The Kernel Classification-Based Metric Learning in Face Verification

Siew-Chin Chong, Thian-Song Ong
Faculty of Information Science & Technology,
Multimedia University, Melaka, Malaysia.
chong.siew.chin@mmu.edu.my

Abstract**—In this paper, a kernel classification distance metric learning framework is investigated for face verification. The framework is to model the metric learning as a Support Vector Machine face classification problem, where a Mahalanobis distance metric is learnt in the original face feature space. In the process, pairwise doublets that are constructed from the training samples can be packed and represented in a means of degree-2 polynomial kernel. By utilizing the standard SVM solver, the metric learning problem can be solved in a simpler and efficient way. We evaluate the kernel classification-based metric learning on three different face datasets. We demonstrate that the method manages to show its simplicity and robustness in face verification with satisfactory results in terms of training time and accuracy when compared with the state-of-the-art methods.**

Index Terms**—Face Verification; Kernel Classification; Metric Learning; Support Vector Machine; Doublets.**

I. INTRODUCTION

Face verification has attracted enormous interest among the computer vision and biometric researchers in the past few decades. The main factor of its popularity is due to the wide range and non-intrusiveness of its practical applications such as the law-enforcement and military applications. Face verification aims to determine whether a given pair of face images is from the same person or different person. It is crucial that the significant variations of a face image caused by varying aging, lighting, pose, expression and others to be handled well in order to satisfy the real-world scenarios.

Metric learning techniques play an important role in many machine learning tasks such as image retrieval, face verification, image identification and activity recognition to improve their performance. Metric learning techniques have been extensively applied in face verification [1]-[4] over the years. A new distance metric is always learned from the training samples to effectively measure the similarity between face samples by enlarging the similarity of similar pairs and reducing the similarity of the dissimilar pairs. There are various types of metrics can be learned, depending on the objective functions of the metric learning algorithms. Not limited to Mahalanobis metrics [5,6], there are also similarity metrics [7,8], nonlinear distance metrics [9] and multiple metrics [10].

Although numerous metric learning algorithms have been proposed and proved to be useful, there are still problems to be further investigated. Certain metric learning methods which require all pairwise distances between points [11] are inefficient to solve large-scale problems. There is also situation where some metric learning methods relying on the additional information might be impractical in some scenarios such as verifying a foreigner who does not have any identity information in data bank, or an intruder who tries to abuse the system repeatedly. In addition, the ability of recasting metric learning as a supervised learning problem remains as an interesting topic to further study.

With considerations of the mentioned issues and inspired by the work of [29], a kernel classification-based metric learning, dubbed Support Vector Machine Metric Learning (SVMML) is modified to fit the face verification pipeline in order to learn a Mahalanobis distance metric of the original face feature space. This framework prepares a unified model to be integrated in the existing metric learning method, such as large margin nearest neighbor (LMNN) [5].

Experiments results are reported based on two types of settings: standard and restricted wild face verification protocol. The former one is implemented by using FERET [13] and AR [14] datasets, where the number of classes, the number of images per class and the class of a particular image belongs to are considered in the learning process. The latter case is implemented by using Labeled Faces in the Wild (LFW) [15] dataset, where only same or not same person labels are used in training and no other information about the person is available.

The paper reviews the related works in Section 2 and lists the contribution of SVMML in face verification in Section 3. Section 4 presents the work in detail. Experimental results are presented in Section 5. Finally, Section 6 concludes the paper.

II. RELATED WORKS

A good metric learning algorithm should equip the ability to emphasize relevant dimensions while reducing the influence of non-informative dimensions [16]. When learning a Mahalanobis matrix, attention should be paid to three criteria. The first criterion is that the learning algorithm should be global. All the useful samples should be used for training as many as possible. However, due to the limitation of algorithm efficiency, not all the samples could be trained. This in turn may cause the overfitting problem. The second criterion is that the labels of the training samples should be as weak as possible. In our real life scenarios, it is always difficult to obtain strict label of the training samples. Compared with class labels, data pair labels (similarity/ dissimilarity) are weaker and more practical in metric learning applications. The third criterion is that the metric learning algorithms should be scalable with respect to the size of the training samples. In another word, the algorithmic efficiency should be high.

