Person Identification using Gait by Combined Features of Width and Shape of the Binary Silhouette

M.K. Bhuyan, *Member, IEEE,* and Aragala Jagan.

Abstract—Current image-based individual human recognition methods, such as fingerprints, face, or iris biometric modalities generally require a cooperative subject, views from certain aspects, and physical contact or close proximity. These methods cannot reliably recognize non-cooperating individuals at a distance in the real world under changing environmental conditions. Gait, which concerns recognizing individuals by the way they walk, is a relatively new biometric without these disadvantages. The inherent gait characteristic of an individual makes it irreplaceable and useful in visual surveillance.

In this paper, an efficient gait recognition system for human identification by extracting two features namely width vector of the binary silhouette and the MPEG-7-based region-based shape descriptors is proposed. In the proposed method, foreground objects *i.e.,* human and other moving objects are extracted by estimating background information by a Gaussian Mixture Model (GMM) and subsequently, median filtering operation is performed for removing noises in the background subtracted image. A moving target classification algorithm is used to separate human being (*i.e.,* pedestrian) from other foreground objects (*viz.,* vehicles). Shape and boundary information is used in the moving target classification algorithm. Subsequently, width vector of the outer contour of binary silhouette and the MPEG-7 Angular Radial Transform coefficients are taken as the feature vector. Next, the Principal Component Analysis (PCA) is applied to the selected feature vector to reduce its dimensionality. These extracted feature vectors are used to train an Hidden Markov Model (HMM) for identification of some individuals. The proposed system is evaluated using some gait sequences and the experimental results show the efficacy of the proposed algorithm.

Keywords—Gait Recognition, Gaussian Mixture Model, Principal Component Analysis, MPEG-7 Angular Radial Transform.

I. INTRODUCTION

TODAY, in metropolitan public transport stations, authentication or verification using conventional technologies is practically infeasible. In such type of applications, biometric authentication methods are more attractive. Biometrics are the unique features of a person. Biometric recognition refers to an automatic recognition of individual based on feature vectors derived from their physiological and/or behavioral characteristic. Biometric systems for human identification at a distance have ever been an increasing demand in various significant applications. Recognition/Identification using gait becomes more attractive in such type of situations. Human gait is a spatio-temporal phenomenon that characterizes the motion of an individual. The definition of gait is *"A particular way or manner of walking on foot"* [7]. Human gait recognition works from the observation that an individual's walking style is unique and can be used for human identification. Depending on feature extraction, gait recognition methods are classified as appearance-based and model-based gait recognition. The appearance-based approaches suffer from changes in the appearance owing to the change of the viewing or walking directions. Model-based approaches extract the motion of the human body by means of fitting their models to the input images. Model-based methods are view and scale invariant.

In [6], a gait recognition system using the extended features which incorporate spatial and temporal information and the combination of Eigen Space Transformation (EST) and Canonical Space Transformation(CST) for feature extraction was proposed. Advantages of this method is, EST and CST has been used to reduce data dimensionality with less computation time. The use of extended features greatly increases the robustness and accuracy of recognition. But, it needs further improvement for large database. In [8], gait recognition method based on statistical shape analysis is discussed. Main drawback of this method is that it is view-dependent. Silhouette analysis based recognition system was proposed in [9]. In this, distance signal was the feature vector, which is obtained by calculating distance between each pixel and centroid of the binary silhouette. Principal Component Analysis (PCA) is used for training of gait sequences and Spatio-Temporal Correlation (STC) is used for similarity measurement. It gives poor results for different clothing varieties , especially at different seasons.

In this paper, some of these limitations are overcome by taking combined features in the form of width and shape information of the binary silhouette of the person to be identified. First step of the proposed method is the extraction of foreground objects *i.e.,* human and other moving objects are extracted from input video sequences. Gaussian mixture model is used for foreground object estimation in which an additional step of filtering by median filter is incorporated to remove noises. Moving target classification algorithm is used separate human being (*i.e.,* pedestrian) from other foreground objects (*viz.,* vehicles). Shape and boundary information is used for this moving target classification. Width vector of outer contour of binary silhouette and MPEG-7 ART (Angular Radial Transform) coefficients are taken as the feature vector. PCA is applied to feature vector to reduce its dimensionality. These extracted feature vectors are used to recognizing individuals. Hidden Maekov Model (HMM) is used for recognizing persons on the basis of gait.

M.K. Bhuyan is with the Department of Electronics and Electrical Engineering, Indian Institute of Technology, Guwahati, India–781039, e-mail: *mkb@iitg.ernet.in.*

Aragala Jagan is with M M University, Mullana, Ambala, Haryana, India-133203 e-mail: *jaganaragala@gmail.com.*

II. PROPOSED APPROACH

Block diagram shown in Fig. 1 describes the steps involved in the proposed gait recognition system. Each step of the proposed system and methods used for the same are explained below.
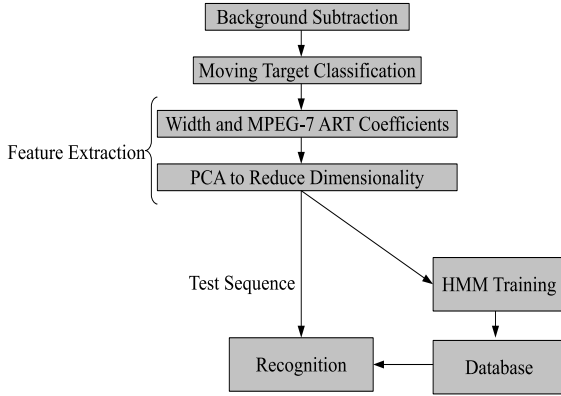

Fig. 1. Block diagram of proposed gait recognition system

A. Background Subtraction

Background subtraction identifies moving objects from background in the scene. Pixels in the current frame that deviate significantly from the stationary background are considered to be moving objects. Gaussian mixture model (GMM) in addition with median filter is proposed for background subtraction. GMM is an adaptive model which uses a mixture of normal distributions to model a multi-modal background image sequences [11]–[13]. Each surface which comes into the view of a given pixel is represented by one of set of states $k \in \{1.2....K\}$, where the number of surfaces k is an assumed constant. The process which generates the state at each frame time $t = 1, 2, 3...$. is simply modeled by a set of k parameters $\omega_k = p(k), k \in \{1, 2, ..., K\}$ each representing the priori probability of surface k appearing in the pixel view, and $\sum_k^K \omega_k = 1$. The density parameter set is defined as $\theta_k = \{\mu_k, \sigma_k\}$ for a given k and total set of parameters becomes $\phi = \{w_1, ..., w_k, \theta_1, ..., \theta_k\}$. The pixel value process X is assumed to be modelled by a mixture of K Gaussian densities with parameters sets, one for each state k :

$$f_{X/k}(X/K, \theta_k) = \frac{1}{(2\pi)^{\frac{n}{2}} |\sum_k|^{\frac{1}{2}}} e^{-\frac{1}{2}(X-\mu_k)^T \sum_k^{-1}(X-\mu_k)} \quad (1)$$

The first step in this is estimating current state. The k which maximizes the value $w_k f_{X/k}(X/k, \theta_k)$ gives the current state.

$$\hat{k} = \arg \max_k w_k f_{X/k}(X/k, \theta_k) \quad (2)$$

To get fast approximation, variance is used to estimate current state. To estimate the current state a *match* defined as, a pixel value falling with in $\lambda = 2.5$ standard deviation of the mean of one of the Gaussian distributions. Next step is the estimation of the foreground objects. It is started by ranking the K states by a criterion w_k/σ_k. In general, background has more probability and less variance. To estimate background, a prior probability T is provided. The first B of the ranked states whose accumulate probability accounts for T are deemed to be background,

$$B = \arg \min_b (\sum_{k=1}^{b} w_k > T) \quad (3)$$

and the rest of of the states are by default foreground. The parametric equations are updated as

$$\omega_{k,t} = (1 - \alpha)\omega_{k,t-1} + \alpha(M_{k,t}) \quad (4)$$

$$\mu_t = (1 - \rho)\mu_{t-1} + \rho X_t \quad (5)$$

$$\sigma_t^2 = (1 - \rho)\sigma_{t-1}^2 + \rho(X_t - \mu_t)^T(X - \mu_t) \quad (6)$$

$$\rho = \alpha\eta(X_t/\mu_k, \sigma_k) \quad (7)$$

Where α is the learning rate constant which controls the speed of updating the density parameters. $M_{k,t} = 1$ for matched state and for unmatched states, this value will be zero.

Median filtering: After background subtraction, some noises may present due to bad background subtraction. Median filter is used to remove such type of noises.

B. Moving Target Classification

Moving objects obtained from background subtraction can be classified as human, vehicle and background clutter. A classification metric operator $ID(x)$ *i.e.,* notion of temporal consistency is used for classification [1]. The metric is based on the knowledge that humans are, in general, smaller than vehicles. In this, dispersedness is taken as classification metric and is given by

$$Dispersedness = \frac{Perimeter^2}{Area} \quad (8)$$

Figure 2 shows some typical dispersedness values for human and vehicles. The first step in this is to record all N_n potential targets $P_n(i) = R_n(i)$. These regions are classified according
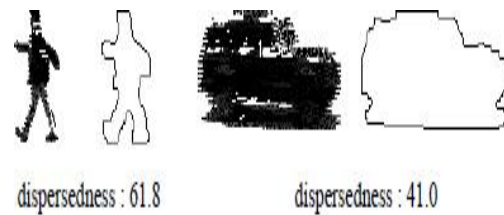

Fig. 2. Typical dispersedness values of human and vehicle [1]

to the classification metric and the result is recorded as classification hypothesis $\chi(i)$ for each one.

$$\chi(i) = \{ID(P_n(i))\} \qquad (9)$$

Each one of these potential targets must be observed in subsequent frames. Each previous motion region $P_{n-1}(i)$ is matched to the spatially closest current motion region $R_n(j)$. Any previous potential targets have not been matched to current regions removed from the list, and any current motion regions R_n which have not been matched are added to the list. At each frame, all these intermediate classifications are used to update the classification hypothesis. After this, a simple application of MLE is employed to classification. Figure 3 shows the some examples of moving target classification.

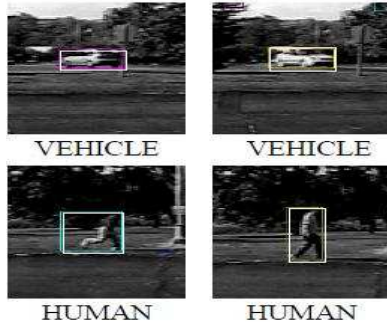$$\chi(i) = \{\chi(i) \cup ID(P_n(i))\} \qquad (10)$$

Fig. 3. Examples of moving target classification [1]

C. Feature Extraction

Feature selection is the crucial step in gait recognition. The feature must be robust to operating conditions and should yield good discriminability across individuals. In our proposed method, combination of width vector of outer contour of the the binary silhouette and MPEG-7 region-based shape descriptors (ART coefficients) are used as the feature vector. After getting feature vectors, Principal Component Analysis (PCA) is applied to reduce the dimensionality of feature vectors. Advantages using region based shape descriptors are:

1) Sometimes during the process of background subtraction object may be split into disconnected sub-regions. Such object can still be retrieved.
2) The descriptor is robust to noise which is presents due poor background subtraction, for example salt and pepper noise.

The step by step procedure involved in feature extraction process is described below:

1) Width of the outer contour of the binary silhouette: Distance between left and right extremities of the silhouette gives the width vector. From the binarized silhouettes, the left and right boundaries are traced. Width calculation is shown in Fig. 4. The width along a given row is simply the difference between leftmost and rightmost boundary pixels in that row

[10]. Figure 5 shows the width vector of a person for one frame. Here, x axis denotes the row index and y axis denotes the width associated with that row.

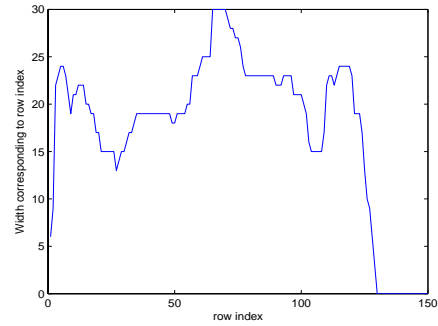Fig. 4. Width vector extraction [10]

Fig. 5. Width vector for one frame

2) ART coefficients: MPEG-7 region-based descriptors used to represent shapes. This descriptor takes into account all pixels constituting the shape, that is both the boundary and interior pixels. The descriptor works by decomposing the shape in to 2-D basis functions (complex-valued), defined by the Angular Radial Transform (ART) [4]. The normalized magnitude of coefficients are used to describe the shape. The ART coefficients are defined by:

$$F_{nm} = \langle V_{nm}(\rho, \theta), f(\rho, \theta) \rangle = \int_0^{2\pi} \int_0^1 V_{nm}^*(\rho, \theta) f(\rho, \theta)$$
$$(11)$$

Where F_{nm} is an ART coefficient of order n and m, $f(\rho, \theta)$ is an image function in polar coordinates and $V_{nm}(\rho, \theta)$ is the ART basis function that are separable along the angular and radial direction, that is

$$V_{nm}(\rho, \theta) = A_m(\theta) R_n(\rho) \qquad (12)$$

In order to achieve rotation invariance, an exponential function is used for the angular basis function,

$$A_m(\theta) = \frac{1}{2\pi} \exp(jm\theta) \qquad (13)$$

The radial basis function is defined by a cosine function,

$$R_n(\rho) = 1, n = 0 \qquad (14)$$

$$R_n(\rho) = 2\cos(\pi n\rho), n \neq 0 \qquad (15)$$

The ART descriptor is defined as a set of normalized magnitudes of complex ART coefficients. Rotational invariance is obtained by using the magnitude of the coefficients. Twelve angular and three radial functions are used ($n < 3, m < 12$). Total thirty six coefficients obtained to represent particular shape. ART coefficients are normalized by dividing with magnitude of ART coefficient of order $n = 0, m = 0$.

3) Principal component analysis (PCA): After getting above two features, PCA is applied (separately) to reduce dimensionality. After reducing the dimensionality, these two features are combined. Here, c number of classes are used for training and each class represents the sequence of feature vectors of one gait sequence of a person (subject). All sequences of one person can be added for training. Let $W_{i,j}$ be the jth feature vector of class i and N_i is the number of such feature vectors in that class. The whole feature vector set is given by $[W_{1,1}, W_{1,2}, ..., W_{1,N_1}, W_{2,1}, ..., W_{c,N_c}]$ and the total number of feature vectors are $N_t = N_1 + N_2 + ... + N_c$. The mean and covariance of feature vector set are given as

$$m_d = \frac{1}{N_t} \sum_{i=1}^{c} \sum_{j=1}^{N_i} W_{i,j} \qquad (16)$$

$$\sum = \frac{1}{N_t} \sum_{i=1}^{c} \sum_{j=1}^{N_i} (W_{i,j} - m_d)(W_{i,j} - m_d)^T \qquad (17)$$

From the covariance matrix, we can calculate N eigen values $\lambda_1, \lambda_2, ..., \lambda_N$ and associated eigen vectors $e_1, e_2, ..., e_N$. Generally first few eigen vectors represents maximum variation in patterns. So, those eigen vectors which correspond to small variation can be neglected. This can be achieved by using the following relation

$$w_k = \sum_{i=1}^{k} \lambda_i / \sum_{i=1}^{N} \lambda_i > T_s \qquad (18)$$

w_k is the accumulated variance of the first k largest eigen vectors. T_s is the threshold value and it is chosen as 0.95. By taking only k eigen vectors, the dimensionality is reduced. The new feature vector set is given by

$$P_{i,j} = [e_1, e_2, ..., e_k]^T W_{i,j} \qquad (19)$$

D. Classification/Recognition

After getting the feature vectors, next step is classification/recognition. In this paper, HMM-based recognition is used. For each person, one HMM model is developed at training stage [2]. The primary HMM parameters used are number of states N, initial probability (π), the transition probability (A), output probability (B). From the training sequence feature vectors HMM model is estimated as $\lambda = (A, B, \pi)$ [5]. The parameters of HMM are described below:

1) N, the number states in the model. Number of states used in this are N=5. HMM states are represented as $S = \{s_1, s_2, ..., s_N\}$.
2) M, the number of distinct symbols (observation features) per each state. Each feature vector is treated as one observation symbol. The number M depends on the number of frames per cycle, the number of states in the

model and how to divide one cycle in to clusters. The observation symbols for one HMM state are denoted as $V = \{v_1, v_2, ..., v_M\}$.

3) A, the transition probability matrix. $A = \{a_{ij}\}$, a_{ij} is defined as

$$a_{ij} = P[q_{t+1} = S_j | q_t = S_i], 1 \le i, j \le N \qquad (20)$$

where q_t is the state at time t. In this, left to right model is chosen, which allows the transition from j^{th} state to either j^{th} or the $(j + 1)^{th}$ state.

4) B, the observation symbol probability matrix. $B = \{b_j(k)\}$, where

$$b_j(k) = P[v_k att | q_t = S_j], 1 \le j \le N, 1 \le k \le M \qquad (21)$$

5) π, the initial probability. $\pi = \{\pi_i\}$, where

$$\pi_i = P[q_1 = S_i], 1 \le i \le N \qquad (22)$$

First frame is assigned to first state only, so the initial probability π_1 is set be 1 and remaining π_i are set to be zero.

The complete parameter set of the HMM can be denoted as

$$\lambda = (A, B, \pi) \qquad (23)$$

- **Initial estimate of HMM parameters**: Each training sequence $\chi = \{f_1, f_2, ..., f_T\}$ is divided in to cycles. Gait cycle is defined as person starts from rest, left foot forward, rest, right foot forward, rest. Figure 6 shows the walking sequence during gait cycle .

Fig. 6.   Walking sequence during a gait cycle

Gait cycle is determined by calculating sum of the foreground pixels. At rest positions this value is low. By calculating number of frames between two rest positions, gait cycle (period) is estimated [3]. Figure 7 shows the sum some of the foreground pixels of two persons. x axis denotes frame index, y axis denotes sum of foreground pixels. Here, valley points represents rest position.
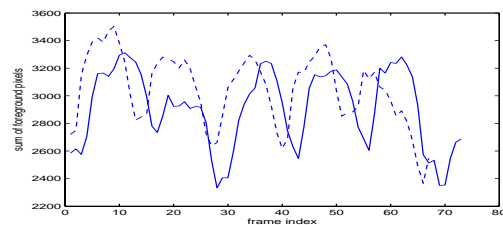

Fig. 7.   Sum of foreground pixels to estimate gait cycle (gait period)

During gait cycle, every person transits across some phases or stances. These stances are used as states. N exemplars (*i.e.,* stances) $\varepsilon = \{e_1, e_2, ..., e_N\}$ are picked up from a pool training features to get states. Exemplars are show in Fig. 8.

$$e1 \quad e2 \quad e3 \quad e4 \quad e5$$

Fig. 8.   Exemplars (stances) during a gait cycle (gait period)

Each gait cycle again divided in to N temporally adjacent clusters of approximately equal size. By taking mean of all feature vectors in n^{th} cluster, the state or exemplar e_n is estimated. Estimate of the states is given by

$$e_n = \frac{1}{N} \sum_{f_t \in C_n} f_t \qquad (24)$$

Where f_t is the feature vector of tth frame, C_n represents the nth cluster, N_n is the number frames in the nth cluster. The states are $\varepsilon = \{e_1, e_2, ..., e_N\}$. The observation probability for each feature vector is obtained by calculating the distance between the exemplars and the feature vector. The observation probability matrix $B = \{b_j(k)\}$ is defined as

$$b_n(f_t) = 0.01 \delta_n e^{-\delta_n XD(f_t, e_n)} \qquad (25)$$

where,

$$\delta_n = \frac{N_n}{\sum_{f_t \in C_n} D(f_t, e_n)} \qquad (26)$$

The parameter δ_n can reflect how much a feature vector belongs to the cluster. Distance D is measured by inner product distance. Inner product distance is defined as

$$D(f, e) = 1 - \frac{f^T e}{\sqrt{f^t f e^T e}} \qquad (27)$$

Initial estimate of transition probabilities has done by trial and error method. In this, initial estimate of transition probabilities are $A_{j,j} = A_{j,j+1} = 0.5$, other $A_{j,K}=0$ and $A_{N,N}=1$, where $k \neq j + 1$. Initial probability for first state is $\pi_1=1$ and for remaining states it is zero.

- **HMM training**: Training is carried out in two steps. For the incoming training feature vectors, Viterbi algorithm is used. Using the current values of exemplars ε^i, transition matrix A^i, Viterbi decoding on the sequence χ yields the most probable path $Q = \{q_1^i, q_2^i, ..., q_T^i\}$. In this, q_t^i is the estimated state at time t and iteration i. After estimating most probable state sequence, feature vectors are clustered according to the most likely state of origination.

States are estimated by calculating the mean of feature vectors belongs to particular cluster (state) and is given as

$$e_n^{i+1} = \frac{1}{N} \sum_{f_t \in C_n} f_t \qquad (28)$$

The transition probability is estimated as

$$a_{ij} = \frac{\left| \{l = 1, ...L : q_k^l = s_i \, and \, q_{k+1}^l = s_j\} \right|}{\sum_{l=1}^{L} (k_l - 1)} \qquad (29)$$

Where, numerator denotes number of transitions from state i to state j and denominator denotes number of feature vectors in that state (cluster) i. Observation probability and δ_n for next iteration is calculated using the equations (25) and (26). Initial probabilities π_i are same as in previous iteration. These estimated parameters are used for next iteration. After few iterations, acceptable HMM model $\lambda = (A, B, \pi)$ is obtained. For each of the persons, one HMM model is estimated.

- **Recognition**: Test sequence of unknown person y is processed in the same way as training sequences. Subsequently, Log probabilities between test sequence and models in the gallery (data base) is calculated using the Viterbi algorithm. The person corresponding to the model yielding the higher probability is considered as the identified individual, and is given as

$$identity(y) = \arg\max_j P\left(\tilde{f}_y / \lambda_j\right), j = 1, 2, ..., N. \qquad (30)$$

Here, N is the number of subjects (persons) in the database, $\tilde{f}_y$ are the feature vectors (test sequence) of unknown person y.

III. EXPERIMENTAL RESULTS

In our method, the first step is background subtraction. Gaussian mixture model in addition with median filter is used for background subtraction. Number of mixtures used in this model are three. Proposed method is tested on three different databases such as, PETS'2001 data set, a surveillance video collected in Brisbane railway station, Australia and the videos collected for different persons in IIT Guwahati, India. It is to be mentioned that, PETS'2001 data set contains more complex scenes with clouds motion, shadows and illumination changes. It has a frame rate of 30 frames per second and with frame size of 768 x 576. The surveillance video has the frame rate of 18 frames per second and frame size of 704 x 576. In this video, illumination variations are quite significant. Videos collected in our institute are having a frame size of 480 x 640. Median filter is used to remove noises obtained from background subtraction. Figures 9-11 shows background subtraction results. First row shows input video frames, second row shows background subtracted frames and third row shows the background subtraction results after median filtering. Experimental results demonstrate the effectiveness of proposed method.

Next step in the proposed system is moving target classification. After background subtraction all moving objects come into picture. Moving target classification is used to

classify humans from other moving objects which are obtained from background subtraction. Using shape information, one classification metric is defined to classify the moving objects. In this, dispersedness is taken as the classification metric. Generally humans have more dispersedness value. Figures 12-13 show the moving target classification results.
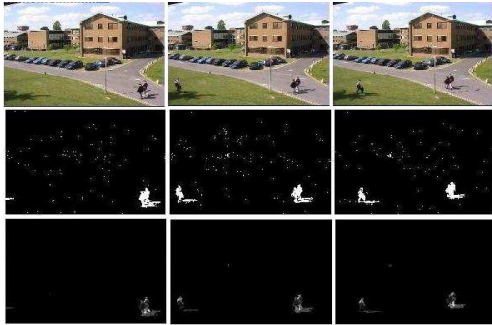

Fig. 9.   Background subtraction results. The first row shows input video frames (Pets2001: *http://www.cvg.cs.rdg.ac.uk/PETS2001/pets2001-dataset.html*). Second row shows background subtracted frames. Third row shows the background subtraction result after median filter.

Fig. 10.   Background subtraction results. The first row shows input video frames (at Brisbane railway station). Second row shows background subtracted frames. Third row shows the background subtraction result after median filter.
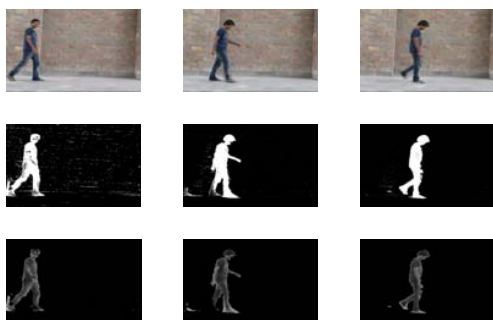

Fig. 11.   Background subtraction results (collected at IIT Guwahati campus). The first row shows input video frames. Second row shows background subtracted frames. Third row shows the background subtraction result after median filter.

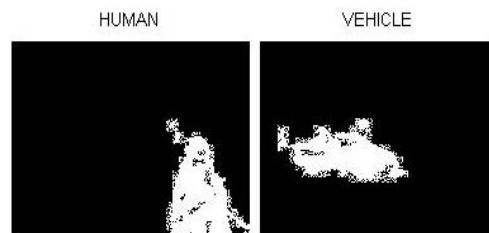Fig. 12.   Input video frame used to show classification results.

Fig. 13.   Moving target classification showing human and the vehicle.

Next step is the classification/recognition of the person. As discussed earlier, features used in this are combination of width vector of the outer contour of the binary silhouette and ART coefficients. Figures 14-15 show the binary silhouettes of different persons of the gait database. Figures 16-17 show the width profiles of two persons. Here, x denotes the frame index, y denotes the index of the width vector (row index). Brightness parts represents maximum width regions such as hands swing, distance between feet. Each sequence is divided into gait cycles and subsequently, each gait cycle is considered as a training sequence. Hidden Markov Model (HMM) is used for recognition. During gait cycle, each person comes into the view of stances (states). This information is used to estimate the states in HMM. Then, each cycle again divided in to N temporally adjacent clusters of approximately equal size. By taking mean of all feature vectors in n^{th} cluster, the state or exemplar e_n is estimated. Number of states used are five. In this paper, gait database of the Institute of Automation, Chinese Academy of Sciences (CASIA) and the database collected at our institute (IIT Guwahati, India.) are used for evaluation of classification algorithm. All these sequences were taken in a single camera-based setup and the walking surface is a plane ground/floor. CASIA database has nine people and each person has sixteen gait cycles and each gait cycle is considered as one sequence. Half of the gait cycles are used for training and the remaining half are used for testing. Our institute database has six persons and each person has fifteen gait cycles. Eight cycles are used for training and the rest seven cycles used for testing. For CASIA database, width vector length is one hundred and fifty. By concatenating the ART coefficients, it becomes one hundred and eighty five and by using PCA, it is reduced to one hundred and sixty seven. For our institute database, each width vector length is two hundred and twenty. After adding ART coefficients, it becomes

two hundred and fifty five and the PCA reduces it to one hundred and eighty five. Some of the input sequences of both databases are shown in the Figures 14-15. First, the recognition process is carried out only by considering the width vector and then, the combination of both width as well as ART coefficients. Experimental results show the improvement in recognition rate of the proposed system. Tables 1 and Table 3 show the recognition rate by considering only the width as the feature vector. It clears that recognition rate is low for width feature vector. For disjoint silhouettes, width alone may not discriminate the binary silhouettes of different persons. In such cases, our proposed feature vector (combination of both width and ART coefficients) approximate the silhouettes in a better way. Table 2 and Table 4 show the recognition rate by considering combination of both width and ART coefficients. Table 5 and Table 6 show the average recognition rate for both the databases. From all these results, it is quite evident that our proposed system gives a good recognition rate.
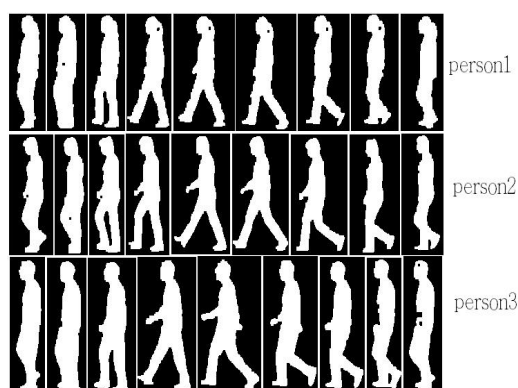

Fig. 14.   Binary silhouettes of different persons (CASIA database).
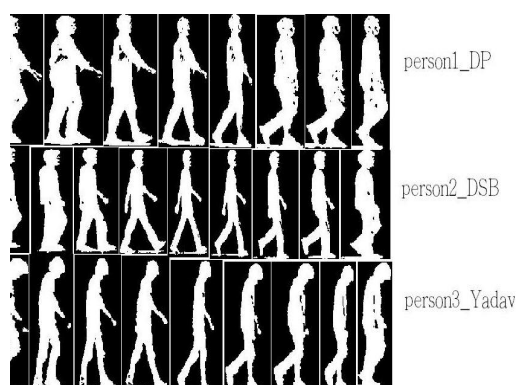

Fig. 15.   Binary silhouettes of different persons (our institute database).
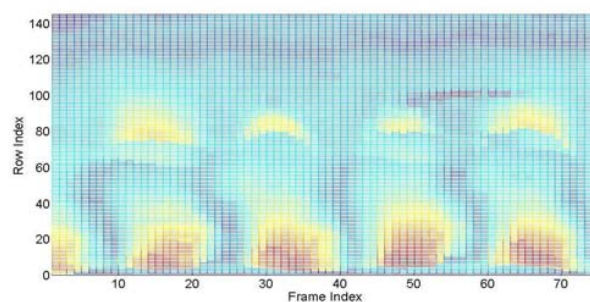

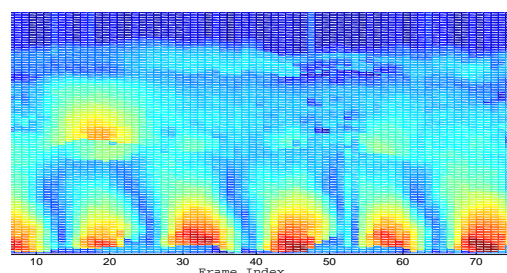Fig. 16.   Width vector profile of person #1.

Fig. 17.   Width vector profile of person #2.

TABLE I
WIDTH VECTOR BASED CLASSIFICATION RESULTS FOR CASIA DATABASE

test sequence	Recognition									Accuracy	error rate
	fyc	hy	ljq	lqf	ls	ml	nhz	rj	syj		
fyc	7	0	0	0	0	0	0	1	0	87.5	12.5
hy	0	7	0	0	0	0	0	1	0	87.5	12.5
ljq	0	0	5	0	0	1	2	0	0	62.5	37.5
lqf	0	0	1	5	0	1	1	0	0	62.5	37.5
lsl	0	0	0	0	8	0	0	0	0	100	0
ml	0	0	1	1	0	5	1	0	0	62.5	37.5
nhz	0	0	0	0	0	0	8	0	0	100	0
rj	2	0	0	0	0	0	0	6	0	75	25
syj	0	0	0	0	0	0	0	0	8	100	0

TABLE II
WIDTH VECTOR AND ART COEFFICIENT BASED CLASSIFICATION RESULTS FOR CASIA DATABASE

test sequence	Recognition									Accuracy	error rate
	fyc	hy	ljq	lqf	ls	ml	nhz	rj	syj		
fyc	8	0	0	0	0	0	0	0	0	100	0
hy	0	7	0	0	0	0	0	1	0	87.5	12.5
ljq	0	7	0	0	0	0	0	1	0	87.5	12.5
lqf	0	7	0	0	0	0	0	1	0	87.5	12.5
lsl	0	0	0	0	8	0	0	0	0	100	0
ml	0	7	0	0	0	0	0	1	0	87.5	12.5
nhz	0	0	0	0	0	0	8	0	0	100	0
rj	2	0	0	0	0	0	0	6	0	75	25
syj	0	0	0	0	0	0	0	0	8	100	0

TABLE III

WIDTH VECTOR BASED CLASSIFICATION RESULTS FOR OUR INSTITUTE DATABASE

test sequence	Recognition						Accuracy	error rate
	DP	SB	MH	KK	SY	CS		
DP	6	0	0	1	0	0	85.7	14.3
SB	0	5	2	0	0	0	71.4	28.6
MH	0	1	5	0	0	1	71.4	28.6
KK	0	0	0	7	0		100	0
SY	0	0	0	0	7	0	100	0
CS	0	1	1	0	0	5	71.4	28.6

TABLE IV

WIDTH VECTOR AND ART COEFFICIENT BASED CLASSIFICATION RESULTS FOR OUR INSTITUTE DATABASE

testsequence	Recognition						Accuracy	error rate
	DP	SB	MH	KK	SY	CS		
DP	6	0	0	1	0	0	85.7	14.3
SB	6	0	0	1	0	0	85.7	14.3
MH	6	0	0	1	0	0	85.7	14.3
KK	0	0	0	7	0		100	0
SY	0	0	0	0	7	0	100	0
CS	6	0	0	1	0	0	85.7	14.3

TABLE V

CLASSIFICATION RESULTS IN TERMS OF AVERAGE RECOGNITION FOR CASIA DATABASE

Feature type	attempts	success	Recognition rate
Width of the contour	72	59	81.9
width+ART coefficients	72	66	91.7

TABLE VI

CLASSIFICATION RESULTS IN TERMS OF AVERAGE RECOGNITION FOR OUR INSTITUTE DATABASE

Feature type	attempts	success	Recognition rate
Width of the contour	42	35	83.3
width+ART coefficients	42	38	90.4

IV. CONCLUSIONS

In this paper, Gaussian mixture model in addition to median filtering steps have been used for better background substraction. MPEG-7 ART coefficients are concatenated with the width vector of the outer contour of the binary silhouette, which is taken as the feature vector for HMM-based recogntion. Using the shape descriptors (ART coefficients), disjoint regions can also be represented. Subsequently, PCA is used to reduce data dimensionality. All these experiments are carried out using single camera-based set up only. The overall performance of the proposed system can be significantly improved by using multiple camera-based setup. Using single camera based setup, it is quite difficult to identify the person in occlusion conditions. Even though proposed system gives satisfactory result, it has to be improved for a very large database and also for the conditions under occlusion.

REFERENCES

[1] H. F. Alan J. Lipton and R. S. Patil, "Moving target classification and tracking from real-time video," *Proceedings of Fourth IEEE Workshop on Applications of Computer Vision (WACV'98)*.

[2] A. N. R. N. P. C. A. K. R.-C. V. K. Amit Kale, Aravind Sundaresan and R. Chellappa, "Identification of humans using gait," *IEEE Transactions on Image Processing*, vol. 13, no. 9.

[3] A. K. R.-C. Aravind Sundaresan and R. Chellappa, "A hidden markov model based framework for recognition of humans from gait sequences," *Proceeding of ICIP*.

[4] M. Bober, "Mpeg-7 visual shape descriptors," *IEEE Transactions on Circuit and Systems for Video Technology*, vol. 11, no. 6.

[5] H. Z. Changhong Chen, Jimin Liang and H. HU, "Gait recognition using hidden markov model," *Lecture Notes on Computer Science*.

[6] P. S. Huang, "Automatic gait recognition via statistical approaches for extended template features," *IEEE Transaction on Systems, Man, and Cybernetics-Part B: Cybernetics*, vol. 31, no. 5, October 2001.

[7] L. Lee and W. E. L. Grimson, "Gait analysis for recognition and classification," *Proceedings of Fifth IEEE International Conference on Automatic Face and Gesture Recognition.*, 2002.

[8] W. H. Liang Wang, Tieniu Tan and H. Ning, "Automatic gait recognition based on statistical shape analysis," *IEEE Transaction on Image processing*, vol. 12, no. 9, September 2003.

[9] ——, "Silhouette analysis-based gait recognition for human identification," *IEEE Transaction on Image processing*, vol. 25, no. 12, December 2003.

[10] D. H. Nikolaos V. Boulgouris and K. N. Plataniotis, "Gait recognition: A challenging signal processing technology for biometric identification," *IEEE sinal processing magazine*, November 2005.

[11] M. Piccardi, "Background subtraction techniques: A review," *IEEE International Conference on Systems, Man and Cybernetics*.

[12] P. W. Power and J. A. Schoonees, "Understanding background mixture models for foreground segmentation," *Proceedings of IEEE International Conference on Image and Vision Computing*.

[13] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time video tracking," *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 82, pp. 246–252.