Gait-Based Gender classification Using Zernike Moments

M. Keerthana Devi, Mrs. R. Newlin Shebiah
Department of ECE, Mepco Schlenk Engineering College

Abstract: Most biometric systems necessitating human identification encompass techniques of finger-print systems, iris sensing, DNA, palm print and odour/scent. With these paramount advantages, Gait is emerging rapidly and gaining wide attention as a biometric identifier, and minor variations in gait features can be used to classify the gender of a person. Human Gait recognition has been a major concern because of its immense applications’ in the field of autonomous video surveillance, video retrieval, which require methods for recognizing human actions in various scenarios. This paper focuses on a novel method of incorporating the techniques of utilizing the Motion History Images and Zernike moments of each person to classify their genders. Exceptional videos solely based on deducing the gait movements of a person were gained from Dgait a Gait database, acquired with a depth camera. This database comprises of videos from 53 subjects walking in different directions. Each of these videos were processed individually to obtain their Motion history Images. From these images, calculations for Zernike moments were built up based on which gender of a person were classified using a SVM classifier. The recognition rate for RGB Zernike moments is 82.3%.

Keyword: Video surveillance, Biometric Identifier, Gait, Kinect, Motion History Images, Zernike moments, SVM classifier.

1. INTRODUCTION

Video surveillance has been posing a ginormous slot of immense privileges, and is rapidly gaining recognition as a supportive tool for monitoring human movements and preventing thefts. Some fatal disadvantages like vandalism deterrence, copper thefts, outer perimeter security, crime in public, protecting the employer from frivolous employee lawsuits can all be outdone with the help of a simple video surveillance mechanism. Human gait is affected by various factors such as walking features, carrying status and several environmental factors [1]; this was proposed by XuelongLi, StephanJ et al. The paper shows the use of 3D sensor Microsoft Kinect that helps in recording depth information in a confined environment. Kinect records RGBD videos at an rate of 30 frames per second with an resolution of 640*480 [2] was proposed by authors, Milos Milovanovic et al. Microsoft Kinect provides wide range of API to identify the key position on human body [2]. For data retrieval color histogram technique and for recognition Hidden Markov model is used [2]. Sensor helps in tracking skeletal of a human body by randomly marking skeletal points of human joints, which was proposed by AlesProchazka et al. [3]. The skeletal data obtained from Kinect are processed for feature extraction and gender of the person was recognized using WEKA tool [4]. Segmented image were processed to obtain a complete gait cycle [1]. Based on shape descriptors, authors Ricard Borras extracted 2D and 3D gait features [5]. Depth information provided by sensor helps in accurate human segmentation as in OPENNI middleware [6]. Apart from D-Gait datasets, Casia datasets provide silhouette that are extracted in various environmental condition [7]. Silhouettes are computed effectively from Kinect data by computing discrete cosine transform as proposed by Hansung Kim et al. [8]. James W. Davis proposed that fusing the results with the statistical template and feature extractions obtained by identifying the Gait cycles helps in isolating gait features in various environmental conditions [9]. Moment invariants based features was used to classify the shape of an object with respect to translation, scale and rotation [10]. For representing action of each individual, pattern of motion over a specific time stamp from...
reference was done by Laura Keyes and Adam Win Stanley [11]. Ju Han and BirBhanu together proposed that shape descriptors used in object reorganization were calculated by using central moments [10]. Past work on finding shape was centroid calculation is replaced by motion patterns which helps in finding spatial distribution of an object by Parul Arora et al. Application of histogram of oriented gradients on Gait Energy Image preserves edge information and reduces the amalgamation of noise [12]. The main difference between genders was proposed by Luciano Santos Constantin Raptopoulos and is distinguished by Kinectamics chain in gender locomotion. Finally gait modes are analyzed using Karhunem –Loeve approach [13]. Information set contained in each frames is treated as a single quantity whereas fuzzy set is formed by intensity values of pixels at each location was concluded by Parul Narula et al. Gait Information image computes two functions one is based on energy and other is based on sigmoid function [14]. Robust identification of object can be obtained by optical flow and by using Spatio-temporal filtering techniques which was proposed by Manorajan Paul [15].

