IMPROVING SPATIAL-TEMPORAL GAIT FEATURES BASED ON REGIONAL ADJACENT PATCHES DESCRIPTORS

By

Golnaz Mahdavikhah

A thesis submitted to the

Department of Computer Science

in conformity with the requirements for

the degree of Master of Science

Bishop’s University

Canada

August 2023

Abstract:

Due to the significant role of recorded images in enhancing public security, this thesis focuses on the challenges of image recognition during walking. The foundation of recent algorithms is based on extracting silhouettes from input videos and averaging them over a walking period. However, this process is vulnerable to noise, temporal ordering of walking, and partial silhouette defects. To address these issues, a novel template based on patch-based analysis is developed to improve the common walking features in local regions by searching for neighboring patches and eliminating noisy or faulty patches. A Histogram of Gradient (HoG) descriptors is computed to capture important features in clustered regions. The gait signatures are then computed based on the center and samples of each cluster. Finally, the gait template is derived by combining Gabor features of averaged silhouettes and corresponding gait signatures. This approach called improved Patch Gait Features (iPGF), demonstrates a 1% improvement in Rank 1 and Rank 5 compared to the standard PGF, as observed through experimental results using the Random Subspace Method (RSM) classifier.

Keywords:

Gait recognition, Gabor filter, Gait classification, Patch Gait Feature, Regional Patch-based

Acknowledgements

I want to take a moment to thank the people who have helped me get to this point in my life that I am grateful for. First, I would like to thank my parents and my brother for their genuine support and kindness and motivating me throughout the journey. Without them, it will not be possible to achieve this goal. Next, I would like to thank my supervisor, Professor Madjid Alili, for his support and guidance in preparing the thesis. Next, I would like to thank Bishop's University and especially the Department of Computer Science for giving me the opportunity to pursue my graduate studies in a field that I am truly passionate about.

Contents

Chapter 1

Introduction..1

1.1. The Concept of Behavioral Biometrics 1

1.2. What is Human Gait Recognition?...5

1.2.1. Gait Identification 6

1.3. Gait Challenges 9

1.3.1. External Factors 9

1.3.2. Internal Factors 9

1.4. Gabor-Base Feature...9

1.4.1. Gabor Functions...10

Chapter 2

Related Works..14

2.1. Video-Based Approaches 16

2.1.1. Model-Based Methods 19

2.1.2. Feature-based Approaches 21

2.2. Gait Energy Image 23

2.3. Chrono-Gait Image 29

2.4. Patch Gait Features (PGF) 35

2.4.1. Local patch extraction 37

2.4.2. The Possible Distribution of Patches 39

Chapter 3

Proposed Method...46

3.1.iPGF Approach..46

3.1.1. iPGF Approach..50

3.2.Gait Classification..55

3.2.1. Random Subspace Sampling...56

3.2.2. Dimensionality Enhancing..57

3.2.3. Final Classification...60

Chapter4

4.1. Implementation and Result...62

4.2. Gait Dataset...63

4.2.1. Algorithm Benchmark...64

4.3. USF dataset Result..66

4.4. Complexities 70

Chapter 5

Conclusion 75

References 77

List of Figures

1.1.Several frames from the experiment of moving light points 6

1.2. Schematic of the behavioral pattern biometric system 7

1.3.The evolution of deep gait recognition methods ……...................................…..……8

1.4. 40 Gabor functions to display GEI according to rotation and different scales…….11

1.5. The plan of extracting three types of behavioral templates based on Gabor functions...13

2.1 Some examples of surveillance images...16

2.2. Several pictures of the person's movement (first row) and the corresponding silhouette...18

2.3. The model of inclinations and corresponding frequencies..20

2.4. Shows behavioral energy templates..24

2.5. Some examples of synthetic images produced from the original example (top left) 26

2.6. General block diagram of behavioral identification algorithm based on GEI...........27

2.7. Alternating signal extracted from behavioral templates..30

2.8. Some examples of CGI templates (first row) and artificial templates (second row) 34

2.9. General overview of three main steps of PGF...36

2.10. (a) The process of calculating X-Y-T histograms based on the location of local extrema..41

2.11. D+t2 position of top 20 X-Y-T histogram indices for three different people...........42

3.1. The main steps of our iPGF method...50

3.2. The general structure of RSM classification. 56

4.1. Computational load (in seconds) for different algorithms in different settings..........73

4.2. Amount of memory consumed (in GB) for different algorithms in different settings...74

List of tables

2.1. GEI synthetic pattern generation pseudocode ……..……..26

2.2. Three Approaches synthetic template generation pseudocode……...................……..44

4.1. The specification of probe sets in the USF gait dataset..............................………..64

4.2. Comparisons of Rank1 CCR (%) of the approaches on the USF dataset.....………..67

4.3. Comparisons of Rank5 CCR (%) of the approaches on the USF dataset...................68

Chapter 1

Introduction

Gait recognition is an essential component in applications like monitoring, security, and traffic management [1, 2, 11, 99]. However, accuracy can be affected by factors like clothing changes and lighting variations. To address this, Spatio-temporal templates have been used to capture both spatial and temporal information. Existing templates, like Pixel-level Gaussian Fitting (PGF), may leave noisy pixels, leading to inaccurate results. This thesis aims to enhance gait recognition systems by improving the temporal template in Spatio-temporal templates. A novel modification to PGF is proposed, involving post-processing to eliminate noisy pixels by selecting more robust patches. Experimental results show significant improvement in recognition rate compared to the original PGF method. Additionally, a feature selection approach called Improved Pixel-level Gaussian Fitting (IPGF) is investigated to enhance recognition accuracy.

This thesis has five chapters. The first chapter introduces the basic concepts. The second chapter examines the works that have been done and studied before, and the proposed model is presented in the third chapter. The fourth chapter introduces the classification method, and the results are presented in the fifth chapter.

1.1. The Concept of Behavioral Biometrics

Behavioral biometrics is the measurement of a person's physiological or behavioral characteristics for identification and authentication. The most famous methods of this type of identification are based on fingerprints, iris, signatures, and walking [8]. In behavioral biometrics, physiological, behavioral, and a combination of these characteristics are analyzed in people. Physiological features such as type of signature or style of walking deal with the things that humans are born with [3, 4].Behavioral traits are the result of our interaction with the environment and nature and can change over time. How to walk, how to speak, and how to write are including the combined features of both. For example, voice recognition can be considered in such a way that the size and shape of the vocal cords, the nostrils and lips, and the shape of the mouth on the one hand, and the age and geographic conditions of a person's life, on the other hand, are involved in this issue [3]. One of the important uses of behavioral biometrics in the field of information security is to identify people using the unique characteristics of activities at a distance in an unobtrusive way [1].

The daily increase in the need for security systems with computer vision methods has caused people to be recognized and identified with high accuracy and speed in surveillance environments. This type of identification is generally done in outdoor environments and without the person's knowledge. People who appear in the environment are identified without prior notice. Therefore, we can boldly consider the identification of "behavioral patterns" as one of the new technologies in this field. This technique is very useful in the “subtle" identification of people who are threats or suspects in security environments. In this research, we focus on human identification using gait features. Gait is defined as the coordinated and periodic movement of the human body, which leads to movement and transfer. Coordinated movements must be repeated according to a regular time pattern for behavior to occur [1, 2].Therefore, people's movement patterns walking is known as an effective type of biometric for human identification in public environments in surveillance and security programs [1]. The reason for this is mainly due to the ease of capturing walking at a distance. However, any biometric system based on gait recognition is affected by some external factors, i.e., covariates, such as clothing conditions, carrying conditions, viewing angle, surface, and the passing of age (and time) [2, 13]. However, it is important to note that gait can still be cited as an irreplaceable biometric. One of the main advantages of gait biometrics is that it can be captured from a distance, without requiring the subject to touch any devices. This makes it a convenient and non-intrusive form of biometric identification. Additionally, gait biometrics can be used in situations where other biometric modalities such as fingerprints or facial recognition may not be feasible or reliable, such as in low-light conditions or when a subject's face is obscured. Therefore, many identification and authentication problems can be solved with its help and used in identification problems [14].

Two general algorithms for gait identification have been developed during the last decade [23] to properly represent the human movement pattern: model-based algorithm and feature-based algorithm. In model-based algorithms, a predetermined structure is considered for the motion model [27, 40]. In these models, for example, the joint parts of the body (limb) are characterized by some parameters, and the recognition process can be transformed to adjust human motion with the predefined model. One of the effective and early research was presented in [26], which two types of parameters are proposed: time-independent parameters (static) and time-varying parameters (dynamic). Static parameters are limb length or body size, while dynamic parameters are the angles between limbs that are constantly changing during movement. For gait recognition, the statistical information of limb movement such as mean and variance of angles, is first calculated. Then, these features are related to previous knowledge in the Bayesian format and framework, and the final model is selected from among different models [26]. Here, although motion detection based on the Bayesian framework has a high performance for walking in the indoor environment, it has a high computational cost and is very vulnerable to noise [15]. Meanwhile, the model parameters can be effectively improved by hidden Markov model (HMM) or Fourier descriptors [41].

Nevertheless, model-based algorithms are vulnerable to some parameters: On the one hand, the initial model is vulnerable to noise and occlusion [15, 24]; on the other hand, the fitting process of the model is very time-consuming, and the complexity of the model increases exponentially with the increase in the number of parameters [11, 29]. According to this hypothesis, it is reasonable to calculate the movement characteristic from the walking appearance. In addition, for better recognition, the sequence of moving silhouettes is collected and converted into a single image called a pattern [1, 24]. For example, the most famous behavior patterns based on appearance are the gait energy image (GEI) [1], "Gait Flow Image" (GFI) [42], and "Gait Entropy Image". (GEnI) [34], all of which have a simple structure with low computational costs. Since the nature of human movement is a Spatio-temporal process, this type of transformation removes the time sequence from the appearance of walking [15, 16, 25]. More precisely, the templates collected and analyzed spatial-based features without using time-based features.

Recently, some gait templates have been proposed to maintain the timing of movement steps in a periodic period [14, 15]. However, none of the developed features adequately characterizes the type and model of human movement. An optimal approach based on local patches, called Patch Gait Feature (PGF) [54], has recently been developed to solve this problem. In the proposed method, local patches are extracted from a motion sequence. Then the local extremum points are extracted, and possible distributions based on the patches are calculated. Finally, the final template will be calculated as a movement feature based on the possible distribution of patches. But this method has limitations in different aspects. Based on the distribution of local patches in human walking, the proposed approach provides an optimal model by focusing on extracting patches efficiently and using them effectively. The proposed model, iPGF or improved PGF, aims to enhance the limitations of the PGF template. To introduce iPGF completely, we will expend in the next parts.

1.2. What is Human Gait Recognition?

Gait recognition is a biometric method aiming to identify people based on how they are walking [6, 11]. Gait, referring to the style of human walking, contains valuable features that are applicable in various fields, including sports science, sentiment analysis, health, and people identification. People's walking patterns can be captured using different sensing methods, such as wearable sensors attached to the body, such as accelerometers, gyroscopes, and force and pressure sensors. Non-wearable gait detection systems, on the other hand, mainly rely on vision and are commonly referred to as vision-based gait detection. These systems utilize imaging sensors to record walking data without the need for subject cooperation, even from a considerable distance [25]. Since 2015, with the introduction of deep learning into this category, gait recognition methods based on deep learning have now made the subject of biometrics more practical using the most advanced technologies [25, 28].

Behavior pattern as a biometric feature has a long history. One of the important and old works in this field is identifying people from "moving light displays" (MDL). This issue was planned and followed by Johanson at Uppsala University [9]. In this research, similar experiments were conducted to investigate how human vision is stimulated by moving light points. In this way, 13 different light points were installed on the tested person’s body. These points were installed on the person's head, shoulders, elbows, chest, thighs, knees and ankles and he was asked to move in the dark environment. Then, the lights received from a stationary person compared to a moving human eye and a moving person compared to a fixed human eye were examined [37]. Figure 1.1 shows several frames of moving light points in this experiment.

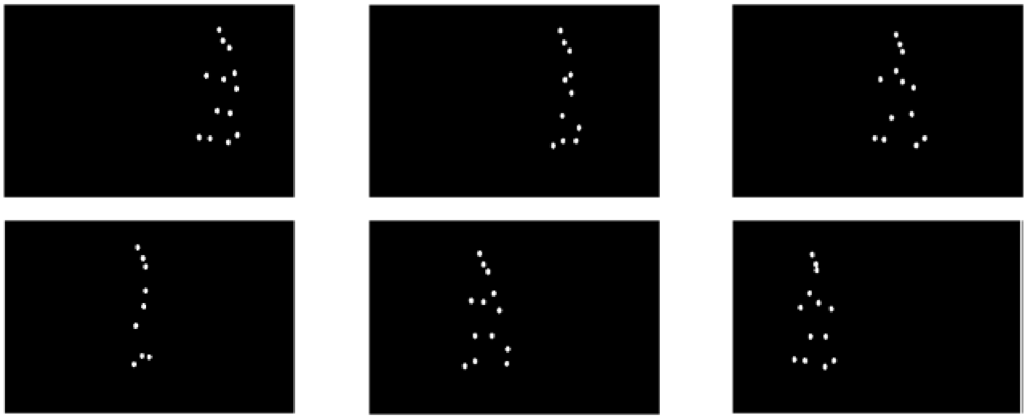


Figure 1.1. Several frames from the experiment of moving light points.

This experiment has an important result from a biomechanical point of view :people have a unique and distinct way of walking. Also, the results of this experiment from a psychological point of view prove that the human visual perception system has three other unique abilities: 1. It can distinguish the human movement pattern Figure 1.1 from other movement patterns, 2. It can recognize the type of walking of its friends and 3. It can recognize the gender, the direction of movement and the conditions of carrying a person's load [8]. The type of human movement is not limited to how he walks, but most of the approaches use it as a case study [2, 9, 10]. Therefore, in the following, the behavioral pattern is considered to mean the "type of walking" of a person [42, 44, 52, 62, 72].

1.2.1. Gait Identification

Identification systems based on biometrics such as behavioral patterns usually have two phases of training and testing. In the training phase, patterns of people are defined and stored, and then in the testing phase, new patterns of people are received and compared with the database. The new models in the test phase refer to the different conditions of the individual from the training phase. For this reason, the training phase is also called learning or registration. If the new template does not match any of the previous templates, it should be stored in the database [22]. The structure of a behavioral biometric system is shown in Figure 1.2.

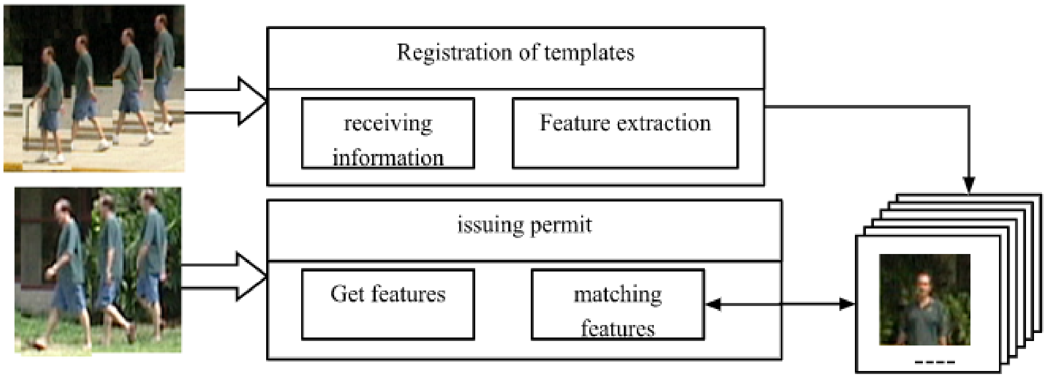


Figure 1.2. Schematic of the behavioral pattern biometric system[8]. In the registration (training) phase, the patterns are received and the features are extracted and stored according to the algorithm. Then, in the licensing (testing) phase, the features received from the new templates are matched with the predefined features.

The licensing department in Figure 1.2 identifies the new person in two general ways. In the first case, a person's biometric information is obtained from input sensors such as a camera along with other characteristics such as an ID card and is matched with the registered information of the person. But in the second case, only the biometric information of the person is received, and the search is done with the database information. Thus, in the first category, we have "one-to-one search" and in the second category "one-to-many search" and the second category we have one-to-many search. Issuing a license in the first case is called confirmation (authentication) and in the second case, it is called authentication [6, 8].

In general, two error criteria are defined to evaluate the algorithm in the test phase (issuance of permission). These errors are calculated in terms of "False Accept Rate" (FAR) and "False Rejection Rate"

(FRR) from a person's model [6]. They turn out the false acceptance rate is an error in which the intruder is recognized as one of the people in the database. But the false rejection rate is the error in which the person in the database is not identified correctly.

In the collection of biometric terms, in addition to the above concepts, two important concepts "gallery" and "probe" are also defined [7]. A gallery is a collection of people whose biometric information is defined in the system. But the probe (or signature) is a set of people who should get a license [8]. Therefore, Gallery and Probe contains a collection of behavioral videos of people under different conditions.

The performance of these systems is affected by various factors, the change which will make the diagnosis process more difficult:1. Changes to the person's appearance, such as carrying a handbag/backpack or wearing clothing such as hats or caps. put on a coat. 2. Changes in camera perspective. 3. occlusion factors, for example, where parts of the subject's body are partially covered by an object or part of the subject's own body known as self-occlusion, and 4- changes in the environment, such as complex backgrounds and high brightness levels or Low [10, 11]. Gait identification can be expressed in two external and internal formats, the details of which are stated in the next two sections.

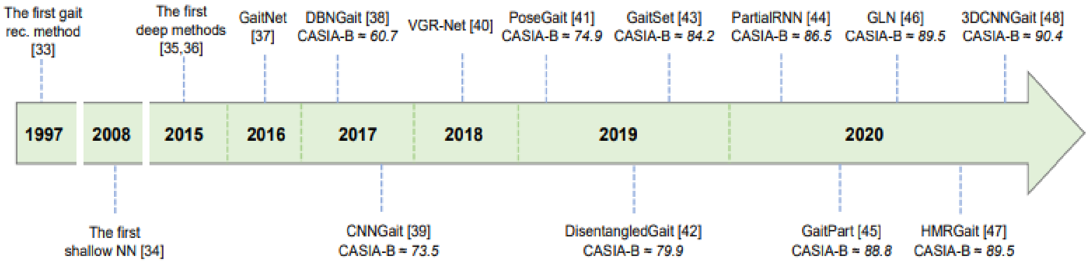

The first gait rec. method [33]
The first deep methods [35,36]
GaitNet [37]
DBNGait [38] CASIA-B = 60.7
VGR-Net [40]
PoseGait [41] CASIA-B = 74.9
GaitSet [43] CASIA-B = 84.2
PartialRNN [44] CASIA-B = 86.5
GLN [46] CASIA-B = 89.5
3DCNNGait [48] CASIA-B = 90.4
1997
2008
2015
2016
2017
2018
2019
2020
The first shallow NN [34]
CNNGait [39] CASIA-B = 73.5
DisentangledGait [42] CASIA-B = 79.9
GaitPart [45] CASIA-B = 88.8
HMRGait [47] CASIA-B = 89.5

Figure 1.3. The evolution of deep gait recognition methods

1.3. Gait Challenges

The subject of identifying humans based on their behavioral patterns, like any other subject, is an ongoing endeavor that continues to face challenges. In general, two factors have continuously created challenges in this field: external factors and internal factors [2, 16]. These factors affect the performance of all three approaches mentioned above. In the following, these external and internal factors are briefly examined.

1.3.1. External Factors

This factor affects the accuracy of identification algorithms more. For example, conditions such as angle 2 and internal view 1 (front, side, etc.), change of light (day and night), and external environments.

Building, change in seasons (rainy or sunny), type of clothing and clothing, type of movement surface (hard or soft, wet, stairs or 4 or 3 level concrete, grass, etc.), type of shoes (sandals, cotton, etc.) and the type of carrying object (backpack or handbag) affect the type and quality of human walking.[2]

1.3.2. Internal Factors

These factors change the movement from its standard state according to the internal state of the person. For example, various types of diseases, from colds and nervous diseases to organ defects, change the type of human movement. Also, psychological changes in the body due to old age, drunkenness, pregnancy, weight gain and weight loss are all internal factors.

