Periocular Recognition using Unsupervised Convolutional RBM Feature Learning

Lei Nie[1, 2]， Ajay Kumar[1], Song Zhan[2]

[1]Department of Computing
The Hong Kong Polytechnic University Technology,
Hung Hom, Kowloon, Hong Kong

[2]Shenzhen Institute of Advanced Technology,
Chinese Academy of Sciences
Shenzhen, China

Abstract— **Automated and accurate biometrics identification using periocular imaging has wide range of applications from human surveillance to improving performance for iris recognition systems, especially under less-constrained imaging environment. Restricted Boltzmann Machine is a generative stochastic neural network that can learn the probability distribution over its set of inputs. As a convolutional version of Restricted Boltzman Machines, CRBM aim to accommodate large image sizes and greatly reduce the computational burden. However in the best of our knowledge, the unsupervised feature learning methods have not been explored in biometrics area *except* for the face recognition. This paper explores the effectiveness of CRBM model for the periocular recognition. We perform experiments on periocular image database from the largest number of subjects (300 subjects as test subjects) and simultaneously exploit keypoint features for improving the matching accuracy. The experimental results are presented on publicly available database, the *Ubripr* database, and suggest effectiveness of RBM feature learning for automated periocular recognition with the large number of subjects. The results from the investigation in this paper also suggest that the supervised metric learning can be effectively used to achieve superior performance than the conventional Euclidean distance metric for the periocular identification.**

Key Word—Periocular Recognition; Biometrics; Unsupervised Feature Learning; CRBM; Supervised Metric learning

I. INTRODUCTION

Automated identification of humans is one of the most sought and challenging task to meet the ever growing demands for biometrics security. Contactless biometrics identification, like those using iris or facial imaging, is increasingly becoming part of infrastructure in e-governance, e-security and e-business. Currently available iris recognition systems operate under highly constrained environment, *i.e.*, require eye images acquired under near infrared illumination from the close distances using stop-and-stare mode of operation. Such systems degrades in performance when are environments are less-constrained, *i.e.* under visible illumination and from a greater standoff distances. Periocular region is referred to as the region surrounding eyes and is inherently acquired under during the conventional iris imaging. The periocular region illustrates complex and diverse set of features which can be utilized to improve the accuracy of iris recognition. The objective of our work in this paper presents is to investigate new approaches for periocular biometric recognition which can be more effective for identifying human subjects in larger databases.

The restricted Boltzmann machines (RBM) is a generative stochastic neural network that can learn a probability distribution over a set of inputs. The convolutional RBM is one variant of RBM to make it capable in deal with large resolution image data. The CRBM has been successfully utilized in various applications, such as the handwriting recognition, image classification, and face verifications, *etc*. Compared with traditional biometrics like fingerprint and palm print, the periocular recognition is an emerging biometrics targeted to enhance iris or face recognition under visible illumination using at-a-distance imaging. In this study, the unsupervised CRBM framework is firstly introduced to the periocular recognition paradigm. Our work described in this paper presents some encouraging results on larger publicly available benchmark periocular databaset.

The advantages of biometrics recognition using periocular images in comparison with iris recognition have been demonstrated in [4]. Its feasibility was comprehensively investigated by using various local descriptors, such as Local Binary Pattern (LBP), Histogram of Oriented Gradients (HOG) and Shift Invariant Feature Transform (SIFT) *etc*. As one of the most representative models in deep learning, the deep belief network (DBN) [5] is a generative graphical model which contains a layer of visible units and multiple layers of hidden units. Each layer encodes correlations in the units of next layer. DBNs and related unsupervised learning algorithms such as auto encoders [12] and sparse coding [13]-[14] have been used to learn higher-level feature representations from unlabeled data. A lot of related works have been successfully applied for the visual recognition and classifications [15]-[17]. Metric Learning is a method to learn a transformation matrix from the training data so that the new metric can perform better than the Euclidean space. Learning this target is identical to the learning of Mahalanobis distance. Metric learning has been applied in the face verification [11] and image recognition [19] domains.

In this paper, we present a new approach for the periocular identification using unsupervised feature learning method based upon the principle of convolutional RBM. A few trained geniune pairs are used as a constraint and the Mahalanobis distance is learned. The trained features are combined for the subsequent metric learning and SVM classification. The key contributions from this paper can be summarized as follows.

a) This paper proposes automated periocular identification using an unsupervised feature learning approach. Our experimental results on (test) database from 300 subjects, which is largest periocular subject's dataset for such

This work is partially supported by the internal competitive research grant no. G-YM09 (PolyU 5175/12E).from The Hong Kong Polytechnic University.

evaluation in our knowledge, achieve state-of-the-art performance when simultaneously combined with the hand-crafted features;

b) We also evaluate the effectiveness and efficacy of supervised metric learning for the biometric identification. The results presented in this paper demonstrate that the supervised Manhalanobis distance can outperform those achieved using the traditional metric space;

c) In the context of periocular biometrics, our experiments suggest that the nonlinear combination of match scores can achieve better separation of genuine pairs than the traditional weighted sum and linear SVM fusion.

