automatic extraction and description of human gait models for recognition purposes. *Computer Vision and Image Understanding*, pages 1–41, 2003.

[4] Tao Ding. A robust identification approach to gait recognition. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on*, pages 1–8. IEEE, 2008.

[5] ChewYean Yam, Mark S Nixon, and John N Carter. Automated person recognition by walking and running via model-based approaches. *Pattern Recognition*, pages 1057–1072, 2004.

[6] Ju Han and Bir Bhanu. Individual recognition using gait energy image. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, pages 316–322, 2006.

[7] Khalid Bashir, Tao Xiang, and Shaogang Gong. Gait recognition using gait entropy image. *IET Conference Proceedings*, pages 2–2(1), 2009.

[8] Md Islam, MA Hossain, MKI Molla, et al. Window based clothing invariant gait recognition. In *Advances in Electrical Engineering (ICAEE), 2013 International Conference on*, pages 411–414. IEEE, 2013.

[9] M Rokanujjaman, MA Hossain, and MR Islam. Effective part selection for part-based gait identification. In *Electrical & Computer Engineering (ICECE), 2012 7th International Conference on*, pages 17–19. IEEE, 2012.

[10] Md Rokanujjaman, Md Hossain, Md Islam, et al. Effective part definition for gait identification using gait entropy image. In *Informatics, Electronics & Vision (ICIEV), 2013 International Conference on*, pages 1–4. IEEE, 2013.

[11] Altab Hossain, Yasushi Makihara, Junqiu Wang, Yasushi Yagi, et al. Clothing-invariant gait identification using part-based clothing categorization and adaptive weight control. *Pattern Recognition*, pages 2281–2291, 2010.

[12] Sabesan Sivapalan, Daniel Chen, Simon Denman, Sridha Sridharan, and Clinton Fookes. Gait energy volumes and frontal gait recognition using depth images. In *Biometrics (IJCB), 2011 International Joint Conference on*, pages 1–6. IEEE, 2011.

[13] Martin Hofmann, Sebastian Bachmann, and Gerhard Rigoll. 2.5 d gait biometrics using the depth gradient histogram energy image. In *Biometrics: Theory, Applications and Systems (BTAS), 2012 IEEE Fifth International Conference on*, pages 399–403. IEEE, 2012.

[14] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*, pages 886–893. IEEE, 2005.

[15] Milos Milovanovic, Miroslav Minovic, and Dusan Starcevic. Walking in colors: Human gait recognition using kinect and cbir. *MultiMedia, IEEE*, pages 28–36, 2013.

[16] James Little and Jeffrey Boyd. Recognizing people by their gait: the shape of motion. *Videre: Journal of Computer Vision Research*, pages 1–32, 1998.

[17] Ralph Gross and Jianbo Shi. The cmu motion of body (mobo) database. 2001.

[18] M Nixon, J Carter, J Shutler, and M Grant. Experimental plan for automatic gait recognition. *Technica l Rep or t. Southampton, UK: University of Southampton*, 2001.

[19] Shiqi Yu, Daoliang Tan, and Tieniu Tan. A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition. In *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*, pages 441–444. IEEE, 2006.

[20] Sudeep Sarkar, P Jonathon Phillips, Zongyi Liu, Isidro Robledo Vega, Patrick Grother, and Kevin W Bowyer. The humanid gait challenge problem: Data sets, performance, and analysis. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, pages 162–177, 2005.

[21] Martin Hofmann, Jürgen Geiger, Sebastian Bachmann, Björn Schuller, and Gerhard Rigoll. The tum gait from audio, image and depth (gaid) database: Multimodal recognition of subjects and traits. *Journal of Visual Communication and Image Representation*, pages 195–206, 2014.

[22] Massimo Piccardi. Background subtraction techniques: a review. In *Systems, man and cybernetics, 2004 IEEE international conference on*, pages 3099–3104. IEEE, 2004.

[23] Enrique J Fernandez-Sanchez, Javier Diaz, and Eduardo Ros. Background subtraction based on color and depth using active sensors. *Sensors*, pages 8895–8915, 2013.

[24] Tracking users with kinect skeletal tracking. `http://msdn.microsoft.com/en-us/library/jj131025.aspx`.

[25] Edouard Auvinet, Franck Multon, and Jean Meunier. Lower limb movement asymmetry measurement with a depth camera. In *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pages 6793–6796. IEEE, 2012.

[26] Gait cycle. `http://hortonworks.com/blog/snapshots-for-hdfs/`.

[27] Wilfrid Taylor Dempster and George RL Gaughran. Properties of body segments based on size and weight. *American Journal of Anatomy*, pages 33–54, 1967.

[28] Yu Guan, Chang-Tsun Li, and Yongjian Hu. Robust clothing-invariant gait recognition. In *Intelligent Information Hiding and Multimedia Signal Processing (IIH-MSP), 2012 Eighth International Conference on*, pages 321–324. IEEE, 2012.

[29] Ning Li, Yi Xu, and Xiao-Kang Yang. Part-based human gait identification under clothing and carrying condition variations. In *Machine Learning and Cybernetics (ICMLC), 2010 International Conference on*, pages 268–273. IEEE, 2010.

[30] Jae Young Choi, Konstantinos N Plataniotis, and Yong Man Ro. Face feature weighted fusion based on fuzzy membership degree for video face recognition. *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, pages 1270–1282, 2012.

[31] Microsoft kinect for xbox 360. `http://www.xbox.com/en-SG/Kinect/`.

[32] Kinect studio. `http://msdn.microsoft.com/en-us/library/hh855389.aspx`.

[33] Tin Kam Ho. The random subspace method for constructing decision forests. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 20(8):832–844, 1998.