2.1 PROPOSED METHODOLOGY

This topic outlines the basic methodologies that are further incorporated in this paper to obtain higher understanding. Here first the input video is converted to frames, from which RGB and depth data is segregated, for which silhouettes were extracted using background subtraction technique. From a group of these silhouettes, Motion History Images were calculated for which Zernike moments were calculated.

The database utilized here is the Dgait database; this can be acquired at the following link Dgait database. This database was obtained from an indoor environment using the Microsoft Kinect. In order to capture videos using Kinect an exclusive Software Development Kit (Kinect SDK) can be used for an effortless recoding of videos using this hardware. The Dgait database consists of 53 subjects, 36 male and 17 female most of them were Caucasian. Microsoft Kinect gives 640*480 images at 30 frames per second with depth resolution of few centimetres. The Database consists of RGBD video per subjects with total number of 11 sequences.

2.1.2 SILHOUETTE EXTRACTION:

The videos obtained from Dgait database were first segregated into depth and RGB videos. These videos were labeled accordingly as male and female subjects, which were then converted to individual frames out of which the foreground was distinguished from the background. Each of the frames was converted to hsv models for better accuracy. Detecting foreground and background image from Dgait datasets and finding an absolute difference between two images is done for extracting silhouettes from the RGB videos. An predefined 2-D filter is used for filtering purpose. After filtering, perfect silhouette of RGB dataset were obtained. In case of depth extraction, the current frame is at particular distance from background image. Suitable threshold value is set to classify the pixel in each frame. Silhouettes are generated from frames by facing the camera in lateral position rather than camera facing in frontal view.

2.1.3 MOTION HISTORY IMAGES:

For recognizing human action motion history image is calculated by accumulating patterns of motion of each individual. Recognizing human action through frames extracted from video clips is done here. Here the approach used is a compact, yet descriptive, captured sequence of motions in a single static image. Silhouette sequence is condensed in to gray scale images where important information is preserved, because these images can easily demonstrates the flow of motion over a period of time. Adjusting the values of decay parameter the information about the motion of an individual can be extended. Moving parts of video clips can be converted into single
image where maximum brightness corresponds to recent changes in images. For every changes in time, absolute difference between the current and previous intensity values are computed and this value is compared with the threshold value. The motion of individual at every location decays over an period of time. Therefore for a fixed timestamp the sequence of motion flow was generated which corresponded to the motion of an individual from their initial to final position.

\[
MHI_{\delta}(x, y) = \begin{cases} 
\delta \rho \phi((I(x, y)) \neq 0) \\
\text{else if } MHI_{\delta}(x, y) < \tau - \delta
\end{cases}
\]  

(2.1.3a)

Where each pixel \((x, y)\) in the MHI is marked with a current timestamp if the function signals object presence in the current video image \(I(x, y)\); the remaining timestamps are removed if they are older than the decay value \(\tau - \delta\). This update function is called for every new video frame analyzed in the sequence. The function \(\psi\) selects a pixel location in the input image for inclusion into the MHI.

### 2.1.4 ZERNIKE MOMENTS:

The moments vary in accordance with translation, scale and rotation. The set of polynomial that are rotational invariant is Zernike moments. Initially, the center of an image is taken as an origin and the pixel coordinates are mapped to an unit circle. The pixels that fall outside the unit circle will not be used for computation. Moments in Zernike computes both magnitude and phase values that are invariant over rotation. Orthogonal moments are suited for computing Zernike values which involves less information redundancy. Zernike values are possible even with fewer data points. The image is subjected to normalization using regular moments to obtain translation and scale invariance. Rotational invariance Zernike features are extracted from scale and translation invariance normalized images. Zernike is highly preferred over other invariants because it poses a higher accuracy with simple rotational invariance. Another advantage is that Zernike moments are accurate descriptors even with few data points.