II. Methodology

The proposed approach for the completely automated periocular recognition is summarized in figure 1. The periocular regions are firstly segmented and preprocessed. This is followed by an unsupervised convolutional RBM training using the available database and the activation of the output pooling layer. Next step is to achieve eliminate redundancy in features, *i.e.*, achieving dimensionality reduction using the Principle Component Analysis (PCA). Finally we employ the supervised metric learning and SVM training to generate the labels for the evaluation. In this work we use unsupervised CRBM model training process and the supervised metric learning is used for the verification. Finally, a binary SVM is trained to classify the genuine pairs. During the evaluation of unseen test images, the CRBM features are generated by the pre-trained CRBM model. For the combination with the hand-crafted features, two SVM scores are employed. The final scores are generated by fusing two scores via non-linear transformation

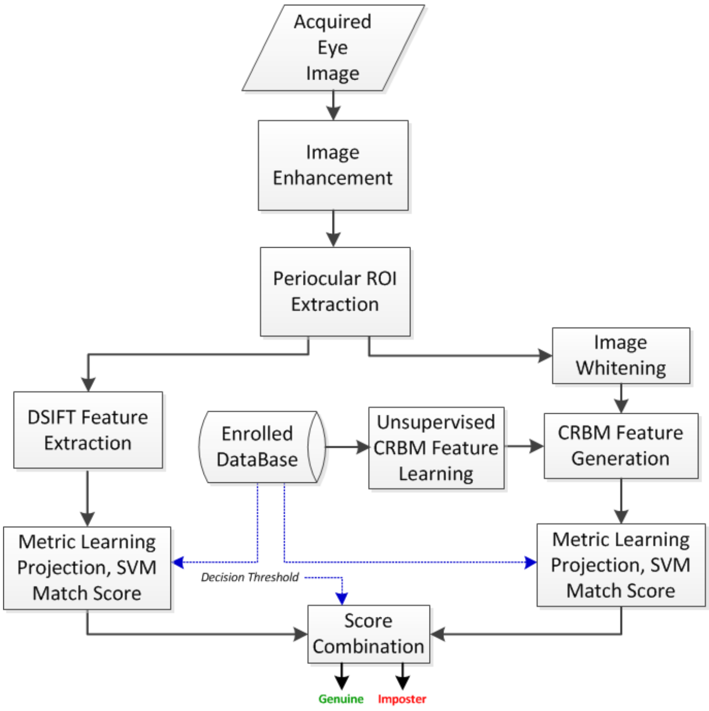


Figure 1: The block-diagram of the proposed approach for the automated periocular image identification.

A. Preprocessing for Recovering Periocular Features

The periocular region is not well-defined in literature but widely believed to represent the region in the vicinity of eyes. Prior work on periocular biometrics used iris center and the width to define the periocular region. However, it is sufficiently accurate to use the pupil center and eye size. Such assumptions can often introduce misalignment and degrade the recognition accuracy. In [1], midpoint of the eye corners is used instead of the iris center point, and better performance is reported. Some sample images are given in Figure 2. In our work, we adopt the midpoint of the eye corners as the center point. In our work, the periocular region proportional to the image size is automatically segmented. The ratio adopted is fixed to *r*. The rotational variation can be minimized by aligning the line connecting two eye corners with the *x*-axes of the image.

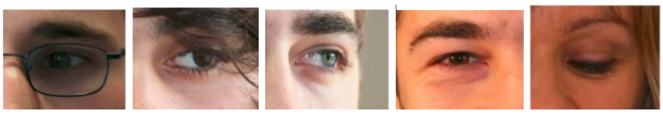


Figure 2: Sample images in “UBriPr” dataset [1].

