Gait Recognition Using Wavelet Packet Silhouette Representation and Transductive Support Vector Machines

Farzin Dadashi*, Babak N.Araabi*, Hamid Soltanian-Zadeh*†
*Control and Intelligent Processing Center of Excellence,
School of Electrical and Computer Engineering, University of Tehran,
P.O.Box 14395/515, Tehran, Iran
†Image Analysis Lab., Dept of Radiology, Henry Ford Hospital, Detroit, MI 48202, USA

***Abstract*—Gait is an idiosyncratic biometric that can be used for human identification at a distance and as a result gained growing interest in intelligent visual surveillance. In this paper, an efficient gait recognition method based on describing subject outer body contour deformations using wavelet packets is proposed. With the use of Matching Pursuit algorithm, *k* bases of wavelet packet tree that have maximum similarity to the signal are selected and corresponding coefficients are used as features. Finally, Transductive Support vector Machine (TSVM) classification is utilized on computed eigengait space for semi-supervised identification. The proposed method of selecting features which uses a complete orthogonal or near orthogonal basis from a wavelet packet library of bases and investigating the correlational structure of gait features for each individual using TSVM, result in encouraging identification performance.**

I. Introduction

Biometrics is physiological or behavioral characteristics which can be used for reliable identification of individuals. The most prominent and flourishing physiological biometrics, derived by direct measurement of a part of human body, include fingerprint and hand geometry scans, iris identification and face recognition. On the other hand, extracted traits based on an action performed by individual form behavioral biometrics. The main feature of this group of biometrics is the use of time as a metric. Established measures include key-stroke scan and speech patterns. All these methodologies are restricted to controlled environments and more importantly they need cooperation of the subject. Therefore, we use gait as an unobtrusive biometric which can be obtained from a distance. Gait is defined as a particular way or manner of moving on feet. Early psychological studies by Murray [1] suggested that gait is a unique personal trait with cadence which is cyclic in nature. Many different gait recognition techniques have been investigated thus far which can clearly broken down into two main categories. Model-based approaches [2,3] use a priori knowledge of the object, being searched for, in each frame of a walking sequence. This information includes kinematics of joint angles, limb length and step cadence. Even though evidence gathering techniques can be used before making a choice on the model fitting, high computational burden, on the basis of complex matching and searching has limited the efficiency of model based methods. Model-free methods [4,5,6,7 and 8], on the other hand, consider gait as a sequence of body posture and usually use silhouette to recognize similarity of body poses; so, most of the existing gait recognition techniques belong to this class owing to its simplicity and efficiency.

The current study aims to propose a genuine automatic gait recognition approach based on describing silhouette using wavelet packet library. Our algorithm detains the spatiotemporal characteristics of gait and projects them in many levels on dictionary atoms. The good localization property of wavelet functions in both time and frequency domains results in a flexible representation of the local features of the silhouette. The new discriminative features generated by our method results in an improved classification rate. The block diagram in Figure 1 summarizes the proposed method. The rest of this paper is organized as follows. Section 2 introduces the walking subject detection and silhouette representation method. Section 3 examines feature extraction by insightful choice of wavelet bases functions using Matching Pursuit algorithm. Section 4 provides a brief overview of TSVM theory. We present our experiment results on two databases in Section 5. Eventually, Section 6 concludes the paper.

II. Walking subject detection and silhouette representation

To analyze the gait, one should first detect the human in the image scene. We made a simplifying assumption that the video sequence is captured by a fixed camera and the only moving object in each video frame is the ambulatory body. Of course, for unconstrained environments the first problem to be solved is difficulties emanate from body occlusion, scene illumination and shadow in parts of the image.