1.4. Gabor-Base Feature

As we have explained, the average gait image is one of the most powerful features for gait recognition tasks. This means that a person can be identified through the average obtained from their gait images under different conditions. Moreover, it can be said that the image analysis based on Gabor functions is biologically related to image understanding and recognition. Hence, we utilize Gabor functions to model average walking images [25, 29].

1.4.1. Gabor Functions

According to the paper presented in [28] and in [10, 11] and in [71], they were able to develop two-dimensional Gabor functions that have good orientation and frequency selectivity for spatial localization. Also, in [41] provided a good definition for image representation using Gabor functions. A Gabor function (wavelet, kernel, or filter) is the product of an elliptic Gaussian envelope and a complex surface wavelet defined as.

The GEI model calculated in the previous section contains two important points:

1. Average behavioral images of a person under different conditions produce similar visual templates and 2. Average behavioral images of different people even in the same conditions create different templates. Therefore, we can identify the behavior of a person by the average of his behavior template. In addition, research shows that Gabor-based functions are very useful for image analysis and pattern recognition. In general, a Gabor kernel is the product of a Gaussian function with a complex wave:

$$\psi_{s,d}(x,y) = \psi_{\bar{k}}(\bar{x}) = \frac{\|\bar{k}\|}{\delta^2}.exp\left(-\frac{\|\bar{k}\|^2.\|\bar{x}\|^2}{2\delta^2}\right).\left[exp\left(i\bar{k}.\bar{x}\right) - exp\left(-\frac{\delta^2}{2}\right)\right] \quad (1\text{-}1)$$

where the variable $\bar{x} = (x, y)$ is the spatial coordinates and $\bar{k}$ is the frequency vector that determines the scale and direction of the Gabor function ($\bar{k} = k_s \exp(i\phi_d)$). Also, in Gabor functions, $k_s = k_{\max} / f^s$ is the value$k_{\max} = \pi/2$. Here we put the values 2f=, 4, 3, 2, 1, 0s and $\phi_d = \pi d/8$ for 0*d*=1,2,3,4,5,6,7. Also, in the Gabor relation, the value of exp(-δ2/2) is the DC component, which subtraction is used to remove this component and stabilize the brightness.

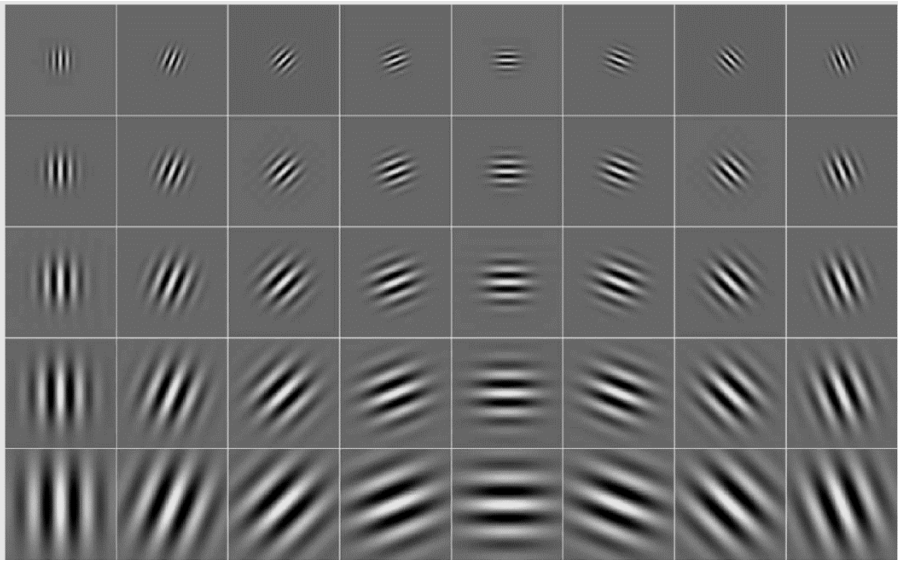


Figure 1.4. 40 Gabor functions to display GEI according to rotation and different scales [25].

Before explaining our method, it is necessary to introduce Gabor filtering. A Gabor filtering was first utilized in the GTDA template. According to GTDA [25], five different Gabor scales with eight different rotation directions are calculated, producing 40 Gabor functions. Each function consists of two real and imaginary parts, and Figure 1.4 shows the real part of these 40 functions.

Gabor's behavioral template is obtained by GEI convolution in Gabor functions. The result of this convolution will be the collection of images in space$R^{h\times w\times 5\times 8}$ (h and w dimensions of each image). Also, the obtained images are mixed, and their size is known as the Gabor template of each person. But calculating and directly using these 40 functions for behavioral identification will be very expensive. To solve this problem in GTDA, three types of behavioral representations are proposed. These three templates are: the sum of Gabor functions on direction (Gabor D), the sum of Gabor functions on scale (Gabor S) and the sum of Gabor functions on direction and scale (Gabor SD). These three functions will greatly reduce the time cost, and complexity of calculations will be greatly reduced. Therefore, Gabor D is the size of the image resulting from GEI convolution with a total of 8 directions (with a fixed scale).

$$GaborSD(x,y) = \left|\sum_{d} I(x,y) * \psi_{s,d}(x,y)\right|$$

$$= \left|I(x,y) * \textstyle\sum_{d} \psi_{s,d}(x,y)\right| \tag{1-2}$$

So, in this type of display, we will have five Gabor D images according to different scales. Similarly, Gabor S is the size of the GEI convolutional image with a sum of 5 scales (with fixed direction).

$$GaborSD(x,y) = \left|\sum_{s} I(x,y) * \psi_{s,d}(x,y)\right|$$
$$= \left|I(x,y) * \textstyle\sum_{s} \psi_{s,d}(x,y)\right| \tag{1-3}$$

Therefore, the number of 8 Gabor S images is obtained according to different directions. Finally, Gabor SD is the size of the template resulting from GEI convolution with the sum of all 40 Gabor functions.

$$GaborSD(x,y) = \left|\sum_{s}\sum_{d} I(x,y) * \psi_{s,d}(x,y)\right|$$
$$= \left|I(x,y) * \textstyle\sum_{s}\sum_{d} \psi_{s,d}(x,y)\right| \tag{1-4}$$

Figure 1.5 schematically shows the steps of extracting three Gabor D, Gabor S and Gabor SD functions. As we can see, three types of Gabor behavior representation will be obtained for each person. These three types of representation form a total of 14 Gabor template images of each person.

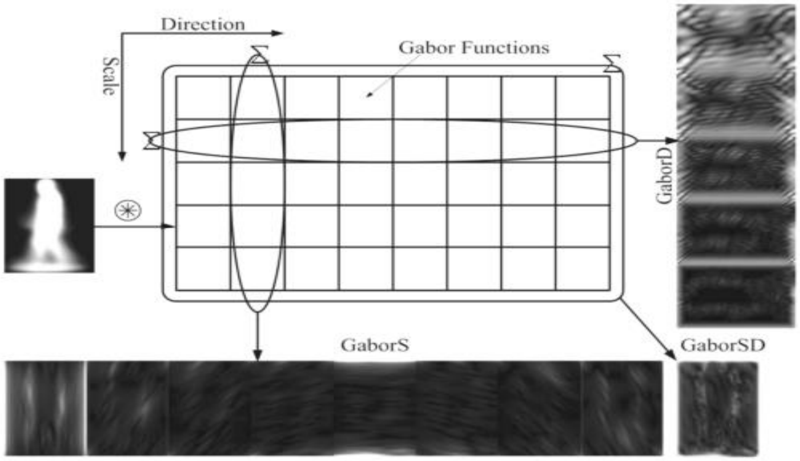


Figure 1.5. The plan of extracting three types of behavioral templates based on Gabor functions [25].

$$D(AS_P, AS_G) = \text{Median}_{i=1}^{N_P}\left(min_{j=1}^{N_G}\|AS_P(i) - AS_G(j)\|\right) \quad (1\text{-}5)$$

GaborSD has been used in PGF to extract patch features. In our method, we use GaborSD as we attempt to enhance PGF by discarding non-essential patches."

Chapter 2

Related Work

The daily increase in the need for efficient security systems has caused people to be recognized and identified with high accuracy and speed in surveillance environments. This type of identification is generally carried out in wide environments without the person's knowledge. People who travel in the environment are identified without prior notice [89].

For this purpose, the behavioral pattern is a coordinated and periodic combination of body movement that leads to human movement and transfer. Coordinated movements must be repeated according to a regular time pattern for the behavior to occur. Therefore, the two "coordinated" and "periodic" natures of human movement have distinguished his behavior. For example, the type of walking, running, walking, or climbing stairs are all types of human behavior [89, 96]. But sitting, lifting an object, or throwing it away are not considered behavior patterns. Because these are regular but non-intermittent movements. Also, "jumping in a person's place" is not considered his behavior. Because this type of movement, despite being regular and intermittent, does not lead to human transfer [90, 96]. Therefore, it is a style behavior model and a human transfer method, which is formed based on the regular and periodic movements of a person. The most common type of behavior pattern is the way a person walks on a flat surface. Another type is running or jogging, which has received less attention. Because in surveillance applications and public environments, the movement of the public is normal walking or walking. Behavioral biometrics should have five features [91, 96]:

1- Comprehensiveness: all people should have it.

2-Uniqueness: no two people have the same properties.

3-Durability and survival: These characteristics should not change over time.

4- Collectable: these properties should be measurable by physical and practical sensors.

5- Quality: the increase and growth of the population,in general , does not have a noticeable effect on its measurement.

It must be accepted that no type of biometric has fully covered one or all the above features. For example, the fifth item "quality" has a great impact on the measurement method. Therefore, the issue of behavioral pattern identification is important from two aspects: first, in terms of developing biometric systems and comparing them with other features such as face and fingerprints. Secondly, in terms of recognizing the behavioral pattern and its functional evaluation compared to similar methods [9].

Biometric systems are functionally divided into helpful and non-helpful categories, and in terms of characteristics, they are divided into physiological and behavioral categories [3, 94]. In cooperative (or shared) systems, the person cooperates with the system to confirm his identity, while in non-participatory systems, the act of identifying and determining the identity is done without the person's knowledge. Also, physiological properties such as fingerprints and hand geometry are physical properties that are measured and calculated at several points in space at a specific time. But behavioral biometrics, such as the way a person walks, the type of signature, or the tone of voice, rely on approaches that begin at a specific time and continue over time. In other words, behavioral biometrics can be learned and recorded over time and is dependent on the personal state of mind [5]. In fact, Physical Biometrics are sufficiently distinguishable if recorded at a specific time. But in behavioral biometrics, each sample alone does not contain useful information about a person's identity, but a period of samples will contain useful information [3, 94, 95].

Today, in most security systems, we need to identify people without their knowledge. Therefore, the identity and biological information of people should be recorded from a distance. But in real conditions, the presence of noise and images with low resolution limits the choice of biometric type. Meanwhile, the behavioral pattern can be identified from far distances. Therefore, one of the advantages of the behavior pattern compared to the face is that it has the capacity for remote and low-resolution identification [16]. Even in some situations, people's faces cannot be measured and distinguished, but how they walk can be measured [4]. For example, Figure 2.1 shows three images of different people captured by surveillance cameras. As we can see, the face image, geometry of the person, or other biometric parameters are ambiguous and cannot be measured. But their walking style is clear is clear in the picture.

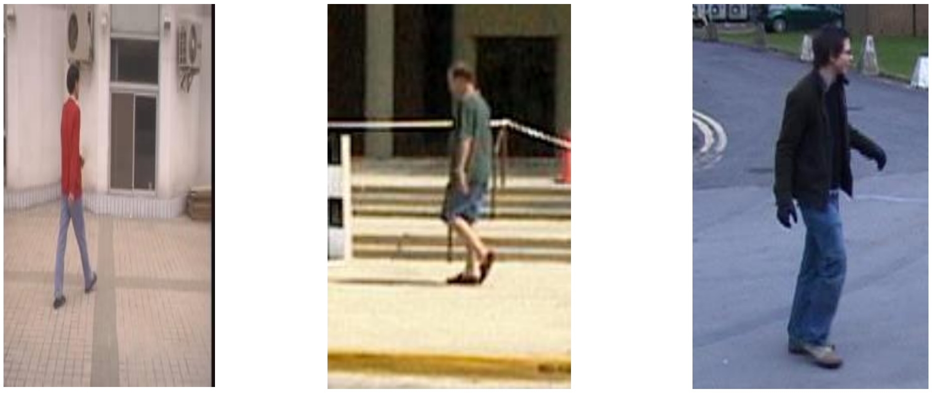


Figure 2.1 Some examples of surveillance images [8].

2.1. Video-Based Approaches

In this type of identification, behavioral patterns are recorded from a distance using surveillance cameras. Image and video processing techniques are then used to extract behavioral features from the footage, which are then used for recognition. Older approaches use step distance and movement rhythm to perform identification [12]. Some models also use static parameters of the body, such as height, distance between the head and waist, maximum distance between the waist and legs, and distance between the legs, for identification [12]. In general, most algorithms in this field perform human identification based on silhouettes. First, the background image is calculated, and then the input video is subtracted to obtain the silhouette image. For instance, Figure 2.2 shows several images of a person's movement along with their silhouette images [13, 14]. With this method, we can calculate the average of the silhouettes over a period and use the Euclidean distance to measure the similarity between these averages [33].

The initial techniques in this field had promising results in identifying behavioral patterns. But their big problem was the limited amount of data [15]. Fortunately, today, with significant progress in pattern recognition, the accuracy of algorithms for a large amount of data has increased. For example, the results on more than 1870 behavioral videos showed a recognition accuracy of about 40%, which today has reached more than 70% [16]. Most of the current methods in gait recognition are based on extracting the motion vector (MV) features from the video sequences [1, 2,5, 6, 35]. In other words, In gait recognition, motion vector refers to a mathematical representation of the direction and magnitude of motion for each pixel or group of pixels in a video frame sequence. These vectors capture the motion information of the human body during walking and are used to extract distinctive features for gait recognition. Motion vectors can be computed using various techniques, such as optical flow, block matching, or gradient-based methods, and they are often used in combination with other image processing and machine learning algorithms to build gait recognition systems [2, 35]. The important fields of application of this method are surveillance systems, criminology, behavioral therapy, and physiotherapy. Although the accuracy of this method is less than other biometrics such as fingerprints, it can be used as a useful tool.

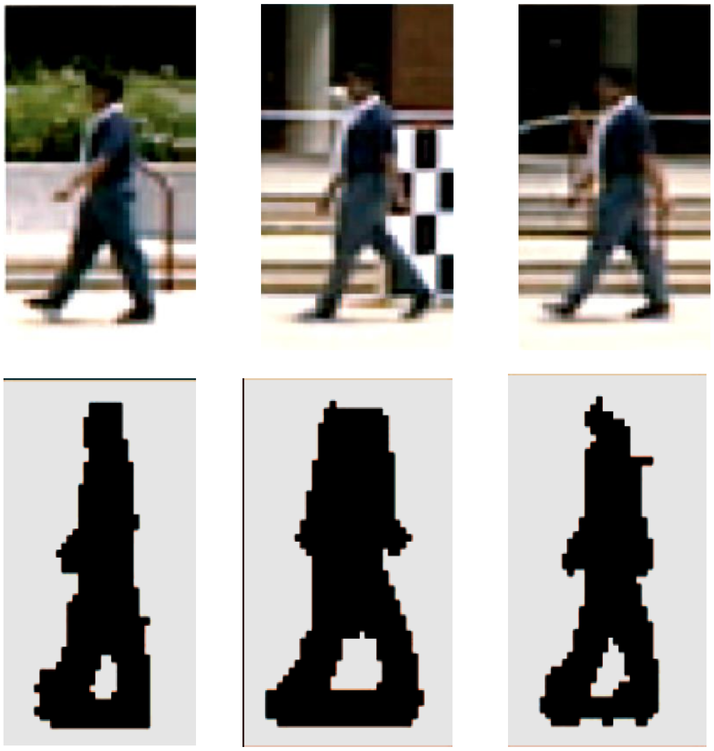


Figure 2.2. Several pictures of the person's movement (first row) and the corresponding silhouette (second row).

For example, in a bank robbery in Denmark in 2007, the court used behavioral pattern information to identify criminals and confirm them [8]. Similar applications in England and other countries indicate the usefulness of video-based approaches in identifying people [8].

2.1.1. Model-Based Methods

For this purpose, they first extracted a contour or edge from the silhouette image and then fitted an inclination model to it. In this method, the initial behavioral model and its parameters are produced in the XYT space. Then, parameters such as the frequency of movement of the bars, their angles relative to each other, or the length of each member (bar) are calculated as characteristics of the individual's movement. For example, Figure 2.3 shows several models of shafts for the lower body (left legs, L, and right legs, R) [8]. Here, the person's movement model consists of several different frequencies, the main frequency of which corresponds to a person's walking period. Also, other frequencies are multiples of the main frequency, which is related to the person's "step-by-step" movement. In model-based approaches, this value is set as the initial value according to the training data [27, 45].

Algorithms in this category try to describe human movement according to a previous model. The goal of this approach is to learn a direct mapping of observation data to the initial structure in terms of training data [3, 17, 27, 28]. For this purpose, a basic model of the structure of the human body is described first. Then the necessary parameters are calculated and extracted from this initial model. Finally, by defining a suitable fitting function, the degree of similarity of the new parameters to the initial model is found. The parameters of the model can be dynamic characteristics such as the length of each step and movement speed or static characteristics such as the ratio of the sizes of different body parts [10, 11, 45]. Older models tried to produce a basic Spatio-temporal volumetric model of human movement [8]. For this purpose, they first extracted a contour or edge from the silhouette image and then fitted an inclination model to it. In this method, the initial behavioral model and its parameters are produced in the XYT space. Then, parameters such as the frequency of movement of the bars, their angles relative to each other, or the length of each member (bar) are calculated as characteristics of the individual's movement. For example, Figure 2.3 shows several models of shafts for the lower body (left legs, L, and right legs, R) [8]. Here, the person's movement model consists of several different frequencies, the main frequency of which corresponds to a person's walking period. Also, other frequencies are multiples of the main frequency, which is related to the person's "step-by-step" movement. In model-based approaches, this value is set as the initial value according to the training data [27, 45].

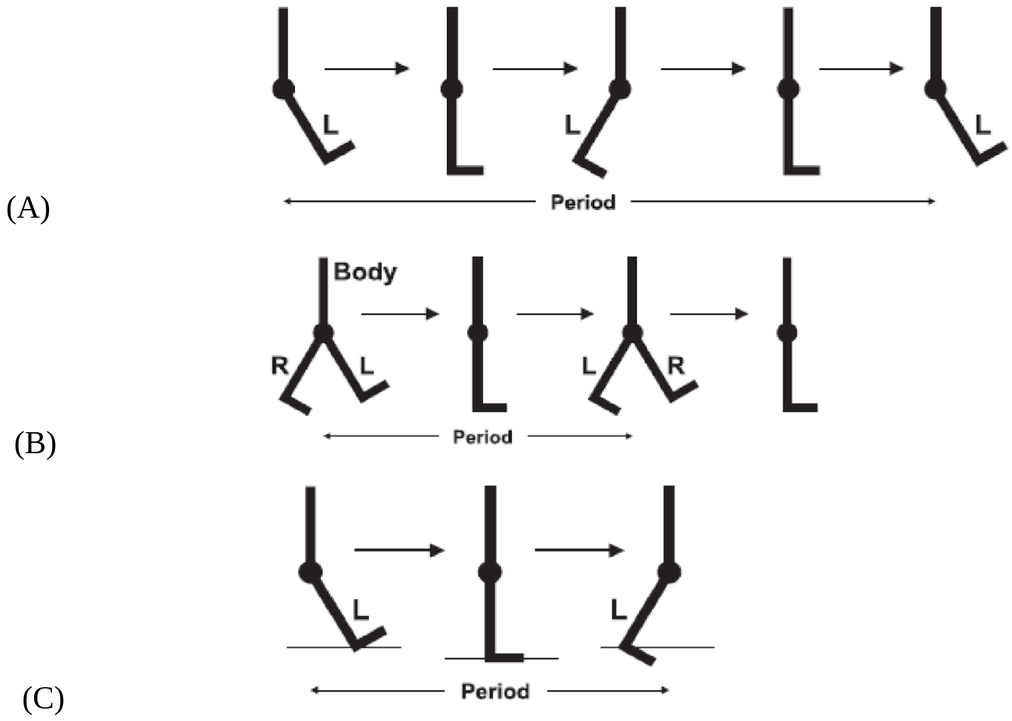


Figure 2.3. The model of inclinations and corresponding frequencies. (a) main frequency (b) and (c) frequency multiplier [8].