\[
V_{nm}(\rho, \theta) = R_{nm}(\rho)e^{jnm\theta}
\]

(2.1.4a)

Where \(V_{nm}(\rho, \theta)\) is known as the Zernike polynomial,

\[
A_{nm} = \frac{m+1}{\pi} \int_{x} \int_{y} f(x, y)[V_{nm}(x, y)]dxdy, x^2 + y^2 \leq 1
\]

(2.1.4.b)

Gives the magnitude.

\[
R_{nm}(\rho) = \begin{cases} 
\sum_{l=0}^{n-m/2} \frac{(-1)^l (n-l)!}{l! \left(\frac{n-m-l}{2}\right)!} \rho^{n-2l} & \text{for } n-m \text{even} \\
0 & \text{for } n-m \text{odd}
\end{cases}
\]

(2.1.4.c)

Where \(R_{nm}(\rho)\) is a radial polynomial.

### 2.1.5 HU MOMENTS:

Moment invariants have been frequently used as features for image processing, remote sensing, shape recognition and classification. Moments can provide characteristics of an object that uniquely represent its shape. Invariant shape recognition is performed by classification in the multidimensional moment invariant feature space.

### 2.1.6 SVM CLASSIFIER:

Support Vector Machines are based on the concept of decision planes that define decision boundaries. A decision plane is one that separates between a set of objects having different class memberships. Classification tasks based on drawing separating lines to distinguish between objects of different class memberships are known as hyper plane classifier.
3 RESULTS AND DISCUSSION:

3.1 SILHOUETTE EXTRACTION:

The absolute difference between current frames gives and blob like image. After suitable application of special filter and perfect silhouette are extracted.

3.2 MOTION HISTORY IMAGES:

The motion of an individual from their initial to final position is determined precisely by increasing its timestamp for each individual.
**3.3 FEATURE EXTRACTION AND SVM CLASSIFIER:**

From motion history image [2.3] and Zernike calculation [2.4], the distinct magnitude and phase of each male and female are extracted. SVM (Support Vector Machine) is a machine learning model with learning algorithms that analyses data and recognize patterns. Here, SVM classifier are used to train and test D-gait datasets. SVM classifier avoids the regression error and maximize the better margin between two different classes. SVM builds a hyperplane that separate the data points into two classes.

Table 1. Tabulation of Classification Rate obtained.

<table>
<thead>
<tr>
<th>Method</th>
<th>Total Number of Samples</th>
<th>Correctly Classified</th>
<th>Recognition Rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>Zernike from RGB</td>
<td>19  16</td>
<td>17  12</td>
<td>82.3%</td>
</tr>
</tbody>
</table>

Here, D-gait datasets contain actions that are taken at a fixed distance from a Kinect. The videos were first segregated to female and male videos. Each of these videos was converted to individual frames which were further processed to obtain their respective silhouettes. These silhouettes were then computed for obtaining Motion History Images which were further processed to calculate Zernike moments. The values obtained were stored in separate Mat files and loaded into classifiers to obtain the gender of a person. The training dataset is independent of testing dataset (i.e., separate persons for each sets of dataset). Separate training and testing are loaded into classifier. There a recognition rate using RGB Zernike moment is 82.3%.

**4 CONCLUSION**

In this proposed method, Gait image is taken into consideration for processing. This processing scheme is based on extracting silhouettes from each image, computing the Motion History Images for each set of silhouettes to find the path traced by a subject and computing its respective Zernike moments which are further loaded to a SVM Classifier. The recognition rate achieved is given in the Table above. From the total number of samples classified, the recognition rate is obtained. The Zernike moments from RGB have a recognition rate of 82.3%.
5. REFERENCES:


