

# Gait Recognition by Cross Wavelet Transform and Graph Model

Sagar Arun More and Pramod Jagan Deore

**Abstract**—In this paper, a multi-view gait based human recognition system using the fusion of two kinds of features is proposed. We use cross wavelet transform to extract dynamic feature and bipartite graph model to extract static feature which are coefficients of quadrature mirror filter (QMF)-graph wavelet filter bank. Feature fusion is done after normalization. For normalization of features, min-max rule is used and mean-variance method is used to find weights for normalized features. Euclidean distance between each feature vector and center of the cluster which is obtained by k-means clustering is used as similarity measure in Bayesian framework. Experiments performed on widely used CASIA B gait database show that, the fusion of these two feature sets preserve discriminant information. We report 99.90 % average recognition rate.

**Index Terms**—Binary sequences, feature extraction, identification of persons, linear discriminant analysis (LDA).

## I. INTRODUCTION

RECOGNIZING human from a distance by gait biometric has attracted researchers in recent years. It has many advantages like; non-invasive, less obscured, unobtrusive, without subject cooperation, ability to work from a distance and with low quality video. Comparatively, gait biometric is newer modality than face, iris and fingerprint. Recognizing someone from a certain distance, where no fine details are available is very difficult task. In such case, gait biometrics in which the person gets recognized by the manner of walking only is useful. Gait is a potential biometric trait where unconstrained person identification is demanded. It is a protocol free biometric technique which does not require willingness of person and hence found application in surveillance. However, commonly used biometric recognition systems usually operate in constrained acquisition scenarios and under rigid protocols. The finger print, iris and face recognition could not be the right choice in the unconstrained environment, where distant data capture is required. Comparatively, gait comprised of motion trajectories of various body parts, have a potential to get captured properly from relatively far distance. It does not need systematic data capture process, where subjects should

necessarily be informed. This makes the identification process protocol free. The extended application of gait recognition can be suspect identification in a sensitive area where security is at the highest priority. These characteristics of gait biometrics lead it to be an attractive modality from the perspective of human recognition from a distance.

In spite of various advantages, co-variate factors like; walking speed, carrying conditions, clothing, the surface of walking, fatigue, drunkenness, pregnancy, injury to feet and the psycho-somatic condition affect the normal walking style. View angle also plays vital role while testing such system. It may be possible that certain view angle provides discriminant information of walking individual while another may not. Hence, an investigation is needed to find robust gait representation which can cope up with these challenges in multi-view scenario.

The main contribution of this paper is the achievement of significant recognition rate in co-variate conditions like carrying bag and cloth variation. There is no need to segment bag from subject to remove co-variate. It does not need any complex model to extract static or dynamic features. This scheme is simple, as it does not need color and texture information of the sequences and innovative in the sense that, the application of cross wavelet transform and graph model is not proposed in fusion approach yet.

Rest of the paper is organized as follows. Section II briefs existing methods of gait recognition. In Section III, proposed method is discussed. Section IV explores feature extraction and feature fusion along with training and testing of the system in details. Experimental results are discussed in Section V followed by conclusion in Section VI.

## II. LITERATURE OVERVIEW

Gait recognition and analysis have been studied heavily in recent past. In this section we will discuss it briefly. The approaches in the literature can be broadly classified into two types viz. model free [1]–[4] and model-based [5]–[8]. In these approaches, various static and dynamic features of gait sequences were extracted by using shape analysis, image geometry transformations, wavelet analysis, so on and so forth. The model free approach extracts features directly from the image plane. Whereas, the model based approach models the human gait and then extracts model parameters as features. In [1], procrustes shape analysis is used to represent gait signature, which is obtained by extracting the mean shape of the unwrapped silhouette. Whereas, [2] is a 3D approach for gait recognition, which constructs 3D silhouette vector of 2D

Manuscript received February 9, 2017; accepted October 12, 2017. Recommended by Associate Editor Wei He. (Corresponding author: Sagar Arun More.)

Citation: S. A. More and P. J. Deore, "Gait recognition by cross wavelet transform and graph model," *IEEE/CAA J. of Autom. Sinica*, vol. 5, no. 3, pp. 718–726, May 2018.

S. A. More and P. J. Deore are with the Department of Electronics and Telecommunication Engineering, R. C. Patel Institute of Technology, Shirpur 425405, India (e-mail: sagar.more27@gmail.com; pjdeore@yahoo.com).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/JAS.2018.7511081

scene by using stereo vision method. In a recent work [4], complete canonical correlation analysis is used to compute correlation between two gait energy image (GEI) features. In another recent paper [3], authors extract different width vectors and combined them to construct gait signature. This feature is then approximated by radial basis function (RBF) network for recognition.

In an earlier model based approach [5], gait pattern was detected in XYT spatio-temporal volume. The bounding contour of walking person is found by the snake. Furthermore a 5-stick human model is controlled by these contours. Various angle signals are then extracted by using this model for recognition. Whereas in [6], each silhouette is first labelled manually. Various features like; area, gravity center, orientation of each body part is then calculated. In [8], the gait cycle is modelled as chain of key poses first, then features like pose energy image and pose kinematic are extracted. View invariant approaches have also been proposed in the regard of gait recognition such as [7]. In this paper, authors estimate marker less joints followed by viewpoint verification.

Either static or dynamic feature alone can perform well for recognition but with some limitations. While dealing with static features, one can not analyze dynamic features and vice-versa. Extracting both features simultaneously improves the recognition rate on the cost of increased computational complexity. Various approaches are proposed in this regard, which extract static and dynamic features simultaneously, either fusing model free and model based approaches [9]–[13] or fusing various features into a single augmented feature vector [14]–[17].

In [14], gait energy image and motion energy images are combined to form feature vector, whereas in [15], the static silhouette template (SST) and dynamic silhouette template (DST) are fused to construct dynamic static silhouette template (DSST). The position of the gravity center of human body may change because of various co-variate factors as aforementioned. This problem is addressed by [16]. In this paper, authors divide the GEI transformed image into three body parts like; head, torso and leg. Furthermore, they compute shifted energy image (SEI) features which are horizontal centres of body parts. Next, gait structural profile (GSP) extracted to capture body geometry. For this, silhouette segmented into four body parts as per the anatomical measurements like; head, torso, left and right leg. The GSP, which is the difference of gravity center of these segmented body parts and entire body is computed. These two features are then used in combination for recognition. In [17], two distinct features namely frieze pattern and wavelet coefficients are extracted. The frieze pattern preserves spatial information and wavelet coefficients preserve low frequency information. Factorial hidden Markov model (HMM) is used to combine these features and parallel HMM facilitates decision level fusion of two individual classifiers for recognition. All these approaches signify that, the fusion of multiple gait features improves the recognition system performance.

