ISSN 1978-5933

The Second International Conference
On Green Technology and Engineering (ICGTE) 2009
Faculty of Engineering Malahayati University

SMOOTH SUPPORT VECTOR MACHINE FOR FACE RECOGNITION USING PRINCIPAL COMPONENT ANALYSIS

[1]Muhammad Furqan, [2]Abdullah Embong, [3]Suryanti Awang, [4] Santi W. Purnami, [5]Sajadin Sembiring

[1,2,3,4,5]University Malaysia Pahang
Faculty of Computer Systems & Software Engineering
Lebuhraya Tun Razak 26300 Kuantan, Pahang
mhdfurqan@gmail.com,{ae,suryanti}@ump.edu.my,{santipurnami,sajadin_sembiring1}@yahoo.com

Abstract

Face is one of the unique features of human body. It has complicated characteristic. Facial features (eyes, nose, and mouth) can be used for face recognition in face detection. Support Vector Machine (SVM) is a new algorithm of data mining technique, recently received increasing popularity in machine learning community. The Smooth Support Vector Machine (SSVM) is a further development of a SVM. The SSVM convert the SVM primal formulation to a nonsmooth unconstrained minimization problem. Since the objective function of this unconstrained optimization problem is not twice differentiable, smoothing techniques will be used to solve this problem. This paper presents Smooth Support Vector Machines (SSVM) for few samples-based face recognition with Principal Component Analysis (PCA) for face extraction called eigenfaces. The Eigenfaces is projected onto human faces to identify features vector. This significant features vector can be used to identify an unknown face using the SSVM to construct a nonlinear classifier by using a nonlinear kernel for classification and recognition. Firstly, a preprocessing is done by taking facial features from various expressions from each individual. Secondly, we obtain the features vector from eigenfaces. Thirdly, we use SSVM to train and test dataset. The proposed system has shown competitive result and demonstrates that methods are available.

Keywords: Smooth Support Vector Machine, Face Recognition, Principal Component Analysis, Feature Vector, Classification

1. Introduction

Pattern Recognition is a part of Computer Science, mapping data onto specific concept which is defined previously. The specific concept mentioned class or category. Application of pattern recognition is very wide, some of them, are voice recognition in security system, iris recognition, face recognition, finger recognition, and diagnosis of disease from medical records. Methods are known in pattern recognition, such as linear discrimination analysis, hidden markov model and artificial intelligent. The latest method is Support Vector Machine (SVM) [1][2]. SVM developed by Boser, Guyon, Vapnik, and first presented in 1992 in Annual Workshop on Computational Learning Theory. The Basic concept of SVM is harmonic combination from computation theories has existed several years before, like margin hyperplane (Duda & Hart in 1973, Cover in (1965), Vapnik (1964), and so on. Kernel published by Aronszajn in 1950, and the other supported concept. Until 1992, has never been effort to unite the component from theories [3][4].

Differences from neural network is to find hyperplane separator between class, SVM is to find the best hyperplane input space. The SVM basic principle is linear classifier and develops to apply non-linear problem with input of kernel trick on high space dimensional. This development gives stimulus in pattern recognition research to investigate the potential of SVM ability theoretical and application. From now on, SVM has success applied to real-word problems. And generally comparable result is better then other methods like artificial neural network [5].

As one of the most successful applications of image analysis and understanding, face recognition has recently received significant attention, especially during the past few years and has own the advantages and disadvantages result. Done to the complex representation of the human face, it is very complicated to develop an ideal computation model.

The earliest work on face recognition can be traced back at least 1950s in psychology and the 1960s in the engineering literature. Some of the earliest studies include work on facial expression by Darwin (1972), and research on automatic machine recognition of faces started in 1970s. Over the 30 years extensive research has been conducted by psychophysicist, neuroscientist, and engineers on various aspects by humans and machines [13].

