



Journal of Applied Sciences

ISSN 1812-5654

science
alert

ANSI*net*
an open access publisher
<http://ansinet.com>

Bayesian Face Detection Scheme Based on Support Vector Machine Random Samples

¹Lu Zhaogan and ²Sun Jiangfeng

¹College of Computer and Information Engineering, Henan University of Economics and Laws,
Zhengzhou, 450002, China

²College of Computer Science and Technology, Henan Polytechnic University, Jiaozuo, China

Abstract: Up to now, the support vector machine theory had been widely used to face detection by reason of its good capability to deal with non-linear high-dimension training problems by small samples. In fact, a large number of facial images were needed to train the support vector machine algorithms, whose complexity were increased by geometric series with training image numbers. So, the hybrid Monte Carlo method and Bayesian support vector machine were combined to find way out for the problems of high-dimension and long training time. The training method of support vector machine was implemented by the hybrid Monte Carlo method. Then, the scheme was verified by experiments that long training time was avoided and good face detection could also be reached.

Key words: Face feature, support vector machine, random samples, feature exaction

INTRODUCTION

Support Vector Machine (SVM) (Cortes and Vapnik, 1995) is one supervised study algorithm which is used widely in the statistical classifications and linear regression analysis. It could map one vector into one higher dimension space, where one super plane with the largest distance could be implemented for linear classification. Based on the statistics learning theory (Vapnik, 1998), the SVM could be constructed by minimizing the upper bound of expectation risk and obeys the rules of Structure Risk Minimization (SRM) other than the classic Experience Risk Minimization (ERM). Its design is refined and it has unique strong point to find the resolutions of the problems with small samples, non-linear and high dimensions. The problems that puzzled human beings such as model selection, dimension tragedy and local minimization in the machine learning field, could be resolved by the SVM theories. What's more, many classic learning models could be equivalent to the SVM learning algorithms. Therefore, the SVM and SLT are thought as the perfect basic frame for current machine theories.

Since, the appearing of the SVM, it's used widely as one effective intelligent machine learning method in the image analysis and disposing, especially for face identification and detections (Zhang *et al.*, 2011; Zeng and Ye, 2008; Zhang and Zhang, 2009). One face identification system was constructed based the SVM by

Zhao *et al.* (2012) in their research experiments (Zeng and Ye, 2008). The Gabor wavelet was used to conduct multiple scale transformation to extract the texture feature of face images (Zhang and Zhang, 2009) and then principal component analysis was done for the face identification based SVM algorithms. However, the face image features consist of one vector with high dimension which would lead to the issues that the computation complexity of the learning algorithms based SVM increased with the training samples (Shu *et al.*, 2011). In fact, a large number of training data must be used to reach the accuracy of face image identification algorithms. Therefore, the random sample method based on Monte Carlo and Bayesian support vector machine are combined to get rid of the issues of high learning dimension, long learning time in the process of face detection and identifications.

PROBLEM DESCRIPTION

In order to find the human information in one image, the face information should be figured out firstly. The common method is to conduct image segmentation and the segmentation area would be check if there contains face information features. In fact, the diversification of face shapes makes it impossible to describe one face by one modality function f and the face with different shape would be presented by different face feature values. If one

data variance x is used to describe face feature, the modality function f is the function of variance x . Obviously, if the sample face images are enough, the value ranges of human face feature vector x and the human face modality function could be determined by the statistics approaches. Thus, the statistical relations between the face feature vector x and the face modality function f could be setup. Here, the non-linearity relations between the feature vector x and modality function f is supposed and one linear relation $\phi(x)$ could be figured out if it was transformed by one non-linear transformation ϕ . Thus, the relations between the modality function f and the face features could be described by the following linear regression problems (Eq. 1):

$$f(x) = w\phi(x) + b \tag{1}$$

Here, the weight parameters w and the threshold b could be determined by the following programming Eq. 2 without restriction:

$$\min_{w,b} \left(\frac{1}{2} \|w\|^2 + C \sum_{n=1}^N L[t_n - f(x_n)] \right) \tag{2}$$

Here, the symbol L is the loss function. Thus, the linear regression problem could be resolved by the SVM with good results, where the loss function could be selected from the functions, i.e., Laplace, Huber and ϵ functions.