A. Segmentation of silhouette

Extraction of human silhouette from the background plays a key role in gait recognition system. Let F represent a sequence of N frames. A simple approach for moving object detection in a video sequence is background subtraction. This method is a pixel based segmentation in which we compute absolute

978-1-4244-4131-0/09/$25.00 ©2009 IEEE

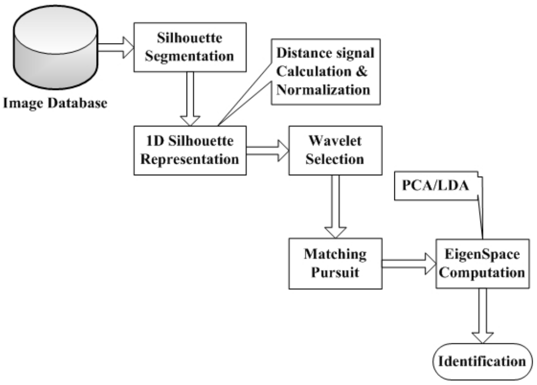

Fig. 1. Overview of the proposed method

difference of each frame form background image to obtain the difference image. If this difference for a given pixel is more than a threshold, the pixel is classified as a pixel of moving object; else this pixel is regarded as a background pixel. Here, background image can be calculated using a modified running average algorithm [9]. First an initial background image is obtained using equation (1):

$$B_1(x,y) = median(f_1(x,y), f_2(x,y), ..., f_N(x,y)) \quad (1)$$

where $f_i(x,y) \in F, i = 1,2,...,N$ and B_1 is background for the first frame. Now a pixel in each frame is marked as foreground pixel if $|f_i(x,y) - B_i(x,y)| > T$ where T is a predefined threshold. The background update rule is as equation (2):

$$B_{i+1}(x,y) = \alpha f_i(x,y) + (1-\alpha)B_i(x,y) \quad (2)$$

where α is adaptation coefficient and kept relatively small to prevent forming spurious tails behind the moving object. The following corrections would be used when necessary:

- The background model at each pixel is determined by its recent history. If a pixel is marked as foreground for more than k of the last K frames, then the background is updated as $B_{i+1}(x,y) = f_i(x,y)$. These way, sudden changes in illumination and the appearance of static new objects are compensated.
- To compensate for fluctuating illumination in a pixel state which changes repeatedly from foreground to background -such as swinging branches in outdoor scenes - it is masked out from inclusion in the foreground.

Now, further filtering using morphological operations and a simple border following algorithm based on connectivity is utilized to obtain the silhouette contour.

B. Silhouette representation

Temporal changes of the subject outer body contour are important sources in determining implicit motion aspects of a walking body. We consider the silhouette contour as a deformable object through time that can be represented by its samples. Moreover, by converting 2D silhouette changes into associated 1D signals, the computational cost reduces significantly. For silhouette representation, we use the method proposed by Wang et al [5]. The contour in each frame is unwrapped by finding the upper most pixel of the contour as the start point and moving clockwise on the boundary with equidistance steps until we select a fixed number of points $p_k(x,y), k = 1,2,r,$ on it. The number of samples, r, should be selected in a way that two different contours from different frames can be distinguishable when represented just by their resampled sequences. Additionally, in each frame, we compute the silhouette centroid. The distance of each resampled contour pixel from the centroid in frame t is calculated then to obtain vector $[d_1^t d_2^t ... d_{r-1}^t d_r^t]^T$. The sequence $D_i = [d_i^1 d_i^2 ... d_i^{t_f -1} d_i^{t_f}]$ constitutes the i^{th} signal. To eliminate the effects of spatial scale, each distance signal is normalized. The vertical component of the centroid $\{y_c^t\}_{t=1}^{t_f}$ is another useful signal that we use for identification. Note that if we keep all contour point distances from the centroid in each frame the contour is completely recoverable. Figure 2 depicts five of such signals for subject number 124 of the CASIA database B.

