A Comparative Evaluation Study of Automated Gait Recognition based on Spatiotemporal Feature and Different Neural Network Classifiers

Noha A. Hikal
Assistant Professor.
Information Technology Dept.
Faculty of Computers and Information System, Mansoura University, Egypt

ABSTRACT

New areas of applications such as: human-computer interaction, access control, surveillance, activity monitoring and clinical analysis depend on hepatic technology. Gait analysis has been explored thoroughly during the last decade as a behavioral biometric feature which doesn't require subject interaction. In this paper, persons can be recognized from their gait regardless of the angle of walking seen. The performance of four artificial neural networks (ANNs) based

classifiers was evaluated and tested, based on spatiotemporal features. The results show that discrete wavelet transforms and support vector machine recognition technique provides a recognition rates up to 94%. Moreover, it is characterized by speed and accuracy compared with other classifiers.

General Terms

Biometric-based recognition. Human-computer interaction. and Biometric Security.

Keywords

Gait energy image, discrete cosine transform, discrete wavelet transform, principle component analysis, support vector machine, multilayer perceptron, radial basis function, generalized feed forward network.

1. INTRODUCTION

Currently, real world applications demand new techniques for person's identification. Using biometrics has a great influence in this field. Unlike conventional recognition techniques of passwords or token cards, Biometric identification techniques are safer and more reliable, since they are almost impossible to be disguised, shared, or misplaced [1, 2]. However, most of the biometric features need a physical interaction between the person and a certain device to gather the features. Thus, gait has recently gained a considerable attention as a promising behavioral biometric feature due to several but great advantages. The benefits of gait-based authentication are [3]: (i) It has unique capability to recognize people at a distance without the subject's knowledge or interaction while other biometrics can't perform the same; (ii) No special equipment is required for image acquisition, it is easy to fix surveillance cameras at the corners of public; (iii) It doesn't require images that have been captured of a very high resolution, and therefore it can be used in situations where face or iris information is not available in high enough resolution for recognition. (iv) Gait of an individual is difficult to alter, people need to walk; and their manner of walking is usually observable and more difficult to obscure or disguise, by trying; the individual will probably looks more suspicious. Furthermore, gait information can be useful in detecting the gender, psychological abnormalities, and many medical applications [4]. However, there are a lot of challenges in gait recognition, such as foreground segmentation, changes in appearance due to clothing variations, walking velocity, carrying objects [5]. But, good recognition technique should be able to understand the basic features of the biometric, regardless of the presence of such factors.

Based on the above discussion, this paper proposing a new gait recognition technique based on cascaded spatial and transform-based feature extraction; to extract spatiotemporal gait features. Then, different supervised neural network classifiers were employed to classify the gaits based on the extracted features. The novelty in this paper is that the person can be recognized from his gait regardless of the angle of walking seen. Hence the supervised learning neural networks are trained using four different gait cycles corresponding to four different view angles; (i) right to left direction with walking angle 0, (ii) left to right direction with walking angle 0, (iii) right to left direction with walking angle 45, and (iv) left to right direction with walking angle 45. Different orthogonal transforms together with different NN's classifiers are simulated and tested. Their performances are evaluated, compared, and discussed. The rest of this paper is organized as follows: section 2 briefly reviews the most recent trends in gait recognition. Section 3 introduces the proposed frame work. Section 4 explains the used feature extraction techniques. Section 5 presents the different NN's classifiers deployed in this paper. Finally, section 6 presents results and conclusion.

2. RELATED WORK

Various techniques have been proposed for gait recognition. These techniques fall into two categories: model-based approaches and model-free approaches [2].

Recent trends in gait recognition researches tend to focus on model free approaches. These approaches use motion information directly extracted from silhouettes without resolving the body pose. Compared to model-based approaches, the model free approaches are characterized by simplicity and speed [2, 3, 4, 5]. Most of these approaches usually perform spatial feature extraction. By extracting silhouettes, a large part of physical appearance features have been removed from the image representation of human. And then match the feature sequence using simple temporal correlation or dynamic time warping, etc. [1, 3, 6]. Nevertheless, a silhouette still contains information about the shape of human body that is vulnerable to changes caused by conditions such as carrying objects, and the variations in clothing which affect the silhouettes dynamics. Therefore, the

classification rate is not satisfying. Model-based approaches mainly focus on the dynamics of gait, while shape features are omitted. These approaches usually require high quality gait sequences. Thus, some researches apply multiple cameras. In addition, these approaches tend to be more complex and computationally extensive than model free approaches [7, 8, 9]. This paper considers model- free approaches but without omitting the dynamic parts of the body.