Corresponding author : Muhammad Furqan Amal, Faculty of Computer Systems & Software Engineering, Lebuhraya Tun Razak 26300 Kuantan, Pahang, E-mail : mhdfurqan@gmail.com

Smooth Support Vector Machine developed by Y.J. Lee and O.L. Mangasarian [10] and its application in medical representation by Santi W.P and A.Embong [8,10]. It is shown as machine learning make better training than other techniques in supervised learning method [5,12].

Face recognition get many attention this day, because many application use it as identification tool, ATM (Automatic Teller Machine), crime, and others. Many features extraction techniques applied as Principal Component Analysis (PCA) or Karhunen-Loeve transform or Hostelling[13].

2. Principal Component Analysis (PCA)

A face as picture can be seen as a vector *X*. If length and width from that picture is *w* and *h* pixels then amount components from vector are *w*h*. Each pixel is coded by one component vector.

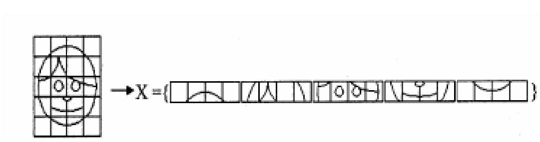

Fig. 1. Face picture vector formation

Face vector mentioned in a space, that is face space from picture has dimension of *w*h* pixel. Each face looks similar to others. Each face has two eyes, one mouth, one nose and others where set at the same place, so all face vector at narrow set of space image. Therefore all space image is not optimal to describe face. In order make a face space which describes face better. The basis vector from space face called principal components[6].

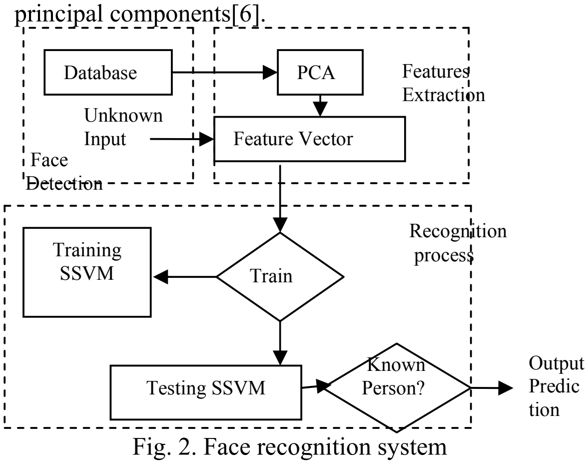

Fig. 2. Face recognition system

2.1. Eigenfaces method

The basic idea of eigenfaces is that all face images are similar in all configurations and it can described to basic face images. Based on this idea, the eigenfaces procedures are as follows (7) :

a. Assume the training sets of images are $\Gamma^1,\Gamma^2,\Gamma^3, \ldots, \Gamma^m$ with each image is *I* (*x,y*). Convert each image into set of vectors and new full-size matrix (*mxp*), where *m* is the number of training images and *p* is *x*y*.

b. Find the mean face by :

$$\Psi = \frac{1}{m}\sum_{i=1}^{m}\Gamma_i \qquad (1)$$

c. Calculate the mean- subtracted face :

$$\Phi_i = \Gamma_i - \Psi, \; i = 1,2, \ldots, m \qquad (2)$$

and a set of matrix is obtained with $A = [\Phi_1, \Phi_2, \ldots, \Phi_3]$ is the mean-substracted matrix vector with its size A_{mp}.

d. By implementing the matrix transformations, the vector matrix is reduced by :

$$C_{mm} = A_{mp} \text{ x } A^T_{pm} \qquad (3)$$

Where *C* is the covariance matrix and *T* is transpose matrix.

e. Find the eigenvectors, V_{mm} and eigenvalues, λ_m from the *C* matrix (using Jacobi method) and ordered the eigenvectors by highest eigenvalues. Jacobi's method is chosen because its accuracy and realibility than other method [15].