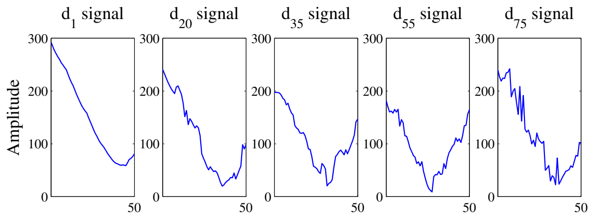

Fig. 2. The distance between contour samples and the centroid

III. Feature extraction using Wavelet Packets

Wavelet transform has been used as an effective method for shape representation due to its robustness against rotation, translation and scale. When the signals obtained in the previous section are projected onto the wavelet and scaling functions in many levels, we often desire to achieve an efficient representation where sparsity of decomposition is of main concern and resolution near the limits as defined by Heisenberg uncertainty principle may be achieved as well. In general, for sampled waveform datasets, linear feature extraction is an effective method of selecting features which best discriminate between different classes of waveform [10]. It consists of finding basis functions $\Phi = [\varphi_1 \varphi_2 ... \varphi_L]$ such that the coefficients of projection $\Gamma = \Phi^T d$ can be used for a pattern recognition system. A variety of attempts have been performed to use wavelet and wavelet packet signal processing as a linear feature extraction method for waveform pattern recognition systems. Wavelet packet system was proposed by Ronald Coifman [11] which provides a finer and regulating resolution of frequencies at high frequencies and also gives a flexible structure that allows adaptation to particular signal classes. Starting from an orthogonal wavelet $\psi(t)$ and two associated filters *g(n)* and *h(n)* of length *2N* the wavelet packet is generated as follows:

$$W_{2n}(t) = \sqrt{2} \sum_{k=0}^{2N-1} h(k) W_n(2t-k) \tag{3}$$

$$W_{2n+1}(t) = \sqrt{2} \sum_{k=0}^{2N-1} g(k) W_n(2t-k) \tag{4}$$

where $W_0(t) = \varphi(t)$ and $W_1(t) = \psi(t)$ are related scaling and wavelet functions respectively and $n = 1, 2, 3, \ldots$

Starting from the W_n , let us consider the three-index family of wavelet packet atoms (instead of two parameters of scale and translation as used in standard multiresolution structure), obtained by dyadic dilation and translation of W_n:

$$W_{j,n,k}(t) = 2^{\frac{-j}{2}} W_n(2^{-j}t - k) \tag{5}$$

Roughly speaking, $W_{j,n,k}(t)$ allows analyzing the fluctuations of a given signal around position $2^j k$ at the j^{th} scale. A wavelet packet dictionary provides an over-complete $L.log_2 L$ time-frequency localized basis functions where L is the length of the signal. Redundancy in wavelet packets provides us with a wider set of basis functions to select suitable discriminatory directions. The main blueprint for a wavelet packet feature extractor is the problem of choosing a subset basis function from the dictionary. Several criteria are discussed in wavelet packet basis selection among which we choose Matching Pursuit Algorithm which is closely related to the projection pursuit strategy developed in [12].

A. Matching Pursuit algorithm

Mallat and Zhang [13] proposed matching pursuit algorithm as a computationally efficient procedure for basis selection in a given dictionary. Successive approximation of the given signal at each stage can be obtained by selecting a single wavelet or scaling function from the dictionary. A basis is selected such that it has the highest matching (resemblance) with the signal at that stage and thus achieves minimal error for signal approximation in a single step. The algorithm is as follows:

1) At the initial stage indexed by 0, the entire signal $S = S^{(0)}$ is projected onto each of the atoms(elements of the dictionary).

$$S = \langle S \mid \varphi_i \rangle \varphi_i + R, i = 1, 2, 3, ... \tag{6}$$

where $\langle S \mid \varphi_i \rangle$ represents inner product of S and ϕ_i.