The famous metric learning algorithm introduced by Weinberger et al. [5] learns a transformation matrix in order to improve the k nearest neighbour (kNN) classification. The objective is to maintain the consistency in the sample's neighbourhood and keeping a large margin at the boundaries of different categories. Kumar et al. [17] proposed an extension version of LMNN, named ILMNN for transformation invariant classification. On the other hand, Information Theoretic Metric Learning (ITML) [18] is designed to deal with general pair-wise constraints, which maximizes the differential entropy of a multivariate Gaussian subject to constraints on the associated Mahalanobis distance. ITML is fast and scalable but the constraints of the model are restricted. An extended ITML [19] is presented by Saenko et al. for visual category domain adaptation. Logistic Discriminant Metric Learning (LDML) [6] learns the metric from a set of labelled image pairs. Hieu et al. [2] implements Cosine Similarity Metric Learning (CSML) which leads to a fast gradient-based optimization algorithm.

Furthermore, Perez-Suay et al. [20] proposed a batch and online scheme for metric learning based on margin maximization. Its metric learning method utilizes the doublet-based constraints but its model is different with [29]. There are also methods proposed for learning the nonlinear distance metric [9],[10] and multiple distance metric [21].

III. CONTRIBUTION

Several issues of the existing metric learning methods as discussed motivate us in adapting the SVMML in face verification problem. The SVMML offers several merits in face verification such as:

- Provide a platform for developing a new metric learning algorithm by adopting the standard SVM solvers.
- Transferrable weak supervised metric learning.
- Scalable to big data.
- Simple structure of learning pipeline promises faster processing time

IV. SYSTEM OVERVIEW

In this section, the overview of the proposed face verification system is explained. Firstly, the cropped images are filtered with DoG filter [22] to enhance the image quality and suppress the noises. Then, each face image is partitioned into several local regions and the face descriptors are extracted from each region independently via OCLBP [24]. Due to the large dimension of the face descriptors, dimension reduction is needed. Two types of dimension reduction techniques: PCA [25] and WPCA [26] are applied separately in order to compare their performance in the flow to obtain the higher accuracy. The reduced features are then processed by the SVMML approach to produce the transform matrix.

A. *Face Descriptor*

OCLBP is an extended version of the original LBP [23], where it is computed with overlapping blocks and repeated with different sizes and radius [24]. The main reason to choose OCLBP as the face descriptor is due to its simplicity and speed in the implementation.

Given an input image and a set of parameters as in Equation (1), local descriptors can be generated:

$$L = \{(m_i, n_i, v_i, h_i, p_i, r_i)\}_{i=1}^{k} \tag{1}$$

where image is divided into the blocks of size $m_i \times n_i$ with vertical overlap of v_i, horizontal overlap of $h_i$using the operator $LBP_{p_i,r_i}^{U2}$with U2, the uniform patterns, the number of points p that are uniformly sampled over a circle of radius r. Fig. 1 illustrates the circular neighbourhoods by (p, r). The computation is repeated for k configurations in L.

All the descriptors are to be concatenated to form a single vector which is the resulted OCLBP descriptor. The resulted OCLBP descriptor will then be processed with PCA or WPCA to reduce its large dimension.

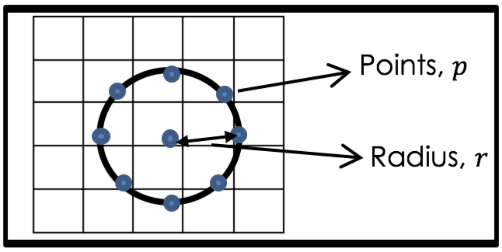

Figure 1: Illustration of circular neighborhoods (8,1)