The approach similar to the rod model estimates the angles between the thigh and the knee from the body contour model by linear regression analysis [10]. Then, a trigonometric polynomial function is used to fit the series of angles over time. Finally, the estimated parameter values are used to identify and identify people [9,8]. In another approach, the human silhouette image is divided into local areas. These local areas correspond to different parts of the human body. Then, by fitting different ellipses, the body structure is described. Therefore, in this method, spectral and spatial characteristics of local areas have been calculated and used for identification.

In general, in these approaches, the accuracy of motion model recovery depends on the quality of silhouette image extraction. Therefore, in the presence of noise, the extracted parameters may not be reliable. To solve this problem, model-based methods have been developed to use tracking estimates in them. For example, Bayesian tracking is used to estimate human motion models [26,27]. The purpose of this approach is to use the advantages of the Bayesian approach and the model-based method.

2.1.2. Feature-based Approaches

The most important limitation of model-based approaches is the dependence of the accuracy of the algorithm on the quality of the initial model. Therefore, if the initial model is somewhat noisy or the model has an error, then the identification accuracy will drop a lot. This issue has led to the development of feature-based approaches. In this approach, suitable features are extracted and collected from the behavioral template during the person's movement, and the calculated features of people (and not the initial templates) are compared [2, 5]. These behavioral features are of different types, the most important of which are: full silhouette image [2, 5, 6, 15] Fourier descriptor [28], Gabor filter [25, 29], Radon transform [30], scatter display [11], color behavioral image (CGI), wavelet transform, capsule neural network (LBC) based feature [86], and patch-based features [53, 62]. Since the feature-based approach works better in different conditions, we use it in this research as well.

Feature-based approaches are divided into two general groups regardless of the feature type of temporal information preservation.

These two approaches are known as time models and format models [2, 11, 24]. In fact, these two models refer to the way of extracting shape and behavioral dynamics. The first strategy stores time information for pattern recognition. Therefore, behavioral videos of human movement are stored in this field. Then, according to the fact that each video contains several cycles of human movement, movement strings are extracted. Finally, in this model, these movement sequences (including an individual movement period) are directly compared [2]. In this field, the Hidden Markov Model (HMM) and its population type (pHMM) [5] or Fourier analysis [28] are widely used. Here, to increase the accuracy, a large amount of data should be used for learning. Therefore, the most important weakness in the direct comparison of disciplines is its computational complexity. In addition, a lot of memory is needed to store data.

In the second strategy, the sequence of images is converted into a single template [1, 24, 31, 32]. For example, the simplest way is to average human silhouettes during movement [1,14], known as "Behavioral Energy Image" (GEI). In this model, a gray image is produced from the average of binary silhouettes. Then, the feature space is reduced by analyzing the basic components. Finally, identification is done by defining the distance in the reduced space. Thus, with the development of GEI, the methods of "General Tensor Discrete Analysis" (GTDA) [25] or "Discrete Analysis with Tensor Representation" (DATER) were obtained for pattern recognition [6]. Another approach is "multilinear tensor-based nonparametric dimensionality reduction" (MTP) [23]. In addition, recently, the MTP method has also been used to identify behavioral templates with low resolution [4]. The most important limitation of this approach is the loss of "temporal order" between human movement phases. Because there is no temporal information in the mean templates. To solve this problem, a multi-channel technique called color behavioral imaging (CGI) has recently been developed [15]. Here, to preserve temporal information, each motion phase is represented by a separate color (channel). Then averaging is done in the RGB color space to get an average color image. In addition, a model was obtained by analyzing statistical information that examines the effect of different characteristics on behavioral templates [33]. In general, regardless of the variety of model-free approaches, the models of this field have progressed in two directions: 1- extraction of more powerful features and 2- more detailed analysis of basic components and more effective reduction of feature dimensions. For example, the extraction of suitable features has led to a behavioral color image [14] or sparse data representation [11]. Also, significant developments have been made in reducing dimensions. For example, classical techniques used "Principal Component Analysis" (PCA) and "Linear Discrete Analysis" (LDA). But newer methods use tensor approaches in which the behavioral energy image is expressed with a second-order tensor (i.e. matrix) to maintain the two-dimensional structure of the data [15, 34, 35].

2.2. Gait Energy Image

As stated, all human motion frames are compared in time characteristics. The accuracy of these models depends on the synchronization of movement phases. Because the beginning, middle and end frames of the movement sequences should refer to the same state of the person's step [14]. For example, in videos with a low frame rate, the movement phases between the gallery set and the probe are reduced, and the probability of moving frames is high. Because the behavioral characteristics in each movement cycle are very scattered and few. To solve the problem of synchronization, the features of the template in behavioral recognition have been proposed and developed. In these models, all images are converted into a single image in one motion cycle. Then the unit images are compared and identified. The simplest approach in this model is the averaging of silhouette images [1, 2].

One of the most useful approaches in extracting the feature of the template and the basis of the approaches in this field is "Behavioral Energy Image" (GEI). This model was proposed and developed in 2006[1]. The results show that this method has good accuracy despite its simplicity in the calculation. Here we represent the sequence of input silhouette images at time t by Bt(x,y). Therefore, the GEI template, G(x,y), will simply be the average of the silhouette images in a motion cycle:

$$G(x, y) = \frac{1}{N}\sum_{t=1}^{N} B_t(x, y)$$

(2-1)

where N is the complete cycle of the behavior template. Figure 2.4 shows two strings of binary images of the motion template and its extracted GEI template.

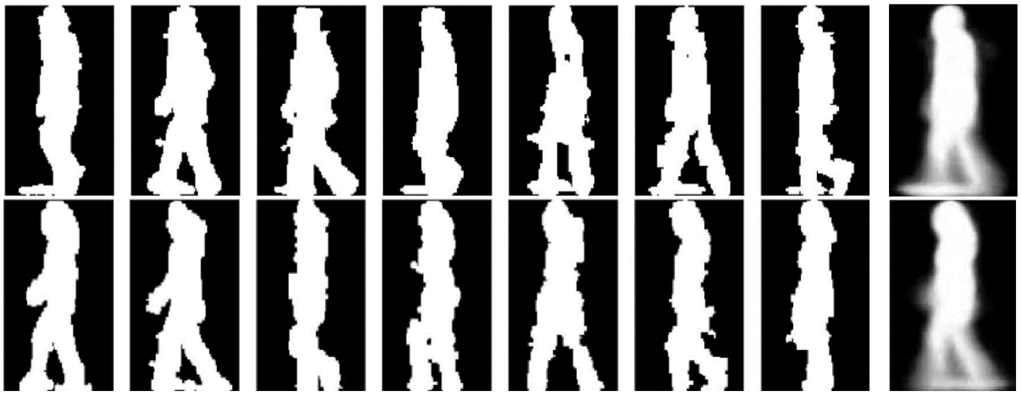


Figure 2.4. Shows behavioral energy templates. Multiple movement threads (left side) and extractive template (right side) [1].

As can be seen, GEI shows the main shape of silhouette images and their movement over time. This template is of the energy type because 1. each silhouette image is a normalized spatial image of the person's movement, 2. GEI is obtained from the accumulation of temporal energies, and 3. a pixel with a higher brightness (or energy) value means that the number of moves at that point has been higher.

The use of this type of display is important from two aspects. First, we prove that GEI is more robust to noise than the silhouette image. Secondly, in this algorithm, a model for extracting artificial images is proposed, which is more effective in increasing the efficiency of the identification algorithm. For behavioral identification, the discrete linear analysis (LDA) approach has been used, and for classification, the Euclidean distance criterion has been used in the reduced space[36].

The second step after extracting the GEI template is to generate synthetic images. In general, the number of images for each person is limited. This issue makes it difficult to accurately identify the behavior template under different conditions. A solution is directly comparing of the sequence of images used by temporal models. But as mentioned, this approach is sensitive to silhouette changes such as size change and displacement. To overcome this problem, we produce two real and artificial sets of behavioral templates. Real formats ({Ri}) are obtained by direct calculation from the string of images in the data [1]. But the artificial templates ({Si}) are obtained from the Ri and using a deviation model [1]. This deviation model is somewhat stable to silhouette changes and similar to the original model. Therefore, from a real form, R0, several synthetic models are developed that have the general properties of the GEI form. To produce this template, we cut and resize the lower parts of the template in different scales. Because most of the silhouette changes in different people happen in the leg area, we use this area to remove the changes. The experiment silhouettes that the height of the bottom of the leg to the top of the ankle in the silhouette image is about 1.24 of the total shadow. To ensure the scale, in the production of synthetic images, 3.24 of the height of each silhouette image is selected for cutting. Finally, with an iterative process, the lower rows of the silhouette are cut and each time the cut image is resized. Table 2.1 shows the pseudocode of the steps of producing synthetic templates [1].

Table 2.1. GEI synthetic template generation pseudocode [1].

```
1. The input GEI template (R0) with dimensions Y×X is known.
2. The h variable indicates the maximum cutting height from the floor. This value shows the maximum allowed deviation from the original template.
3. We initialize k=2 and i=1.
4. Remove r=k*i number of rows from the input template.
5. Resize the obtained template with dimensions Y×(X-r) to size (XY/X-r)×X.
6. Cut the left and right edges equally to get the Si templatewith Y×X dimensions.
7. i=i+1
8. If k*i≤h go to step 4, otherwise, stop the process of generating artificial images.
```

With this process, several synthetic templates are produced from each real template. Figure 2.5 shows 8 synthetic templates produced from real templates. The image on the left shows the original GEI. Also, the first row of the cutting steps and the second row of Figure 2.5 shows the fitting steps. Also in the GEI, it has been applied nearest neighbor interpolation for fitting. These synthetic templates are generated for all individuals in the gallery and probe sets.

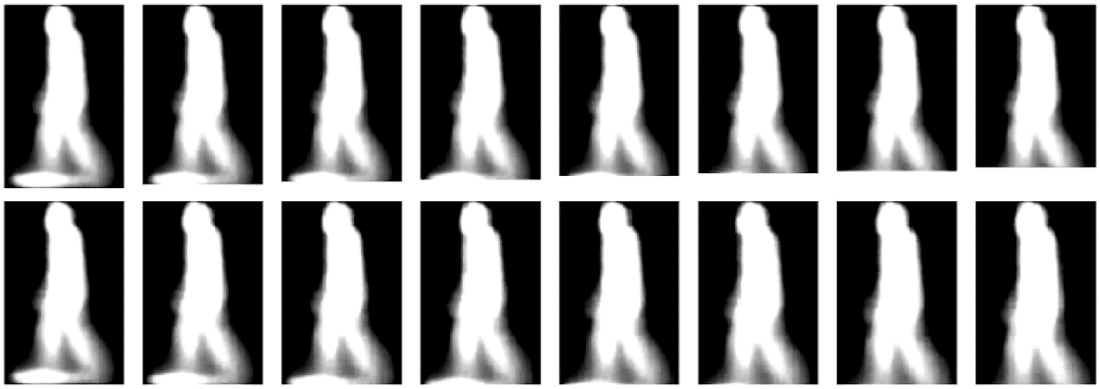
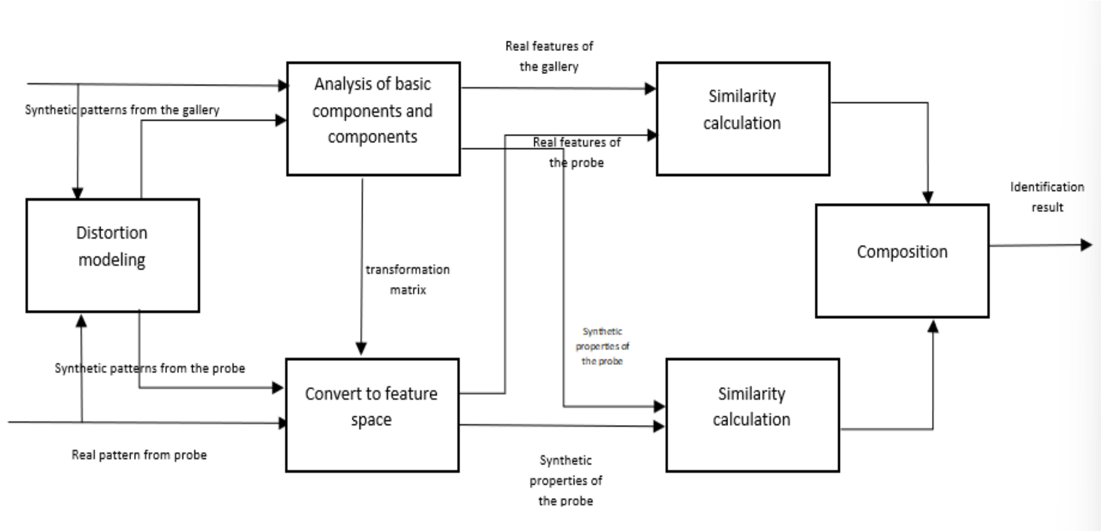


Figure 2.5. Some examples of synthetic images produced from the original example (top left) [1].

After producing these images, a comparison and matching between the gallery collection and the probe should be made. For this purpose, in GEI, examine the real patterntemplates together and the artificial templates separately and finally combine the results.

To identify, first, a transformation matrix is generated using the real data by the LDA method. Then, by this transformation matrix, the feature space of artificial data is also reduced. After reducing the space, we compare the real and artificial sets using the similarity criterion and finally combine the results. Figure 2.6 shows the general block diagram of the identification algorithm based on GEI.



Real features of the gallery
Analysis of basic components and components
Similarity calculation
Synthetic patterns from the gallery
Real features of the probe
Identification result
Distortion modeling
Composition
transformation matrix
Synthetic properties of the probe
Synthetic patterns from the probe
Convert to feature space
Similarity calculation
Real pattern from probe
Synthetic properties of the probe

Figure 2.6. General block diagram of behavioral identification algorithm based on GEI [1]

To measure the similarity, we use the Euclidean distance measure in the reduced space. Here, the sets {ri} and {si} are the real and artificial samples in the i-th class, c is the number of classes, and the average of the real and artificial samples in the i-th class. With this assumption, to calculate the distance of the probe templates to the average of each class, D(Rp, Ri), which is:

$$D(R_p, R_i) = \frac{1}{n_p} \sum_{j=1}^{n_p} \|r_j - m_{ri}\|, \quad i = 1, \dots, c \qquad (2\text{-}2)$$

where rj is the transformed matrix of the probe template and np is the number of probe samples. Therefore, the similarity of the probe template to my class will be obtained by minimizing this distance:

$$D(R_p, R_k) \min_{\mathrm{i=1 \to c}} D(R_p, R_i) \qquad (2\text{-}3)$$

Similarly, to calculate the distance between artificial templates, we have:

$$D(S_p, S_i) = \sum_{j=1}^{n_s} \|S_j - m_{si}\|, \quad i = 1, \dots, c$$

$$D(S_p, S_k\ ) = \min_{\mathrm{i=1 \to c}} D(S_p, S_i) \qquad (2\text{-}4)$$

where sj is the transformed matrix of the probe template and np is the number of probe samples. By combining real and synthetic distances, it is:

$$D(\{R_p, S_p\}, \{R_i, S_i\}) = \frac{D(R_p, R_i)}{N_{RR}} + \frac{D(S_p, S_i)}{N_{SS}} \quad , i = 1, \dots, c \qquad (2\text{-}5)$$

where $N_{RR} = 2 \sum_{i=1}^{c} \sum_{j=1, j \neq i}^{c} \frac{D(R_i, R_j)}{c(c-1)}$ and $N_{SS} = \sum_{i=1}^{c} \sum_{j=1, j \neq i}^{c} \frac{D(S_i, S_j)}{c(c-1)}$ are the normalization coefficients of real and artificial templates, respectively. Finally, the distance between the probe pa and the gallery ka will be obtained by minimizing this distance:

$$D\left(\{R_p, S_p\}, \{R_k, S_k\}\right) = min_{i=1}^{c} D\left(\{R_p, S_p\}, \{R_i . S_i\}\right) \qquad (2\text{-}6)$$

2.3. Chrono-Gait Image

One of the most recent ideas developed in the field of behavioral recognition is the Chrono (color image) algorithm. The initial model of this algorithm was proposed in 2010 and standardized in 2012 [14]. The purpose of this algorithm is to display the average behavioral image in color to describe it more appropriately. Because the most important limitation of template models is the removal of information and time sequence from the individual's behavior. Thus, chrono-behavioral imaging (CGI) is a type of multi-channel coding in which the behavioral string is converted into a color multi-channel image. This conversion will preserve the temporal information of the behavioral template [14]. Here, if the number of channels is 3, then the multi-channel image will become an RGB color image, each channel of which is a function of the individual's behavior in time. This algorithm is based on three different steps:

1.calculation of behavioral periodicity, 2. multi-channel mapping, and 3.CGI production. In the following, we briefly review these three stages.

The first step in CGI extraction is to calculate the behavioral periodicity. This is because the proposed method aims to determine the position and state of each frame relative to the initial frame of the period, and time information is obtained in terms of the distance from the beginning. In this process, as in the basic algorithm, the distance between the legs is used to calculate the frequency. However, factors such as bags, silhouettes, and surfaces can cause changes in efficiency and errors in calculating frequency. Therefore, in CGI, the effective distance of the legs (W) in the Ith image is calculated as follows:

$$W = \frac{1}{\beta h-\alpha h+1}\sum_{i=\alpha h}^{\beta h}(R_i - L_i), \quad 0 \le \alpha \le \beta \le 1 \qquad (2\text{-}7)$$

where h is half of the odd height and R_i and L_i are the locations of the left and right pixels of the i-th line in the foreground image, respectively. Pay attention that in relation (2-7) the height of the person is used instead of the length of the whole image. Because anatomical studies show that the points of the body in most people have a specific proportion to the person's height [14]. Here, α and β parameters are used to limit the area of the legs and reduce the influence of external factors in calculating the frequency, Figure 2.7 shows the alternating signal calculated by the CGI method. (Blue line) verses baseline method (red line). As we can see, the CGI approach produces sharper peaks than the basic algorithm. This issue makes comparison and period extraction more accurate.

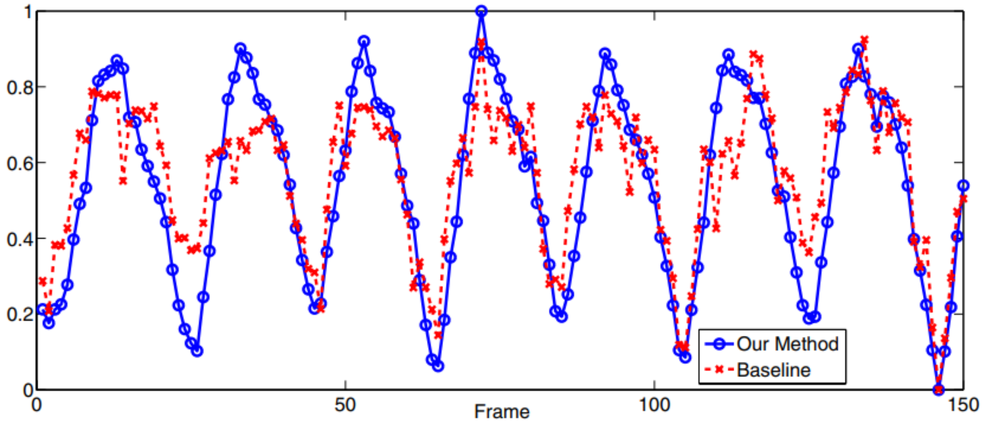


Figure 2.7. Alternating signal extracted from behavioral templates [15].