Finally, the segmented images are rescaled to a fixed size *w***h*. The retinex image enhancement is adopted to neutralize the illumination factor. In order to eliminate the correlation of adjacent pixels, we firstly transform the image to the Fourier domain. The power spectrum is then flattened and then we transform the image back to the spectral domain. The images

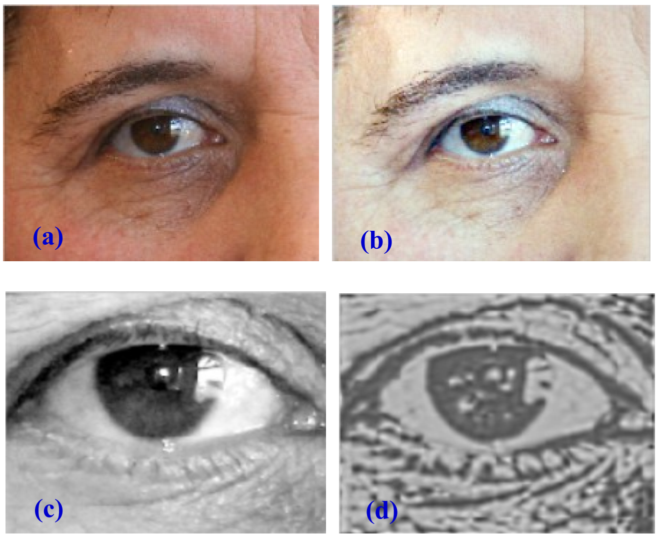


Figure 3: Preprocessed a typical image in UBiPr dataset: (a) original eye image, (b) the enhanced retinex image, (c) the segmented eye region, and (d) image whitening result depicting the periocular features.

after preprocessing are similar to as samples shown in Figure 3. We can see that the cropped images contain entire periocular region, and the eye corners are aligned to the center.

B. Convolutional RBM

In the CRBM model [17], all nodes in the convolution and visible layers share one CRBM weight. The model consists of two layers: an input layer D and a convolution layer C. The input layer consists of an $\mathrm{N_d} \times \mathrm{N_d}$ array of real-value units. The convolution layer consists of K groups, where each group is an $\mathrm{N_c} \times \mathrm{N_c}$ array of binary units, resulting in $\mathrm{N_h^2 K}$ hidden units. Each of the K group is associated with a $\mathrm{N_w} \times \mathrm{N_w}$ filter $\mathrm{N_c} = \mathrm{N_d} - \mathrm{N_w} + 1$. The filter weights area across all the hidden units within the group. In addition, each hidden group has a bias $\mathrm{e_k}$ and all visible units share a single bias *b*.

The number of hidden nodes is far more than the visible layers. In order to reduce the computation burden and also be tolerant to small translational misalignment, the pooling stage is included. In the pooling stage, the convolution layer $\mathrm{C^k}$ is partition to blocks of $\mathrm{m} \times \mathrm{m}$ and each block a is connected to exactly one binary unit $\mathrm{p_\alpha^k}$ in the pooling layer. The resulting structure of the CRBM model is as illustrated in figure 4.

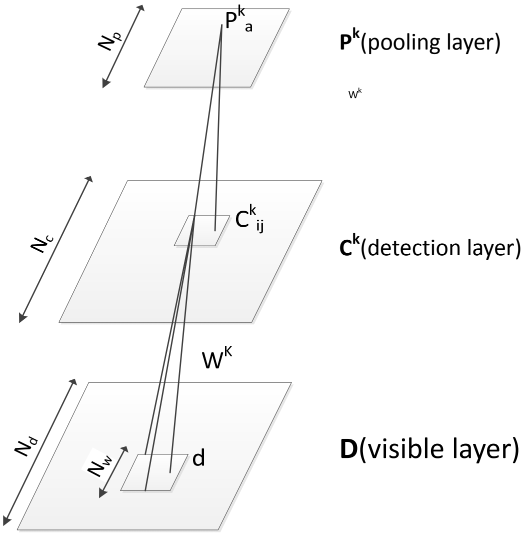


Figure 4: The interaction between visible layer, detection layer and pooling layer.

The energy function of CRBM is defined as follows:

$$E(d,c) = -\sum_{k=1}^{K} c^k (W_k * d) - \sum_{k=1}^{K} e_k \sum_{i,j} c_{i,j}^k - b \sum_{i,j} d_{ij} \quad (1)$$

Thus we can conclude the conditional probability similar with the standard RBMs as:

$$P(d_{ij}|c) = N\left(\sum_k (W^k * c^k)_{ij} + b, \sigma\right) \quad (2)$$

$$P(c_{ij}^k = 1|d) = \sigma\left(\sum_k (\widetilde{W}_k * d)_{ij} + e_k\right) \quad (3)$$

We can introduce the constraint that the at most one unit in the block layer are on, thus we can derive the final hidden unit probability and the pooling unit conditional probability after pooling stage.

$$P(c_{i,j}^k = 1|d) = \frac{exp\left(I(c_{i,j}^k)\right)}{1 + \sum_{(i',j')\in B} exp\left(I(c_{i',j'}^k)\right)} \quad (4)$$

$$P(p_a = 0|d) = \frac{1}{1 + \sum_{(i',j')\in B} exp\left(I(c_{i',j'}^k)\right)} \quad (5)$$

Since the feature response is much larger than the input layer, the sparsity constraint must be involved in order to learn reasonable results. The objective function (log-likelihood) is regularized to encourage each hidden unit group to have a mean activation close to a small constant.