2) A waveform with the highest coefficient (corresponding to best match) is chosen.

$$\begin{array}{l} S = \langle S \mid \varphi_0 \rangle \varphi_0 + R^{(0)} \\ |\langle S \mid \varphi_0 \rangle| \geq \sup \langle S \mid \varphi_i \rangle, i = 1, 2, 3, ... \end{array} \tag{7}$$

3) At this stage indexed by 1, the residual is evaluated.

$$R^{(1)} = S^{(0)} - \left\langle S^{(0)} \mid \varphi_1 \right\rangle \varphi_1 \tag{8}$$

The norm $\left\| R^{(1)} \right\|_2$ is used as the approximation error of the first stage, i.e., approximating S by $\langle S \mid \varphi_1 \rangle \varphi_1$. The basis of the first stage ϕ_1 is chosen such that $\left\| R^{(1)} \right\|_2$ is minimum.

4) Now, selection procedure is repeated where at each stage the residual signal is projected onto the elements of the entire dictionary and an additional atom is selected. In general, at stage *k* we can write:

$$S^k = S^{(k-1)} + \alpha_k \varphi_k, \alpha_k = \left\langle R^k \mid \varphi_k \right\rangle \tag{9}$$

where $S^{(k)} = R^{(k)}$ and ϕ_k is chosen such that the error of approximation at the second stage $\left\| R^{(k)} \right\|_2$ is minimal, i.e., $\left\langle S^k \mid \varphi_k \right\rangle \geq \sup \left\langle S^k \mid \varphi_i \right\rangle, i = 1, 2, 3, ...$ A new residual *R* with $\left\| R^{(k)} \right\|_2 < \left\| R^{(k-1)} \right\|_2$ is constructed.

5) The procedure is repeated until the approximation error (norm of the residual of the last stage) is less than a given threshold.

It can be shown that the algorithm converges in a countable number of steps and error approaches zero as the number of iterations tends to infinity, i.e., $\lim_{k \to \infty} \left\| R^{(k)} \right\|_{(2)} = 0$ is the residual of signal *S* at stage *k*. Matching pursuit algorithm in literature is referred to as a "greedy algorithm" since at each step it chooses an atom that best correlates with the signal (or residuals at different stages) and the best approximation (minimal error) for signal representation is achieved. Choosing $S = D_i$ and selecting *m* first coefficients obtained by $\left\langle S^k \mid \varphi_k \right\rangle$, we represent each distance signal with *m* features.

Once the gait features - which are wavelet packet atoms coefficients- are extracted for each individual, we face a high dimensional measurement space that should be mapped onto a low dimensional eigenspace. Therefore, Principlal Component Analysis (PCA) is used to get rid of extensive dimensionality by eliminating the correlation between features and possibly removing noise.

IV. The principle of Transductive support vector machines

Support vector machines - first proposed by Vapnik et al [14]- are categorized as generalized linear classifiers. SVM learner aims to compute transformation $f : X_{train} \rightarrow \{-1, 1\}$ from the training data $S_{train} = \{(x_i, c_i) | x_i \in X_{train}, c_i \in \{-1, 1\}\}_{i=1}^{l}$ that at the same time maximizes the geometric margin and minimizes the empirical risk of misclassification as:

$$R = \frac{1}{l} \sum_{i=1}^{l} |f(x_i) - c_i| \tag{10}$$

The primary concern is finding out an optimal separating hyperplane with low generalization error[15]. The optimal hyperplane that separates the convex hull of the two classes, should satisfy the following Quadratic programming optimization problem:

$$\begin{array}{l} \text{minimize(in w,b): } \frac{1}{2} w^T w \\ \text{subject to: } c_i(w^T.x_i + b) \geq 1, i = 1, 2, ..., l \end{array} \tag{11}$$

For the linearly non-separable case, introducing a new set of slack variables ζ_i (i = 1,2,...,l), to penalize violation of constraints is necessary.