The second step in CGI calculation is multichannel mapping. For this purpose, the outer contour information of the silhouette image is calculated first. Because the contour is a suitable spatial feature of the silhouette information. But to calculate the contour, there are various methods such as gradient operator, LoG and local entropy. In CGI, the use of local entropy information expresses more suitable properties of the edge [14]. Therefore, using this method, edge information is extracted here, which is beyond the scope of the discussion. After calculating the contour, a linear interpolation between the frames is done to obtain the position of each contour relative to the reference. First, by using a function, the position of each frame of the motion sequence is mapped to the interval [0,1] in 1.4 behavior intervals.

$$r_t = \frac{W_t - W_{min}}{W_{max} - W_{min}} \qquad (2\text{-}8)$$

where W_t is the average leg width signal (relationship (2-8)), W_{max} and W_{min} are the upper and lower limits of leg width in ${}_4/^1$ behavioral frequency. Then, according to the position of each frame, a different weight (C_i (r_t)) is assigned to it to form the color of the channel. When the number of channels is one (k=1), the strategy will be the same as the GEI template; That is, each frame has the same weight in color composition.

But when k>1 all frames are divided into k-1 equal parts during ${}_4/^1$ interval. The border between these parts is determined by the points 1 - $p_i={}^i/_{(k-1)}$, i=0, 1 ,…,k . Then we assign a certain weight to the i-th channel of the image to describe the (i-i th to i-1 part of this time interval:

$$C_i(r_t) = \begin{cases} \left(\frac{r_t - p_{i-2}}{p_{i-1} - p_{i-2}}\right) I & p_{i-2} < r_t < p_{i-1} \\ \left(1 - \frac{r_t - p_{i-1}}{p_i - p_{i-1}}\right) I & p_{i-1} < r_t < p_i \\ 0 & Others \end{cases} \qquad (2\text{-}9)$$

where I is the maximum brightness (or value of 255). For a proper description of the channels, their number is chosen equal to 3 (K = 3) to correspond to the weight in red, green and blue channels.

Therefore, the weight of human movement in ${}_4/^1$ of its behavioral frequency is mapped to the RGB space as follows:

$$B(r_t) = C_1(r_t) = \begin{cases} (1-2r_t)I & 0 \leq r_t \leq 1/2 \\ 0 & 1/2 < r_t \leq 1 \end{cases}$$

(2-10)

$$G(r_t) = C_2(r_t) = \begin{cases} 2r_t I & 0 \leq r_t \leq 1/2 \\ (2-2r_t)I & 1/2 < r_t \leq 1 \end{cases}$$

(2-11)

$$R(r_t) = C_3(r_t) = \begin{cases} 0 & 0 \leq r_t \leq 1/2 \\ (2r_t - 1)I & 1/2 < r_t \leq 1 \end{cases} \tag{2-12}$$

In the third step and before calculating the CGI template, a multi-channel behavioral contour image (C_t) is obtained by multiplying the RGB weights in the input image (h_t).

$$C_t(x,y) = \begin{pmatrix} h_t(x,y) * C_1(r_t) \\ h_t(x,y) * C_2(r_t) \\ \vdots \\ h_t(x,y) * C_k(r_t) \end{pmatrix} \tag{2-13}$$

Then, by summing the C_t values in the direction of all the channels (n_i), the middle image of PGI ($PGI_i(x,y) = \sum_{t=1}^{n_i} C_t(x,y)$) is obtained in every interval of ${}_4/^1$ intervals. Finally, by adding PGIs and averaging them, a CGI template is produced.

$$CGI(x,y) = \frac{1}{p} \sum_{i=1}^{p} PGI_i(x,y) \tag{2-14}$$

Where p is the number of ${}_4/^1$ behavioral intervals. Pay attention that we have used " saturated addition " in calculating PGI. Thus, whenever the total value exceeds the maximum brightness value (ı), the result will be the value of ı. But in the calculation of the CGI template, the values are added normally, Because the sum operation is just a simple averaging over the color channels.

After extracting CGIs in the set of probe and gallery images, the similarity between templates should be measured. Here are the steps of behavioral identification similar to the GEI model [14]. In this way, first a set of artificial images is produced, and the templates are not compared directly. Because firstly, the number of CGI templates is very small and cannot model the behavioral characteristics of training data (sample reduction problem). Secondly, if each pixel is considered as a dimension in the feature space, the feature space is very large, and we face the problem of the " curse of dimensionality ".

To solve these problems, in addition to producing artificial images, the dimensions of the feature are reduced, and the distance criterion is calculated in the new space. To produce artificial templates like GEI, we repeatedly cut the bottom lines of the image and resize the result. Also, the PCA+LDA combination is used to overcome the dimensionality limitation problem. Therefore, the reduction of dimensions is done in two steps. First, the feature space is reduced by PCA, and then by LDA again, the dimensions of the reduction and the new space are obtained. Figure 2.8 shows some examples of real CGI images (in the first row) and artificial templates (in the second row).

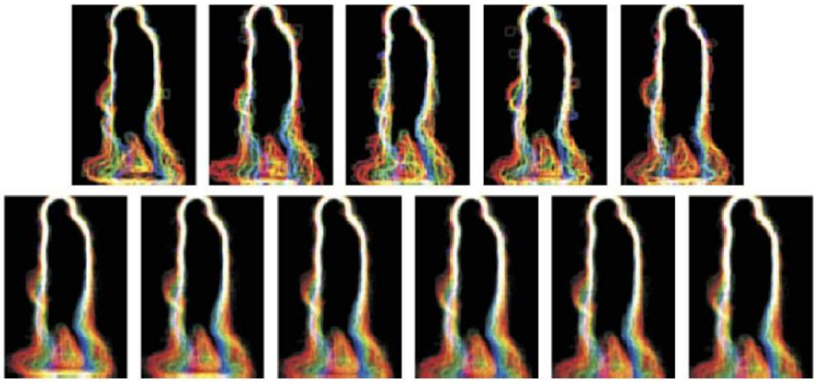


Figure 2.8. Some examples of CGI templates (first row) and artificial templates (second row) [15].

The distance criterion in the new space is also calculated in the same way as the equation (2-9). Thus, if d(R_p, R_i) is the distance between R_p and R_i (real) templates, d(S_p, S_i) is the distance between S_p and S_i artificial templates, S_p and S_i and C is the number of gallery classes; Then to calculate the total distance we will have:

$$d(R_P, S_P, R_i, S_i) = \frac{d(R_P, R_i)}{\min_j d\left(R_P, R_j\right)} + \frac{d(S_P, S_i)}{\min_j d\left(S_P, S_j\right)},$$

$$i, j = 1, \cdots, C \qquad (2\text{-}15)$$

Therefore, similar to the relation (2-12), the input probe template is assigned the k class label whenever its overall distance is minimized:

$$k = \arg\min_i d\left(R_P, S_P, R_i, S_i\right), \quad i = 1, \dots, C \qquad (2\text{-}16)$$

2.4. Patch Gait Features (PGF)

Recently, an approach based on local patches as mentioned in the previous chapter has been developed. In this method, local patches are extracted from movement sequences. Then the local extremum points are extracted, and the probability distribution based on the patches is calculated. Finally, the final template will be formatted and calculated as Patch Gait Features [50]. Since patch-based methods have been developed in recent years to identify the movement template [11, 47, 48, 51, 52], its use will be the ability to effectively display human movement. Because with the help of these patches, we can obtain Spatio-temporal and local information about the human movement process in periodic intervals. Figure 2.9 shows the structure of PGF [53] whit three main steps. Considering the PGF approach, some movement patches are more important than others. But the major limitation of the PGF approach is that all patches in the Spatio-temporal space are given the same importance. Only its histogram is calculated in the formation of the probability distribution (second step). In other words, the location of some patches may indicate local noise distorting the probability distribution. This issue will decrease the quality of the final feature display. To solve this limitation in this thesis, we add a condition that patches are monitored (refined) after being selected in the first step. Then we will use the monitored patches for possible distribution in the second step. The proposed algorithm is called the improved Patch Gait Feature (iPGF).

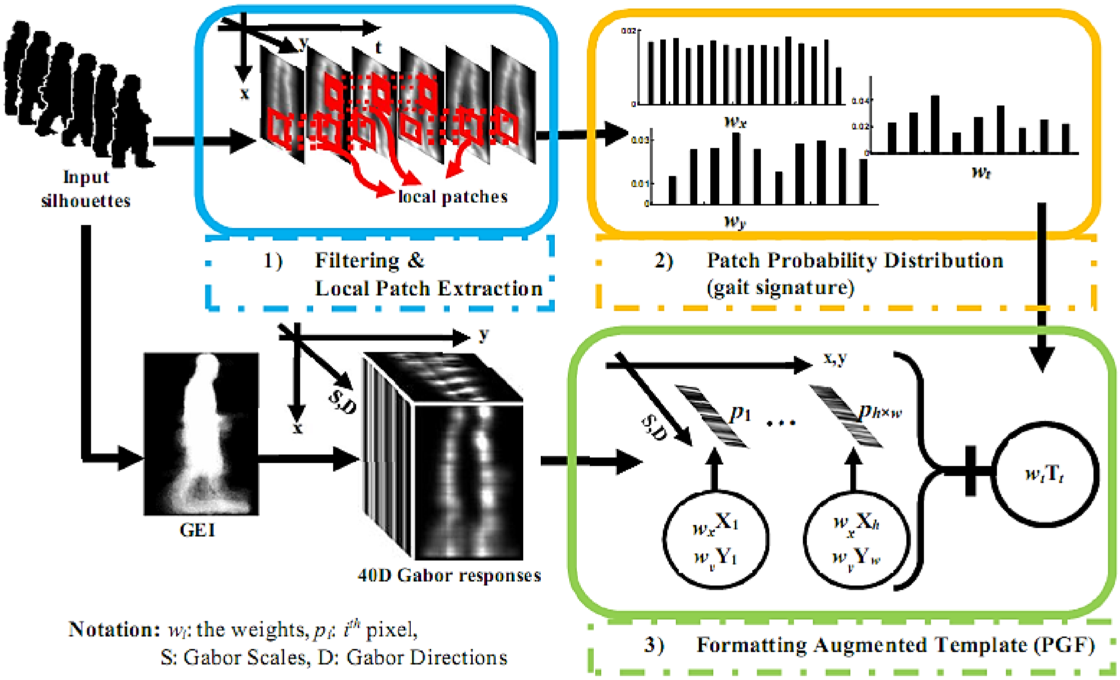

Input silhouettes
x
y
t
local patches
1) Filtering & Local Patch Extraction
w_x
w_y
w_t
2) Patch Probability Distribution (gait signature)
GEI
y
x
S,D
40D Gabor responses
x,y
S,D
p_1 ...
$p_{h\times w}$
$w_x X_1$
$w_y Y_1$
$w_x X_h$
$w_y Y_w$
$w_t T_t$
3) Formatting Augmented Template (PGF)
Notation: w_i: the weights, p_i: i^{th} pixel,
S: Gabor Scales, D: Gabor Directions

Figure 2.9. General overview of three main steps of PGF [53].

The PGF approach is based on three main steps [53]:

A.1. preprocessing and silhouette extraction of Spatio-spatial patches

A.2. calculation of the probability distribution of patches

A.3. calculation of PGF

In the following, we briefly explain these steps.

This section describes a different algorithm. In the proposed method, local patches are extracted from a motion sequence, and the local extremum points are identified. The probability distribution based on the patch is then calculated, and the final template is formatted and computed as a motion feature based on the possible distribution of patches [50]. The use of patch-based methods has been developed in recent years to identify movement templates [11, 47, 48, 51, 52]. By utilizing these patches, the proposed method is able to effectively display human movement and obtain Spatio-temporal and local information about the human movement process in periodic intervals.

Since the extraction of patches is based on the use of Gabor filters, the proposed method will briefly explain the relevant relationships before proceeding with the extraction of the patches. These patches will then be utilized to describe the movement template. To describe the motion template based on local patches, two different algorithms are presented in this section. First, the "patch-based motion features" (PGF) algorithm will be described, in which Spatio-temporal motion information will be added to the final templates [50]. Then, in the second part, "Gabor energy weighted template" (wGbEI) is stated, in which the information of patches and their density will be used in the description of the final template. Both features stated in this section are more accurate than filter-based features and are more suitable for describing spatial and temporal information [11, 47, 48, 51, 52].

2.4.1. Local patch extraction

Recently, the Gabor filter feature has been recognized as an effective feature for modeling human movement templates [11, 25, 35]. A set of Gabor filters in a default pixel z=(x,y) is defined as a complex exponential function (relation (2-7)), which briefly consists of:

$$\phi_{\tau,v}(z) = \frac{\|r_{\tau,v}\|^2}{\delta^2} e^{-\frac{\|r_{\tau,v}\|^2\|z\|^2}{2\delta^2}}\left[e^{ir_{\tau,v}.z} - e^{-\frac{\delta^2}{2}}\right], \qquad (2\text{-}17)$$

where it $r_{\tau,v} = \theta_\tau e^{i\varphi_v}$ represents the scale and direction of the Gabor kernel function. In the above relationship similar to [25], the value of u is equal to {0, 1, 2, 3, 4} and v is equal to {0, 1, 2, 3, 4, 5, 6, 7}. As a result, 40 Gabor kernel functions are obtained in 5 scales and 8 different directions (according to Figure 2.8). Also, the value $\exp(-(\delta^2/2))$ in relation (2-17) is subtracted so that the functions are independent of the DC value and are resistant to changes in brightness.

In the proposed approach, all motion frames are channelized with the above filters to extract the patches. For each image with dimensions $N_2 \times N_1$, 40 Gabor filter responses are obtained after convolution. Of course, according to relations (2-8) to (2-10) [25], the absolute value of each response is calculated to obtain real functions. Also, to reduce the computational burden, the dimensions of the Gabor images are reduced to $[{}^{N_1}/_2]\times[{}^{N_2}/_2]$, where $[{}^{N_1}/_2]$ and $[{}^{N_2}/_2]$ are the largest integers smaller than or equal to ${}^{N_1}/_2$ and ${}^{N_2}/_2$ [11]. It has been proven that this dimension reduction lowers the computational load without reducing the identification accuracy [11].

$$R_{SD} = \left|\sum_{\tau}\sum_{v} I(z) * \phi_{\tau,v}(z)\right|$$

$$= \left|I(z) * \textstyle\sum_{\tau}\sum_{v}\phi_{\tau,v}(z)\right| \qquad (2\text{-}18)$$

where I(z) is the input image with reduced samples and "*" is the convolution operator. In patch-based approaches [11], the spatial information of the pixels is added to the value of the pixels obtained from the Gabor filter. Therefore, the enhanced Gabor filter (ρ = 42) $p_h = \left[q_h^T, X_h, Y_h\right]^T \in R^{\rho}$ is obtained by adding pixel information. Also, q_h (h=1...H) is the value of Gabor pixels, where it $\left[\frac{N_1}{2}\right]\times\left[\frac{N_2}{2}\right]$ is the value of all filter pixels.

But in the proposed approach, instead of adding the coordinates of the pixels, first the appropriate extremes are extracted from the response of the filters and the Spatio-temporal coordinates will be added to the filter values.

Suppose all the silhouette images are filtered according to the equation (2-18) in a motion interval and their responses are calculated. Then, all of them are collected in a standard format (called image stack). Then consider a window with dimensions wt × wy × wx in x, y and t directions (3×3×3 in this research). Now, by moving the window in the original space, the maximum values are found within each 3D window. Suppose the location of the local maximum point is at the point (Xi, Yj, tk), then by searching for local points in the entire image space, the coordinate values of the local maxima are stored. Obviously, by completing this process, the desired values will be in the range of i=1... $\left[\frac{N_1}{2}\right]$, j=1... $\left[\frac{N_2}{2}\right]$ and t=1...T (T time period) [50]. The distribution of these extremes (or representative of local patches) will be used to represent the features of the motion template.

2.4.2. The Possible Distribution of Patches

After collecting the extremum points (or patches), the next step is to use the Spatio-temporal statistical distribution of the patches. These points and their distribution in the stack of images are the Spatio-temporal characteristics of the desired movement template. The PGF has been used histograms to describe this dispersion in the image stack space. Suppose the function is a function that assigns the location of the local maximum Xi to the index b(Xi). The probability distribution function (histogram) of the coordinates of extremes in three dimensions can be calculated as follows:

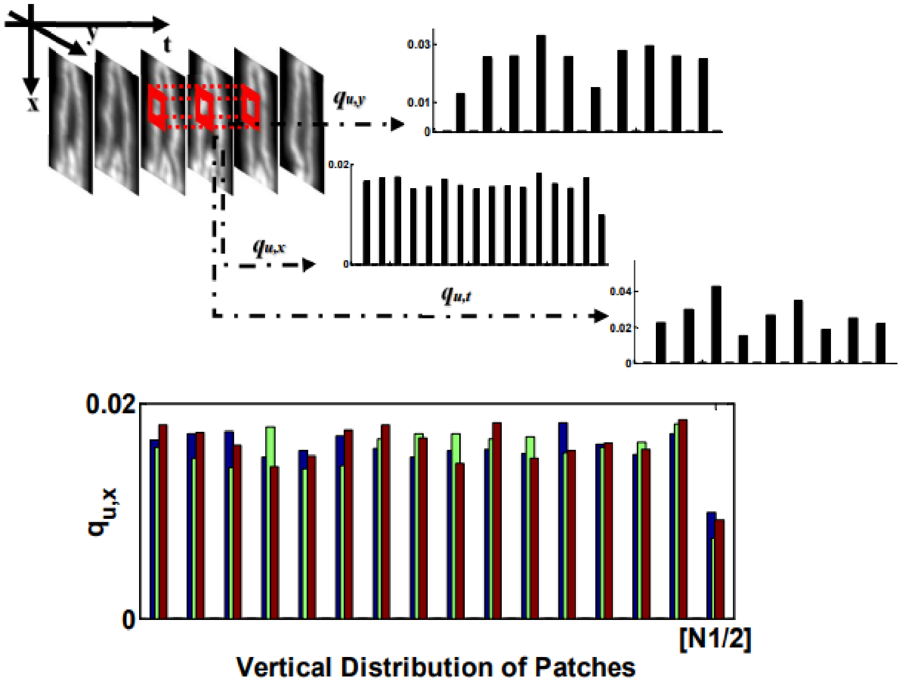
$$q_{u,x} = C_x \sum_{i=1}^{\left[\frac{N_1}{2}\right]} \delta[b(X_i) - u]$$

$$q_{u,y} = C_y \sum_{j=1}^{\left[\frac{N_2}{2}\right]} \delta\left[b(Y_j) - u\right],$$

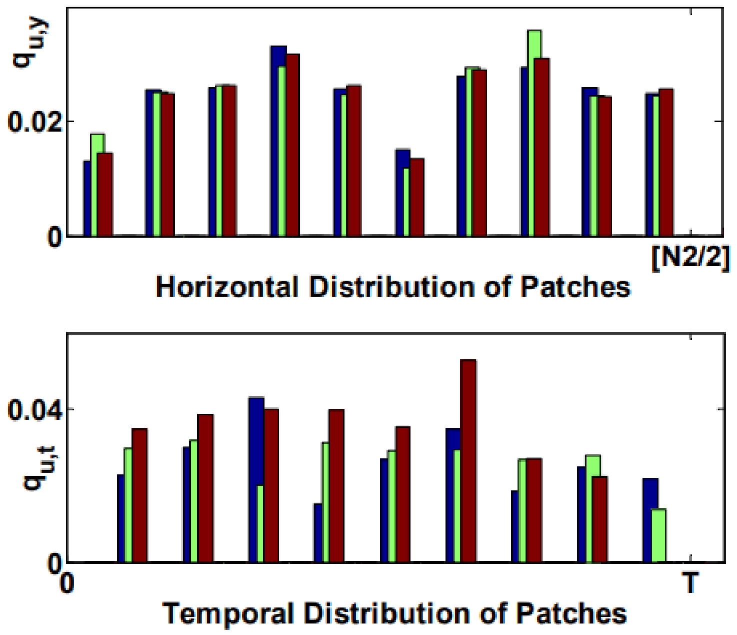
$$q_{u,t} = C_t \textstyle\sum_{k=1}^{T} \delta[b(t_k) - u] \qquad (2\text{-}19)$$

where Cx, Cy and Ct coefficients are $\sum_{u=1}^{m} q_u = 1$normalization coefficients so that by calculating the above histograms, we will have three possible distributions corresponding to the horizontal, vertical and time directions. In the next section, it has been used these histograms to form enhanced patches.