3. THE PROPOSED FRAMEWORK

In order to improve the detection accuracy, the silhouettes are extracted from infrared (IR) thermal imaging. Since the human body is a nature emitter of infrared ray, the temperature of the human body is different from that of background or package. Consequently, the silhouettes will be cleaned from the effect of background, carrying, and clothing. Fig. 1 gives the comparison between conventional image, Fig.1 (a), and infrared image, Fig.1(c), when subject walks with a package. The package changes human body silhouette in the conventional image as shown in Fig. 1(b) and the impact from package is difficult to remove. However, Fig. 1(c) reveals that the package is invisible in infrared image because of different temperature between human body and the package. The clean silhouette can be generated for the infrared image as illustrated in Fig. 1(d) [11]. Therefore infrared pattern facilitates silhouette extraction and eliminates the impact of carrying and clothing. In addition, thermal imaging is suitable for night surveillance.

(a) (b)

(c) (d)

Fig.1: Comparison of silhouette extracted from conventional image and IR image [10]: (a) conventional image, (b) silhouette in conventional image, (c) IR image, (d) silhouette in IR image.

The architecture of the proposed frame work is as shown in Fig. 2, it is divided into five main phases: (i)video acquisition using a IR video surveillance device; (ii) frames pre-processing; (iii) feature extraction; (iv) machine learning, and (v) classification.

Through the pre-processing phase, each frame is applied to: size normalization, horizontal alignment, and then silhouettes are extracted. The feature extraction phase combines two types of features; spatial feature and transform-based feature. Now, the dimension of the data has been reduced and thus retaining the most salient features. Finally, the machine learning phase based on supervised ANN's is employed to accomplish the learning process from these features and get the final recognition result. The next subsections are focusing on the feature extraction, the automated learning, and the classification process.

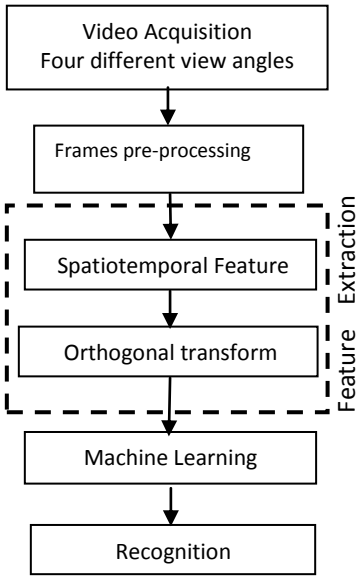


Fig. 2: The proposed frame work

4. FEATURE EXTRACTION

Feature extraction phase has a great necessity to reduce data dimensionality. In this paper, two features are cascaded in two phases. The first phase extracts spatial feature which is the gait energy image (GEI). The benefit of using GEI is to reduce the gait cycle dimension along the sequence of frames while preserving information about body shape and stance. The output of the first phase, which is the GEI, is feed as an input to the second phase. Different transform based feature extraction techniques are employed to obtain a transformed coefficients highly defining for GEI.

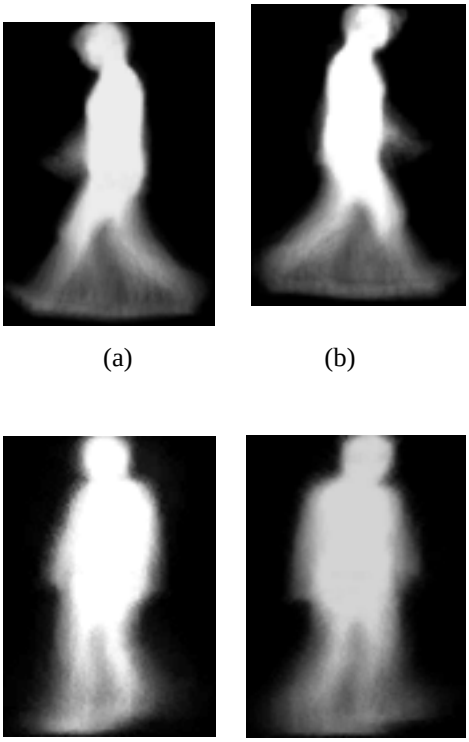
4.1 Spatiotemporal feature

Given a gait cycle, the silhouettes are extracted after applying the size normalization and horizontal alignment to each extracted silhouette [6]. The GEI is then computed as:

$$\mathrm{GEI} = \mathrm{G}(x, y) = \frac{1}{T}\sum_{t=1}^{T} \mathrm{I}(x, y, t) \qquad (1)$$

Where: T is the number of frames in a complete gait cycle, I is a silhouette image whose pixel coordinates are given by x and y, and t is the frame number in the gait cycle. The GEI reflects major shapes of silhouettes and their changes over

gait cycle. Examples of GEI at different walking angles are shown in Fig. 3. The original gait cycles are obtained from the standard CASIA database [10]. As seen from Fig.3, the static area of the GEI (that corresponds to body parts that move little during walking) contains helpful information about the body shape. Moreover, it can be used in gender classification [4]. Whilst the dynamic parts of the GEI (correspond to body parts that move constantly during moving like legs and arms) tells us more about how people move during walking. The benefit of using GEI is to reduce the gait cycle dimension while preserving the static and the dynamic features of the gait. However, it is a spatial feature and it is sensitive to changes in various conditions. Therefore, GEI itself needs to go through a feature extraction phase.



(a) (b)

(c) (d)

Fig. 3: Different GEI for the same subject: (a) right to left direction with walking angle 0, (b) left to right direction with walking angle 0, (c) right to left direction with walking angle 45, and (d) left to right direction with walking angle 45.

4.2 Transform-based feature extraction

Regarding the previous observation about the GEI, transform-based feature extraction is needed. In this paper, we examine the most three basic orthogonal transforms used in classification problems and deduced successful results. The three famous orthogonal transforms are: (i) principle component analysis (PCA), (ii) discrete cosine transforms (DCT), and (iii) discrete wavelet transform (DWT). The orthogonal coefficients obtained are used to learn different NN's classifier for gait recognition.

4.2.1 Principle Component Analysis (PCA)

Principal-component analysis (PCA) is a useful technique can be used to reduce the dimensionality of large data sets. The method generates a new set of variables, called *principal components*. Each principal component is a linear combination of the original variables. All the principal components are orthogonal to each other, so there is no redundant information [12, 13].

The PCA transform of a given set of *n* input vectors (variables) with the same length *K* formed in the *n*-dimensional vector $x = [x_1, x_2 \ldots xn]^T$ into a vector y according to [12]:

$$Y=A(x-m_x) \quad (2)$$

The vector m_x in Eq. (2) is the vector of mean values of all input variables defined by relation:

$$m_x = E\{x\} = \frac{1}{K}\sum_{k=1}^{K} x_k \quad (3)$$

Matrix *A* in Eq. (2) is determined by the covariance matrix C_x. Rows in the *A* matrix are formed from the eigenvectors of C_x ordered according to corresponding eigenvalues in descending order. The evaluation of the C_x matrix is possible according to relation:

$$C_x = E\{(x-m_x)(x-m_x)^T\} = \frac{1}{K}\sum_{k=1}^{K} x_k x_k^T - m_k m_k^T \quad (4)$$

As the vector *x* of input variables is *n*-dimensional it is obvious that the size of C_x is *n x n*. The elements C_x *(i, i)* lying in its main diagonal are the variances of x:

$$C_x(i,i) = E\{(x_i - m_i)^2\} \quad (5)$$

And the other values C_x *(i, j)* determine the covariance between input variables x_i, x_j, as:

$$C_x(i,j) = E\{(x_i - m_i)(x_j - m_j)\} \quad (6)$$

The principal components as a whole form an orthogonal basis for the space of the data that ensures high data decorrelation.

4.2.2 Discrete Cosine Transform (DCT)

The DCT, and in particular the 2D-DCT, is often used in signal and image processing .A DCT expresses a sequence of finitely data points in terms of a sum of cosine functions oscillating at different frequencies. 2D-DCT tends to concentrate the information in a number of orthogonal coefficients. The use of cosine functions are much more efficient; due to: (i) energy compaction; (ii) decorrelation; (iii) separability; (iv) symmetry; and (v) orthognality.

The 2D-DCT is given according to the formula [12]:

$$P(u,v) = \left(C(u).C(v).\sum_{x=0}^{N}\sum_{y=0}^{M} I(x,y).\cos\frac{\pi(2x+1)u}{2N} . \cos\frac{\pi(2y+1)v}{2M} \right) \quad where \quad 0 \le u \le N-1, \; 0 \le v \le M-1 \tag{7}$$

Where: *P* is the DCT coefficients, *I* is the original image of dimension *N*x*M*. *C(u)* and *C(v)* are constants computed according to the formula:

$$c(u) = \begin{cases} 1/\sqrt{N} & u = 0 \\ \sqrt{2/N} & 1 \le u \le N-1 \end{cases} \tag{8}$$

$$c(v) = \begin{cases} 1/\sqrt{M} & v = 0 \\ \sqrt{2/M} & 1 \le v \le M-1 \end{cases} \tag{9}$$

4.2.3 Discrete Wavelet Transform (DWT)

The wavelet transform has gained wide spread acceptance in signal and image processing; because of their inherent multiresolution nature. Wavelet-coding schemes are especially suitable for applications where scalability and tolerable degradation are important. The most commonly used wavelets are Daubechies [13]. This formulation is based on the use of recurrence relations to generate progressively finer discrete samplings of an implicit mother wavelet function; each resolution is twice that of the previous scale [13, 14]. The DWT of a signal (x) is calculated by passing it through a series of filters. First the samples are passed through a low pass filter with impulse response g; the signal is also decomposed simultaneously using a high-pass filter with impulse response h.

The outputs are then divided into: (i) detail coefficients; from the high-pass filter, and (ii) approximation coefficients; from the low-pass. It should be noted that, the two filters are related to each other and they are known as a quadrature mirror filter. However, half the frequencies of the signal have now been removed and half the samples can be discarded according to Nyquist's rule [14]. The filter outputs are then sub sampled by 2 as:

$$y_{low}[n] = \sum_{k=-\infty}^{\infty} x[k]g[2n-k]$$

$$y_{high}[n] = \sum_{k=-\infty}^{\infty} x[k]h[2n+1-k] \tag{10}$$

This decomposition has halved the time resolution since only half of each filter output characterizes the signal. However, each output has half the frequency band of the input so the frequency resolution has been doubled.

This decomposition is repeated to further increase the frequency resolution and the approximation coefficients decomposed with high and low pass filters and then down sampled. This is represented as a binary tree with nodes representing a sub-space with different time-frequency localization. The tree is known as a filter bank as shown in Fig.4. For the GEI the proposed coefficients are extracted using the technique showed in Fig.4.

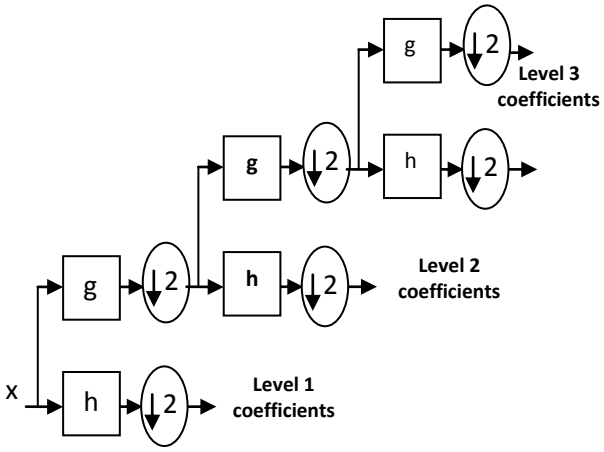


Fig.4: 3 level DWT filter bank

5- CLASSIFICATION TECHNIQUES:

An institutive goal of classification in this paper is to discriminate between known and unknown persons upon their gait. A more ambitious goal is to recognize the person himself. There are a large number of ANN's based classifiers, regardless of their many enhancement techniques. However, this paper focuses on the supervised ones for GEI recognition. The performance of four famous supervised learning NN's will be briefly explained to show how they can be deployed in gait recognition. These major techniques are :(i) Multi-layer perceptron, (ii) Support vector machine, (iii) Generalized feed forward network, and (iv) Radial basis function neural network.

5.1 Multi-Layer Perceptron networks:

MLP is a layered feed forward networks typically trained with static back propagation [15]. These networks have found their way into countless applications requiring static pattern classification. Their main advantage is that they are ease to use, and that they can approximate any input/output map. The key disadvantages are that they train slowly, and require lots of training data. Fortunately, many NN specialized software [16] automatically include speeding up learning methodologies in its learning phase The MLP configuration deployed for GEI recognition has three layers; (1)an input layer of length equals to the number of the significant transformed coefficients, (2) one hidden layer, and (3) an output layer of one output value corresponds to agree or disagree decision.

5.2 Support Vector Machine networks:

Support vector machines (SVMs) offer an extremely powerful method of deriving efficient models for multidimensional function approximation and classification. SVM classifiers are a close cousin to classical multilayer perceptron neural networks. Using a kernel function, SVM's apply an

alternative training method in which the weights of the network are found by solving a quadratic programming problem with linear constraints, rather than by solving a non-convex, unconstrained minimization problem as in standard neural network training. The goal of SVM modeling is to find the optimal hyperplane that separates clusters of vector in such a way that cases with one category of the target variable are on one side of the plane and cases with the other category are on the other size of the plane. Support Vector Machine (SVM) is implemented using the kernel Adatron algorithm [15]. The kernel Adatron maps inputs to a high-dimensional feature space, and then optimally separates data into their respective classes by isolating those inputs which fall close to the data boundaries. Therefore, the kernel Adatron is especially effective in separating sets of data which share complex boundaries.

5.3 Generalized Feed-Forward (GFF) networks:

GFF networks are a generalization of the MLP such that connections can jump over one or more layers. In theory, a MLP can solve any problem that a generalized feed forward network can solve. In practice, however, GFF networks often solve the problem much more efficiently. It suffices to say that a standard MLP requires hundreds of times more training epochs than the GFF network containing the same number of processing elements [15].

5.4 Radial Basis Function NN

Radial basis function (RBF) networks are nonlinear hybrid networks typically containing a single hidden layer of processing elements. This layer uses Gaussian transfer functions, rather than the standard sigmoid functions employed by MLPs. The centers and widths of the Gaussians are set by unsupervised learning rules, and supervised learning is applied to the output layer. These networks tend to learn much faster than MLPs. For RBF NN, the number of cluster centers is by definition equal to the number of GBF used at output layer [15].

6-EXPERIMENTAL RESULTS

For comparison purpose, four different measures have been used to evaluate the performance of the ANN-based classification techniques:

(i)Mean-Square-Error (MSE):

The mean squared error of an individual case *(i)* is evaluated by the equation:

$$MSE = \sum_{j=1}^{n} (O_{ij} - T_j)^2 / n$$

Where O_{ij} is the value predicted by the ANN based classifier, j is the length of the output vector; and T_j is the target value for fitness case j.

(ii) Recognition rate (R %):

Recognition rate measures the percent of correct recognition cases. It is calculated as the percentage between the successful recognition cases and the overall true recognition cases.

(iii) False positive rate (F_P%)

F_p is the ratio between the numbers of false GEI's that are incorrectly classified as true ones to the total number of true GEI.

(iv) False negative rate (F_N %)

F_N is the ratio between the numbers of true GEI's that are incorrectly classified as false ones to the total number of false GEI.

A successful detection algorithm should achieve high *R*, low F_N and low F_P.