$$\begin{aligned} &\text{minimize: } \frac{1}{2}w^T w + m\Big(\sum_{i=1}^{l} \zeta_i\Big)^k \\ &\text{subject to: } c_i(w.\varphi(x_i) + b) \geq 1 - \zeta_i \end{aligned} \qquad (12)$$

where *m* and *k* are used to weight the slack variables and $\varphi(.)$ is a nonlinear function which maps the input space to a higher dimension space. TSVM takes into account the structural properties of data, by giving the learner an i.i.d sample set of $S^* = \{x_i^* | x_i^* \in R^n\}_{i=1}^{k}$ of test samples to be classified, in addition to the training set S_{train}[16]. The TSVM requires to satisfy the following constrained minimization as:

$$\begin{aligned} &\text{minimize(in w,b,}c^*\text{): } \frac{1}{2}w^T w \\ &\text{subject to: } \begin{cases} c_i(w.x_i + b) \geq 1, i = 1, 2, ..., l \\ c_j^*(w.x_j^* + b) \geq 1, j = 1, 2, ..., k \end{cases} \end{aligned} \qquad (13)$$

Intuitively, the algorithm is to start with small c^*, and adding labels to some unlabeled data based on classifier prediction. Then c^* is slowly increased and labeling unlabeled data and re-running the classifier is repeated. The efficacy of TSVM on multi-class classification in high dimensional input space and investigating correlational structures between gait signatures of each subject, made it a suitable tool for our purpose.

V. Experiment result and analysis

In order to evaluate the performance and efficacy of our proposed method, we tested it on two databases. The first database is CASIA Gait Database (Data Set B) containing 124 subjects (93 males and 31 females) and 11 views. For each view, there exist six sequences for each subject. We used just the first six views (0° to 90°). Besides, we used CASIA Infrared night gait database [17] of 153 subjects which encompasses four walking modes: normal walking, slow walking, fast walking and normal walking with a bag. For each subject there exist 10 sequences of gait. Four sequences are allotted to normal walking while all three other cases include two sequences for each subject. To choose the wavelet for signal decomposition and feature extraction, several tests were carried out. Experiment shows to the degree that related to gait signals, low frequency components of the silhouette turn out to be less sensitive to walking style variations. Therefore, we used energy matching scheme for wavelet selection. The wavelets that have a high portion of their energy content in frequency band similar to that of input gait signals are selected in the first step. The energy spectrum of such signals and 4 wavelets are shown in Figure 3 (a) to (i). It can be seen from Figure 3 that the majority of energy distribution of extracted signals are held in frequencies that are relatively low; thus the wavelet with similar frequency content would be Daubechies number 2 ('db2').

As stated in Section 3, a sparse representation of signal is often desirable for classification purposes. Therefore, a signal can be represented using a few coefficients. To this end, we decomposed the gait signals from different subjects using discrete wavelet transform into 3 levels. After resampling the silhouette, 75 distance signals are preserved; Augmenting the distance matrix with the signal obtained from the vertical variation of centroid, 76 signals are kept for each individual. Each signal is then projected onto atoms of the wavelet packet and 5 biggest coefficients using Matching Pursuit algorithm are kept to represent each distance signal. Therefore, each gait sequence is mapped in a space of 380 dimensions.

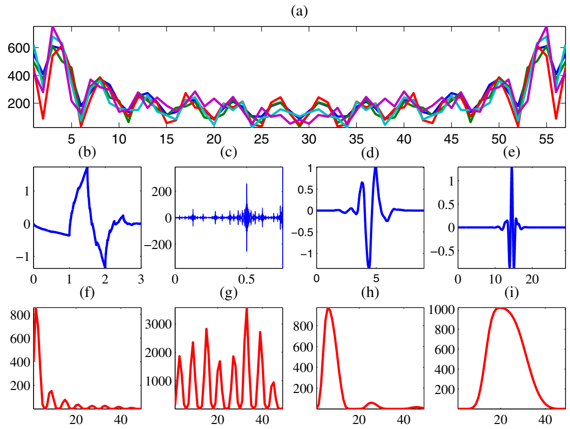

Fig. 3. Energy matching between gait signals and mother wavelet bases: (a) Energy spectrum for the signals in Fig.3, (b),(f) Daubechies 2 wavelet and its Energy Spectrum, (c),(g) Bior 3.1wavelet and its Energy Spectrum, (d),(h) Symlet 5 wavelet and its Energy Spectrum, (e),(i) Coiflet 5 wavelet and its Energy Spectrum

To reduce the dimension of the feature space and eliminate redundant features, PCA and LDA applied and the first 18 eigenvalues and their associated eigenvectors to form our eigenspace transformation matrix are kept for CASIA database to retain at least 95% of information after exerting dimension reduction contrivances. Table 1 compares correct classification ratio for two different signal projection frameworks. Note that implementing LDA after PCA yields a superior performance. Noteworthy, dimension reduction does not change the separability of the extracted features in our experiment, and on the other hand, reduces the TSVM identification computational cost remarkably.

TABLE I
Classification Performance of the proposed method on the CASIA database B for different views

	View	0°	18°	36°	54°	72°	90°
CCR	PCA	81.1	79.9	84.3	83.3	87.5	92.9
	PCA +LDA	84.4	81.2	85.1	82.9	90.3	96.2

The best correct classification rate under lateral walking view reinforces that the multi-resolution representation of data shows less sensitivity to silhouette variations compared with the other views. To assess the proposed method while walking mode changes, four experiments on CASIA Infrared Night Gait dataset were performed. Table 2 presents these experiments. The performance measure used here is Cumulative

TABLE II
EXPERIMENTS ON CHANGING WALKING MODES

Experiment	Gallery	Probe	$\#Gallery$	$\#Probe$
A	Normal	Normal	459	153
B	Normal	Slow	459	306
C	Normal	Fast	459	306
D	Normal	Bag	459	306

Match Score (CMS) [18]. The similarity measure is calculating the output of each SVM and finding out which one puts the prediction the furthest into the positive region.

Figure 4 discloses that recognition accuracy declines due to changes in the walking speed. the accuracy drop off is insignificant as a result of preserving the frequency content of the gait sequences in the wavelet domain. However, appearance variation diminishes the recognition performance extensively and seems to be still an unsolved challenging problem.

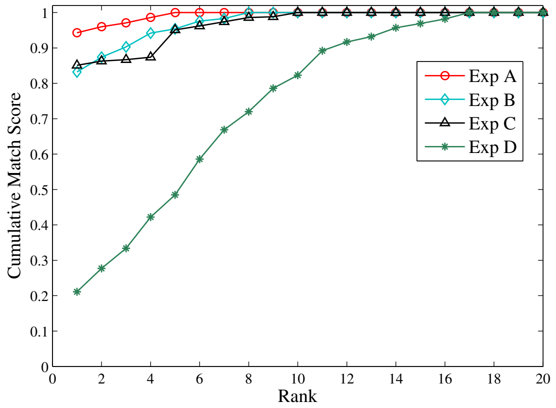

Fig. 4. The CMS curve of the proposed method on the CASIA Night Infrared Dataset

VI. Conclusion and Future Work

In this paper, we proposed a gait recognition method based on one dimensional representation of silhouette changes and analyzing it using the wavelet packet transform. The Matching Pursuit algorithm, which uses maximum similarity to choose suitable bases among all bases, allowed us to obtain flexible multi-resolution signal representation along with new features for identification. Transduction is an intrinsically easier task than first learning a general inductive rule and then applying it to the test samples. Acquiring the correlational structure of gait features of each individual using TSVM, besides preserving the most informative coefficients of the walking sequence expansion in the wavelet packet domain with the aid of matching pursuit algorithm, made the proposed method robust to walking speed variations.

Even though high identification accuracy was obtained, much work still remains to be done in the proposed direction. Further evaluation on much larger databases with more subjects is necessary. To select optimal views in multi-camera systems a sensitivity analysis of features extracted from different views is needed. Also, view-invariant silhouette representation based on planar curve matching methods is worthy of paying more attention in the future work.

References

[1] M.P. Murray, ”Gait as a total pattern of movement.,” American Journal of Physical Medicine, vol. 46, 1967, p. 290.

[2] D. Cunado, M.S. Nixon, and J.N. Carter, ”Using gait as a biometric, via phase-weighted magnitude spectra,” Audio-And Video-Based Biometric Person Authentication: First International Conference, Avbpa’97, Crans-Montana, Switzerland, March 12-14, 1997: Proceedings, Springer, 1997, p. 95.

[3] L. Wang, H. Ning, T. Tan, and W. Hu, ”Fusion of static and dynamic body biometrics for gait recognition,” Ninth IEEE International Conference on Computer Vision, 2003. Proceedings, 2003, pp. 1449-1454.

[4] J.B. Hayfron-Acquah, M.S. Nixon, and J.N. Carter, ”Automatic gait recognition by symmetry analysis,” Pattern Recognition Letters, vol. 24, 2003, pp. 2175-2183.

[5] L. Wang, T. Tan, H. Ning, and W. Hu, ”Silhouette analysis-based gait recognition for human identification,” . IEEE Transaction on Pattern Analysis and Machine Intelligence, vol.25 ,2003,pp.1505-1518.

[6] J. Lu, E. Zhang, and C. Jing, ”Gait Recognition Using Wavelet Descriptors and Independent Component Analysis,” LECTURE NOTES IN COMPUTER SCIENCE, vol. 3972, 2006, p. 232.

[7] J. Man and B. Bhanu, ”Individual recognition using gait energy image,” IEEE transactions on pattern analysis and machine intelligence, vol. 28, 2006, pp. 316-322.

[8] B. Ye and Y. Wen, ”A New Gait Recognition Method Based on Body Contour,” Control, Automation, Robotics and Vision, 2006. ICARCV’06. 9th International Conference on, 2006, pp. 1-6.

[9] J. Heikkil and O. Silvn, ”A real-time system for monitoring of cyclists and pedestrians,” Image and Vision Computing, vol. 22, 2004, pp. 563-570.

[10] G. Rutledge and G. McLean, ”Comparison of several wavelet packet feature extraction algorithms,” Submitted to IEEE Trans. on Pattern Recognition and Machine Intelligence, 2000.

[11] R.R. Coifman and M.V. Wickerhauser, ”Entropy-based algorithms for best basis selection,” IEEE Transactions on Information Theory, vol. 38, 1992, pp. 713-718.

[12] J.H. Friedman and W. Stuetzle, ”Projection pursuit regression,” Journal of the American statistical Association, 1981, pp. 817-823.

[13] S.G. Mallat and Z. Zhang, ”Matching pursuits with time-frequency dictionaries,” IEEE Transactions on Signal Processing, vol. 41, 1993, pp. 3397-3415.

[14] C. Cortes and V. Vapnik, ”Support-vector networks,” Machine learning, vol. 20, 1995, pp. 273-297.

[15] N. Kasabov and S. Pang, ”Transductive support vector machines and applications in bioinformatics for promoter recognition,” Neural Information Processing-Letters and Reviews, vol. 3, 2004, pp. 31-38.

[16] T. Joachims, ”Transductive inference for text classification using support vector machines,” Sixteenth International Conference on Machine Learning, 1999.

[17] D. Tan, K. Huang, S. Yu, and T. Tan, ”Efficient night gait recognition based on template matching,” The 18th IEEE International Conference on Pattern Recognition(ICPR’06), Proceedings, 2006, pp. 1000-1003.

[18] P.J. Phillips, H. Moon, S.A. Rizvi, and P.J. Rauss, ”The FERET evaluation methodology for face-recognition algorithms,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 22, 2000, pp. 1090-1104.