Figure 2.10(a) shows the process of processing and calculating Spatio-temporal histograms. Each moving window in D+t2 space has a local maximum whose coordinates are mapped to an index and placed in the corresponding histogram. The location of extremes in the entire D+t2 space represents the distribution of Spatio-temporal patches for each person. Figure 2.10(b, c, and d) shows sets of X, Y, and T histograms for three different people, each of which has a unique color bar.



A,B



C,D

Figure 2.10. (a) The process of calculating X-Y-T histograms based on the location of local extrema, (b), (c), and (d) the summary of vertical, horizontal and time histograms for different people expressed by blue, green, and red bars [53].

As can be seen, each graph of desires for people has different distributions and they have little similarity to each other. Therefore, they have a suitable capacity to display local patches or movement templates. In other words, the sum of these three histograms is known as the signature of the motion template [50]. For the three-dimensional visualization of the distribution of patches in the stack of images and in the D+t2 space, Figure 2.10(a) shows the position of the 20 most histogram values in Figure 2.10(b-d).

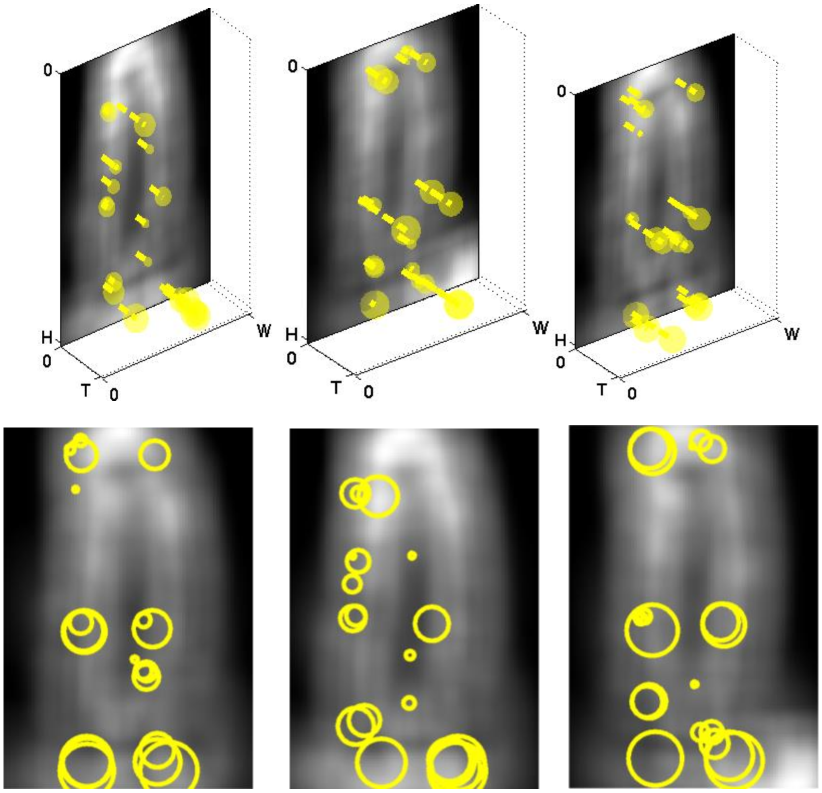


Figure 2.11. D+t2 position of top 20 X-Y-T histogram indices for three different people, (a) 3D view and (b) 2D view [53].

The higher the value of the corresponding histogram index, the larger the circle's radius. Then, the two-dimensional view of these yellow circles on the X-Y plane is shown in Figure 2.10(b). In Figure 2.11, it can be said that the larger radius is related to the higher density of local patches in the motion space. Because the higher the density of patches in a region of 2D+t space, the higher the histogram index [50]. It is clear from Figures 2.10 and 2.11 that a set of X-Y-T histograms can be used as a motion template signature to describe people's walking type. In fact, for each walking condition, there are some essential points whose probability in 2D+t space can be used to describe the type of movement. It should be noted that most of the patches are in the areas around the feet, hands and shoulders, in other words, the proposed algorithm highlights these areas for better identification. The calculated histograms are used to form the final template. The final template is an enhanced Gabor feature where the 3D coordinates are added to the Gabor filter values.

Approches	references	principle	invariance	Used Metric	Complexity if it is Available	Computer Configuration if it is available	Dataset	Scores
CGI	[14]	a colorful template in which the rhythm of walking in time domain is represented in color spectrum. It has been proved that such colorful template can represent gait in different conditions more accurately	Check if there is invariance to light, rotation, etc. If there is no invariance, please mention the used preprocessing to achieve it	Rank1 & rank2	the time complexity of generating all CGI templates for each training and test data is ðNtrTWHk þ NteTWHkÞ, w	run a Matlab code on a machine with an Intel Core2 Duo CPU T9600 2.80 GHZ and 3 GB of DDR3 memory	CASIA database (Dataset B).	Obtained scores on each dataset
PGF	[11,47, 48,50, 51,52 ,53]	local patches are extracted from a motion sequence. Then the local extremum points are extracted, and possible distributions based on the patches are calculated. Finally, the final template will be calculated as a movement feature based on the possible distribution of patches	Shoes type Viewing angle Walking surface Time + Shoes +Surface +Clothing	Rank1 & rank2	the general complexity of all the PGF templates is in the order of O(40(T + 1) (ngl + npr)WHwh)	the time of computing a PGF for an individual is 820 ms using MATLAB 8.3.0 (2014a) running on an Intel (R) Core (TM) i7 processor with 8 GB RAM working at 2.39-GHz for USF dataset	USF HumanID dataset (dataset version 2.1), CASIA Dataset (Dataset B) and OU-ISIR (Dataset B) The OU-ISIR dataset includes of 48 individuals walking on a treadmill with 32 types of different clothing The USF dataset consists of 122 individuals walking in elliptical paths in	

							front of the camera.	
GSI	[24, 97]	A pattern called gait salience image (GSI), which encodes relevant Spatio-temporal features into a single pattern. One of the great strengths of GSI is the extraction of motion-based features by applying a filter scheme to walking silhouettes. In other words, a Spatio-temporal impulse response is adapted to compute the local walking features in a video sequence. Since there are two steps with the same walking mode in each walking period, therefore, the filtering process is repeated in each step, i.e. half of the period, separately.	-grass, C-concrete, A-shoe A, B-shoe B, R-right view, L-left view, NB-no briefcase, BF-briefcase, T-time and avg. period-the average period of the individuals in given set	Rank1 & rank2	total complexity of GSI templates will be O(4(ntr + nte) NgWHwh) 1-NN classifier	The processing time is measured with MATLAB code running on a machine with Intel Core2 Duo CPU P8400 2.20 GHz and 2 GB of DDR2 memory.	USF (Sarkar et al., 2005) and CASIA (Yu et al., 2006) filtering is optimised to measure the motions captures from 90 degrees	
GEI	[14]	It only considers individual recognition with activity-specific human motion, for example, regular human walking, which is used in most current approaches to individual recognition with walking.	clothing, shoes, or environmental context	Rank1 & rank2	1-Nearest Neighbor (1-NN)	run a Matlab code on a machine with an Intel Core2 Duo CPU T9600 2.80 GHZ and 3 GB of DDR3 memory	CASIA Gait Database	

Table 2.2. Three Approches synthetic template generation pseudocode[1].

Chapter 3

Proposed Method

3.1. Proposed Method

In this section, we describe the proposed method in detail. Our method is a contribution of PGF that is trying to keep the most useful valuable patches and withdraw noisy patches, and this is done by adding two more steps to original PGF.

This is achieved by first creating an efficient template, which requires the use of robust gait templates. Based on this, templates are divided into two categories: spatial templates and Spatio-temporal templates. The difference between the two categories is that the latter captures both the spatial and temporal features of the walking template. Spatial templates only represent the spatial characteristics of walking templates, meaning they only consider the features of the person's silhouette or body shape in a given moment. They are extracted from individual frames of a video or series of images that capture the person's walking motion. On the other hand, Spatio-temporal templates represent the changes in the person's silhouette or body shape over time, taking into account the dynamics of the walking motion. They are typically extracted from a sequence of images or video frames that capture the person's walking motion.

For better gait detection, an improved Spatio-temporal template is needed, especially in clothing conditions. Due to limitations in complexity, the parameters of features should be decreased, and only efficient features should be kept. After reviewing recent Spatio-temporal templates, including CGI, GSI, GSTI, and PGF, I chose PGF for its simplicity, robustness, and low complexities [97]. However, some noisy pixels remain in the final PGF template.

The heart of PGF is the patch extraction process, and the most important part is to select the most significant features. Thus, I proposed a method to select the most important features called iPGF.

In summary, spatial templates only capture the static appearance of a person's walking template, while Spatio-temporal templates capture both the appearance and the dynamics of the walking template.

The basis of the development of current algorithms is the baseline approach to identifying the movement template. The initial idea of template-based templates was presented by Mr.Han et al. [1] under the title "Gait Energy Image" (GEI). In addition, features of the enhanced template have been developed based on the Gabor filter. For example, Mr.Tao et al. [7] presented a GTDA algorithm based on the Gabor filter in which the calculated features, like the GEI algorithm, are collected in one image. Recently, an improved Gabor algorithm based on patch statistical patch (Gabor-PDF) has been developed, solving many matching problems [5, 6].

Certainly, here's a summary of the role of Gabor filters in walking detection compared to other filters, based on the proposed Spatio-temporal walking models using regional adjacent patch descriptors:

Gabor filters are a specific type of linear filter widely used in image processing to analyze textures and edges. They excel at capturing information from various orientations within the same image, making them highly effective in gait detection compared to other filters like Fourier.

Orientation Sensitivity: Gabor filters are sensitive to specific orientations in an image, a crucial trait for gait detection. Body parts like legs and arms move in different directions during walking, and Gabor filters automatically capture these directional features, whereas Fourier transforms might struggle to do so.

Frequency and Spatial Localization: Gabor filters simultaneously handle frequency and spatial information, making them suitable for detecting local patterns in walking, such as the movement of different body parts. In contrast, Fourier transforms excel in capturing frequency but may miss gait-specific attributes.

Analysis Flexibility: Gabor filters can adapt to various detections and analyze characteristics in different ways. This is essential for gait recognition, which involves both fine-grained details and overall movement patterns. Fourier transforms lack this versatility.

Robustness to Change: Gabor filters' ability to capture texture and orientation makes them robust in handling changes due to lighting, clothing, and other environmental factors, unlike Fourier-based sources.

Multi-Directional Features: Gabor filters diversify responses in different directions, making them crucial in capturing features irrespective of orientation. This is vital for analyzing body parts moving in diverse directions during walking.

Texture and Edges: Gabor filters excel in capturing fine texture details and edges, crucial in identifying individuals based on clothing and body parts. Fourier filters primarily focus on frequency data.

Localization and Selectivity: Gabor filters can localize in both domain and frequency, pinpointing specific attributes in a spatial region while considering frequency characteristics. This is important for recording localized movement patterns of body parts.

In contrast, Fourier transforms emphasize frequency components and lack the same level of directional sensitivity and multi-orientation feature detection as Gabor filters. Gabor filters provide a comprehensive representation of gait features, capturing both directional and localized details. Their ability to analyze diverse orientations and scales makes them particularly effective for gait detection, especially when features exist in various directions. This thorough analysis distinguishes Gabor filters from Fourier-based methods, enhancing the accuracy of gait detection techniques.

As mentioned, the main weakness of template-based systems is the elimination of movement parameters. For example, the time sequence of human movement is removed in the final image. Various features have been proposed to calculate the time sequence and provide the template based on it [5, 8, 9, 10] (e.g. CGI, GSTI, etc.) to solve this problem.

Two steps are suggested in this thesis for refining local patches and throwing away noisy patches. In the first step, a local feature vector is calculated and collected from the region around each patch. Like this feature, vector has already been developed in SIFT and SURF pairs. Then, in the second step, the local patch feature vectors are clustered using the k-means technique, and the segmented information is used to calculate the histogram in the second step of Figure 3.1 Finally, the rest of the steps are calculated according to Figure 3.1, and the final template is obtained. Figure 3.1 shows the two proposed steps.

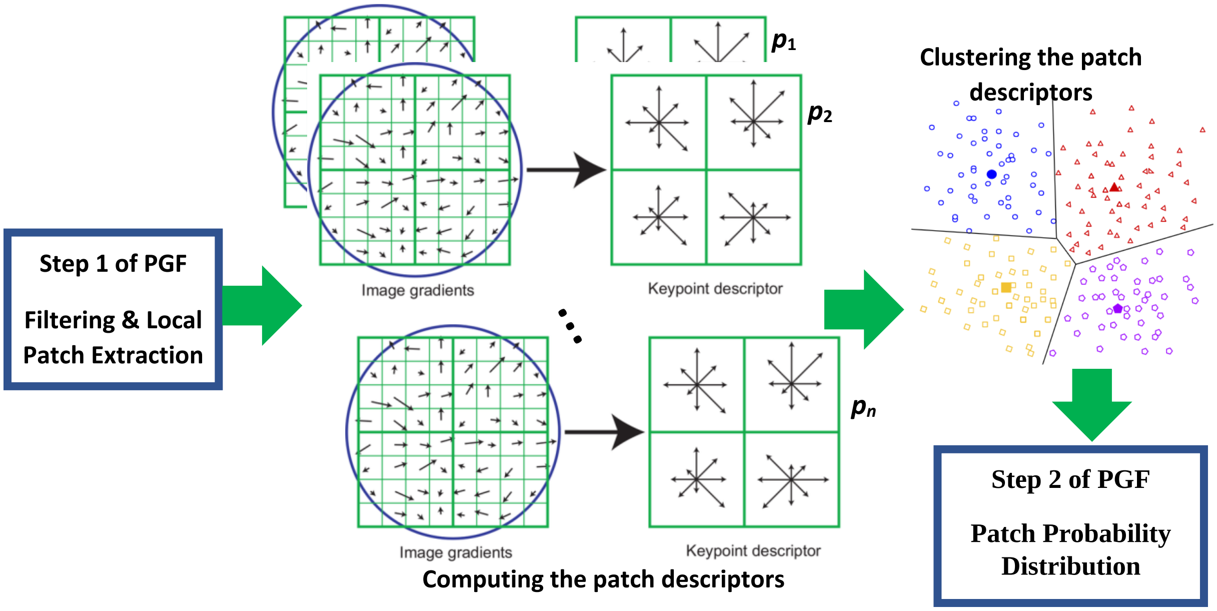

Step 1 of PGF
Filtering & Local Patch Extraction
Image gradients
Keypoint descriptor
p_1
p_2
p_n
Image gradients
Keypoint descriptor
Computing the patch descriptors
Clustering the patch descriptors
Step 2 of PGF
Patch Probability Distribution

Figure 3.1. The main steps of our iPGF method. In this step, first, a feature vector is calculated from the surrounding regions of each patch. Then, three data clustering distributions are calculated along the x, y, and z directions. The histogram of the patches is obtained in the second stage of iPGF based on the clustered data so that the center of each cluster will be a bin of the histogram.

3.1.1. iPGF Approach

The generated patches in the first stage allowed us to obtain the template distribution of each individual in space-time. But due to the patch calculation process, local noise may be generated in the feature vector, and the accuracy of the description will decrease. A two-step solution was proposed to solve this problem.

Step1. computing HOG features

In the first step, a local region around each patch is defined and weighted by a Gaussian function. In other words, it gives more weight to the areas near the patch and less to surrounding areas. Then the gradient histogram vector (Histogram of Gradient - HoG) is calculated from all these local areas and is the basis of the description of the area. In the second step, all the gradient vectors of the areas are collected and using the k-means clustering algorithm, the states are identified and grouped. In other words, the data obtained from clustering is given to the second stage of PGF and based on that, two cluster distributions are obtained in X and Y directions, and the PGF template will be calculated according to Figure 3.1.

Suppose the coordinates of the local patches are in the x and y directions, and w and h are the width and height of the input image. A local area around this patch is defined as follows:

$$Region_i \equiv R_\sigma\left(p_{i,j,k}\right) = \{(x,y) : x-\sigma \leq j < x+\sigma;\ y-\sigma \leq k < y+\sigma\} \qquad (3\text{-}1)$$

where σ is the standard deviation of the Gaussian function in the area around each patch. $Region_{\,i}$ represents a region of the image that surrounds the patch $p_{i,j,k}$.

First, we define a two-dimensional Gaussian function as follows:

$$G_{jk}(x,y) = \frac{1}{\sqrt{2\pi\sigma^2}} exp(-\frac{\left(x-p_i(x)\right)^2}{2\sigma^2} - \frac{\left(y-p_i(y)\right)^2}{2\sigma^2}) \qquad (3\text{-}2)$$

where pi(x) and pi(y) are the coordinates of the patch in the x and y directions, respectively. In this way, the local area is smoothed by this Gaussian function:

$$Region_{i\_norm} = Region_i \times G_{jk} \qquad (3\text{-}3)$$

To calculate the features in this area, we use the gradient histogram (HoG) vector. For this purpose, first, the magnitude of the gradient, m (x, y), and its angle, θ (x, y), are calculated in the above area:

$$m(x.y) = \sqrt{\left(R(x+1.y) - R(x-1.y)\right)^2 + \left(R(x.y+1) - R(x.y-1)\right)^2}$$

$$\theta(x.y) = atan2\left(R(x.y+1) - R(x.y-1). R(x+1.y) - R(x-1.y)\right)$$

(3-4)

where R is the pixels of the smooth region (Relation (3-4)). After calculating the magnitude of the gradient and its angle, the histogram of the gradient is calculated similar to equation (3-4):

$$g_{u.x} = C_h \sum_{\theta=0}^{180} \delta\left[b\left(m_{x.y}\right) - u_\theta\right] \quad (3\text{-}5)$$

where g_(u.x).u=1...180 is the histogram index (bin) value corresponding to the x,y area. This histogram is used to describe each area. Then all these histograms are placed next to each other to form the final description vector, GX,Y. This final vector represents all the areas in the motion space, and with its help we can model the distribution of patches.

Step 2. Clustering the features

In the second step of the proposed approach, this feature vector is clustered to obtain the weights. For this purpose, we use the k-means technique to group similar features (related to the same areas). The number of clusters, k, in this method will be equal to $\left[\frac{N_1}{2}\right] \times \left[\frac{N_2}{2}\right]$ (the number of image pixels). More precisely, the number of generated data related to each pixel will be categorized using k-means. If we call the label of each pixel, xi, IDX, then by a multi-step process and updating over several time steps, (t), the labels, k, will be obtained as follows:

$$IDX^{(t)} = \left\{g_u : \left\|g_u - x_i^{(t)}\right\|^2 \leq \left\|g_u - x_j^{(t)}\right\|^2 .\ \forall j. 1 \leq j \leq k\right\} \quad (3\text{-}6)$$

Finally, we get help from these labels to produce weights. The weight of the pixels in this step will be obtained simply by counting the number of labels generated for each pixel. After calculating the weights, the rest of the steps will be calculated according to the PGF approach. These weights will replace qu, x, and qu, y in equations (3-12).