A simulation computer program using matlab was used to gather the GEI from gait sequence of images. A series of experiments are carried out for testing and evaluating the performance of different NN's classification techniques. For fairness, all tests were done using the same gait database [10]. Although identifying the person from its GEI is considered a challenging work, acceptable recognition results of the different transform - NN classifier pairs are presented in Table 1 and Table 2. Regarding the results obtained, the recognition parameters and the MSE values both indicate that using 2D-DWT for feature-based recognition is the most suitable method. Moreover, the combination between the 2D-DWT and SVM has achieved the minimum MSE, high recognition rate, low F_P and low F_N . In comparing with previous literature [11, 17], this combination is capable of highly recognizing the subjects from their GEI from different four walking view angles. Moreover, the SVM learning curves combined with the three suggested transforms DCT, PCA, and DWT are shown in Figures 5,6, and 7 respectively. From these figures, it can be noticed that the learning curve of the DWT-SVM technique reaches the most minimum value of MSE compared with the other cases considering the number of iterations. Hence, SVM has achieved efficient hyper plane classification process based on the DWT, which can highly decorrelate the gait energy image.

It can be concluded that, based on different three individual gait view angles, applying GEI followed by DWT-SVM technique has shown the superiority in performance for individual's recognition. This work can be integrated by constructing a multi-camera surveillance system on which long-term video surveillance would be performed and the proposed activity recognition algorithm could be tested.

Table 1. Transform-classifier error comparative results

Transform	Classification technique	MSE
2D-DCT	MLP	0.19011
	SVM	0.14381
	GFF	0.182339
	RBF	0.19088
PCA	MLP	0.26506
	SVM	0.220069
	GFF	0.184731
	RBF	0.449770
2D-DWT	MLP	0.167345
	SVM	0.079286

	GFF	0.182339
	RBF	0.151776

Table 2. Transform-classifier recognition comparative results

Trans-form	Classifier	R%	FP%	FN %
2D-DCT	**MLP**	76.92	36.12	47.22
	SVM	77.77	38.25	47.11
	GFF	84.61	40.33	43.56
	RBF	69.23	36.12	49.11
PCA	**MLP**	73.07	29.34	44.23
	SVM	72.22	36.44	40.22
	GFF	61.11	34.26	36.12
	RBF	69.64	36.25	46.12
2D-DWT	**MLP**	84.61	23.12	32.56
	SVM	94.44	22.56	23.15
	GFF	84.61	27.77	23.53
	RBF	83.45	29.35	25.33