B. *Dimension Reduction*

PCA is a conventional dimensional reduction technique that forms the basis of numerous studies in face recognition literature. The use of PCA was proposed by Turk et al. [25]. PCA-based algorithms are popular because of the ease of implementing them and their reasonable performance level [27]. The PCA extracts the eigenvectors corresponding to the largest eigenvalues which serve as the principal components by computing the covariance matrix of the feature set. Suppose that there are N training samples of n dimension for each vector, $\{x_i\}_{i=1}^{N} \in \mathbb{R}^n$ and m is the mean of the total training samples. The covariance matrix of the training can be defined as follow:

$$C = \frac{1}{N}\sum_{i=1}^{N}(x_i - m)(x_i - m)^T \tag{2}$$

Recently Weighted PCA (WPCA) is a famous tool for dimension reduction among the researchers. WPCA is an extended version of PCA, which it considers the weighted coefficient to suppress the responses from larger eigenvalues. It emphasizes on the training samples that are very close to the test sample and reduces the influence of the other training samples. Suppose that there are N training samples of n dimension for each vector, $\{x_i\}_{i=1}^{N} \in \mathbb{R}^n$ and let t be the test sample. The weighted covariance matrix of WPCA is as follows:

$$C_W = \frac{1}{N}(w_i x_i)(w_i x_i)^T \tag{3}$$

where $w_i = exp\left(-\frac{maxd - dist(x,t)}{\mu}\right)$, $maxd$ is the maximum value of the distance between $x_i \dots x_N$ and t, $dist(x, t)$ is the distance between x_i and t, while μ is a positive constant. w_i is called as weight coefficient. WPCA takes the eigenvectors corresponding to the first d largest eigenvalues of $C_W$as projection axes and exploits these projection axes to transforms the sample into a d-dimensional space.

C. Doublets and Pairwise

SVMML considers a set of constraints imposed on the doublets or pairwise of training face samples to learn the distance metric. There are two face verification settings in our experiments: standard face verification and wild face verification. For standard face verification, SVMML is operated on the doublets. Doublets of a training sample are composed of a nearest similar neighbour m_1 and a nearest dissimilar neighbour m_2. Let $D = \{(x_i, y_i)|i = 1,2, \dots, n\}$ be a training dataset, where vector $x_i \in \mathbb{R}^d$ represents the ith training sample and scalar $y_i$denotes the class label of x_i. A doublet $(m_1 + m_2)$ is built from any two samples extracted from D and a label e is given to this doublet where $e = 1$ if $y_i = y_j$ and $e = -1$ if $y_i \neq y_j$. By combining all the doublets constructed from all training samples, a double set is formed by $\{z_1, \dots, z_{N_d}\}$, where $z_l = (x_{l,1}, x_{l,2}), l = 1,2,.., N_d$. The label of doublet of z_l is denoted by e_l.

For the wild face verification, SVMML learns the pairwise of the training samples. Due to the lack of information on the class label and the number of classes in the LFW setting, we could only generate the pairwise based on the restricted protocol of the LFW. Among the matched pairs, we set $e = -1$; while for the mismatched pairs, we set $e = 1$. Same with the standard face verification setting, all the label of the pairwise generated from the training samples are pooled and denoted by e_l.

D. Kernel Classification-based Metric Learning

Distance metric learning can be readily formulated as a kernel classification problem by incorporating the degree-2 polynomial functions which can be operated on the pairs of doublets/ pairwise.

Let x_i and x_j be the two training samples, degree-2 polynomial kernel can be defined as:

$$\begin{aligned} K(x_i, x_j) &= tr(x_i x_i^T x_j x_j^T) \\ &= tr(x_i^T x_j)^2 \end{aligned} \tag{4}$$

where $tr(\cdot)$ is the trace operator of a matrix. From here, it is said to fulfil the Mercer's condition [28].

In order to apply the kernel function as defined in Eq. (4) to a pair of doublets/ pairwise, we can extend the degree-2 polynomial kernel as:

$$\begin{aligned} K_D(z_i, z_j) &= tr\Big(\big(x_{i,1} - x_{i,2}\big)\big(x_{i,1} - x_{i,2}\big)^T \\ &\quad \big(x_{i,1} - x_{i,2}\big)\big(x_{i,1} - x_{i,2}\big)^T\Big) \\ &= \Big(\big(x_{i,1} - x_{i,2}\big)^T\big(x_{j,1} - x_{j,2}\big)\Big)^2 \end{aligned} \tag{5}$$

where $z_i = (x_{i,1}, x_{i,2})$ and $z_j = (x_{j,1}, x_{j,2})$ are the pair of doublets/ pairwise. With Equation (5), a decision function is learnt to decide whether the two samples of a doublet have the same class label.

With the introduction of the degree-2 polynomial kernels, the task of metric learning can be solved. Any kernel classification method can be used to learn the kernel classifier as follows:

$$g_d(z) = sgn\Big(\sum\nolimits_l e_l \propto_l K_D(z_l, z) + b\Big) \tag{6}$$

where $z_l, l = 1,2,.., N$ is the doublet of the training dataset, $z = (t_i, t_j)$ is the test doublet, $\propto_l$ is the weight and b is the bias.

By substituting Eq. (5) into Eq. (6) for doublets,

$$\begin{aligned} &\sum\nolimits_l e_l \propto_l tr\Big(\big(x_{l,1} - x_{l,2}\big)\big(x_{l,1} - x_{l,2}\big)^T \\ &\quad \big(t_i - t_j\big)\big(t_i - t_j\big)^T\Big) + b \\ &= \big(t_i - t_j\big) W \big(t_i - t_j\big)^T + b \end{aligned} \tag{7}$$

where W is the matrix of Mahalanobis distance metric.

$$W = \sum\nolimits_l e_l \propto_l \big(x_{l,1} - x_{l,2}\big)\big(x_{l,1} - x_{l,2}\big)^T \tag{8}$$

The kernel decision function in Equation (6) can be used to determine whether the test doublets are similar to each other or not.

On the other hand, the SVM-like model can be adopted to learn the distance metric:

$$\min r(W) + p(\xi) \tag{9}$$

$$s.t. f_l^{(d)}\Big(\big(x_{l,1} - x_{l,2}\big)W\big(x_{l,1} - x_{l,2}\big)^T, b, \xi_l\Big) \geq 0 \tag{10}$$

$$\xi_l \geq 0 \tag{11}$$

where $r(W)$ is the regularization term, $p(\xi)$is the margin loss term, the constant $f_l^{(d)}$can be any linear function. If the Frobenius norm is applied to regularize W and the hinge loss penalty ξ, the model in (9) will become the standard SVM [12].

To build SVMML, Frobenius norm regularizer is set as $r_{KCML}(W) = \frac{1}{2}\|W\|_F^2$, and the margin loss term is set as $p_{KCML}(\xi) = C \sum_l \xi_l$. The SVMML can be redefined as:

$$\min \frac{1}{2}\|W\|_F^2 + C\sum\nolimits_l \xi_l \tag{12}$$

$$\text{s.}t. e_l\Big(\big(x_{l,1} - x_{l,2}\big)W\big(x_{l,1} - x_{l,2}\big)^T + b \geq 1 - \xi_l\Big) \tag{13}$$

$$\xi_l \geq 0, \forall l \tag{14}$$

where $\|\cdot\|_F$ is the Frobenius norm. The Lagrange dual problem of the proposed SVMML is as below:

$$\max_{\alpha} -\frac{1}{2}\sum\nolimits_{i,j} \alpha_i \alpha_j e_i e_j K_D(z_i, z_j) + \sum\nolimits_i \alpha_i \tag{15}$$

$$s.t. 0 \leq \alpha_l \leq C, \forall l \tag{16}$$

$$\sum\nolimits_l \alpha_l e_l = 0 \tag{17}$$

By using the existing SVM solvers, the problem can be easily solved. A two-step greedy strategy is applied for metric learning. The positive semi-definite constraint is neglected and the LibSVM is used to learn a preliminary matrix W, which is mapped onto the space of positive semi-definite matrices.

V. EXPERIMENTAL RESULTS

Experiments are conducted based on two types of settings: standard and wild face verification, to evaluate the performance of SVMML in constrained and unconstrained setting for face verification. FERET [13] and AR [14] datasets are used in the standard face verification settings, while LFW [15] is used in the wild face verification setting.

For FERET dataset, 2000 images are randomly selected from the original dataset. There are 200 classes in the subset, where each class consists of 5 training images and 5 testing images.