The results as shown by Eq. 2 presents the SRM approximate rules, where the first item is the VC item, the second item is the experiment risk item and the regular factor C is the bridge of the two items. In order to find the resolution of Eq. 2, the Lagrange multiplier α_n and α_n^* are used and it could be transformed into one secondary programming problem which could be referred (Pereira *et al.*, 2011). Now, the last SVM regression could be presented as the following Eq. 3:

$$f(x) = \sum_{n=1}^N (\alpha_n - \alpha_n^*) K(x, x_n) + b \tag{3}$$

THEORIES OF BAYESIAN SUPPORT VECTOR MACHINE

In fact, the Bayesian Support Vector Machine (BSVM) is one effective form of the classic support vector machine. According to the BSVM theories, the weight vector w is considered as the weight parameters and the notability frame (Zhang and Zhang, 2009). Another approach is that the modality function $f_N = (f(x_1), f(x_2), \dots, f(x_N))^T$ was used as the weight parameter. Here, the support vector machine that used f_N as weight

function is called as the Bayesian support vector machine, i.e., GSVM. The following gives the details about the theories of BSVM.

Deduction with w as weight parameters: Based on the conclusions of notability frames (Zhang and Zhang, 2009), suppose the transcendent probability could be described as following Eq. 4:

$$P(w) \propto \exp\left(-\frac{\alpha}{2} \|w\|^2\right) \tag{4}$$

Here, the symbol \propto is used without $=$, as it could ignore the normalized factors. Then, the likelihood probability is given as Eq. 5:

$$P(D|w) = \prod_{n=1}^N P(t_n | x_n, w) = \exp\left[-\frac{\beta}{2} \sum_{n=1}^N L[t_n - f(x_n, w)]\right] \tag{5}$$

According to the Bayesian formula (Shu *et al.*, 2011), the posteriori distribution of weight vector w could be evaluated on the Eq. 4-5:

$$P(w|D) \propto \exp(-M(w)) \tag{6}$$

Here:

$$M(w) = \frac{1}{2} \left(\alpha \|w\|^2 + \beta \sum_{n=1}^N L[t_n - f(x_n, w)] \right) \tag{7}$$

When the posteriori distribution $P(w|D)$ is maximized, the equation $-\log P(w|D)$ is minimized. With considering $-\log P(w|D) \propto M(w)$, the maximization $P(w|D)$ is equal to minimize $M(w)$, i.e.:

$$\min_{w,b} \frac{1}{2} \|w\|^2 + \frac{\beta}{2\alpha} \sum_{n=1}^N L[t_n - f(x_n)] \tag{8}$$

After the comparison of Eq. 8 and 2, it's obvious that they are equivalent. This means that the optimal value w_{MP} from the BSVM deduction is equivalent to the results from the secondary programming based on the SVM theories.

According to the context above, the expected output t distribution of BSVM, could also be given as Eq. 9:

$$P(t|x,D) = \int P(t|x,w)p(w|D)dw \tag{9}$$

From the Eq. 5, the item $P(t|x, w)$ in Eq. 9 is in the form of as following Eq. 10:

$$P(t|x, w) = \exp\left(-\frac{\beta}{2}L[t - f(x, w)]\right) \quad (10)$$

When it is put into the Eq. 9 with 6, we could get $P(t|x, D)$.