f. Apply the eigenvectors matrix, V_{mm} and adjusted matrix, Φ_m. These vectors determine linear combinations of the training set images to form the eigenfaces, $U_{k'}$ by :

$$U_k = \sum_{n=1}^{m}\Phi_n v_{kn}, \; k = 1,2,\ldots,m' \qquad (4)$$

Instead of using *m* eigenfaces, $m' < m$ which we consider the image provided for training are more than 1 for individuals or class. m' is the total class used.

g. Based on the eigenfaces, each image have its face vector by:

$$W_k = U_k^T(\Gamma - \Psi), k = 1,2,\ldots,m' \quad (5)$$

And mean subtracted vector of size *(px1)* and eigenfaces is $U_{pm'}$. The weights form a feature vector

$$\Omega^T = [w_1, w_2, \ldots, w_{m'}]$$

h. A face can reconstructed by using its feature, Ω^T vector and previous eigenfaces, $U_{m'}$ as :

$$\Gamma' = \Psi + \Phi_f \qquad (6)$$

Where $\Phi_f = \sum_{i=1}^{m'} w_i U_i$

3. Support Vector Machines

SVM concept a simple describe as effort to find the best hyperplane as separator 2 class in space input. Hyperplane in vector space *d* dimension is affine subspace to *d-1* dimension divided vector space into two parts where each corresponds with different class. The figure shown some pattern cluster from two class : +1 and -1. Classification problem interpreted to find the hyperplane line as separator with many alternative discrimination boundaries.

The best separator hyperplane between two class found by measure hyperplane margin and finding the maximum point. Margin is distance between the hyperplane and the near pattern from each pattern. The nearest pattern called support vector.

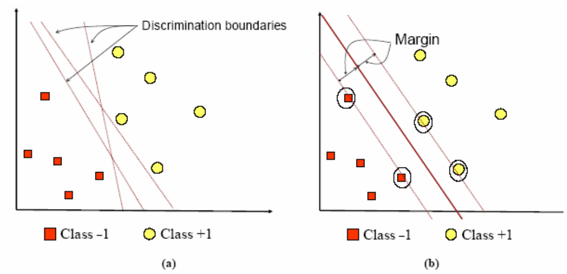

Fig.3. SVM Finding optimal hyperplane to separate between two class -1 and +1

4. The Standard Support Vector Machines

We begin with the simple linear support vector machines formulation as follows [1]:

$$\min_{w,\gamma,y} \upsilon e' y + \|w\|$$

$$\text{s.t. } D(Aw - e\gamma) + y \geq e \quad (1)$$

$$y \geq 0$$

Here, υ is a positive weight, $\|\cdot\|$ is an arbitrary norm and the *m x n* matrix A represents m given in R^n which belong to class 1 or -1 depending on whether the corresponding elements of the given *m x m* diagonal matrix *D* are 1 or -1 respectively.

We consider the problem of classifying *m* points in the *n*-dimensional real space R^n, represented by the m x n matrix A, according to membership of each point A_i in the classes 1 or -1 as specified by a given *m x m* diagonal matrix *D* with ones or minus ones along its diagonal. For this problem the standard SVM with a linear kernel is given by the following for some $v > 0$ [2]:

$$\min_{(w,\gamma,y)\in R^{n+1+m}} \quad ve'y + \frac{1}{2}w'w \quad (2)$$

$$\text{s.t. } D(Aw - e\gamma) + y \geq e$$

$$y \geq 0$$

Here *w* is the normal to the bounding planes :

$$x'w - \gamma = +1$$

$$x'w - \gamma = -1 \quad (3)$$

And γ determines its location relative to the origin. The first plane above bounds the class 1 points and the second plane bounds the class -1 points when the two classes are strictly linearly separable, that is when the slack variable $y = 0$. The linear separating surface is the plane

$$x'w = \gamma \quad (4)$$

Midway between the bounding planes (3). If the classes are linearly inseparable then two planes bound the two classes with a *soft margin* determined by a nonnegative slack variable *y*, that is:

$$x'w - \gamma + y_i \geq +1, \ for\ x' = A_i\ and\ D_{ii} = +1,$$

$$x'w - \gamma - y_i \leq -1, \ for\ x' = A_i\ and\ D_{ii} = -1, \quad (5)$$

4. The Smooth Support Vector Machines

4.1. SSVM with a linear kernel

In this smooth approach, the square of 2-norm of the slack variable *y* is minimized with weight $\frac{v}{2}$ instead of the 1-norm of y as in (2). In addition the distance between the planes (3) is measured in the (n+1)-dimensional space of $(w,\gamma) \in R^{n+1}$, that is $\frac{2}{\|(w,\gamma)\|_2}$. Thus using twice the reciprocal squared of the margin instead, yields our modified SVM problem as follows :

$$\min_{(w,\gamma,y)\in R^{n+1+m}} \frac{v}{2}y'y + \frac{1}{2}(w'w + \gamma^2)$$

$$\text{s.t. } D(Aw - e\gamma) + y \geq e \quad (6)$$

$$y \geq e$$

The constraint in (6) can be written

$$y = (e - D(Aw - e\gamma))_+ \quad (7)$$

Thus, we can convert the SVM problem (6) into an equivalent SVM which is an unconstrained optimization problem as follows :

$$\min_{w,\gamma} \quad \frac{v}{2}\|(e - D(Aw - e\gamma))_+\|_2^2 + \frac{1}{2}(w'w + \gamma^2) \quad (8)$$

This problem is a strongly convex minimization problem without any constraints. It is easy to show that it has a unique solution. However, the objective function in (8) is not twice differentiable which precludes the use of a fast Newton Method. We thus apply the smoothing techniques and replace x_+ by a very accurate smooth approximation that is given by $p(x,\alpha)$, the integral of the sigmoid function $\frac{1}{1+\varepsilon^{-\alpha x}}$ of neural networks, that is

$$p(x,\alpha) = x + \frac{1}{\alpha}\log(1 + \varepsilon^{-\alpha x}), \ \alpha > 0 \quad (9)$$

This p function with a smoothing parameter α is used here to replace the plus function of (8) to obtain a smooth support vector machine (SSVM) :

$$\min_{(w,\gamma)\in R^{n+1}} \Phi_\alpha(w,\gamma) := \min_{(w,\gamma)\in R^{n+1}}$$

$$\frac{v}{2}\left\| p\left(e-D(Aw-e\gamma),\alpha\right) \right\|_2^2 + \frac{1}{2}(w'w+\gamma^2) \quad (10)$$

We will now show that the solution of problem (6) is obtained by solving problem (10) with α approaching infinity. We take advantage of the twice differentiable property of the objective function of (10) to utilize a quadratically convergent algorithm for solving the smooth support vector machine (10).

Lee, et al [11] explained Newton-Armijo Algorithm for SSVM as follows :

Start with any $(w^0,\gamma^0)\in R^{n+1}$. Having (w^i,γ^i), stop if the gradient of the objective function of (8) is zero, that is $\nabla\Phi_\alpha(w^i,\gamma^i)=0$. Else compute (w^{i+1},γ^{i+1}) as follows:

i). **Newton Direction** : Determine direction $d^i \in R^{n+1}$ by setting equal to zero the linearization of $\nabla\Phi_\alpha(w\ ,\gamma\ )$ around (w^i,γ^i) which gives $n+1$ linear equations in $n+1$ variables :

$$\nabla^2\Phi_\alpha(w^i,\gamma^i)d^i = -\nabla\Phi_\alpha(w^i,\gamma^i)' \quad (11)$$

ii). **Armijo Stepsize** : Choose a step size $\lambda_i \in R$ such that :

$$(w^{i+1},\gamma^{i+1}) = (w^i,\gamma^i)+\lambda_i d^i \quad (12)$$

Where $\lambda_i = \max\left\{1,\frac{1}{2},\frac{1}{4},...\right\}$ such that :

$$\Phi_\alpha(w^i,\gamma^i)-\Phi_\alpha((w^i,\gamma^i)+\lambda_i d^i) \ge -\delta\lambda_i\nabla\Phi_\alpha(w^i,\gamma^i)d^i \quad (13)$$

Where $\delta \in \left(0,\frac{1}{2}\right)$.