For AR dataset, a total of 1980 images for 99 classes have been randomly selected from the original dataset. For each class, there are 10 training images and 10 testing images.

We followed the standard LFW face verification "Restricted View 2" protocol. LFW consists of a total of 13,233 face images from 5,749 individuals. There are 6,000 different face image pairs arranged randomly from the sets to form 5,400 pairs (2,700 matched pairs and 2,700 mismatched pairs) for training and 600 pairs (300 matched pairs and 300 mismatched pairs) for testing.

In our experiments, all the original face images are cropped into 73 x 61 pixels. In order to enhance the quality of the face image and to suppress the Gaussian noises, DoG filter is applied on each face image. To be fair in comparison, our proposed system does not make use of any outside training data. Yet, none of the further type of preprocessing such as pose estimators or 3D modeling is being used in the experiments.

Experiments are conducted using different face datasets with different dimension reduction techniques, different dimensions and different number of training sets. Performance is evaluated based on accuracy in percentage and training time in seconds. In Table 1, we compare SVMML with the state-of-the-art metric learning methods in face verification based on the standard and wild face settings. There is very less similar framework in standard verification setting to solve the supervised problem. Here we compare LMNN with our SVMML and it is interestingly shown that SVMML is able to boost the accuracy rate up to 97% on FERET while LMNN can achieve at 89.89%. For wild face verification setting, SVMML is able to outperform the listed state-of-the-art methods with 77.81% of accuracy rate. Incorporating the classification power in the metric learning process helps in improving the performance compared to the well-known methods such as LDML, NOWAK, ITML which achieve 72.8%, 73.93% and 76.18% respectively.

Table 1
Comparisons of accuracy for various metric learning methods using standard and wild face verification settings.

Standard Face Verification Setting	
Learning Method	Accuracy (%)
LMNN [5]	89.89
SVMML	97

Wild Face Verification Setting	
Learning Method	Accuracy (%)
CSML [2]	71.12
LDML [6]	72.8
NOWAK [30]	73.93
ITML [19]	75.2
MERL+NOWAK [31]	76.18
SVMML	77.81

Figure 2 illustrates the performance of SVMML with the assistance of different dimension reduction techniques: PCA or WPCA, on the standard face datasets, FERET and AR with different reduced dimensions. The best result of FERET can be obtained at 96.2% and 97% by applying PCA and WPCA respectively, when reducing the feature length to 100. On the other hand, AR dataset also achieves the best accuracy at the length of 100 for both the PCA and WPCA, with 86.77% and 88.38% respectively. From the experiments, it is proven that WPCA may assist better in the verification process since it suppresses the responses from larger eigenvalues.

Figure 3 shows the performance of SVMML with WPCA using different number of training sets from the LFW datasets on different dimensions (number of blocks=49; reduced dimension={2,5,10,12,15,20}). Each training sets consists of 300 pairs of images. The highest accuracy rate can be achieved at 77.81% with the dimension of 735 with 6 training sets. This dimension 735 is based on the block-based concept produced by OCLBP, which consists of 49 blocks and 15 feature length of each block. The WPCA is applied on each of the block separately and the reduced features are concatenated before sending to SVMML for learning.

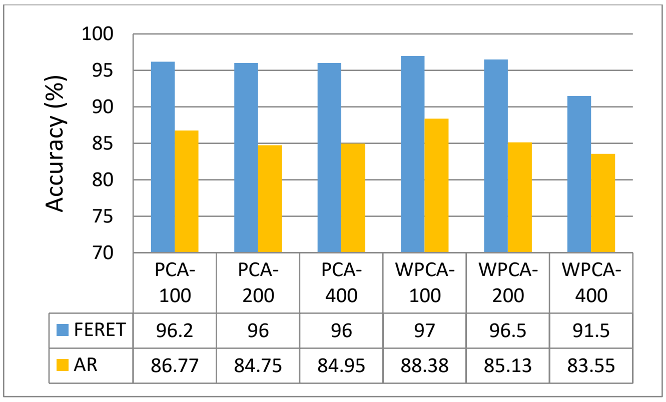

Figure 2: The Performance of SVMML with PCA or WPCA on FERET and AR datasets with different dimensions.