There are certain methods, which explore static and dynamic characteristics of the human body. They fuse static and dynamic features for improvement in performance of

gait recognition system. In [13], features like; centroid, arm swing, stride length, mean height were extracted from the binary silhouette. Further, they fit ellipse on each region and compute its aspect ratio and orientation. These features are then combined and transformed by discrete cosine transform (DCT) and applied to generalized regression neural network for recognition. Whereas in [9], mean shape is extracted by using Procrustes shape analysis as a static feature. The dynamic features are extracted by modelling human body parts by truncated cone, head by sphere and computing joint angles of this model. A human skeleton model is adopted in [10] to extract dynamic features and computing various angles of key body points. The static feature is denoted by wavelet descriptor, which is obtained by applying wavelet transform to the boundary-centroid distance.

In [18], HMMs are used to extract static and dynamic gait features, without using any human body model. The static features are extracted by conventional HMM and dynamic features by hierarchical HMM. After labelling, they extract three features namely; component area, component center and component orientation. First HMM represents general shape information while the second HMM extracts detailed sub-dynamic information. Whereas in [19], local binary pattern is used to denote the texture information of optical flow as the static feature. Dynamic feature is represented by HMM with Gaussian mixture model. In [11], the GEI is transformed by dual tree complex wavelet transform (DTCWT) with different scales and orientations. A two stage Gaussian mixture model denote the patch distribution of each DTCWT based gait image. Further, to model the correlation of multi-view gait feature, a sparse local discriminant canonical correlation model is used. In a recent paper [12], the dynamic feature is extracted by Lucas-Kanade based optical flow image. The mean shape of head and shoulder are then extracted by using Procrustes shape analysis, which is the static feature. The fusion is done on score level.

It can be noted that not all the aforementioned methods adopt human body model such as skeleton to extract dynamic features. Most often, authors prefer mathematical modelling, as it is efficient to extract different kinds of features and also facilitate lower computational complexity.

### III. PROPOSED GAIT RECOGNITION METHOD

This paper aims to develop a method, which is the fusion of both approaches, viz., model free and model based, without using human body model such as skeleton. It facilitates to extract static and dynamic feature sets simultaneously. Dynamic feature set is obtained by computing cross wavelet transform among dynamic body parts like hand and leg from each gait sequence. Further, to extract static feature set, the bipartite graph is used to model gait silhouette, as the graph is a powerful tool to represent an image on the basis of pixel adjacency. We apply quadrature mirror filter (QMF)-graph wavelet filter bank proposed by [20] to each gait sequence. Only analysis filter bank is used for this task. The feature vector (FV) represented by fusion of these two feature sets.

The static and dynamic feature sets extracted from all the sequences with 11 view angles and 10 co-variate conditions.

These features are combined on feature level. The centroid of clusters of these augmented feature vectors then obtained by using k-means clustering. The Euclidean distance computed between each feature vector and centroid of clusters which is linearly classified in linear discriminant analysis (LDA) space. For identification, we use Bayesian framework. These steps are depicted in Fig. 1.

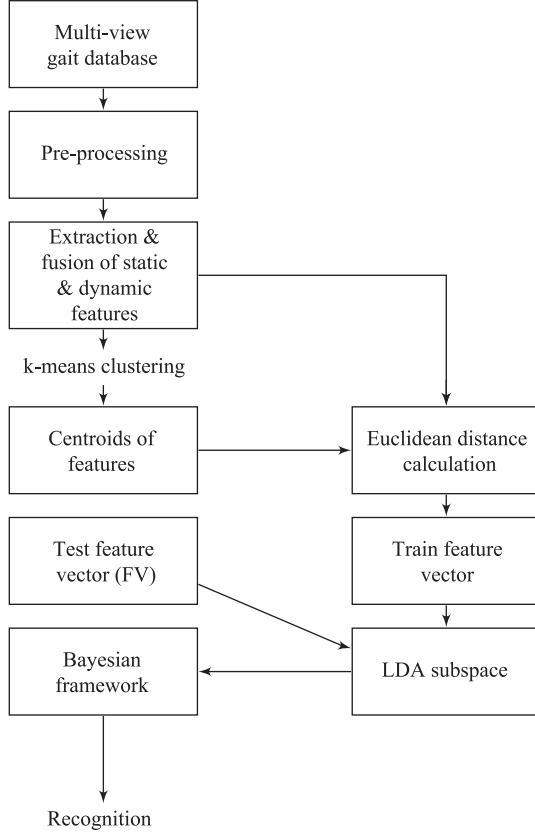


Fig. 1. Proposed method.

#### IV. FEATURE FORMATION

In this work, we use CASIA B multi-view gait database [21], which consists of 124 persons. Each person is depicted in 10 sequences with various co-variate ( $C_V$ ) like; normal/slow walking ( $nm$ -01 to  $nm$ -06), with bag ( $bg$ -01,  $bg$ -02), with coat ( $cl$ -01,  $cl$ -02). The sequences are captured at 11 different viewing angles ( $V_A$ ) ( $0^\circ, \dots, 180^\circ$ ). Table I shows the view angle and co-variate with serial numbers which we use in experiments. Thus, the database consists of  $124 \times 10 \times 11 = 13640$  gait sequences. The feature space consists of two different kinds of feature sets. We use cross wavelet transform to extract dynamic feature set and QMF-graph wavelet filter bank to compress the sequence, which is static feature set of a complete gait sequence at an arbitrary view angle and co-variate condition.

##### A. Pre-processing

The silhouettes which are available readily in CASIA B database have holes as shown in Fig. 2(a) and breaks in successive frames as shown in Figs. 2(b) and 2(c). In order to extract meaningful features, we do morphological operations like;

dilation, erosion, opening and closing. Further, it is required to divide silhouette which contains body parts like hands and legs. A bounding box technique is applied after cropping the divided silhouette from complete silhouette to get horizontal width of cropped silhouette. This width varies in each frame as hands/legs displaces. We divide entire silhouette into three equal parts, viz., the portions containing head and shoulder, hands, legs. We processed only the portions containing hand and legs by first cropping and then applying bounding box on both the portions separately. The width of the bounding box is saved as 1D width vector. Here, in this work we consider the movement of hand and leg only for the computation of dynamic feature using cross wavelet transform.

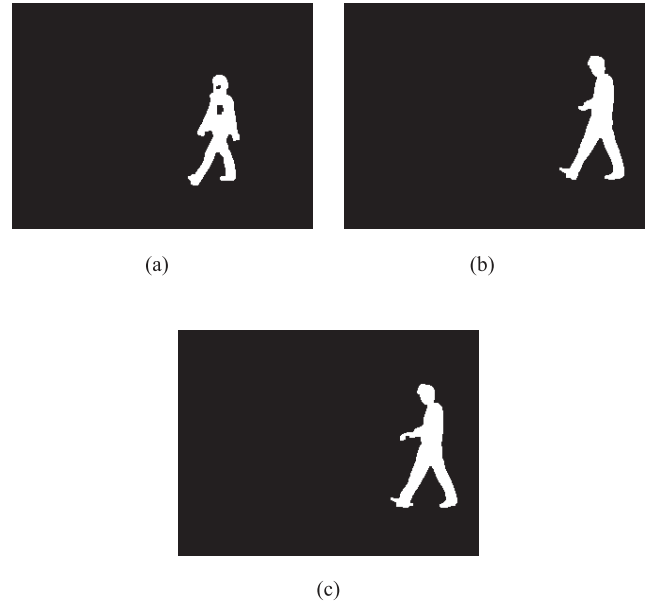


Fig. 2. Inferior silhouette. (a) Hole in silhouette  $t$ . (b) Break in silhouette  $t-1$ . (c) Break in silhouette  $t$ .

Let  $X_n$  and  $Y_n$  be the 1D signals generated due to dynamic movement of hands and legs respectively for  $n$  sequences. Then, we can write

$$X_n = \{x_1, x_2, x_3, \dots, x_t\} \quad (1)$$

$$Y_n = \{y_1, y_2, y_3, \dots, y_t\} \quad (2)$$

where  $x_t$  and  $y_t$  are the width vector computed from bounding box and  $t$  is number of frames in a sequence. The width of the bounding box is saved as 1D width vector as shown in Fig. 3.

##### B. Dynamic Feature Extraction ( $FV_{dynamic}$ )

After pre-processing, the width vectors extracted from each silhouette which represents the dynamic movements of hands and legs throughout the entire gait sequence. Cross wavelet transform then applied to these 1D signals. Morlet wavelet (with  $\omega_0 = 6$ ) is used as it better suits for such nature of signals with regard to time and frequency localization.

$$\psi_0(\eta) = \pi^{-\frac{1}{4}} e^{i\omega_0\eta} e^{-\frac{\eta^2}{2}} \quad (3)$$

where  $\omega_0$  is dimensionless frequency and  $\eta$  is dimensionless time.

TABLE I  
CO-VARIATES AND VIEW ANGLES IN CASIA B DATABASE

Serial number	1	2	3	4	5	6	7	8	9	10	11
$C_V$	<i>nm</i>	<i>nm</i>	<i>nm</i>	<i>nm</i>	<i>nm</i>	<i>nm</i>	<i>bg</i>	<i>bg</i>	<i>cl</i>	<i>cl</i>	-
	01	02	03	04	05	06	01	02	01	02	-
$V_A$	0°	18°	36°	54°	72°	90°	108°	126°	144°	162°	180°

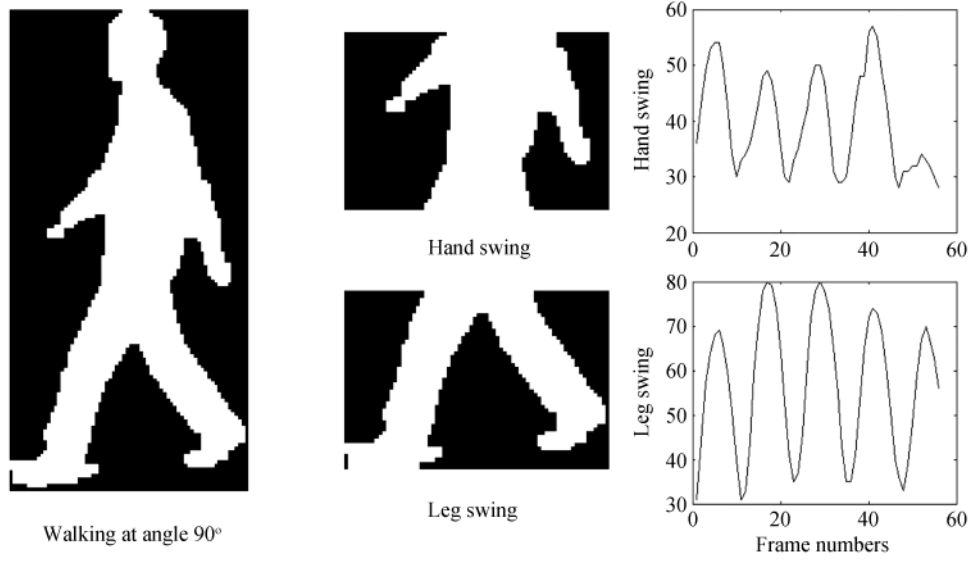


Fig. 3. 1D signals extraction.

1) *Cross Wavelet Transform*: The cross wavelet transform is defined over two time series, which reveal an area of common higher power and relative phase in the time-frequency domain. Cross wavelet transform of two time signals  $X_n$  and  $Y_n$  is expressed as [22]:

$$W(X_n, Y_n) = W(X_n) \cdot W(Y_n)^* \quad (4)$$

where  $W(X_n)$  is continuous wavelet transform and  $*$  is complex conjugation. The cross wavelet power can be defined as

$$W_p = |W(X_n, Y_n)|. \quad (5)$$

The local relative phase between  $X_n$  and  $Y_n$  can be expressed as complex argument

$$\Phi_n = \arg(W(X_n, Y_n)). \quad (6)$$

The complete representation of wavelet cross spectrum is

$$W(X_n, Y_n) = |W(X_n, Y_n)|e^{i\phi_n} \quad (7)$$

where  $\Phi_n$  is the phase at time  $t_n$ .

2) *Wavelet Coherence*: More significant coherence between two continuous wavelet transformed signals is found even when common power is low. This relationship is expressed as wavelet coherence (WCOH). Wavelet coherence denotes the relationship between two independent time series signals expressed in terms of area of the common frequency band at a certain time interval, across which these two signals vary simultaneously. Following [22], [23], wavelet coherence between two signals  $X_n$  and  $Y_n$  can be written as:

$$WC(X_n, Y_n) = \frac{|\zeta[W(X_n, Y_n)]|}{\sqrt{\zeta[W(X_n)] \times \zeta[W(Y_n)]}} \quad (8)$$

where  $\zeta = S \cdot s^{-1}$ ,  $S$  is smoothing parameter and  $s$  is the scale.

For demonstration purpose, we show 1D signals extracted at angle 90° with bag carrying co-variate as shown in Figs. 4–6 denote wavelet cross spectrum, wavelet coherence and phase relationship between extracted signals, respectively. These Figs. are color-coded spectrograms. They denote variations in spectral and coherence components. We choose scale ( $s$ ) up to 75 as it is optimum choice after visual inspection. The smoothing ( $S$ ) is done in both the time and scale directions to compute meaningful coherence. The time smoothing uses a filter given by the absolute value of the Morlet wavelet function at each scale and normalized to have a weight of unity. The scale smoothing is done by using a boxcar filter of a certain width which is 0.60 for Morlet wavelet [24].

The dynamic feature set includes the mean value of cross wavelet spectrum along with wavelet coherence and phase across the entire gait sequence. The length of dynamic features varies with the number of frames present in that particular sequence, hence zero padding is done to individual feature to make the feature set of fixed length.

### C. Static Feature Extraction ( $FV_{static}$ )

Various static features like; mean height, centroid, mean shape, wavelet descriptor have been presented earlier, which are already discussed in literature overview. We do not extract such features. Instead, first we represent gait silhouette by a bipartite graph and further use QMF-graph filter bank to compress the entire gait sequence. A brief about graph and QMF-graph filter bank is discussed next.

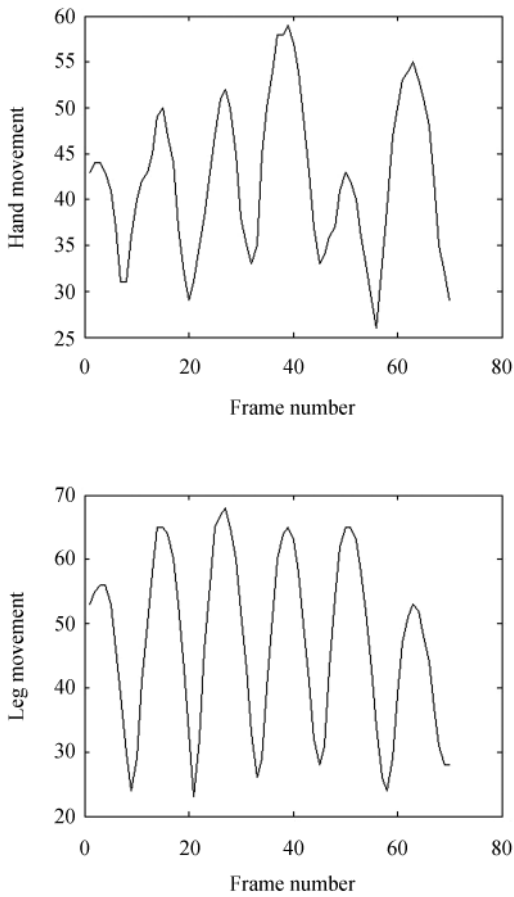


Fig. 4. 1D signals at view angle  $90^\circ$  with bag.

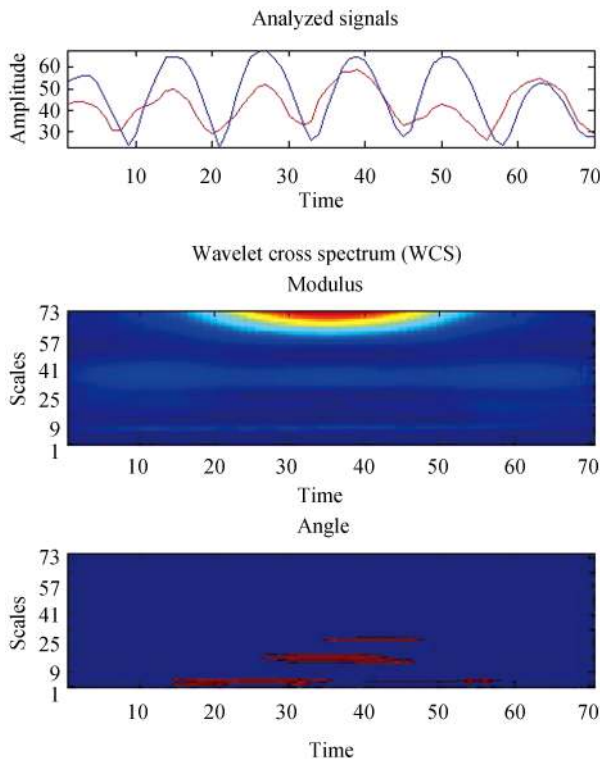


Fig. 5. Wavelet cross spectrum at view angle  $90^\circ$  with bag.

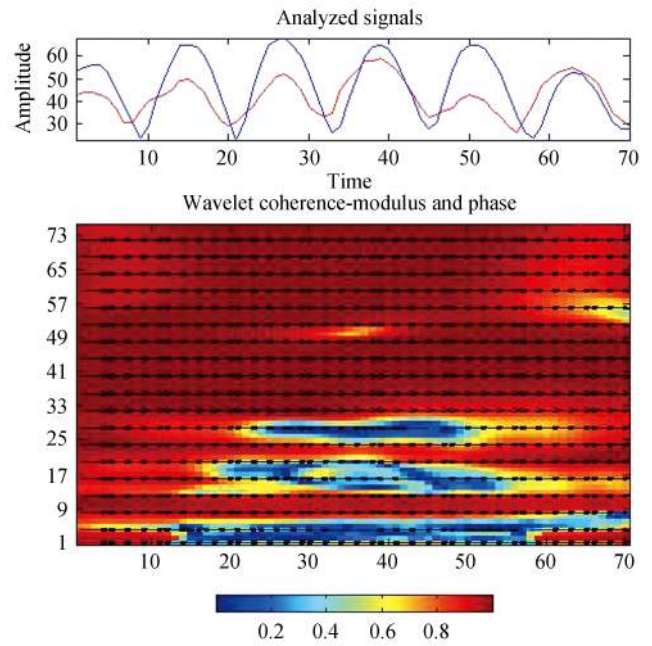


Fig. 6. Wavelet coherence at view angle  $90^\circ$  with bag.

1) *Graph Model*: A graph  $G = (V, E)$ , in which  $V$  and  $E$  are the vertex and edge, respectively, is a powerful tool for modelling image and video signals, as it offers flexibility in adjacent pixel relationship. The 2D images can be represented as graph using various pixel connectivities such as: rectangular, vertical, horizontal, diagonal, 8 connected neighbours. This flexibility leads to different down sampling patterns for filters. Various concepts from signal processing like Fourier decomposition and filtering can be extended to graph domain. These functions are called as graph signals. In this work, we use the bipartite graph to model the silhouette image. The bipartite graph is expressed as  $G = (L, H, E)$ . They are also called as two-colourable graphs as their vertices can be coloured into two colours. The connected two vertices are not of the same colour. The decomposition of bipartite graph produces edge disjoint set of sub-graphs. The vertices  $V$  are divided into two disjoint sets  $L$  and  $H$ . Each vertex in  $L$  is connected to each vertex in  $H$  by a link as shown in Fig. 7. We model the silhouette by such undirected bipartite graph, which is without self loop and considering each pixel as an individual node to form 8 connected image graph  $G$ . An adjacency matrix  $A$  is defined over the graph and  $A(i, j)$  is weight between node  $i$  and  $j$ .  $D = \text{diag}(d_i)$  denote diagonal degree matrix, where  $d_i$  is degree of node  $i$ . Furthermore, the Laplacian matrix of graph is expressed as,  $L = D - A$ , where,  $\mathcal{L} = I - D^{-1/2}AD^{-1/2}$  is the normalized Laplacian.

Here, we apply only analysis filter of perfect reconstruction two channel critically sampled QMF-graph wavelet filter bank as shown in Fig. 8 to compress the graph-structured data. As suggested in [20], the colouring of the vertices is done by using BSC algorithm [25], followed by decomposition of the graph into the set of bipartite graphs using Harary's algorithm [26]. Each sub-graph is down sampled by  $\beta_L$  and  $\beta_H$ , which are down sampling operators. The nodes in  $L$  preserve output

of low pass channel, whereas nodes in  $H$  preserves output of high pass channel. Thus  $H$  and  $L$  facilitates bi-partition of graph. The analysis filters  $H_0$  and  $H_1$  can be written as,

$$H_0 = h_0(\mathcal{L}) = \sum_{\lambda \in \sigma(G)} h_0(\lambda) P_\lambda \quad (9)$$

$$H_1 = h_1(\mathcal{L}) = \sum_{\lambda \in \sigma(G)} h_1(\lambda) P_\lambda \quad (10)$$

where,  $h_0(\lambda)$  and  $h_1(\lambda)$  are spectral kernels,  $\mathcal{L}$  is normalized Laplacian,  $\lambda$  is eigenvalue,  $\sigma(G)$  is spectrum of graph which is set of eigenvalues and  $P_\lambda$  is the projection matrix of eigen space  $V_{(\lambda)}$ . The low pass analysis kernel  $h_0(\mathcal{L})$  is computed by using Chebychev approximation of Meyer kernel. The other spectral kernels can be computed by using QMF relations. The static feature set includes the mean wavelet coefficients of entire gait sequence. The length of static feature is of 256 after second level decomposition.

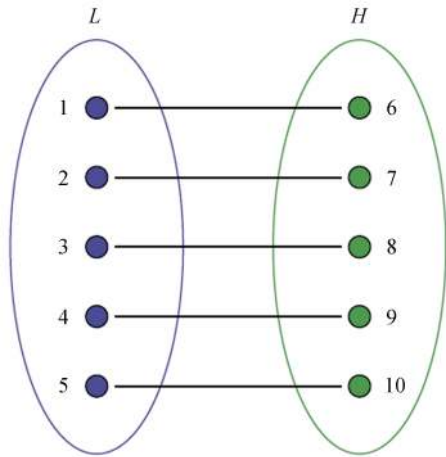


Fig. 7. Bipartite graph.

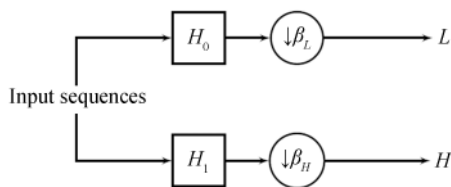


Fig. 8. Graph wavelet filter bank (analysis).

#### D. Feature Fusion

As discussed in previous subsections, static and dynamic features are extracted, which have different discriminating power. We concatenated these two features to construct a single augmented feature vector. Since, these features are not directly comparable, we first normalize them using min-max normalization method [27]. The length of dynamic features varies with the number of frames present in that particular sequence, hence zero padding is done to individual feature to make the dynamic feature set of fixed length. The static feature set is of fixed length, hence zero padding is not required. Further, weights for normalized feature vectors are computed using the mean-variance method.

The normalized static and dynamic feature vectors are

$$\overline{FV}_{\text{static}} = \frac{FV_{\text{static}} - \min(FV_{\text{static}})}{\max(FV_{\text{static}}) - \min(FV_{\text{static}})} \quad (11)$$

$$\overline{FV}_{\text{dynamic}} = \frac{FV_{\text{dynamic}} - \min(FV_{\text{dynamic}})}{\max(FV_{\text{dynamic}}) - \min(FV_{\text{dynamic}})}. \quad (12)$$

The weights for normalized features vectors are

$$w_{\text{static}} = \frac{m_{\text{static}}}{\sigma_{\text{static}}} \quad (13)$$

$$w_{\text{dynamic}} = \frac{m_{\text{dynamic}}}{\sigma_{\text{dynamic}}} \quad (14)$$

where,  $w$  is weight,  $m$  is mean and  $\sigma$  is variance of the feature vector.

Finally, these two features are concatenated as shown in (15) to form a single augmented vector for representation and further processing,

$$FV = [(w_{\text{static}} \cdot \overline{FV}_{\text{static}}), (w_{\text{dynamic}} \cdot \overline{FV}_{\text{dynamic}})]. \quad (15)$$

For demonstration of the proposed fusion approach, we computed dynamic and static features of a person walking normally with view angle of  $90^\circ$ . Figs.9–11 denote zero padded normalized weighted dynamic features. Fig.9 shows mean value of wavelet cross spectrum (WCS), Fig.10 denotes wavelet coherence (WCOH) and Fig.11 shows local relative phase ( $\phi$ ) relation between dynamic body parts. The static feature is shown in Fig.12, which are coefficients of graph wavelet filter bank decomposed up to second level. The augmented feature vector after fusion of static and dynamic feature set is shown in Fig.13. It is the representation of a person walking normally at view angle  $90^\circ$  in feature space.

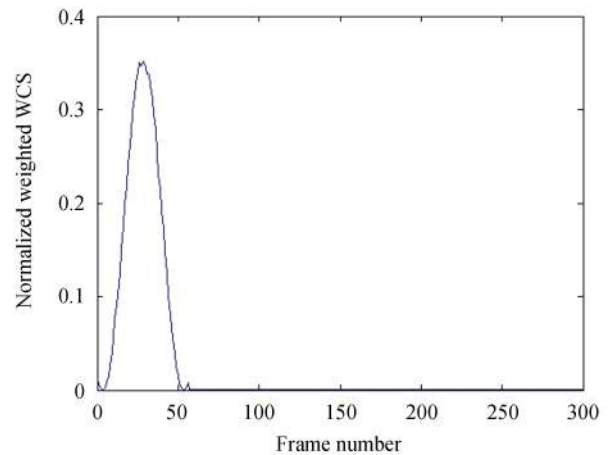


Fig. 9. Normalized WCS feature.

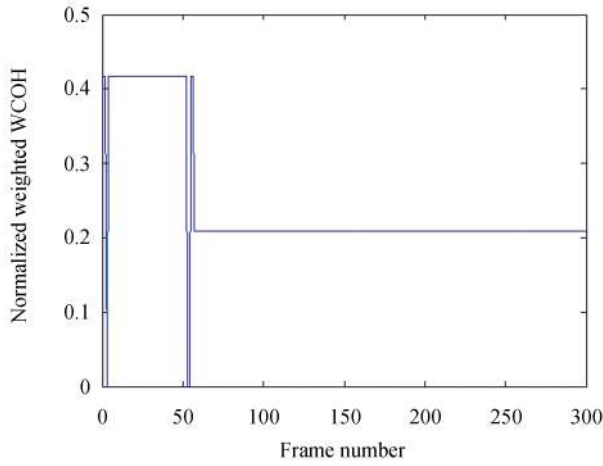


Fig. 10. Normalized WCOH feature.

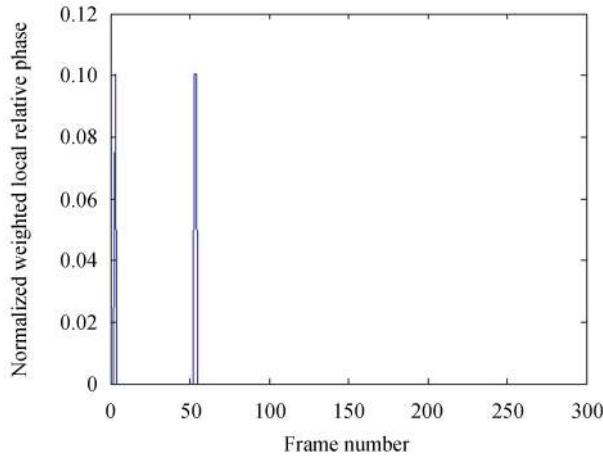


Fig. 11. Normalized phase feature.

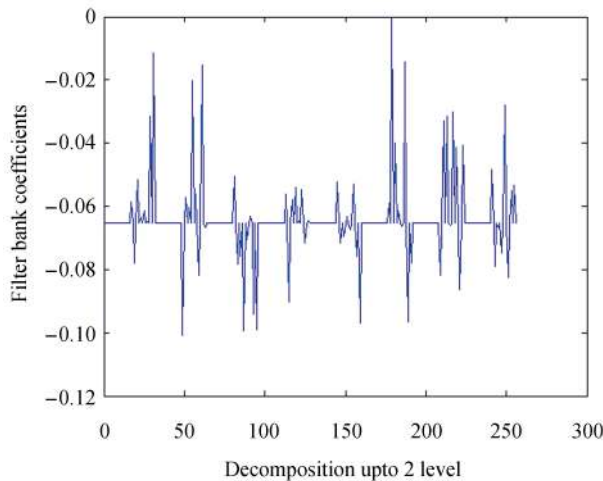


Fig. 12. Normalized filter bank coefficients.

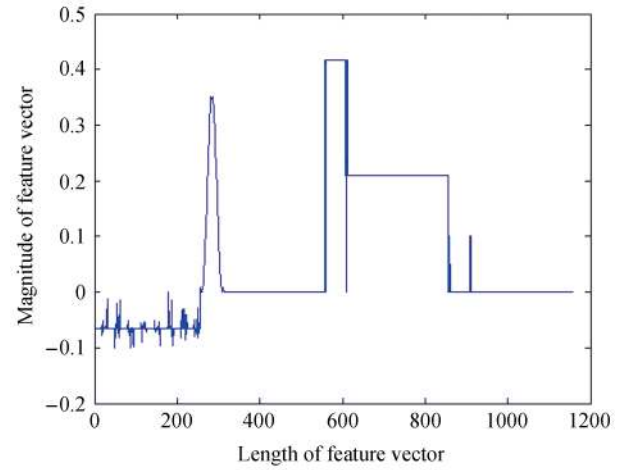


Fig. 13. Augmented feature vector.

### E. Training

For each gait sequence we extract feature vector  $FV^{(P_N, V_A, C_V)}$ , which is fusion of cross wavelet coherence and QMF-graph wavelet filter bank coefficients. Considering  $P_N = 1, \dots, 124$ ,  $V_A = 1, \dots, 11$ ,  $C_V = 1, \dots, 10$ , where,  $P_N$  is number of subjects (persons),  $V_A$  is view angle and  $C_V$  is co-variate. The centroid of clusters  $p_q$  is then obtained as a result of k-means clustering. It clusters training  $FV^{(P_N, V_A, C_V)}$  vectors to  $Q$  clusters to minimize the within-cluster distance

$$\sum_{q=1}^Q \sum_{i=1}^{P_N} \alpha_{iq} \|FV^{(P_N, V_A, C_V)} - p_q\|^2 \quad (16)$$

where  $\alpha_{iq} = 1$ , if  $FV^{(P_N, V_A, C_V)}$  is assigned to the cluster  $q$  and  $\alpha_{iq} = 0$  otherwise. The centroid of clusters  $p_q$ ,  $q = 1, \dots, Q$  are the centres of clusters. The optimal number of clusters is determined by using cross-validation procedure.

The  $FV^{(P_N, V_A, C_V)}$  vector describes the feature vector of  $P_N$ th person, walking at viewing angle  $V_A$  and having co-variate condition  $C_V$ . This feature vector of each training subject is then mapped by Euclidean distance to the centroid of clusters  $p_q$  as follows

$$d_{Ed} = \|FV^{(P_N, V_A, C_V)} - p_q\|. \quad (17)$$

Other distances can also be used but Euclidean distance exhibit simple representation hence preferred. Each distance vector is then finally represented as  $D = [d_1, d_2, \dots, d_Q]^T$ . Further, for the final representation, the Euclidean distances are normalized to get membership vector

$$R_{FV} = \frac{d_{Ed}}{\|d_{Ed}\|}. \quad (18)$$

To make this membership vector duration invariant, as we use multi-period gait sequences, the mean of  $R_{FV}$  is taken. For  $i = 1, \dots, C_V$  and all  $t_j; j = 1, \dots, p_q$  membership vectors

$$v_i = \frac{1}{t_j} \sum_{k=1}^{t_j} R_{FV}^{ik}. \quad (19)$$

Further, linear discriminant analysis is applied to  $v_i$  to project it into low dimensional discriminant subspace. Each

person can be linearly separable in this subspace. In LDA, an optimum projection matrix  $W_{\text{opt}}$  is derived to minimize the Fisher criterion.

$$W_{\text{opt}} = \arg \min \frac{(W^T S_w W)}{(W^T S_b W)} \quad (20)$$

where  $S_w$  and  $S_b$  are the scatter matrices of within class and between class of  $C$  classes

$$S_w = \sum_{n=1}^C S_i \quad (21)$$

$$S_i = \sum_{x \in \omega_i} (x - \mu_i)(x - \mu_i)^T \quad (22)$$

$$S_b = \sum_{i=1}^C n_i (\mu_i - \mu)(\mu_i - \mu)^T \quad (23)$$

where

$$\mu = \frac{1}{n} \sum_{\forall x} x \quad (24)$$

$\mu_i$  is the mean vector of class  $\omega_i$  and  $\mu$  is the mean vector of the training set.

#### F. Testing

The test feature vector is obtained using similar steps as used to obtain the training feature vector. For identification, we use probabilistic model such as Bayesian framework. The Bayesian probabilistic decision theory is a fundamental approach in pattern classification. Assuming equiprobable classes and all probabilities are known, let  $P(j)$  is the *a priori* probability of occurrence of  $j$ th person in the database of  $P_N$  classes. The class conditional probabilities can be expressed as  $P(j|P_N, V_A, C_V)$ , where  $P_N$  are total subjects in databases,  $V_A$  are viewing angles considered and  $C_V$  are co-variate conditions while training. The *a priori* can be estimated during training and *aposteriori* can be estimated by the following equation

$$\begin{aligned} & P(j|P_1, V_1, C_1, \dots, P_N, V_N, C_N) \\ &= \frac{P(P_1, V_1, C_1, \dots, P_N, V_N, C_N|j) \cdot P(j)}{\sum_{n=1}^{N_P} P(P_1, V_1, C_1, \dots, P_N, V_N, C_N|n) \cdot P(n)} \end{aligned} \quad (25)$$

#### V. EXPERIMENTAL RESULTS AND COMPARISON

Experiments are performed on CASIA B multi-view gait database, considering all view angles and co-variate conditions in MATLAB environment. All the training classes are equiprobable as Bayesian framework is used for identification. Training includes all view angles and co-variate factors considering only hand and leg movement. The mean value of wavelet cross spectrum (WCS) along with wavelet coherence and phase ( $\phi$ ) are preserved to construct the dynamic feature set. Each sequence is compressed by QMF-graph wavelet filter bank to get wavelet coefficients as the static feature. The feature fusion is done by min-max rule. The Euclidean distance between each training vector and centroid of the clusters are then preserved. While testing, we divide the

database into three sets. Set A is of sequences *nm-05* and *nm-06*, set B is of *cl-01* and *cl-02* and set C is of *bg-01* and *bg-02*.

We compare our work with [16], [11] and [28], even though it is not straight forward. The rank 1 results are shown in Table II. In [16], first they extract side view gait cycle and extract two kinds of features namely; shifted energy image and gait structural profile. They performed experiments by considering; normal walking sequences *nm-01* to *nm-04* as gallery set, *nm-05*, *nm-06* as set A, *cl-01*, *cl-02* as set B, *bg-01*, *bg-02* as set C. We extract two features considering all view angles and co-variate conditions as gallery set. The probe sets are taken similarly to perform experiments as given in [16]. It is found that our method slightly works better for set A, but outperforms for sets B and C. In [11], authors consider seven angles (from  $36^\circ$  to  $144^\circ$ ) and all co-variate conditions. They performed experiments considering similar probe sets as aforementioned. Here, they use single angle for testing on multi-view gallery each time, e.g., probe angle is  $126^\circ$  and gallery angles are  $36^\circ$  to  $144^\circ$ . In our case, we train and test the system by multi-view sequences. Our method outperforms for all probe sets.

TABLE II  
WORK COMPARISON RANK 1 RECOGNITION RATE (%)

Probe set	SIES [16]	DTCWT-PDF [11]	Hybrid [28]	Our method
Set A	99	76.76	97.98	99.82
Set B	64	73.23	86.02	100
Set C	72	71.71	85.22	100

Whereas in [28], authors extract static and dynamic features. The histogram distribution of optical flow vector is used as dynamic feature and Fourier descriptor is used as static feature. This work is similar to ours in this sense but they consider only three view angles, viz.,  $72^\circ$ ,  $90^\circ$  and  $108^\circ$ . They used rank based fusion, whereas ours is feature based fusion. Experimental results shows that our method performs better for probe sets A, B and C than [28].

The proposed method outperforms the above methods as the features we extracted are robust and found invariant to co-variate conditions especially carrying bag and cloth variations.

#### VI. CONCLUSION

In this work, we use cross wavelet transform and bipartite graph model for gait based human recognition in multi-view scenario. The experimental results show that, the fusion of different kinds of features represent gait pattern of an individual significantly. Table II shows that our method outperforms others in co-variate conditions also. The average recognition rate considering all view-angles and co-variate conditions in Bayesian framework is 99.90%. It has been observed that, the recognition rate decreases if the probe sequence is not included in the gallery while training. In future, we will concentrate on our research work to investigate different gait features which can improve performance of system in various co-variate conditions.



## ACKNOWLEDGMENT

Portions of the research in this paper use the CASIA gait database collected by Institute of Automation, Chinese Academy of Sciences. Authors are grateful to Prof. Yogesh Ratnakar Vispute, for proof reading of the paper.

## REFERENCES

- [1] L. Wang, T. N. Tan, W. M. Hu, and H. Z. Ning, "Automatic gait recognition based on statistical shape analysis," *IEEE Trans. Image Process.*, vol. 12, no. 9, pp. 1120–1131, Sep. 2003.
- [2] H. T. Liu, Y. Cao, and Z. F. Wang, "Automatic gait recognition from a distance," in *Proc. Chinese Control and Decision Conf.*, Xuzhou, China, 2010, pp. 2777–2782.
- [3] W. Zeng and C. Wang, "View-invariant gait recognition via deterministic learning," *Neurocomputing*, vol. 175, pp. 324–335, Jan. 2016.
- [4] X. L. Xing, K. J. Wang, T. Yan, and Z. W. Lv, "Complete canonical correlation analysis with application to multi-view gait recognition," *Pattern Recogn.*, vol. 50, pp. 107–117, Feb. 2016.
- [5] S. Niyogi and E. Adelson, "Analyzing and recognizing walking figures in XYT," in *Proc. 1994 IEEE Computer Society Conf. Computer Vision and Pattern Recognition*, Seattle, WA, USA, pp. 469–474.
- [6] X. X. Huang and N. V. Boulgouris, "Model-based human gait recognition using fusion of features," in *Proc. 2009 IEEE Int. Conf. Acoustics, Speech and Signal Processing*, Taipei, China, pp. 1469–1472.
- [7] M. Goffredo, I. Bouchrika, J. N. Carter, and M. S. Nixon, "Self-calibrating view-invariant gait biometrics," *IEEE Trans. Syst. Man Cybern. B Cybern.*, vol. 40, no. 4, pp. 997–1008, Aug. 2010.
- [8] A. Roy, S. Sural, and J. Mukherjee, "A hierarchical method combining gait and phase of motion with spatiotemporal model for person re-identification," *Pattern Recogn. Lett.*, vol. 33, no. 14, pp. 1891–1901, Oct. 2012.
- [9] L. Wang, H. Z. Ning, T. N. Tan, and W. M. Hu, "Fusion of static and dynamic body biometrics for gait recognition," *IEEE Trans. Circ. Syst. Video Technol.*, vol. 14, no. 2, pp. 149–158, Feb. 2004.
- [10] D. Ming, C. Zhang, Y. R. Bai, B. K. Wan, Y. Hu, and K. D. K. Luk, "Gait recognition based on multiple views fusion of wavelet descriptor and human skeleton model," in *Proc. 2009 IEEE Int. Conf. Virtual Environments, Human-Computer Interfaces and Measurements Systems (VECIMS)*, Hong Kong, China, pp. 246–249.
- [11] H. F. Hu, "Multiview gait recognition based on patch distribution features and uncorrelated multilinear sparse local discriminant canonical correlation analysis," *IEEE Trans. Circ. Syst. Video Technol.*, vol. 24, no. 4, pp. 617–630, Apr. 2014.
- [12] S. M. Jia, L. J. Wang, and X. Z. Li, "View-invariant gait authentication based on silhouette contours analysis and view estimation," *IEEE/CAA J. Autom. Sinica*, vol. 2, no. 2, pp. 226–232, Apr. 2015.
- [13] L. Rustagi, L. Kumar, and G. N. Pillai, "Human gait recognition based on dynamic and static features using generalized regression neural network," in *Proc. 2nd Int. Conf. Machine Vision*, Washington, DC, USA, 2009, pp. 64–68.
- [14] S. J. Hong, H. Lee, and E. Kim, "Fusion of multiple gait cycles for human identification," in *Proc. ICCAS-SICE*, Fukuoka, Japan, 2009, pp. 3171–3175.
- [15] Y. Pratheepan, J. V. Condell, and G. Prasad, "The use of dynamic and static characteristics of gait for individual identification," in *Proc. 13th Int. Machine Vision and Image Processing Conf.*, Washington, DC, USA, 2009, pp. 111–116.
- [16] X. X. Huang and N. V. Boulgouris, "Gait recognition with shifted energy image and structural feature extraction," *IEEE Trans. Image Process.*, vol. 21, no. 4, pp. 2256–2268, Apr. 2012.
- [17] C. H. Chen, J. M. Liang, H. Zhao, H. H. Hu, and J. Tian, "Factorial HMM and parallel HMM for gait recognition," *IEEE Trans. Syst. Man Cybern. C Appl. Rev.*, vol. 39, no. 1, pp. 114–123, Jan. 2009.
- [18] N. V. Boulgouris and X. X. Huang, "Gait recognition using HMMs and dual discriminative observations for sub-dynamics analysis," *IEEE Trans. Image Process.*, vol. 22, no. 9, pp. 3636–3647, Sep. 2013.
- [19] M. D. Hu, Y. H. Wang, Z. X. Zhang, D. Zhang, and J. J. Little, "Incremental learning for video-based gait recognition with LBP flow," *IEEE Trans. Cybern.*, vol. 43, no. 1, pp. 77–89, Feb. 2013.
- [20] S. K. Narang and A. Ortega, "Perfect reconstruction two-channel wavelet filter banks for graph structured data," *IEEE Trans. Signal Process.*, vol. 60, no. 6, pp. 2786–2799, Jun. 2012.
- [21] S. Zheng, J. G. Zhang, K. Q. Huang, R. He, and T. N. Tan, "Robust view transformation model for gait recognition," in *Proc. 18th Int. Conf. Image Processing*, Brussels, Belgium, 2011, pp. 2073–2076.
- [22] A. Grinsted, J. C. Moore, and S. Jevrejeva, "Application of the cross wavelet transform and wavelet coherence to geophysical time series," *Nonlin. Processes Geophys.*, vol. 11, no. 5–6, pp. 561–566, Nov. 2004.
- [23] E. K. W. Ng and J. C. L. Chan, "Geophysical applications of partial wavelet coherence and multiple wavelet coherence," *J. Atmos. Oceanic Technol.*, vol. 29, pp. 1845–1853, Dec. 2012.
- [24] C. Torrence and G. P. Compo, "A practical guide to wavelet analysis," *Bulletin of The American Meteorological Society*, vol. 79, no. 1, pp. 61–78, Jan. 1998.
- [25] W. Klotz, "Graph coloring algorithms," *Math. Bericht*, vol. 5, pp. 1–9, 2002.
- [26] F. Harary, D. Hsu and Z. Miler "The biparticity of a graph," *J. Graph Theory*, vol. 1, no. 2, pp. 131–133, 1977.
- [27] A. A. Ross, K. Nandakumar, and A. Jain, *Handbook of Multibiometrics*. New York, USA: Springer-Verlag, 2006.
- [28] M. F. Ho, K. Z. Chen, and C. L. Huang, "Gait analysis for human walking paths and identities recognition," in *Proc. Int. Conf. Multimedia and Expo*, New York, NY, USA, 2009, pp. 1054–1057.



**Sagar Arun More** was born in Maharashtra, India. He received B.E. degree in electronics from the North Maharashtra University, Jalgaon, India, in 2000 and the M.E. degree in electronics from Sant Gadage Baba Amravati University, Amravati, India, in 2008. He is currently pursuing Ph.D. from North Maharashtra University Jalgaon. His current research interests include image and video processing, biometrics.



**Pramod Jagan Deore** was born in Maharashtra, India. He received the B.E. degree in electronics from the North Maharashtra University, Jalgaon, India, in 1997. He received the M.E. and Ph.D. degrees from Shri Guru Gobind Singhji (SGGS) Institute of Engineering and Technology, Swami Ramanand Teerth Marathwada University, Nanded, in 1999 and 2007, respectively. He is a Professor of Electronics and Telecommunication Engineering Department at the R. C. Patel Institute of Technology, Shirpur, India.

His research interests include interval arithmetic operations applications in robust control, image processing, bio-medical signal processing, microwave circuits and antennas. He has published 40 papers in national/international conferences/journals and he has co-authored two books. He is Life Member of ISTE.