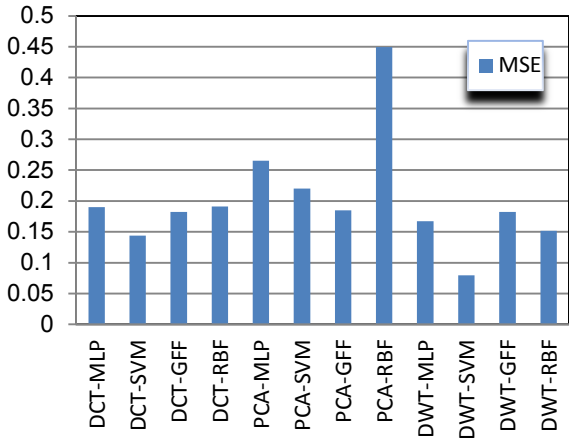


Fig.5: 3 level DWT filter bank

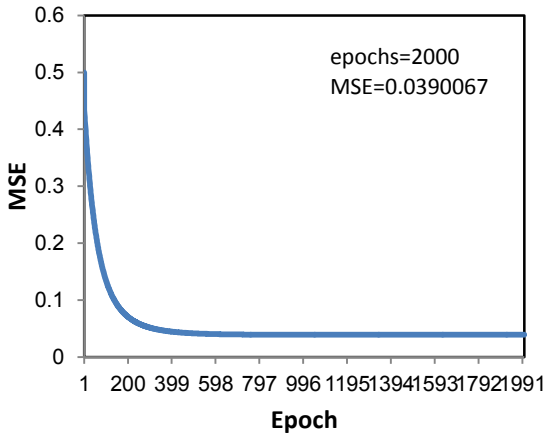


Fig.6: learning curve of PCA-SVM technique

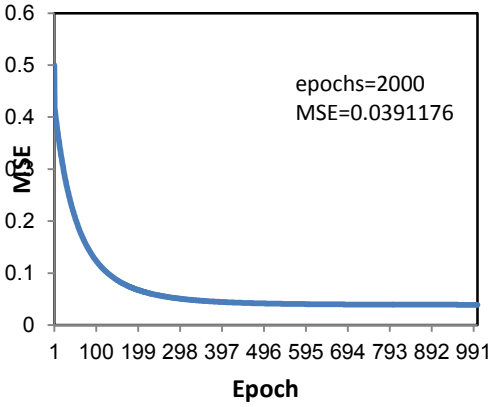


Fig.8: learning curve of DCT-SVM technique

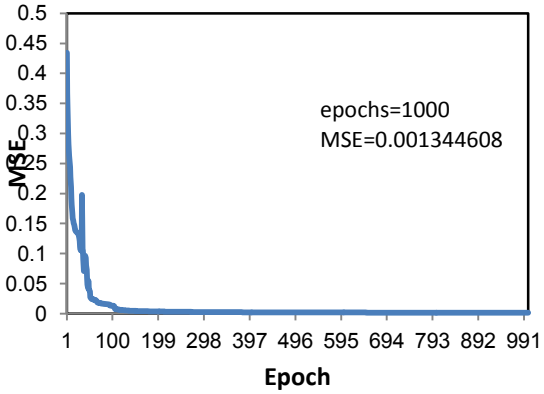


Fig.9: learning curve of DWT-SVM technique

6. ACKNOWLEDGMENTS

This work was supported by information technology department, faculty of computer science and information system, Mansoura university. Egypt.

7. REFERENCES

[1] K. Delac, M. Grgic, A survey of biometric recognition methods, 46th International Symposium Electronics in Marine, 2004, pp. 184-193.

[2] A.K. Jain, Technology: biometric recognition, Nature 449(6) (2007) 38-40.

[3] S. Sarkar, et al, The human ID gait challenge problems: data sets, performance and analysis, IEEE Trans. Pattern Analysis. Machine Intelligence 27(2) (2005), pp. 162-177.

[4] S. Yu, et al, A study on gait based gender classification, IEEE transaction on image processing,vol.18,No.8 august 2009

[5] Wei Zeng, Cong Wang, Human gait recognition via deterministic learning, Elsevier Science, Journal of neural network, Volume 35, November, 2012 Pages 92-102

[6] R. Hu, W. Shen, H. Wang, Recursive spatiotemporal subspace learning for gait recognition, ELSEVIER, neurocomputing 73 (2010),pp. 1892-1899

[7] H. Lu, P. Venetsanopoulos, A layered deformable model for gait analysis, 7th International Conference on |Automatic Face and Gesture Recognition, April 2006, pp. 249-254

[8] G. Zhao, G. Liu, H. Li, 3D gait recognition using multiple cameras, 7th International Conference on Automatic Face and Gesture Recognition. April 2006, pp.529-534.

[9] F. Tafazzoli, R. Safabakhsh, Model-based human gait recognition using leg and arm movements, ELSEVIER, Engineering Application of Artificial Intelligence (2010), doi:10.1016/j. engappai.2010.07.004

[10] http://www.cbsr.ia.ac.cn/english/Gait%20Databases.asp CASIA database. Mars 2013

[11] Z. Xue, et al, Infrared gait recognition based on wavelet transform and support vector machine, ELSEVIER, pattern recognition 43 (2010), pp. 2904-1910.

[12] John C.Russ, The Image Processing Handbook, 3rd Ed., CRC press ISBN:0849325323, 1998

[13] R. C. Gonzales and R. E. Woods. Digital Image Processing. Prentice Hall, second edition. ISBN 0-201-18075-8. 2002

[14] C. Torrence, & G. Compo, A practical guide to wavelet analysis, American Meteorological Society, Vol.79, No.1, January 1998.

[15] Lakhmi Jain, Anna Maria, Recent advances in artificial neural networks design and applications. 2000 by CRC press LLC

[16] MATLAB 2011.b, Neuro-Solutions 5

[17] S. Yu, D.Tan, T.Tan, A framework for evaluating the effect of view angle clothing and carrying condition on gait recognition, 18th International Conference of Pattern Recognition, vol.4,Hong Kong , China,2006, pp.441-444.

[18] K. Bashir, T.Xiang, S. Gong, Gait recognition without subject cooperation, ELSEVIER, pattern recognition letters 31 (2010). Pp.2052-2062.