In the second step of the proposed approach, K-means clustering is employed to group similar features, which correspond to specific areas in the patch distribution. The primary objective of K-means clustering is to identify coherent patterns among the gradient vectors derived from local regions around patches. While the total number of clusters (k) is indeed equal to $\left[\frac{N_1}{2}\right] \times \left[\frac{N_2}{2}\right]$, where N_1 and N_2 represent the dimensions of the image, not every pixel becomes a cluster. Instead, the clustering process involves categorizing the gradient vectors obtained from specific local areas around patches. Each gradient vector corresponds to a specific pixel within that area. The clustering groups together similar gradient vectors, implying that similar local patterns in the input image result in shared cluster assignments. The purpose of clustering is to establish a meaningful grouping of gradient vectors based on their similarity, allowing for the identification of coherent motion patterns in different areas of the image. K-means clustering is performed on the gradient vectors extracted from local regions around patches, and the clusters represent similar patterns of gradient changes. This technique enables the identification of distinctive motion characteristics in the gait patterns, contributing to the accuracy of the subsequent steps in the iPGF approach. In the context of the proposed iPGF approach, the goal of clustering using the K-means algorithm is not to create a cluster for each individual pixel in the image. Instead, the clustering is performed on the gradient vectors extracted from specific local areas around patches. Here's how the clustering process works: Local Area Selection: The image is divided into a grid of patches, and each patch represents a localized region of pixels. For each patch, a local region around it is defined, typically using a Gaussian weighting function to emphasize the central area and attenuate the influence of pixels farther from the center. Gradient Computation: Within each local area, the gradient vectors (both magnitude and angle) are calculated based on the intensity changes in the pixels. These gradient vectors represent how the intensity values change across neighboring pixels within that area. Feature Vector Creation: Each local area's gradient vectors are combined to create a feature vector that describes the gradient information within that area. This feature vector captures the local texture and edge information. Applying K-means: The K-means algorithm is then applied to these feature vectors, not individual pixels. The goal of K-means is to group similar feature vectors together in a way that minimizes the variation within each group and maximizes the difference between groups. Cluster Centers: K-means identifies cluster centers, which are representative points within the feature space. These cluster centers are determined by iteratively updating them to minimize the sum of squared distances between the feature vectors in the same cluster and the cluster center. Assigning Pixels: Once the K-means algorithm has converged and the cluster centers are determined, each local area's feature vector is assigned to the nearest cluster center. This assignment indicates which cluster the local area's gradient information is most similar to. In summary, the clustering process involves grouping together similar gradient feature vectors that describe specific local areas around patches, not individual pixels. This allows the method to capture coherent motion patterns and texture variations within these localized regions of the image. The outcome of the clustering is a set of cluster centers that represent distinct patterns of gradient changes. These cluster centers are then used to compute weights that contribute to the subsequent steps of the iPGF approach, enhancing the accuracy of gait detection and analysis.In the next part, we will explain how to classify people, and then classify the input templates with the help of it.

3.2.Gait Classification

The iPGF format, introduced in the previous part, is calculated for each sequence in a data set. In this section, I have reviewed various classification methods for my project, such as 1NN and PCA LDA. The reason for choosing these two methods is their good performance in gait recognition. However, using only these two methods does not increase classification accuracy in noisy conditions. For this reason, we have examined other methods such as Ensemble Learning, in which thousands of Week Classifiers need to be created. And the basis of this method is 2D PCA and 2D LDA. In the next step, the output of these thousand Week Classifiers is Mager T voting. Then from these 1000 observations, 600 are assigned to class one and 400 to class two. I choose class one and one of the most effective and result-oriented methods mentioned in an article explaining PCA and LDA methods. According to this article, 1D PCA and 1D LDA methods vectorize the image well. This method is not desirable for us because it messes up the image structure. For this reason, I went to the two-dimensional image and chose a Rubos classifier. Like in Spatio-temporal templates, linear transformation of two-dimensional images is necessary for suitable classification. For example, 2 D PCA or 2D LDA ENSEMBLE, but using them does not increase classification accuracy in noisy conditions. For this reason, I went to methods in which we can create 1000 learnings. And put it on 2D PCA and 2D LDA base [97].

For classifying generated templates, Principal Component Analysis (PCA) or Linear Discriminant Analysis (LDA), are two well-known methods for reducing dimensions and template recognition [2, 3]. The main idea of PCA or LDA is to calculate a set of correlated variables in a low-dimension space. However, in conventional PCA or LDA, two-dimensional (2D) walking templates are transformed into one-dimensional feature vectors. This projection removes the image's two-dimensional (2D) structure and converts it into one-dimensional (1D) structures. To deal with this problem, several tensor-based classification tools have been developed to represent input feature vector spaces in the last decade. For example, in classifications based on tensor, we can say the Random Subspace Method (RSM) [15,35], which overcomes covariance variates and impressively improve diagnosis. This is to create a weak classifier with a random sampling of tensor-based feature vectors for better decisions. Then the final decision is made by majority voting of weak classifiers.

RSM classification includes three main steps [3]: spatial random sampling, generating weak classifications, and final voting. These steps are shown graphically in Figure 3.2, which are discussed in more detail in the following subsection.

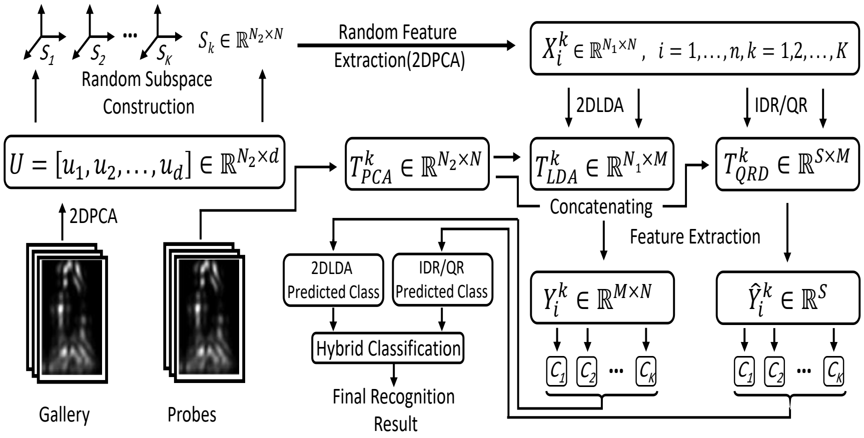

S_1
S_2
...
S_K
$S_k \in \mathbb{R}^{N_2 \times N}$
Random Subspace Construction
Random Feature Extraction(2DPCA)
$X_i^k \in \mathbb{R}^{N_1 \times N}$, $i = 1,\dots,n, k = 1,2,\dots,K$
2DLDA
IDR/QR
$U = [u_1, u_2, \dots, u_d] \in \mathbb{R}^{N_2 \times d}$
$T_{PCA}^k \in \mathbb{R}^{N_2 \times N}$
$T_{LDA}^k \in \mathbb{R}^{N_1 \times M}$
$T_{QRD}^k \in \mathbb{R}^{S \times M}$
Concatenating
Feature Extraction
2DPCA
2DLDA Predicted Class
IDR/QR Predicted Class
Hybrid Classification
Final Recognition Result
$Y_i^k \in \mathbb{R}^{M \times N}$
$\hat{Y}_i^k \in \mathbb{R}^S$
C_1 C_2 ... C_K
C_1 C_2 ... C_K
Gallery
Probes

Figure 3.2. The general structure of RSM classification.

3.2.1. Random Subspace Sampling

Assume there are n gait templates A_i (i=1…n) (or GSTI) in the training set (gallery) with a dimension of $N_1 \times N_2$ pixels. In RSM, we compute 2DPCA projection matrix based on the 2D scatter matrix:

$$S = \frac{1}{n}\sum_{i=1}^{n}(A_i - M)^T \times (A_i - M) \quad (3\text{-}7)$$

where $M = \frac{1}{n}\sum_{i=1}^{n} A_i$ is the global mean of the samples from all classes of the training set. Afterwards, the eigenvectors of the scatter matrix S are computed leading to d eigenvectors with non-zero eigenvalues $U = [u_1, u_2, \dots, u_d] \in \Re^{N_2 \times d}$.,. The K random subspace $\left\{T_{PCA}^{i} \in \Re^{N_2 \times N}\right\}_{i=1}^{K}$ can be computed by random selection of N ($N \leq d$) unique eigenvectors from subsets U and repeating the process K-times. As a result, the random feature sets will be generated in lower dimension space as follows [16]:

$$X_i^k = A_i T_{PCA}^k, \qquad i = 1, \dots, n, k = 1,2, \dots, K. \quad (3\text{-}8)$$

It can be proved that random sampling of eigenvectors can preserve the covariate factors in lower dimension feature space efficiently [3]. However, some redundant information remains in the feature vector X_i^K that may affect the performance of the final decision. To improve the recognition rate, another classification step will be applied in RSM.

3.2.2. Dimensionality Enhancing

The random features in Equation (3-4) have still redundant information that may affect the quality of the decision. To obtain more discriminant features for weak classifiers, an additional dimensionality reduction method should be performed. Here, two known techniques, i.e. 2D Linear Discriminant Analysis (2DLDA) [22] and Incremental Dimension Reduction algorithm via QR decomposition (IDR/QR) [22] can be used alternatively [3]. In this section, both the 2DLDA and IDR/QR methods are being reviewed. The features for final decision are then extracted from each method separately.

In 2DLDA, the class labels are considered to get between-class scatter matrix S_B^k and within-class scatter matrix S_W^k :

$$S_B^k = \sum_{i=1}^{c} n_i(m_i^k - M^k) \times (m_i^k - M^k)^T \tag{3-9}$$

$$S_W^k = \sum_{i=1}^{c} \sum_{X_j^k \in D_i^k} (X_j^k - m_i^k) \times (X_j^k - m_i^k)^T \tag{3-10}$$

where M^k is the global mean of all samples in k^{th} subspace and D_i^k is the index of i^{th} class with sample number n_i and mean of m_i^k . By computing the eigenvectors of $\left(S_W^k\right)^{-1} S_B^k$ in each subspace and selecting M leading eigenvectors, we obtain K transition matrix $T_{LDA}^k = \{\varphi_i\}_{i=1}^M$ (k=1…K) where each one has M random-selected eigenvectors.

An alternative solution to the 2DLDA approach is the IDR/QR technique which applies QR decomposition to maximize the separability of between-class features [3]. Unlike the 2DLDA, the 1D vectors are processed rather than 2D matrices. Therefore, the extracted random features $\left\{X_i^k \in \Re^{N_2 \times N}\right\}$ should be vectorized before training IDR/QR model. By setting $N_v = N_2 N$, the vectorized random features can be represented as$\left\{\hat{X}_i^k \in \Re^{N_v}\right\}$. Now for each subspace, the set of within-class centroids $C = \left[\widehat{m}_1^k, \widehat{m}_2^k, \dots, \widehat{m}_c^k\right]$ is first computed and QR decomposition will be performed on C, as $C = QR$, and $Q \in \Re^{M \times c}$ [3]. With setting,$e_j = (1,1, \dots ,1)^T \in \Re^{n_j}$ two predefined matrices H_W^k and H_B^k will be derived as follow:

$$H_W^k = \left[\widehat{D}_1^k - \widehat{m}_1^k e_1^T, \widehat{D}_2^k - \widehat{m}_2^k e_2^T, \dots, \widehat{D}_c^k - \widehat{m}_c^k e_c^T\right] \tag{3-10}$$

$$H_B^k = \left[\sqrt{n_1}\left(\widehat{m}_1^k - \widehat{M}^k\right), \sqrt{n_2}\left(\widehat{m}_2^k - \widehat{M}^k\right), \dots, \sqrt{n_c}\left(\widehat{m}_c^k - \widehat{M}^k\right)\right] \tag{3-11}$$

Similar to 2DLDA, $\hat{M}^{k}$ is the global centroid (mean of all samples) of k^{th} subspace and $\hat{D}_i^k$ is index of the i^{th} class with centroid $\hat{m}_i^k$ and sample number n_i. Here, within-class and between-class scatter matrices can be computed as:

$$S_W^k = \left(\left(H_W^k\right)^T Q\right)^T \left(\left(H_W^k\right)^T Q\right) \tag{3-12}$$

$$S_B^k = \left(\left(H_B^k\right)^T Q\right)^T \left(\left(H_B^k\right)^T Q\right) \tag{3-13}$$

For each subspace, the given transformation matrix T_{QRD}^k is computed from eigenvectors of $(S_W^k)^{-1} S_B^k$and selecting the M leading eigenvectors$U^k = \{\phi_i\}_{i=1}^M$::

$$T_{QRD}^k = U^k Q \tag{3-13}$$

Therefore, K IDR/QR-based transformation matrix $\left\{T_{QRD}^k \in \Re^{S\times M}\right\}$ will be used to extract more discriminant features.

Now considering two mentioned techniques, three sets of transformation matrices are achieved in the training phase as: 2DPCA matrix, $\left\{T_{PCA}^i \in \Re^{N_2 \times N}\right\}_{i=1}^K$, , 2DLDA ($T_{LDA}^k$) and IDR/QR ($T_{QRD}^k$). For each subspace, dimensionality of our gait templates (GSTI) can be reduced by applying the 2DPCA, 2DLDA, and IDR/QR projection matrices written here:

$$X_i^k = A_i T_{PCA}^k, \qquad i = 1, \dots, n, k = 1,2, \dots, K, \tag{3-14}$$

$$Y_i^k = (T_{LDA}^k)^T X_i^k, \qquad i = 1, \dots, n, k = 1,2, \dots, K, \tag{3-12}$$

$$\hat{Y}_i^k = (T_{QRD}^k)^T \hat{X}_i^k, \qquad i = 1, \dots, n, k = 1,2, \dots, K \tag{3-13}$$

where X_i^k, Y_i^k, and $\hat{Y}_i^k$ are 2DPCA random, 2DLDA enhanced, and vectorized IDR/QR features sets, respectively. Once the mentioned projection matrices are computed, two different methods known as 2DPCA+2DLDA (or 2DLDA) and 2DPCA+IDR/QR (or IDR/QR) is being used for *K* random feature extraction. The hybrid decision level is then achieved based on the outputs of random classifiers which are discussed in the following subsection.

3.2.3. Final Classification

The basic idea behind the ensemble methods is to find a robust classifier based on the performance of the weak classifiers. Here, each feature set in *k*th subspace can make weak decisions according to the covariate factors. The final decision will be taken based on the sub-decisions in each subspace. Suppose there are *c* classes in the training set (gallery) and each has n_i (i=1,...,c) samples. For the *k*th subspace, let $m_i^k(i = 1, \dots, c)$ be the mean of the samples in each class and R^k be the feature samples of the probe set (including n_p gait samples) [35]. The Euclidean distance between R^k and the mean of *i*th class of the gallery m_i^k can be expressed as:

$$d\left(R^k m, m_i^k\right) = \frac{1}{n_p}\sum_{j=1}^{n_p}\left\|R_j^k - m_i^k\right\|, \quad i = 1, \dots, c\ . \tag{3-14}$$

Now, the minimum distance of a given probe template to each class, $\{\omega_i\}_{i=1}^{c}$ in the gallery set is considered as weak decision:

$$\Omega^k(R^k) = \arg\min_{\omega_i} d\left(R^k, m_i^k\right), \quad i = 1, \dots, c. \tag{3-15}$$

Here, we have two sets of weak classifiers based on two feature sets (2DLDA and IDR/QR). A Hybrid Decision-level Fusion (HDF) among K subspace in each set of weak classifiers can be achieved simply by majority voting of all K classifiers [3]. More precisely, for a probe gait query $R = \{R^k\}_{k=1}^{K}$, the mode of K labels in all subspaces is considered the final decision. The correctness of this consideration can be represented as a binary function:

$$\theta_{\omega_i}^{k} = \begin{cases} 1, \ if & \Omega^k(R^k) = \omega_i \\ 0, & Otherwise \end{cases} , i \in [1, c] \tag{3-16}$$

and final classification by majority voting can be expressed as:

$$\Omega(R) = \underset{\omega_i}{argmax} \sum_{k=1}^{K} \theta_{\omega_i}^{k}, \quad i \in [1, c] \tag{3-17}$$

where $\Omega(R)$ is the final class assigned to the given probe templates.

To further improve the performance of the decision, a hybrid strategy has been applied by fusion of the results from two different classifiers [3]. Let $\Omega_{LDA}(R)$ and $\Omega_{QRD}(R)$ be the final decision corresponding to the 2DPCA+2DLDA and 2DPCA+IDR/QR -based features for a query gait R. The hybrid classifier (HC) can be performed by [3]:

$$\Omega_{HC}(R) = \begin{cases} \omega_{i,} & if \quad \Omega_{\mathrm{LDA}}(R) = \\ \omega_{i,} & if \quad \Omega_{\mathrm{QDR}}(R) = \omega_i \ , i \in [c] \\ 0, & otherwise \end{cases} \tag{3-18}$$

It is inferred from Equation (3-18) that HC decision is guaranteed if one of the corresponding classifiers recognizes given individual correctly.

Chapter 4

Results

4.1. Implementation and Result

In this section, we will describe the results related to the Gait classification with the help of the proposed approach. For this purpose, we will just use the USF [2] database. In comparison with other well-known dataset such as CASIA [25], the USF provide more competitive conditions while the quality of silhouettes is noisier and guiltier [53].

The selected algorithms for evaluation and comparison are Baseline algorithm [2] LGSR [11], GEI+RSM [35], Gabor+RSM [35], VI-MGR [13], LPSELA [49], GSTI [16] and PGF [54] will be. Then, in the final part, the proposed approach to PGF will be evaluated from the perspective of computational complexity and memory issues.

The process for pi one involves using 100 templates in the probe and 122 in the gallery. The algorithm receives each person one by one and produces a classification output based on similarity or distance, which is denoted as D. This process is repeated for all 122 people in the gallery. After calculating the distance, we sort them and compare the labels to find the best match. We consider P1 to be the same as P1 and repeat this process for the rest of the labels.

The distances are sorted from lowest to highest, and the lowest distance for pi one indicates that the classifier has recognized pi one. We then match the labels to determine the most similar person. This process is repeated for the entire probe set, and we check which person has the highest matching rank, which is known as rank one.

For rank five, we sort the distances and check whether matching has occurred in the top five distances. We determine whether the minimum distance falls within the specified interval and whether matching has occurred.

Overall, the process involves sorting the distances and matching labels to determine the most similar person. We repeat this process for the entire probe set to determine the rank of the matching person.

The implementations have been done in Google Collab and Python, and the features have been calculated locally in the laptop due to the limited space of Google Collab.

Tools:

1-simulation software: Python

2-Hardware:

-Core i7

-Processor11thGenIntel(R)Core(TM)i7-11800H@2.30GHz,2304Mhz,8 core(s)

-installed physical memory (RAM) 32.0GB

3-This project will use USF [3] and CASIA (SET B) datasets [7].

4.2. Gait Dataset

The USF database consists of 122 subjects moving in an elliptical path in front of the camera [2, 16, 24]. Briefly, walking conditions are movement level (S), shoe type (H), viewing angle (V), bag carrying requirements (C) and, elapsed time (T). Considering these five challenging factors, in this database, the motion sequence with the condition "grass surface, shoe type A, shooting from the right side of the camera, no bag, and recorded at time t1 (May)" is selected for the gallery collection. are Then, 12 different and distinct tests have been selected for the probe set.

Table 4.1. The specification of probe sets in USF gait dataset [2, 15].

Experiment	A	B	C	D	E	F	G
Covariates	V	H	VH	S	SH	SV	SHV
Num. of People	122	54	54	121	60	121	60
Variance	Shoes/View			Surface+Shoes/View			

Experiment	H	I	J	K	L
Covariates	B	BH	BV	THC	STHC
Num. of People	120	60	120	33	33
Variance	Briefcase + Shoes/View			Time+ Shoes +Surface+Clothing	

V-View, H-Shoe, S-Surface, C-Carriage, T-Time, C-Clothing

The conditions in the gallery and each probe set are unique, and there is no commonality between motion conditions in the probe set. All tests in the database can be divided into four distinct groups. The difference between each group to the gallery set, the difference in groups, and the list of probe sets belonging to each group are also shown in table 4.1. In this database, the sequence of normalized silhouettes (Sequence of normalized silhouettes) is also presented.