Deduction with f_N as weight parameters: When the f_N was used as weight parameters, the output $f(x, w)$ of BSVM is actually reformed as $f(x)$. Firstly, $f(x)$ is supposed as one Gauss progress with zero average and then its sample sequence $f_N = (f(x_1), f(x_2), \dots, f(x_N))^T$ are also obeying to the Gauss distribution, i.e.:

$$P(f_N) \propto \exp\left(-\frac{\alpha}{2}f_N^T K_N^{-1} f_N\right) \quad (11)$$

Here, the item K_N is one kernel matrix and the corresponding like hood function is given by Eq. 12:

$$P(D|f_N) \propto \exp\left(-\frac{\beta}{2}\sum_{n=1}^N L(f(x_n) - t_n)\right) \quad (12)$$

Then, f_N is used as weight parameter and the obvious posteriori probability is presented by Eq. 13:

$$P(f_N|D) \propto \exp(-M(f_N)) \quad (13)$$

Where:

$$M(f_N) = \frac{1}{2}\left(\alpha f_N^T K_N^{-1} f_N + \beta \sum_{n=1}^N L[t_n - f(x_n)]\right) \quad (14)$$

According the conclusion by Zhang *et al.* (2011) $f_N^T K_N^{-1} f_N = \|w\|^2$, the Eq. 14 is actually equivalent to Eq. 7. This means that the two deduction methods are consistent to each other. Thus, when the f_N is used as weight parameter, the distribution of the expected output t from BSVM would be Eq. 15:

$$P(t|x, D) = \int P(t|x, f_N)P(f_N|D)df_N \quad (15)$$

IMPLEMENTED BSVM BY MONTE CARLO RANDOM SAMPLING

According to the theories of Hybrid Monte Carlo (HMC) random sampling, it could be used to implement the BSVM algorithm as one simple way. Here, the Eq. 9 should be rewritten as Eq. 16:

$$P(t|x, D) \approx \frac{1}{N_c} \sum_{n=1}^{N_c} P(t|x, w_n) \quad (16)$$

Here, the item $w_n (n = 1, 2, \dots, N_c)$ could be considered as the N_c samples according to the posteriori distribution $P(w|D)$. When the square function was supposed, the expression $P(t|x, w_n)$ could be presented by Eq. 17:

$$P(t|x, w_n) = \exp\left(-\frac{\beta}{2}[t_n - f(x, w_n)]^2\right) \quad (17)$$

According to the two equations above, the expected output distribution $P(t|x, D)$ of BSVM could be evaluated approximately. However, other three numerical methods, i.e., calculus of variations, required to obtain the approximate posteriori distribution $P(w|D)$ before the approximate the expected output distribution $P(t|x, D)$. This process presents the strong points of the hybrid Monte Carlo methods.

Now, the key problem is how to conduct the samples from the posteriori distribution $P(w|D)$. Its expression is given by Eq. 6, where there are two undefined parameters a and b . So, the optimal values α_{MP} and β_{MP} for the two parameters α and β should be evaluated before conducting sample operations. The following content would present the method to obtain α_{MP} and β_{MP} and the details about the practical sample operations:

- **Optimal super parameters for posteriori distribution:** Firstly, we would consider α and β as super parameters, then the values α_{MP} and β_{MP} were computed by the secondary deduction by the notability frame and the flow chart could be presented by Fig. 1. According to the Fig. 1, the optimal values α_{MP} and β_{MP} are obtained by iteration methods, where the symbol M_n is one constant threshold. The definitions of $E_w(w_i)$ and $E_D(w_i)$ are given as the following two equations:

$$E_w(w_i) = \frac{1}{2}\|w_i\|^2, E_D(w_i) = \frac{1}{2}\sum_{n=1}^N [t_n - f(x_n, w_i)]^2$$

- **Monte Carlo sample process:** When the optimal values α_{MP} and β_{MP} for the super parameters α and β , the sample operation could be done on the posteriori distribution $P(w|D)$. In fact, the Metropolis algorithm could be used for the sampling but the classical random walk issue troubles the performance. In order to avoid this issue, the hybrid Monte Carlo method is used to conduct sample operation. Here, the hybrid Monte Carlo method could be considered as one special Monte Carlo method