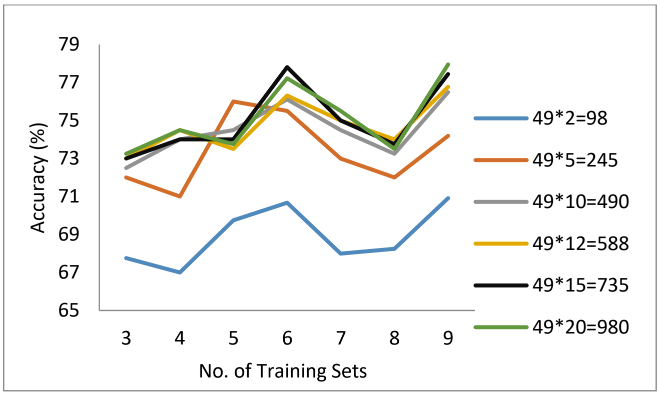

Figure 3: The Performance of SVMML with WPCA on LFW datasets with different dimensions.

Table 2 demonstrates the training time in seconds of SVMML on different face datasets with different number of training images and different dimensions of features. It is obviously shown that the training time for SVMML is much faster than LMNN, even with large training samples in LFW. For LFW, it takes around 5 to 32.5 seconds to train 10800 images. In addition, the training time for the FERET and AR datasets fall within the range of 0.14 seconds to 1.45 seconds. This is considerably fast to train around 1000 images. The short training time proves the simplicity and efficiency of SVMML.

Table 2
Training time of SVMML on difference face datasets with different number of training images and dimensions.

Method	Face Datasets	No. of Training Images	Feature Dimensions	Training Time (Sec)
LMNN [5]	YALE-B	1690	300	480
	ORL	280	200	66
SVMML	FERET	1000	100	0.359
		1000	200	0.547
		1000	400	1.454
	AR	990	100	0.141
		990	200	0.451
		990	400	0.844
	LFW	10800	98	4.938
		10800	245	5.531
		10800	490	11.816
		10800	588	13.706
		10800	735	21.518
		10800	980	32.49

VI. Conclusions

In this paper, the ability for applying the kernel classification concept as distance metric learning in face verification is analysed and evaluated. By coupling a degree-2 polynomial kernel with the kernel methods, SVMML is able to act as a unified model for the metric learning approach. SVMML, which is implemented by the standard SVM solvers on the doublets/pairwise, is able to achieve a satisfactory result which is comparable to the state-of-the-art methods in terms of verification rate. The simple structure of the SVMML learning process, not only guarantee faster processing time, may also encourages the weak supervised learning. The efficiency and effectiveness of the kernel classification-based metric learning is worth to be further investigated with supervised, semi-supervised and even nonlinear metric learning algorithms.

Acknowledgement

The authors would like to thank Fundamental Research Grant Scheme (FRGS) of Malaysia for supporting the research under grants MMUE/140026.