4.2.1. Algorithm Benchmark

Every biometric system operates in two working modes. The first mode is responsible for the identity measurement and the second mode is responsible for the authenticity measurement of people. Based on this, two criteria, "cumulative match characteristics" (CMC) and "receiver operating characteristics" (ROC) have been developed to evaluate these two modes [2, 5]. In the CMC criterion, the behavior template of each person from the probe set is measured with all the templates of people in the gallery set and a similarity score is given to the people. This score indicates the similarity of a person from the probe to all the people in the gallery. Then, the goal of CMC is to measure the answer to the question, "Is the person in question in the probe among the k people with the highest score from the gallery?" If the desired person from the probe was among the top k people from the gallery, then we will have a correct identification rate in the kth row. In general, suppose that in a probe set, P is the total number of individuals to be scored and, Rk is the number of individuals among the top k matches. In this case, the identification rate is equal to:

$$\mathrm{CMC}(k) = \frac{R_k}{P} \tag{4-1}$$

By calculating this value for different values of k, the CMC curve is obtained [2]. In articles, CMC results for two values of k=1 and k=5 are usually expressed as a measure of biometric system accuracy. Therefore, the results are called "Correct Classification Rate" (CCR) Rank 1 (Rank1) and Rank 5 (Rank5) in the articles [11, 15]. This model of identification is called "set-package" and it means that the correct answer is always present in the gallery set [2, 7]. In other words, the person tested in the probe set is one of the selected people in the gallery set, whose movement conditions (such as shoes, ground surface, etc.) are different. For example, in the USF behavioral data set, the number of people in the gallery set is 122 and the number of people selected in one of the probes (for example, B or C) is 54 people. With this assumption, in a biometric system, the Rank1 value equal to 95% means that 95% of the people in that probe have been correctly placed and identified in the first row of scoring. Also, the value of Rank5 is equal to 98%, which means that 98% of the people in the probe have been correctly identified except for the first five people. Finally, in some articles, the CMC curve is used to express accuracy [63]. The horizontal axis in this curve is different values of k and the vertical axis is the CMC values of relation (4-2).

Now, in the USF dataset, because the number of people in each probe set is different, the weighted average of the identification rate of Rank1 and Rank5 is also expressed as a quantitative measure. The value of the weighted average identification rate (W-AvgI) can be calculated according to the following equation:

$$W - AvgI = \frac{\sum_{i=1}^{g} w_i R_k}{\sum_{i=1}^{g} w_i} \qquad (4\text{-}2)$$

Where, wi is the number of people in each probe set and g is the number of tests (g=12). In this section, we will use the value of W-AvgI to calculate the average accuracy of Rank1 and Rank5.

4.3. USF dataset Result

In 2005, extensive research was conducted on the effects of five external factors on behavioral patterns. These factors include two shooting angles (L and R), two types of shoes, two types of surfaces (concrete and grass), type of carrying object (without a bag and with a bag) and shooting time (May and November). The results were compiled into the USF standard data, which consists of 1870 motion videos of 122 individuals.

The set of experiments was defined by combining these five factors in two change modes, resulting in 32 different test conditions. The purpose of this research was to provide challenging conditions for remote behavioral recognition in surveillance applications. Researchers aimed to develop algorithms that could handle various environmental challenges such as variable background silhouettes and different lighting conditions.

The USF data encompass changes in surface type, shoe type, and carrying object, as these factors were hypothesized to influence both walking behavior and selection characteristics of individuals. Researchers from CMU, MIT, Sampton Land MIT, and Georgia Tech agreed that these five variable factors present significant challenges in the field.

The test conditions were designed to be neither too easy nor too difficult, allowing the evaluation of behavioral recognition algorithms effectively. Among the 1870 videos, 12 different tests were defined, consisting of 7 probes (input data) and one fixed test as a gallery (care list).

Overall, the USF data with its 32 modes and 1870 motion videos provided a suitable and challenging dataset for testing and evaluating algorithms in the realm of behavioral recognition.

The results of Rank1 and Rank5 are presented in tables 4.2 and 4.3, respectively. It should be mentioned that the RSM algorithm has been used to classify of features for an accurate and fair evaluation. Also, the results of recent methods have generally used the same or similar classification algorithm.

Table 4.2. Comparisons of Rank1 CCR (%) of the approaches on USF dataset.

Exp.	A	B	C	D	E	F	G	H	I	J	K	L	W–AvgI
Method	Rank5 Performance												
LGSR	95	93	89	51	50	29	36	85	83	68	18	24	70.07
GEI+RSM	98	95	88	54	60	37	44	90	93	83	33	21	70.16
Gabor+RSM	100	95	94	73	73	55	64	97	99	94	41	42	81.15
VI-MGR	95	96	86	54	57	34	36	91	90	78	31	28	68.13
LPSELA	95	91	78	66	59	46	52	93	88	69	30	27	70.49
GSTI	97	95	93	53	49	41	46	96	97	92	33	21	72.25
PGF	100	96	98	62	59	43	46	100	99	94	28	30	76.01
iPGF (ours)	100	96	94	65	61	46	46	100	99	94	33	30	76.84

By evaluating the performance of the proposed algorithms in table 4.2, the following results are obtained:

1) Average Rank1 in the proposed algorithm is very close to the average of the PGF algorithm and comparable with other approaches. The reason is a slight decrease in accuracy in some probes.

2) The proposed system had the best results in 6 out of 12 experiments (probes D, E, F, G, K and L), compared to the PGF

approach. Also, in the remaining six tests (probes A, D, G, H, I, L), the results are very close to each other and are very close to the top value of Rank1. Therefore, in general, we have improved in most of the results.

3) In the conditions of change of level and time (probes D, E, F, G, K and L) where the results of the algorithms are relatively low, the results of the proposed algorithm have been somewhat improved due to the use of local area information. In these probes, there are noisy conditions, and the quality of silhouettes is deficient.

4) The results of the proposed patch-based algorithm (PGF and iPGF) are relatively superior to the filter-based approaches (GSTI), and the reason for this is the removal of extra information from the images and the extraction of local features.

In general, by checking the accuracy of Rank1 identification, it can be concluded that by combining the information of the areas around the patch and using the gradient vectors of the regions, the quality of the extracted features will be improved, and its performance will be better. We evaluate the Rank5 results in table 4.3 for a more detailed review.

Table 4.3. Comparisons of Rank5 CCR (%) of the approaches on USF dataset.

Exp.	A	B	C	D	E	F	G	H	I	J	K	L	W–AvgI
Method	Rank5 Performance												
LGSR	99	94	96	89	91	64	64	99	98	92	39	45	85.31
GEI+RSM	99	99	97	71	68	49	56	98	97	91	40	38	79.01
Gabor+RSM	100	98	98	85	84	73	79	98	99	98	55	58	88.59
VI-MGR	100	98	96	80	79	66	65	97	95	89	50	48	83.75
LPSELA	100	96	93	84	83	73	74	95	96	89	64	52	86.09
GSTI	100	96	97	78	76	72	74	99	99	99	42	36	85.64
PGF	100	98	98	80	77	77	60	100	100	99	48	45	86.59
iPGF(ours)	100	98	95	80	78	83	60	100	100	98	46	48	87.14

By evaluating the performance of the proposed algorithms in table 4.3, the following results are obtained:

1) The average Rank1 in the proposed algorithm is very close to the average of the PGF algorithm and comparable to other approaches. The reason is a slight decrease in accuracy in some probes.

2) The proposed system had the best results in 6 out of 12 experiments (probes D, E, F, G, K and L), compared to the PGF approach. Also, in the remaining six experiments (probes A, D, G, I, L), results are very close to each other and are very close to the top value of Rank1. Therefore, in general, we have improved in most of the results.

3) In the conditions of change of level and time (probes D, E, F, G, K and L) where the results of the algorithms are relatively low, the results of the proposed algorithm have been relatively improved due to the use of local area information. In these probes, the conditions are noisy, and the quality of silhouettes is very low.

4) The results of the proposed patch-based algorithm (PGF and iPGF) are relatively superior to the filter-based approaches (GSTI) and the reason for this is the removal of additional information from the images and the extraction of local features.

From Table 4.3, the proposed gait identification system has much better results and better performance compared to recent approaches (especially PGF). In other words, the average rank of 5 in iPGF process is improved overall methods (except Gabor+RSM). In the GEI+RSM and Gabor+RSM algorithm, a more significant number of random classifiers (parameter K=1000 in RSM from Chapter4 are used to improve performance, which increases the computational load and memory of the algorithm. However, a smaller number of categories has been used in the proposed approach due to maintaining the computational load and increasing the calculation speed [15].

According to the proposed methods in tables 4.3 and 4.4, it is vulnerable to the change of surface and time, but this performance is still comparable with other methods. In addition, the detection rate ranked 5 in Table 4.4 for the proposed methods has improved performance over PGF in some tests. More precisely, in 3 out of 12 tests (probes E, F and L) the accuracy of the proposed algorithms has improved and in the rest of the cases (except for probe C and K) it has been equal to PGF. In these tests, the Rank 5 values of the iPGF method are close to the highest values in the table.

The results obtained in Tables 4.2 and 4.3 state that the patch-based augmented feature can perform better than the conventional patch-based methods, namely LGSR [11] and LPSELA [49], and PGF [53]. In addition, the interesting point in table 5.4 is that in 3 tests (probes A, H, and I) the accuracy of Rank 5 was 100% and in fact, the patch-based approach was able to recognize all the people correctly.

Also, the obtained results confirm that the use of advanced classification can create a motion template detection system with an average detection rate of about 1% compared to methods such as LGSR [11], GEI+RSM [35], VI-MGR [13], LPSELA [49] and PGF [53]).

4.4. Complexities

In the final part, the proposed system will be evaluated from the point of view of processing time and memory consumption. These two criteria together, with the identification accuracy will show a system's efficiency. The obtained results will be compared with similar designs to evaluate the improvement rate.

In evaluating of iPGF, its computational complexity is similar to PGF, with the difference that in the second part, the weight coefficients of the patches are extracted from local information. In the iPGF algorithm, if we assume the computational volume of the triple step is c1, c2, and c3, the total computational volume will equal ctot = c1 + c2

+ c3. But in the first part, the number T (T movement period) of the 40 Gabor filter response is calculated. If we denote the time complexity of each filter by the symbol O(Idn-filt), then the time complexity of the algorithm for one movement period is equal to O(40TIdn-filt(nte+ntr)) (ntr and nte data training and testing). Of course, it goes without saying that the computational burden of O(Idn-filt) is equal to 0.25 of the input image filter O(Ifilt); Because, to maintain the computational load, the number of input image samples has been reduced by half along the length and width [16]. But the time spent to calculate a filter with dimensions w*h in an image with dimensions W*H is equal to O(Ifilt) ~ O(WHwh) [25]. Therefore, the time complexity of the first stage of iPGF in a data set will be equal to c1 ~ O(10(nte+ntr)TWHwh) (W,H image dimensions and w,h Gabor filter dimensions). Also, in the third step of iPGF, due to the fact that a Gabor filter is applied again to the average images, the computational complexity will be approximately equal to c3 ~ O(10Ifilt) [24]. But in the second step, the main calculations are related to the k-means clustering calculation. Because the steps of calculating the gradient histogram of the areas will not be very time-consuming. The time to figure k-means for k clusters, and the number of n data and I iterations (for convergence) is approximately equal to c2 ~ O(I*k*n) [24]. The total calculation time of the iPGF algorithm for a dataset is calculated according to the following equation:

$$O(iPGF) \approx c1 + c2 + c3 \approx$$

$$O(10(nte + ntr)TWHwh) + O\big(I \times k \times n(nte + ntr)\big) + O(10(nte + ntr)WHwh) \qquad (4\text{-}3)$$

By simplifying the formulas, the calculation time of the proposed algorithm is obtained according to the following formula.

$$O(aPGF) \approx O((\mathrm{nte} + \mathrm{ntr})\{(10TWHwh) + (I \times k \times n)\} \qquad (4\text{-}4)$$

In the USF data set in this research, the value of I×k×n is ≈0.1*TWHwh, and therefore the computational burden of the proposed algorithm is equal to O(iPGF)≈O(10.1(nte+ntr)TWHwh). Compared to the PGF algorithm, the computational load will increase by about 10%. Also, the complexity of CGI, with the number of k channels (that is, 3) will be of the order of O(k(ntr+nte)TWH) [14]. Therefore, the calculation time of iPGF compared to CGI will be approximately O(10.1wh/3)≈O(10.1wh/k). Finally, the calculation time of Gabor+RSM [35] will be equal to O(40T(nte+ntr)WHwh). Therefore, the proposed systems will perform well in describing the movement template while maintaining the computational load. Implementing the algorithm in standard Google Colab account shows that the calculation time of iPGF will be 4.5 frames per second.

Also, the memory required to calculate k-means is O ((I + k)*n). This amount of memory will be almost four times more in the settings of the proposed algorithm. With this assumption and considering the PGF [24] algorithm, the total memory required in the entire database will be O(50(T+1) WH(nte+ntr)). More precisely, if we assume that the input image dimensions of the USF dataset are 88x128, the filter dimensions are 39x39, the average periodicity of the entire base is T ≈ 32, and ntr + nte = 1080, the total amount of memory required to calculate the iPGF feature It will be equal to 14.75 GB.

Now, let's evaluate the calculation load graph and the memory consumption compared to the USF and CASIA databases. Suppose there are three variable parameters in each database that have a direct impact on the performance of the proposed algorithm: 1- the size of the input image (H×W), 2- the number of people in the gallery and test set (ntr+nte), and 3- Average periodicity (T). If we consider the dimensions of the input image to be the same (W=H=N), then the order of changes in the size of the input image will be equal to H = N2×W. In addition, the number of people in a database (according to the evaluated data) is two default values ntr + nte = n = {1000, 1300}. The period value should also be defined as two default values T={30, 20}. The filters' time and memory increase factor is equal to wh = 39*39 = 1521. With these assumptions, the graph of the changes in computational load and memory in the proposed algorithm compared to similar algorithms will be shown in Figure 4.1.

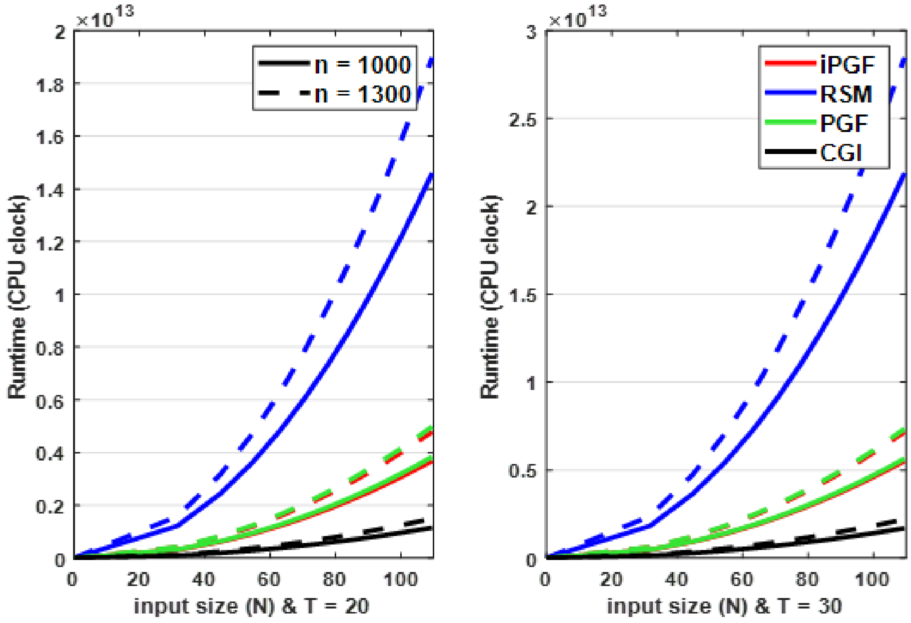


Figure 4.1. Computational load (in seconds) for different algorithms in different settings.

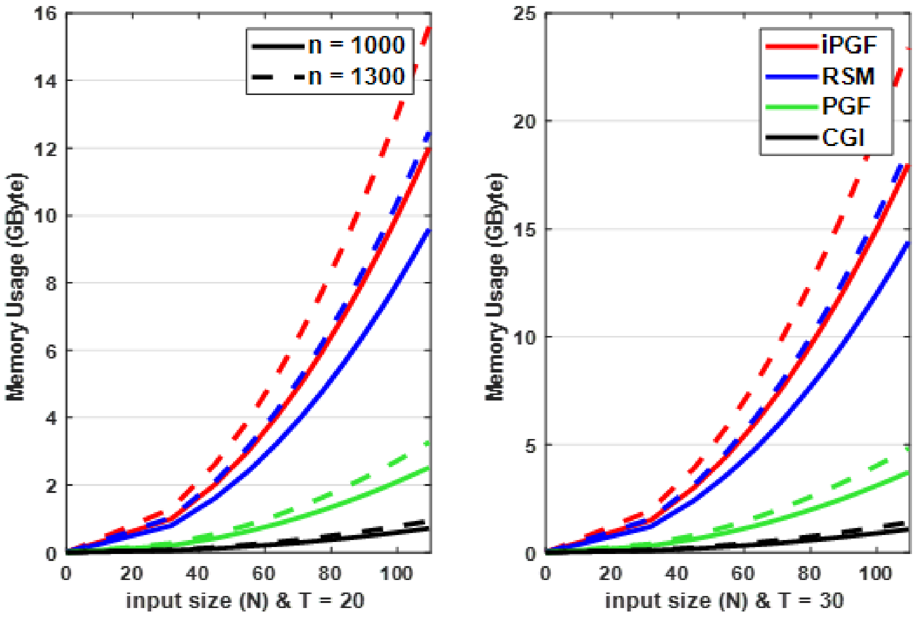


Figure 4.2. Amount of memory consumed (in GB) for different algorithms in different settings.

According to Figures 4.1 and 4.2, we can see that the calculation time of the proposed algorithm is very close to PGF and much faster than Gabor+RSM [35]. Also, the computational load of the iPGF algorithm is not much slower than CGI, and the speed reduction can be compensated by using the techniques of fast algorithm implementation and optimal functions. In addition, the memory consumption for iPGF calculations is close to Gabor+RSM, which can be solved using modern data storage techniques. In short, the proposed algorithm can identify the movement template in relatively more complex conditions with better accuracy. This issue has caused a slight increase in the computational load and an increase in the amount of memory used. But compared to similar algorithms, the computational load is competitive. Therefore, as a suitable algorithm, we can use it to identify the movement template.

Chapter 5

Conclusion

We propose an approach to enhance Spatio-temporal walking characteristics using regional adjacent spot descriptors and investigate a new approach for re-identifying people. We introduce the basic idea of an improved template derived by the Gabor filter and discuss the weaknesses of advanced algorithms. The PGF method can extract human motion information effectively but is vulnerable to local noise and mismatch. We solve this problem by refining local chunks and weighting them based on their importance. Our method, iPGF, is combined with RSM for gait template recognition in the USF database, resulting in a 1% improvement in rank 1 and 5.

The proposed iPGF can compete with well-known gait detection methods, but it can also be further improved by using stronger filtering or deep learning techniques to handle gait problems in real-life scenarios. Classification and feature extraction can be combined, but the amount of processing becomes very high, and normal computers cannot handle this level of calculation and processing. Therefore, we used classical machine learning due to the hardware and resource limitations in this project. Additionally, under normal conditions, our accuracy was about 100%, and this method solves our need by choosing optimal features.

One potential method for improvement in the future is to use wavelets instead of Gabor filters in the PGF method, as this may result in more optimal features from the image. Another method is to have an adaptive expectation mechanism for upper body patches, where if a person wears a long raincoat in the image, their movement behavior in the upper body does not change, but in the lower part of the body, there is a change in walking. We can give more weight to the upper body, such as the arms, head, and neck, to accommodate clothing conditions. The same approach can be taken for bagging, where the bag is seen as noise in the image that we must manage by weighting other parts of the image and reducing the weight of the parts of the image that have noise. In addition, while the current investigation resulted in an accuracy of 100%, the use of deep learning may be a viable alternative for a different database.

References

[1] J. Han, B. Bhanu, “Individual recognition using gait energy image,” *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 28, no. 2, pp. 316-322, 2006.

[2] S. Sarkar, P. J. Phillips, Z. Liu, I. R. Vega, P. Grother, K. W. Bowyer, “The human gait challenge problem; data sets, performance, and analysis,” *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 27, no. 2, pp. 162-177, 2005.

[3] Y. Dupuis, X. Savatier, P. Vasseur, “Feature subset selection applied to model-free gait recognition,”, in Mach. Vis. Conf., 2008.

[4] J. Zhang, J. Pu, C. Chen, R. Fleischer, “Low-resolution gait recognition,” IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, vol. 40, no. 4, pp. 986-996, 2010.