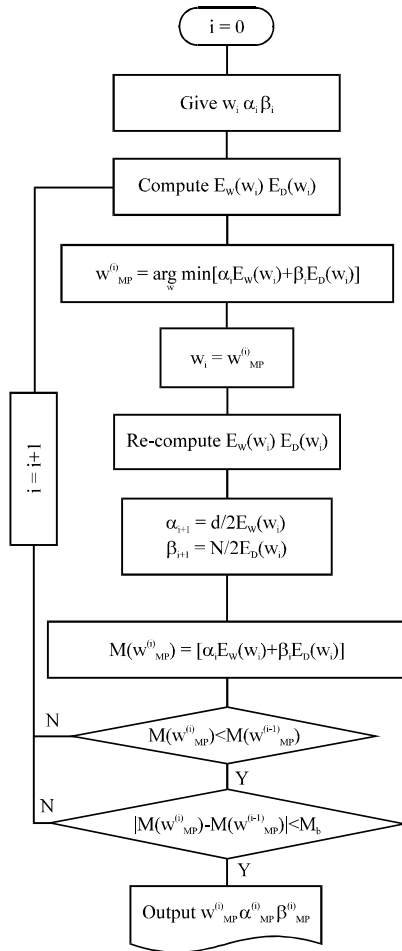


Fig. 1: Process to evaluate the optimal values α_{MP} and β_{MP} of the super parameters of posteriori distribution

Compared to the Metropolis algorithm, the HMC method improves two aspects. Firstly, one more ‘flog leap’ step is used to avoid the random walk phenomenon. Secondly, the supplement variance u is introduced and the posteriori distribution $P(w|D)$ was extended to $P(w, u|D)$. This would lead to easier sample operation from $P(w, u|D)$ other than $P(w|D)$. Thus, the samples $\{(w_i, u_i)\}_{i=1}^N$ could be obtained from $P(w, u|D)$ and its parameters u_i would be ignored.

The HMC sample flow chart could be displayed by Fig. 2a and it’s similar to that of Metropolis algorithm except for one more ‘frog leap’ step.

When the start point of ‘frog leap’ step is given, the values x_M and u_M would be generated after M steps of ‘frog leap’ and they would be used as the $i+1$ sample according to the probability A , i.e., x_{i+1} and u_{i+1} . The step length r of the frog leap should also be considered carefully. If the step length r is too long, the received

probability A is too small while the small r value, the frog leap steps should be increased correspondingly. According to the Gualdi’s research results in his report (Gualdi *et al.*, 2012), the frog leap step length r should change once when the value i is increased once. Thus, the various frog leaps other than constant step should be adopted. Here, the constant step length is used in order to simplify the computations.

NUMERICAL RESULTS

The training database with about 200 face images was constructed and these images were disposed where only the face features were conclude, i.e., eyes, brows, nose, mouth and so on. We could identify the face images on the base of these basic features when the given images are segmented into different areas. Twenty human face images were displayed in Fig. 3 for the constructed face image database. From these typical face images, they have different features for their colors, emotions and shapes (i.e., squares, flanks and incline) with common typical feature of features.

In fact, the effects of human face color could be get ride by gray disposing on these images. However, the pivotal problem is how to evaluate the effects of human emotions and shapes on the human face feature extractions. This indicates actually that the face features extracted by digital image processing algorithms contains the face shape information other than face features. But it’s still one issue to be resolved for any possible effective algorithms. Therefore, the circle degree values of human face images are used to measure the face shape information quantities. Here, the Gabor texture features and Hu invariant variables are used as the face features and the hybrid Monte Carlo random sample operator and the BSVM algorithms are combined to figure out the statistical rules between the human face features and face shapes. Thus, these features could be used to justify the given images whether could be fit for the extracted statistical rules. If these images are fit for the rules, they would be considered as the face images, vice versa. Subsequently, for the about 40 randomly extracted images from the training databases, their Gabor texture features and seven element invariants are extracted for the description, where the Gabor texture features consists of several filter banks with about three different scales and four rotation orientation. These Gabor filter banks are used to filter on the face images and their average values and variances are obtained, correspondingly. While, the Hu invariant moments are obtained by the combination of zeros, one, two, three moments.