4.2. Kernel Trick and Nonlinear Classification

In the real world problem, the training data cannot be linearly separated in the original space but may be linearly separated in a higher dimensional space after applying some nonlinear map. In new vector space, hyperplane separated the two class can be constructed. Cover theory declare "*when transformation is a nonlinear and dimension from feature space is high, then data in input space can mapping onto new feature space which the patterns in huge probability can separated linearly*". Using the kernel techniques we can achieve this goal without knowing the nonlinear map. There are many variants of kernel function, eg. polynomial kernel, radial basis function (RBF), sigmoid function. [5,12].

Table 1. Kernel function using in the system, the parameters p,σ,υ,c are given beforehand

Dot product	$k(x,x^{'}) = x.x^{'}$
Polynomial	$k(x,x^{'}) = (x.x^{'}+1)^p$
RBF	$k(x,x^{'}) = \exp(-\frac{\lvert x-x^{'}\rvert^2}{2\sigma^2})$
Sigmoid	$k(x,x^{'}) = \tanh(v(x.x^{'})+c)$

4.3. SSVM with a Nonlinear Kernel

We now describe how to construct a nonlinear separating surface which is implicitly defined b a kernel function. We briefly describe now how the generalized support vector machine (GSVM) [9] generates a nonlinear separating surface by using a completely arbitrary kernel. The GSVM solves the following mathematical program for a general kernel $K(A,A^{'})$:

$$\min_{(u,\gamma,y)} \quad ve'y + f(u) \quad (14)$$

$$\text{s.t.} \quad D(K(A,A')Du - e\gamma) + y \ge e$$

$$y \ge 0$$

Here $f(u)$ is some convex function on R^m which suppresses the parameter u and v is some positive number that weights the classification error $e'y$ versus the suppression of u. A solution of this mathematical program for u and γ leads to the nonlinear separating surface

$$K(x',A')Du = \gamma \quad (15)$$

to obtain the SSVM with a nonlinear kernel $K(A',A)$:

$$\min_{u,\gamma} \frac{v}{2}\left\| p(e-D(K(A,A')Du-e\gamma),\alpha)\right\|_2^2 + \frac{1}{2}(u'u+\gamma^2) \quad (16)$$

Where $K(A,A')$ is a kernel map from $R^{mxn}xR^{nxm}$ *to* R^{mxm}. We note that this problem, which is capable of generating highly nonlinear separating surfaces, still retains the strong convexity and differentiability properties for any arbitrary kernel. All of the results of the previous sections still hold. Hence we can apply the Newton-Armijo Algorithm directly to solve (16).

5. Experiment Result and Discussion

A. Face Database

The Code for eigenfaces is developed using C and Matlab, SSVM code developed using C++.

The first experiment is performed on the Cambridge Olivetti Research Lab (ORL) face database. Samples taken on two person randomly. Each person has ten different images, taken at different times. We show two individuals (in two rows) in the ORL face images in Fig. 4. There are variations in facial expressions such as open/closed eyes, smiling/nonsmiling. All the images were taken against a dark homogeneous background with the subjects in an up-right, frontal position, with tolerance for some side movements. There are also some variations in scale.

Fig. 4. Two Individuals (each in one row) in ORL face database. There are 10 images for each person.