References

[1] Ying Y. and Li P. 2012, Distance metric learning with eigenvalue optimization, *J. Mach. Learn. Res*., 13:1-26.
[2] Nguyen H. V. and Bai L. 2010. Cosine similarity metric learning for face verification, *ACCV*, 709–720
[3] Huang C., Zhu S. and Yu K. Large Scale Strongly Supervised Ensemble Metric Learning, with Applications to Face Verification and Retrieval, *NEC Technical Report TR115*, 2011.
[4] Cao Q., Ying Y., and Li P. Similarity Metric Learning for Face Recognition. *ICCV*, 2013.
[5] Weinberger K.Q., Blitzer J., and Saul L. K. 2005, Distance Metric Learning For Large Margin Nearest Neighbour Classification, *NIPS*, 1473-1480.
[6] Guillaumin M., Verbeek J., and Schmid C., 2009. Is That You? Metric Learning Approaches for Face Identification, *ICCV*.
[7] Balcan M. F., Blum A., and Srebro N., 2008. A Theory Of Learning With Similarity Functions, *Machine Learning*, 72(1-2):89-112.
[8] Fu Y., Yan S., and Huang T. S., 2008. Correlation metric for generalized feature extraction, *IEEE Trans. Pattern Anal. Mach. Intell*., 30(12):2229-2235.
[9] Kedem D., Tyree S., Weinberger K., Sha F., and Lanckriet G., 2012. Non-Linear Metric Learning, *Adv. Neural Inf. Process*. Syst., 2582-2590.
[10] Wang J., Do H., Woznica A., and Kalousis A., 2011. Metric Learning with Multiple Kernels, *NIPS*.
[11] Globerson A. and Roweis, S. 2006. Metric Learning By Collapsing Classes, *NIPS*.
[12] Change C. C. and Lin C. J., 2011. Libsvm: A library for support vector machines, *ACM Trans. Intelligent System Technology*, 2:1-27.
[13] Phillips P. J., Moon H., Rauss P. J., and Rizvi S., 2000. The FERET Evaluation Methodology For Face Recognition Algorithms, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:10.
[14] Martinez A.M. and Benavente R. 1998, The AR Face Database, CVC Technical Report #24, June.
[15] Huang G.B., Ramesh M., Berg T., and Learned-Miller E., 2007. Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments, *University of Massachusetts, Amherst, Technical Report* 07-49.
[16] Liu, M. and Vemuri, B., 2012. A Robust And Efficient Doubly Regularized Metric Learning Approach, *Computer Vision – ECCV 2012*, 646-659.
[17] Kumar M. P., Torr P. H. S., and Zisserman A. 2010, An Invariant Large Margin Nearest Neighbor Classifier, *ICCV*, 213-226.
[18] Davis J., Kulis B., Jain P., Sra S. and Dhilon I. 2007. Information-theoretic Metric Learning. *ICML*
[19] Saenko K., Kulis B., Fritz M., and Darrel T., 2010. Adapting Visual Category Models To New Domains, ECCV, 213-22.
[20] Perez-Suay A., Ferri F. J., Arevalillo-Herraez M., and Albert J. V., 2013. Comparative Evaluation Of Batch And Online Distance Metric Leraning Approaches Based On Margin Maximization, *IEEE Int. Conf. Syst. Man. Cybern*., 3511-3515,.
[21] Cui Z., Li W., Xu D., Shan S. and Chen X. 2013. Fusing Robust Face Region Descriptors via Multiple Metric Learning for Face Recognition in the Wild. *IEEE CVPR*.
[22] Anila S. & Devarajan N., 2012. Preprocessing Technique for Face Recognition Applications Under Varying Illumination Conditions, Global Journal of Computer Science and Technology.
[23] Ahonen T., Hadid A., Pietikainen M., 2006. Face Description with Local Binary Patterns: Application to Face Recognition, *IEEE Trans. Pattern Analysis and Machine Intelligence*, 28(12):2037-2041.
[24] Barkan, O., Weill, J., Wolf, L., Aronowitz, H., 2013. Fast High Dimensional Vector Multiplication Face Recognition, *ICCV*, 1960-1967.
[25] Turk M. and Pentland A. 1991. Eigenfaces for Recognition. *J. Cognitive Neuroscience*, 3(1):71-86.
[26] Fan Z., Liu E., Xu B., 2011. Weighted Principal Component Analysis, *AICI*, 569-574.
[27] Phillips P. J., McCabe R. M., Chellappa, R. 1998. Biometric Image Processing And Recognition, In *Proceedings, European Signal Processing Conference.*
[28] Shawe-Taylor J. and Cristianini N., 2004. Kernel Methods For Pattern Analysis, *Cambridge University Press,*
[29] Wang F., Zuo W., Zhang L., Meng D., and Zhang D. 2013. A Kernel Classification Framework For Metric Learning. arXiv:1309.5823,.
[30] Nowak E. and F. Jurie, 2007. Learning Visual Similarity Measures For Comparing Never Seen Objects, *CVPR*.
[31] Huang G. B., Jones M. J., and Learned-Miller E. 2008. LFW Results Using a Combined Nowak Plus MERL Recognizer, *ECCV*.