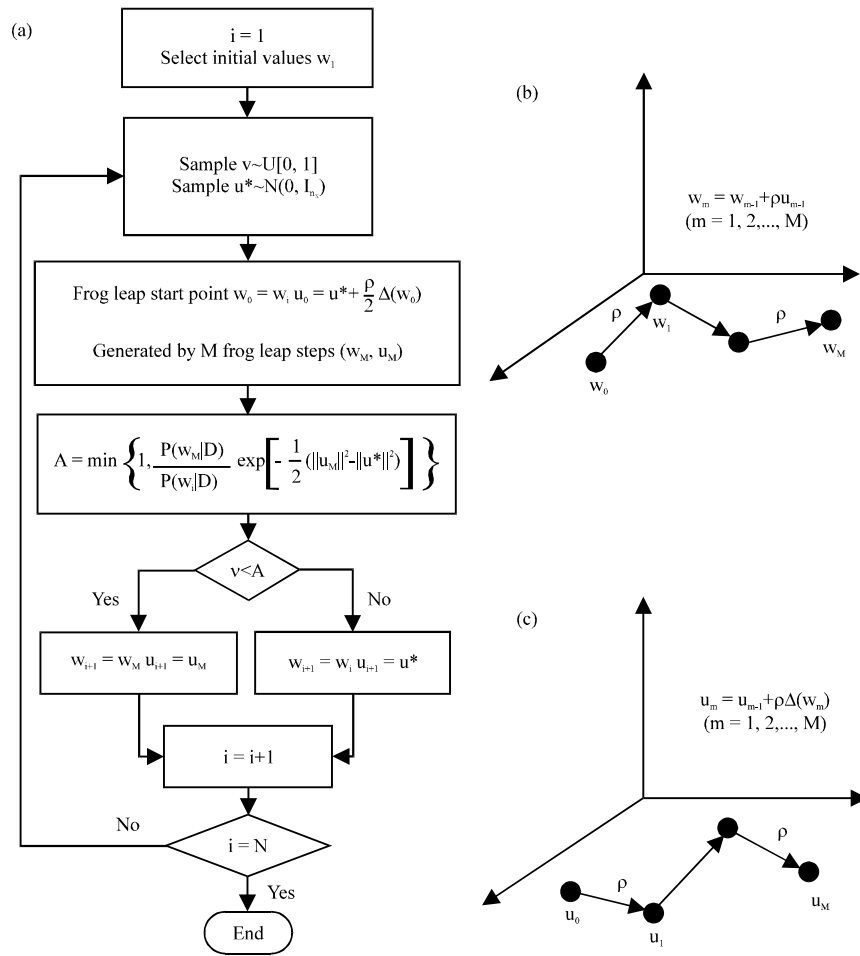


Fig. 2(a-c): HMC flow chart and frog leap and its sketch map (a) Flow chart, (b) M step frog leap to generate w_M from w_0 and (c) M step frog leap to generate u_M from u_0