B. Preprocessing

After we get the raw image, we must preprocess them. Figure 5 expresses the procedures. The procedures consisted of auto locating the centres of the facial features (eyes, nose, and mouth), translating, scaling, and rotating the face to place the center on specific pixel. Preprocessing has done to remove background and hair, histogram equalizing the non-masked facial pixels, and scaling the non-masked facial pixel to zero mean and unit variance, then we get the normal face [12]. Image size 100x100 pixels, and each image has 10.000 vectors. Image in greyscale area and saved in PGM format. Then Eigenfaces step to determine the mean image Ψ shown in figure 6.

Fig. 5. Image after Preprocessing

Fig. 6. Mean Image

C. Weight

The weight *w* is a representation images as a vector which is unit has a direction and value. We got the values from eigenfaces method (2.1).

D. Normalisation

The features vectors used into Smooth Support Vector Machines for classification and recognition for human faces. Before the learning phase, the previous features vector Ω^T is normalize to a range [-1,1] to input value for SSVM requirement, avoid computational problems and to facilitate learning[16].

E. Input SSVM

Input value after normalization is initialisation data as data training will processed. We implemented binary classification described in previous. Each SSVM was trained to distinguish between images from a single person/women (labelled +1) and other images from other person/man (labelled -1).

F. Processing SSVM

Processing shown in the flowchart as follows:

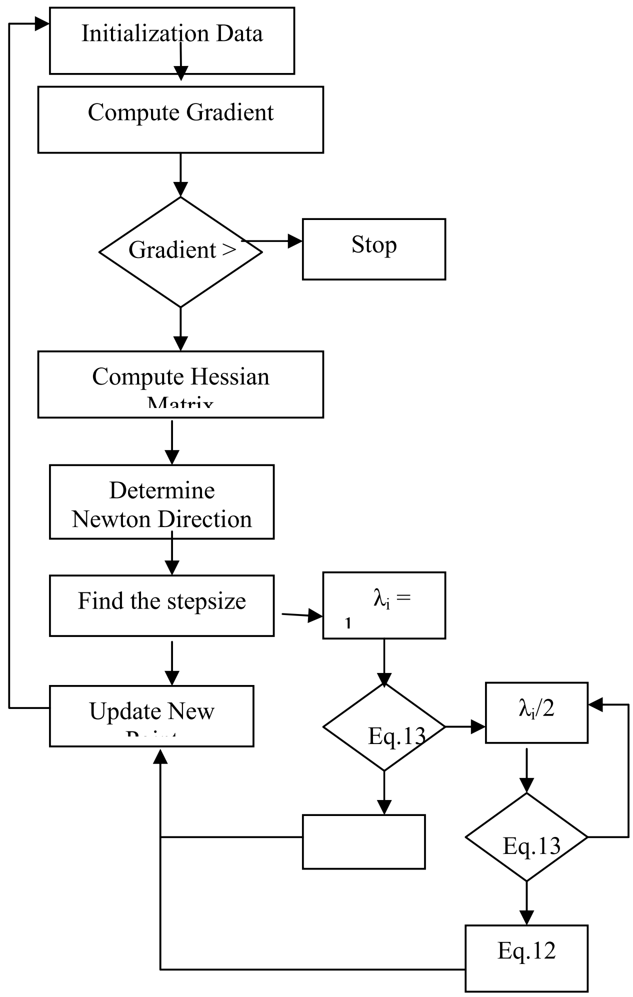

Fig. 7. Flowchart of SSVM

G. Parameter Selection

The performances of SVM depend on the combination of several parameters. They are capacity parameter *v*, the kernel type K and its corresponding parameters. We used RBF kernel function, since of its good general performance and the few number of parameters (only two : υ and μ). After the (υ,μ) is found, the whole training set is trained again the generate the final classifier. C.W.Hsu, et al [14] recommend a "grid-search" on υ and μ using cross validation to choosing good parameter. The range of (υ,μ) are $\upsilon = [2^{-5}, 2^{15}]$ and $\mu = [2^{-15}, 2^{3}]$.