[5] Z. Liu, S. Sarkar, “Improved gait recognition by gait dynamics normalization,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 6, no. 28, pp. 863-876, 2006.

[6] D. Xu, S. Yan, D. Tao, L. Zhang, X. Li, H.-J. Zhang, “Human gait recognition with matrix representation,” IEEE Transactions on Circuits and Systems for Video Technology, vol. 16, no. 7, pp. 896-903, 2006

[7] P. J. Phillips, H. Moon, S. A. Rizvi, P. J. Rauss, “The FERET Evaluation Methodology for Face-Recognition Algorithms,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, no. 10, 2002.

[8] J. E. Boyd, J. J. Little, “Biometric Gait Recognition,” *Advanced Studies in Biometrics, Lecture Notes in Computer Science,* Berlin Heidelberg, Springer-Verlag , pp. 19-42, 2006.

[9] N. V. Boulgouris, K. N. Plataniotis, E. M. Tzanakou, Biometrics: Theory, Methods, and Applications, IEEE Press & John Wiley, 2009

[10] S. A. More, P. J. Deore, “A survey on gait biometrics,” *World Journal of Science and Technology,* vol. 2, no. 4, pp. 146-151, 2012.

[11] D. Xu, Y. Huang, Z. Zeng, X. Xu, “Human gait recognition using patch distribution feature and locality-constrained group sparse representation,” *IEEE Transactions on Image Processing,* vol. 21, no. 1, pp. 316-326, 2012.

[12] D. Gafurov, “A Survey of Biometric Gait Recognition; Approaches, Security and Challenges,” *NIK Conference,* 2007.

[13] S. D. Choudhury, T. Tjahjadi, “Robust view-invariant multiscale gait recognition,” *Journal of Pattern Recognition,*vol. 48, no. 3, pp. 798-811, 2015.

[14] C. Wang, L. W. J. Zhang, J. Pu, X. Yuan, “Human identification using temporal information preserving gait template,” *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 34, no. 11, pp. 2164-2176, 2012.

[15] M. H. Ghaeminia, S. B. Shokouhi, "On the Selection of Spatio-Temporal Filtering with Classifier Ensemble Method for Effective Gait Recognition," *Journal of Signal, Image and Video Processing,* vol. 13, no. 1, pp. 43-51, Feb. 2019.

[16] Z. Liu, S. Sarkar, "Effect of Silhouette Quality on Hard Problems in Gait Recognition," *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics,* vol. 35, no. 2, 2005.

[17] S. Zheng, J. Zhang, K. Huang, R. He, T. Tan, "Robust View Transformation Model for Gait Recognition," *Proceedings of the 18th IEEE International Conference on Image Processing,* 2011.

[18] R. Gross, J. Shi, "The CMU Motion of Body (MoBo) Database," Technical Report, Carnegie Mellon University, 2001.

[19] T. Chalidabhongse, V. Kruger, and R. Chellappa, "The UMD Database for Human Identification at a Distance," Technical Report, University of Maryland, 2001.

[20] J. Shutler, M. Grant, M. Nixon, J. Carter, "On a large sequence based human gait database," *Proceedings of 4th International Conference on Recent Advances in Soft Computing,* 2002.

[21] S. Yu, D. Tan, T. Tan, "A framework for evaluating the effect of view angle, clothing and carrying condition on gait recognition," *Proceeding of 18th International Conference on Pattern Recognition (ICPR),* Hong Kong, 2006.

[22] M. A. Hossain, Y. Makihara, J. Wang, Y. Yagi, "Clothing-invariant gait identification using part-based clothing categorization and adaptive weight control," *Pattern Recognition,* vol. 43, no. 6, pp. 2281-2291, 2010.

[23] C. Chen, J. Zhang, R. Fleischer, "Multilinear tensor-based non-parametric dimension reduction for gait recognition," *Proceedings of 3rd International Conference on Biometrics*, 2009.

[24] M. H. Ghaeminia, S. B. Shokouhi, "GSI: Efficient Spatio-Temporal Template for Human Gait Recognition," *International Journal of Biometrics (IJBM),* vol. 10, no.1, pp. 29-51, 2018.

[25] D. Tao, X. Li, X. Wu, S. J. Maybank, "General tensor discriminant analysis and Gabor features for gait recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 9, no. 10, pp. 1700-1715, 2007.

[26] Z. Zhou, A. P. Bennett, R. I. Damper, "A Bayesian framework for extracting human gait using strong prior knowledge," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 28, no. 11, pp. 1738-1752, 2006.

[27] W. Zeng, C. Wang, Y. Li, "Model-based human gait recognition via deterministic learning," *Journal of Cognitive Computation,* vol. 6, no. 2, pp. 218-229, 2013

[28] Y. Makihara, R. Sagawa, Y. Mukaigawa, T. Echigo, Y. Yagi, "Gait recognition using a view transformation model in the frequency domain," 9*th European Conference on Computer Vision*, Graz, 2006.

[29] H. Haifeng, "Enhanced Gabor feature-based classification using a regularized locally tensor discriminant model for multiview gait recognition," *IEEE Transactions on Circuits and Systems for Video Technology,* vol. 23, no. 7, pp. 1274-1286, 2013.

[30] N. V. Boulgouris, Z. X. Chi, “Gait Recognition Using Radon Transform and Linear Discriminant Analysis,” *IEEE Transactions on Image Processing,* vol. 16, no. 3, 2007.

[31] S. Huang, A. Elgammal, J. Lu, D. Yang, “Cross-Speed Gait Recognition Using Speed-Invariant Gait Templates and Globality–Locality Preserving Projections,” *IEEE Transactions on Information Forensics and Security,* vol 10, no. 1, pp. 2071-2083, Oct. 2015.

[32] T. W. Yeoh, F. Daolio, H. E. Aguirre, K. Tanaka, “On the effectiveness of feature selection methods for gait classification under different covariate factors,” *Journal of Applied Soft Computing,* vol. 61, pp. 42-57, 2017.

[33] M. Hu, Y. Wang, Z. Zhang, D. Zhang, J. J. Little, “Incremental learning for video-based gait recognition with LBP flow,” *IEEE Trans. Syst., Man, Cybern. B, Cybern.*, vol. 43, no. 1, pp. 77-89, 2013.

[34] D. L. Fernandez, F. J. M. Cuevas, A. C. Poyato, R. M. Salinas, R. M. Carnicer, “Entropy Volumes for Viewpoint-Independent Gait Recognition,” *Machine Vision and Applications,* 2015.

[35] Y. Guan, C.-T. Li , F. Roli, “On reducing the effect of covariate factors in gait recognition: A classifier ensemble method,” *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 37, no. 7, pp. 1521-1528, 2015.

[36] M. Li, B. Yuan, “2D-LDA: A statistical linear discriminant analysis for image matrix,” *Pattern Recognit. Lett.,* vol. 26, pp. 527-532, 2005.

[37] Y. Huang, D. Xu, T. Cham, “Face and human gait recognition using image-to-class distance,” *IEEE Trans. Circuits Syst. Video Technol.,* vol. 20, no. 3, pp. 431-438, 2010.

[38] S. M. Darwish, “Design of adaptive biometric gait recognition algorithm with free walking directions,” *IET Biometrics,* vol. 6, no. 2, pp. 53-60, 2017.

[39] Y. Guan, C.-T. Li, Y. Hu, “Random subspace method for gait recognition,” *IEEE Int. Conf. Multimedia Expo Workshops*, 2012.

[40] A. Kale, A. Sundaresan, A. Rajagopalan, N. Cuntoor, A. K. R. Chowdhury, V. Kruger, R. Chellappa, “Identification of humans using gait,” *IEEE Transactions on Image Processing,* vol. 13, no. 9, pp. 1163-1173, 2004.

[41] S. Hong, H. Lee, E. Kim, “Probabilistic gait modeling and recognition,” *IET Computer Vision,* vol. 7, no. 1, pp. 56-70, 2013.

[42] Z. Lai, Y. Xu, Z. Jin, D. Zhang, “Human gait recognition via sparse discriminant projection learning,” *IEEE Trans. Circuits Syst. Video Technol.,* vol. 24, no. 10, pp. 1651-1662, 2014

[43] G. Ma, L. Wua , Y. Wang, “A general subspace ensemble learning framework via totally-corrective boosting and tensor-based and local patch-based extensions for gait recognition,” *Pattern Recognition,* vol. 66, pp. 280-294, 2017.

[44] Y. Chai, Q. Wang, a. R. Z. J.P. Jia, “A Novel human gait recognition method by segmenting and extracting the region variance feature,” *Proceedings of International Conference on Pattern Recognition*, 2006.

[45] I. Bouchrika, M. Nixon, “Model-based feature extraction for gait analysis and recognition,” *Proceedings of Third International Conference on Computer Vision/Computer Graphics Collaboration Techniques and Applications*, 2007.

[46] T. H. W. Lam, K. H. Cheung, J. N. K. Liu, “Gait flow image: a silhouette-based gait representation for human identification,” *Journal of Pattern Recognition*, vol. 44, no. 4, pp. 973-987, 2011.

[47] E. H. Adelson and J. R. Bergen, "Spatio-temporal energy models for the perception of motion," *Journal of Optical Society of America*, vol. 2, no. 2, pp. 284-299, 1985.

[48] A. H. Shabani, J. S. Zelek, D. A. Clausi, “Human action recognition using salient opponent-based motion features,” *Proceedings of IEEE Canadian Conference on Computer and Robot Vision*, Ottawa, 2010.

[49] A. B. Watson, A. J. Ahumada, “Model of human visual-motion sensing,” *Journal of Optical Society of America*, vol. 2, no. 2, pp. 322-342, 1985.

[50] A. H. Shabani, D. A. Clausi, J. S. Zelek, “Improved Spatio-temporal salient feature detection for action recognition,” *Proceedings of the British Machine Vision Conference*, Dundee, 2011.

[51] H. Haifeng, “Multiview Gait Recognition Based on Patch Distribution Features and Uncorrelated Multilinear Sparse Local Discriminant Canonical Correlation Analysis,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 4, pp. 617-630, 2014.

[52] Y. Huang, D. Xu, F. Nie, “Patch Distribution Compatible Semisupervised Dimension Reduction for Face and Human Gait Recognition,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 3, pp. 479-488, 2012.

[53] M. H. Ghaeminia, S. B. Shokouhi, A.Badiezadeh, “A New Spatiotemporal Patch-Based Feature Template for Effective Gait Recognition,” *Multimedia Tools and Applications, vol. 79, pp. 713-736,* 2020.

[54] M. Liu, S. Yan, Y. Fu, and T. Huang, “Flexible X-Y patches for face recognition,” *IEEE Int. Conf. Acoust., Speech, Signal Process*, 2008.

[55] Y. Zhang, K. Shang, J. Wang, N. Li, Monica, M. Zhang, “Patch strategy for deep face recognition,” *IET Image Processing*, vol. 12, no. 5, pp. 819-825, 2018.

[56] S. Lucey, T. Chen, “A GMM parts-based face representation for improved verification through relevance adaptation,” *Proc. IEEE Int. Conf. Comput. Vis. Pattern Recog.*, 2004.

[57] W. Kusakunniran, “Recognizing gaits on Spatio-temporal feature domain,” *IEEE Transactions on Information Forensics and Security* vol. 9, no. 9, pp. 1416-1423, 2014.

[58] H. Iwama, M. Okumura, Y. M., and Y. Yagi, “The OU-ISIR Gait Database Comprising the Large Population Dataset and Performance Evaluation of Gait Recognition,” *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 5, pp. 1511-1521, 2012.

[59] Y. Makihara, H. Mannami, A. Tsuji, M. Hossain, K. Sugiura, A. M. a. Y. Yag, “The OU-ISIR Gait Database Comprising the Treadmill Dataset,” *IPSJ Trans. on Computer Vision and Applications,* vol. 4, pp. 53-62, Apr. 2012.

[60] P. Dollar, V. Rabaud, G. Cottrell, S. Belongie, “Behavior recognition via sparse Spatio-temporal filters,” *IEEE International Workshop VS-PETS,* Beijing, China, August 2005.

[61] Y. Wang, C. [70] Song, Y. Huang, Z. Wang, L. Wang, “Learning view-invariant gait features with Two-Stream GAN,” *Journal of Neurocomputing,* vol. 33, pp. 245-254, 28 Apr. 2019.

[62] A. Ghebleh, M. E. Moghaddam, “Clothing-invariant human gait recognition using an adaptive outlier detection method,” *Journal of Multimedia Tools and Applications (MTAP),* vol. 77, no. 7, pp. 8237-8257, 2018.

[63] Z. Xu, W. Lu, Q. Zhang, Y. Yeung, X. Chen, “Gait recognition based on capsule network,” *Journal of Visual Communications and Image Representation,* vol. 59, pp. 159-167, Feb. 2019.

[64] P. Grother and E. Tabassi, “Performance of Biometric Quality Measures,” *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 29, no. 4, pp. 531-543, 2007.

[65] R. F. S. Teixeira, N. J. Leite, “A New Framework for Quality Assessment of High-Resolution Fingerprint Images,” *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 39, no. 10, pp. 1905-1917, 2017.

[66] F. Alonso-Fernandez, J. Fierrez, J. Ortega-Garcia, “Quality Measures in Biometric Systems,” *IEEE Security & Privacy,* vol. 10, no. 6, pp. 52-62, 2012.

[67] R. M. Bolle, J. H. Connell, S. Pankanti, N. K. R. a. A. W. Senior, “The relation between the ROC curve and the CMC,” Fourth *IEEE Workshop on Automatic Identification Advanced Technologies (AutoID'05),* Buffalo, NY, USA, 2005.

[68] I. Rida, N. Almaadeed, S. Almaadeed, “Robust gait recognition: a comprehensive survey,” *IET Biometrics,* vol. 8, no. 1, pp. 14-28, 2019.

[69] M. Koohzadi, N. M. Charkari, “Survey on deep learning methods in human action recognition,” *IET Computer Vision,* vol. 11, no. 8, pp. 623-632, 2017.

[70] J. Ye, Q. Li, H. Xiong, H. Park, R. Janardan, V. Kumar, “IDR/QR: An incremental dimension reduction algorithm via QR decomposition,” *IEEE Trans. Knowl. Data Eng.,* vol. 17, no. 9, pp. 1208-1222, 2005.

[71] J. Yang, D. Zhang, A. F. Frangi, J. Yang, “Two-dimensional PCA: A new approach to appearance-based face representation and recognition,” *IEEE Trans. Pattern Anal. Mach. Intell,* vol. 26, no. 1, pp. 131-137, 2004.

[72] D. Xu, S. Yan, D. Tao, S. Lin, H. Zhang, “Marginal Fisher analysis and its variants for human gait recognition and content-based image retrieval,” *IEEE Trans Image Process.,* vol. 16, no. 11, pp. 2811-2821, 2007.

[73] G. Willems, T. Tuytelaars, L. V. Gool, “An efficient dense and scale-invariant Spatio-temporal interest point detector,” *European Conference on Computer Vision*, Marseille, France, October 2008.

[74] X. Wang, X. Tang, “Random sampling for subspace face recognition,” *Int. J. Comput. Vis.*, vol. 70, no. 1, pp. 91-104, 2006.

[75] A. Veeraraghavan, A. Chowdhury, R. Chellappa, “Matching shape sequences in video with applications in human movement analysis,” *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 27, no. 12, pp. 1896-1909, 2005.

[76] G. Somasundaram, A. Cherian, V. Morella’s, N. Papanikolopoulos, “Action recognition using global Spatio-temporal features derived from sparse representations,” *Computer Vision and Image Understanding,* vol. 13, pp. 1-13, 2014.

[77] A. H. Shabani, J. S. Zelek, D. A. Clausi, “Multiple scale-specific representations for improved human action recognition,” *Pattern Recognition Letters,* vol. 1, no. 1, pp. 1771-1779, 2013.

[78] S. Rahman, J. See, “Spatio-Temporal Mid-Level Feature Bank for Action Recognition in Low-Quality Video,” *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP),* 2016.

[79] Y. Makihara, A. Suzuki, D. Muramatsu, X. Li, Y. Yagi, “Joint Intensity and Spatial Metric Learning for Robust Gait Recognition,” IEEE *Conference on Computer Vision and Pattern Recognition (CVPR),* Honolulu, HI, 2017.

[80] J. Lu, G. Wang, P. Moulin, “Human identity and gender recognition from gait sequences with arbitrary walking directions,” *IEEE Trans. Inf. Forensics Security,* vol. 9, no. 1, pp. 51-61, 2014.

[81] T. Lindeberg, D. Fagerstrom., “Scale-space with causal time direction,” *European Conference on Computer Vision,* Cambridge, UK, 1996.

[82] I. Laptev, B. Caputo, C. Schuldt, T. Lindeberg, “Local velocity-adapted motion events for Spatio-temporal recognition,” *Computer Vision and Image Understanding,* pp. 207-229, 2007.

[83] T. K. Ho, “The random subspace method for constructing decision forests,” *IEEE Trans. Pattern Anal. Mach. Intell.,* vol. 20, no. 6, pp. 832-844, 1998.

[84] S. Tong, Y. Fu, X. Yue, H. Ling, “Multi-View Gait Recognition Based on a Spatial-Temporal Deep Neural Network,” *IEEE Access,* vol. 6, pp. 57583-57596, 2018.

[85] E. Fendi, I. Chtourou, M. Hammami, “Gait-based person re-identification under covariate factors,” *Journal of Pattern Analysis and Application (PAA),* 11 Feb. 2019, https://doi.org/10.1007/s10044-019-00793-4.

[86] X. Ben, C. Gong, P. Zhang, X. Jia, Q. Wu, W. Meng, “Coupled patch alignment for matching cross-view gaits,” *IEEE Transactions on Image Processing,* vol. 28, no. 6, pp. 3142-3157, June 2019.

[87] Wang, L., & Tan, T. (2003). Recent developments in human motion analysis. Pattern recognition, 36(3), 585-601.

[88] Han, J., Bhanu, B., & Roy-Chowdhury, A. K. (2005). A comprehensive survey of recent advances in video-based human behavior analysis. Pattern Recognition Letters, 26(9), 1167-1179.

[89] Nixon, M. S., Carter, J. N., & Prathalingam, J. (2009). On the recognition of human gait from optical motion capture data. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 39(1), 50-62.

[90] Zhang, J., & Lu, H. (2010). Review of gait recognition. Systems Engineering Procedia, 2, 268-275.

[91] Bashir, K., & Xiang, T. (2016). Human gait recognition: A review. Computer Vision and Image Understanding, 145, 1-20.

[92] Rattani, A., Nappi, M., & Tucci, M. (2015). Biometric systems: An overview. Handbook of Biometrics for Forensic Science, 1-26.

[93] Ross, A., Jain, A. K., & Nandakumar, K. (2010). Handbook of multibiometrics (Vol. 6). Springer Science & Business Media.

[94] Tse, K. H., & Zhang, D. (2011). Biometric recognition: Challenges and opportunities. World Scientific Publishing Co Inc.

[95] Naseem, I., Togneri, R., & Bennamoun, M. (2014). Linear regression for face recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 38(10), 2066-2073.

[96] Zhang, J., & Wu, Y. (2019). Gait recognition: An overview of recent advances and challenges. IEEE Signal Processing Magazine, 36(5), 51-63.

[97] Mohammad H. G, Shahriar B. Sh,Abdollah A. (2021). Biometric Gait Identification Systems: From Spatio-Temporal Filtering to Local Patch-Based Techniques. Springer Science & Business Media.

[98] Gait recognition for security applications using deep learning" ،Mohammad A. AlZu'bi ، Journal of Ambient Intelligence and Humanized Computing2020 ،.

[99] "Intelligent Traffic Control System Based on Image Processing and Machine Learning" ، Rongbing Zhang ،IEEE Access2020 ،.

[100] "Real-time gait recognition system for surveillance" ،Uğur Karabiyik ،Signal, Image and Video Processing2020 ،.

[101] "Gait Recognition System for Security and Surveillance Applications" ،Z. Wu ،International Journal of Digital Crime and Forensics2019 ،.

[102] "Vehicle and Pedestrian Detection for Smart Traffic Management System" ،Oğuzhan Urhan ، IEEE Transactions on Intelligent Transportation Systems2019 ،.