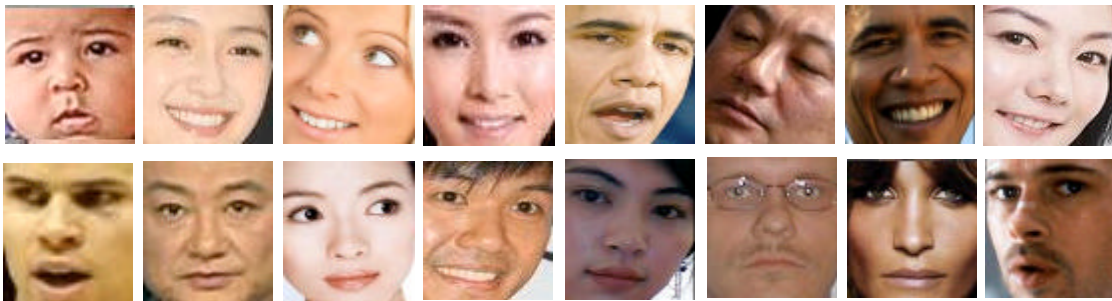


Fig. 3: Typical face images in the constructed training databases

In order to verify the proposed face identification methods, about 50, 75, 100, 125, 150, 175 and 200 face images were used to train the corresponding BSVM algorithms and the statistical rules between the face

features and face shapes are extracted. Here, there are two training methods, i.e., the hybrid Monte Carlo method and the data regress methods. The training method based data regress is only supported by the

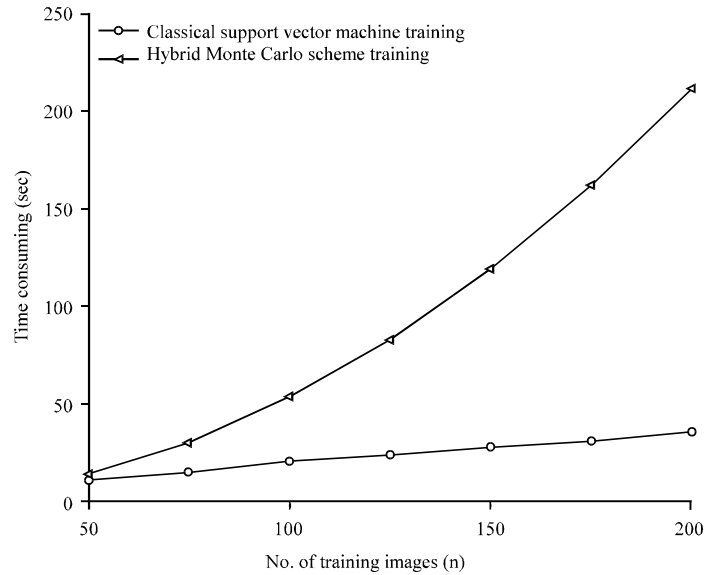


Fig. 4: Time cost comparison of hybrid Monte Carlo random sampling and common training method

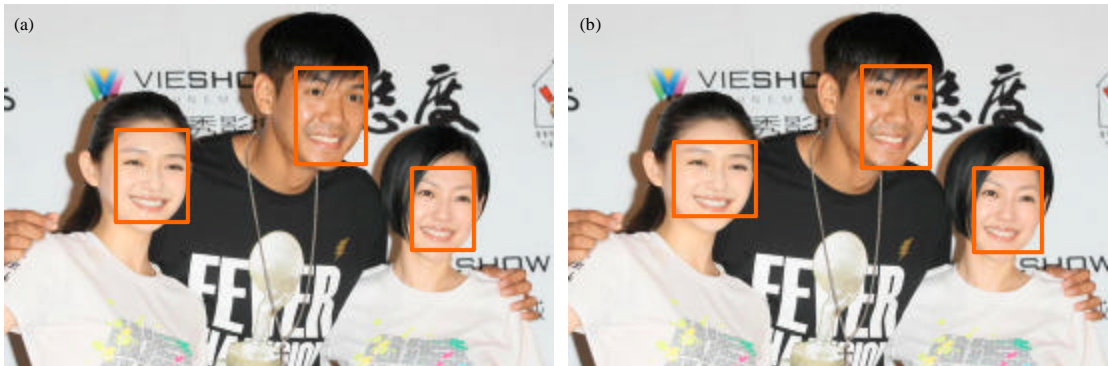


Fig. 5(a-b): Face identification results for the (a) Hybrid Monte Carlo random sample and (b) Common training method