H. Recognition/Prediction

Experiment carried out in two steps: training and testing. Training conducted in 100, 90, 80, 70, 60, 50 of percentage from data set. And testing in difference then calculate the average value to know how machine works effectively.

Table 2. Experiment result with RBF Kernel function

Training set (face)	Number of iteration	Testing set (face)
20	6	-
18	5	2
16	5	4
14	4	6
12	4	8
10	3	10

The experiment accuracy rate worked perfectly because the hyperplane completely separate linearly and experiment conducted in a few faces sample. And SSVM machine useful to solve problem for training in large data [10].

6. Conclusion and Future Work

In this study, we used eigenfaces to represent the features vector for human faces. The features are extracted from the original image represents unique identity used as inputs to present a new formulation in face recognition experiments using nonlinear Smooth Support Vector Machine with a classifier one-against-one approach. The proposed system shown competitive result, and demonstrates that methods are available.

In this paper the experiment result used face recognition in binary classification or two class. We can expand the work into multiclass classification by having a bigger database and more effort put in the training of dataset. Also we can compare this technique (SSVM) with other face recognition techniques in supervised learning such as Neural Network.

References

[1] Byun H., Lee S.W., " A survey on Pattern Recognition Application of Support Vector Machines", International Journal of Pattern Recognition and Artificial Intelligence, Vol.17, No.3, 2003, pp.459-486.

[2] Tsuda K., "Overview of Support Vector Machine" Journal of IEICE, vol.83, No.6, 2000, pp.460-466

[3] Vapnik V.N., "The Nature of Statistical Learning Theory", 2nd edition, Springer-Verlag, New York Berlin Heilderberg, 1999.

[4] Cristianini N., Taylor J.S., "An introduction to Support Vector Machines and other Kernel-Based Learning Methods", Cambridge Press University, 2000.

[5] Shigeo Abe, "Support Vector Machines for Pattern Recognition". Advances in Pattern Recognition Series Book. Springer-Verlag London Limited, 2005.

[6] Turk, M.,Pentland, A.: Eigenfaces for Recognition. Journal of Cognitive Neuroscience, Vol. 3, 1991.pp 72-86.

[7] M.Rizon, M. Firdaus, et., "Face Recognition using Eigenfaces and Neural Networks", American Journal of Applied Sciences 2 (6) : 1872-1875, 2006.

[8] Santi W.P, A.Embong., "Smooth Support Vector Machine for Breast Cancer Classification", IMT-GT Conference on Mathematics, Statistics and Applications(ICMSA), 2008

[9] O.L Mangasarian, "Generalized Support Vector Machines. In Smola, A., Bartlett, P., Scholkopf, B., and Schurrmans, D. editors, Advances in large Margin Classifiers, Cambridge , 2000, pp. 135-146.

[10] Y.J. Lee, and O.L. Mangasarian, "A Smooth Support Vector Machine" Journal of Computational Optimization and Applications 20, 2001, pp. 5-22.

[11] Guoqin Cui, Wen Gao., "Support Vector Machines for Face Recognition with two-layer Generated Virtual Data. "Proceedings of the 17th International Conference on Pattern Recognition (ICPR), 2004.

[12] Ethem alpaydin, Introduction to Machine Learning, The MIT Press Cambridge, Massachusetts London, England, 2004.

[13] Zhao et al. Face Recognition: A Literature Survey, ACM Computing Surveys, Vol.35, N0.4, December 2003, pp. 399-458.

[14] C.W.Hsu, C.C.Chang, and C.J. Lin, "A Practical Guide to Support Vector Classification", Department of Computer Science and Information Engineering, National Taiwan University, last updated, 2008.

[15] Demmel, J. and K.Veselic, Jacobi's method is more accurate than Q.R. Technical Report: UT-CS-89-88. 1-60,, 1989.

[16] Saad, P.,et al., A comparison of feature normalization techniques on complex image recognition. Proc. 2nd Conf. Information Technology in Asia, pp: 397-409