CHAPTER 3
DATA SET DESCRIPTION AND PREPROCESING

3.1. INTRODUCTION

This chapter presents a description of dataset considered in this study and pre-processing of data which is carried out for data cleaning and data selection.

3.2. DATA SOURCE

The datasets which are commonly used by researchers in gait recognition system are Body's CMU Motion (MoBo) database, The USF dataset, soton database, CASIA dataset, TUM-IITKGP gait database. Gait recognition has been an active topic of research in the recent years. The Institute of Automation, Chinese Academy of Sciences (CASIA) has provided the CASIA database to carry out research in the field of gait recognition. The CASIA database consists of three datasets namely Dataset A, Dataset B (multi view dataset) and Dataset C (infrared dataset) and Dataset D. As an experimental study for evaluating the performance of classification algorithms, in this study, the CASIA B dataset is used.

3.3. DATA SET DESCRIPTION

CASIA dataset B (Zheng et al., 2011) was developed by Institute of Automation Chinese Academy of Sciences in January 2005. It includes gait sequences of 124 subjects and the gait data was captured from 11 views. For each of the subject's three variations with respect to view angle, clothing and carrying conditions were recorded. The format of the video filename in Dataset B is 'xxx-mm-nn-ttt.avi', where

xxx : subject id, from 001 to 124.

mm : walking status, can be 'nm' (normal), 'cl' (in a coat) or 'bg' (with a bag).

nn : sequence number.

ttt : view angle, can be '000', '018', ..., '180'.


(a) Walking sequence from eleven camera views (Sivapalan 2014)

Figure 3.1. Example images from the CASIA dataset B.

(a) Walking sequence from eleven camera views, (b) normal walking (nw), (c) walking with different clothing (cl) and (d) walking with a bag (bg).

The figure 3.1. shows the video files available in CASIA dataset B with videos related to normal walking, carrying conditions like bag and clothing conditions. The gait recognition system consists of two parts. The first one is training part and the next is testing part. In the training phase, an individual’s walking style is recorded and stored in the database. In testing phase the input videos are compared with sequences stored in database. System will identify unauthorized individual and compare his gait with stored sequences and recognize him. This system finds wide application in factories and defence systems etc., where security is a major concern.

The system contains N video clips in the training set \{X_1, X_2, ..., X_N\}. Separate each video into k subsets along the particular time axis, each of which contains M frames including at least one walking cycle. The video sequences are taken as an input in testing phase. Each video clip contains Mi frames and there are totally N= Mi frames.
3.4. PREPROCESSING

The pre-processing step in gait recognition mainly involves silhouette extraction. Silhouette is defined as a dark shape seen against a light surface or the region of pixels of a walking person. The process of silhouette extraction mainly involves segmenting the human body and fig 3.2 shows the steps involved in the extraction process.

**Figure 3.2. Silhouette extraction process**

In the figure 3.2, the input image is subjected to background subtraction. In this method if the pixel value of each frame is not equal to the pixel value of background, then the pixel is taken as region of silhouette (Chen et al., 2006). The shadow removal is done by using a threshold value which is applied to the difference image. In order to have an effective and clear pixel distribution along the image, the intensity histogram of the image is generated according to the applied threshold value. Morphological filters are used to remove noise and finally the output image which is the silhouette is obtained. The figure 3.3 shows sample of extraction silhouette process.
(a) Process of differencing (Hayder et al., 2011)

(b) Extracted silhouette

Figure 3.3. Sample of silhouette process (a) Process of differencing.
(b) Extracted silhouette.

3.5. CHAPTER SUMMARY

This chapter gave a detailed description about the dataset used in this work and also explained in detail about the preprocessing steps done in this work.