support vector machines. In fact, if about $n \times N$ images were used to train the Bayesian support vector machines, all the $n \times N$ face images would be used to training process by the data regression method with its computation complexity exponent increase with its training image numbers. However, for the hybrid Monte Carlo random sample methods, about N images were extracted stochastically from $n \times N$ images to train the Bayesian support vector machines for about n times. Then, the results of the BSVM with n time training would be averaged to obtain the Bayesian support vector machines. And, the summed times for all the n times are the total times for all the hybrid Monte Carlo random sample methods. At last, the cost time to train the BSVM with about 50, 75, 100, 125, 150, 175 and 200 face images, was

plotted in the Fig. 4 in the computer with Intel(R) Core(TM)2 Duo CPU T5800 at 2.00 GHz and 1 GB memory.

With the increase of training samples of human face images, the hybrid Monte Carlo random sample method could reduce the cost time of the Bayesian support vector machines as the Fig. 4 indicated. Thus, by comparing the training results of the two different Bayesian support vector machine algorithms, two human being pictures are used to face identifications after the two pictures were segmented beforehand. Their face identification accuracies are showed by the Fig. 5-6, respectively. The images with large face with small image complexity could be identified with good accuracy. However, for the images with small faces and several human beings, the hybrid Monte Carlo



Fig. 6(a-b): Human being face identification results based (a) Hybrid Monte Carlo random sample and (b) Common training methods

Bayesian SVM method could reach to good identification effect. The common training method could not detect the small human being faces.

CONCLUSION

With considerable special advantages of the support vector machine in the resolution of problems in cases of small samples, non-linearity and high dimension learning issues, the SVM algorithm has been used widely in the human being face identification and detections. However, a large number of face images should be used to conduct the training on the support vector machines with the exponent increased complexity with the training samples. Therefore, the statistical learning theories and the support vector machines are integrated to conduct face image identifications. Here, the hybrid Monte Carlo random sample is implemented to train the Bayesian support vector machines and good image identification efficiency are achieved with low training times.

REFERENCES

Cortes, C. and V. Vapnik, 1995. Support-vector networks. *Mach. Learn.*, 20: 273-297.
Gualdi, G., A. Prati and R. Cucchiara, 2012. Multistage particle windows for fast and accurate object detection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 34: 1589-1604.

Pereira, E., H. Gomes, E. Moura, J. Carvalho and T. Zhang, 2011. Investigation of local and global features for face detection. *Proceedings of the IEEE Symposium on Computational Intelligence for Multimedia, Signal and Vision Processing*, April 11-15, 2011, Paris, pp: 114-121.
Shu, S., J. Luo, C. Xu and B. Sun, 2011. A new method of face recognition base on support vector machine. *Comput. Simul.*, 28: 280-283.
Vapnik, V., 1998. *Statistical Learning Theory*. Wiley, New York, USA.
Zeng, Y. and B. Ye, 2008. Face recognition based on support vector machines. *Comput. Eng. Appl.*, 44: 182-184.
Zhang, G. and J. Zhang, 2009. Multi-scale PCA support vector machine algorithm based on gabor wavelet for face recognition. *J. Data Acquisition Process.*, 24: 437-444.
Zhang, G. Y., L. Y. Guo, J. H. Wu, H. M. Li and G. Q. Guo, 2011. Face detection using Gabor wavelet and SVM. *Proceedings of the International Conference on Computer Application and System Modeling*, Volume 2, November 18-20, 2011, XiaMen, China, pp: 588-590.
Zhao, H., Y. Wang, X. Zhang and N. Xue, 2012. Human being faces identification with the combination of wavelet transformation and feature weighting. *J. Image Graphics*, 17: 1522-1527.