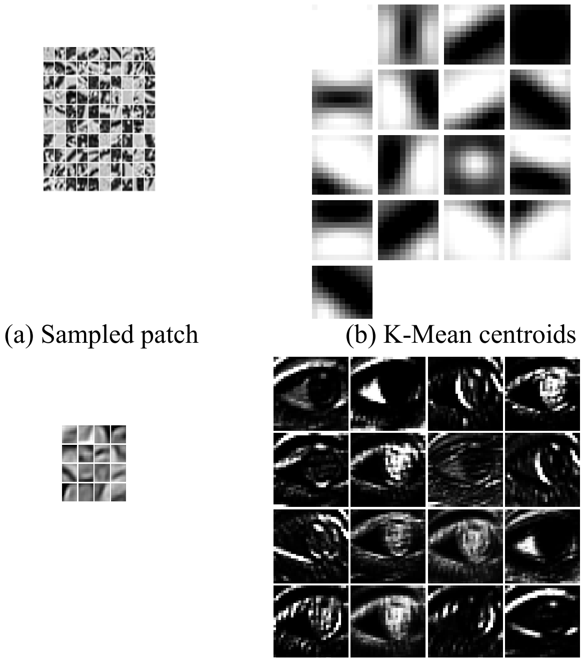
$$\Delta b_k^{sparsity} \propto p - \frac{1}{N_H^2} \sum_{i,j} P(c_{ij}^k = 1|d) \quad (6)$$

The training process of the CRBM is similar as for the conventional RBM, which uses the contrastive divergence to update the parameter. We use CD-1 to estimate the model expectation. In order to reduce the variance, we directly use the expectation value instead of sampling from the model during the training procedure.

After the processing of the image, employed CRBM model is identical to the traditional CRBM training procedure. As shown in [6], the sparse RBM is identical with the Gaussian Mixture Model (GMM), thus we can apply a K-means and GMM model to provide the initial parameters. A one-layer CRBM is trained and the pooling unit response is output. The initialization procedure of the CRBM is as follows.

a) Sample the patch, sample patch size is identical to the CRBM weight size $\mathrm{N_w}$
b) Using K-mean to cluster the patch, the centroids number is set to K+1, K is the number of the filter numbers.
c) Using the K-mean results $Kmeans(c_k, n_k)$, where $\mathrm{c_k}$ is the K-means centroids, and $\mathrm{n_k}$ is the cluster sample numbers to initialize the shared covariance $GMM(\pi_k, \mu_k, \sigma^2 I)$, where all components share a single covariance matrix $\sigma^2 I$, and $\mu_k = c_k$, $\pi_k = \frac{n_k}{\sum_i n_k}$,and $\sigma = \frac{1}{n}\sum_i (x_i - c_i)^2$, c_i is the nearest centroid.
d) Using E-M algorithm to optimize the GMM model alternatively. In expectation step, the weight matrixes of samples are calculated according to the model parameters. In the maximization step, the parameters of π_k, μ_k, σ are updated.
e) Parameter of convolutional RBM can be derived from $GMM(\pi_k, \mu_k, \sigma^2 I)$, i.e. $b = \mu_0$, $w_j = \frac{1}{\sigma}(\mu_j - c), j = 1, \ldots K$, and $e_j = log\frac{\pi_j}{\pi_0} - \frac{1}{2}\|w_j\|^2 - \frac{1}{\sigma} w_j^T c$.

The learned filters for the periocular matching are illustrated in figure 5 (c), which are quite consistent with those learned for the LFW face recognition dataset in the literature. The learned features are visually illustrated in Figure 5 (b). We can see that the feature contains the silhouette of the eye image without uncorrelated noise port, the eyelid, the eyeball, the iris part are clearly exhibited.



(a) Sampled patch (b) K-Mean centroids

(c) Learned filters (d) Generated features

Figure 5: Illustration of features generated using CRBM.

C. Metric Learning and SVM training

The score generation process can be summarized as shown in figure 6. In order to improve the computational efficiency, a whitening PCA compression with normalization is applied [10]. We retain the first k=500 eigenvectors and then normalize every component.

A supervised linear metric learning is conducted to do feature reweighting for the calculated features. Since the goal of unsupervised RBM learning and the discriminated goal of the periocular recognition is different. We adopt the Information Theoretic Metric Learning [9] method for the metric learning. Given genuine pairs S and impostor pairs D, the distance metric learning problem can be expressed as follows:

$$min_A KL(p(x;A_0)\|p(x;A))$$

$$\text{subject to } d_A(x_i,x_j) \le u,\quad (i,j) \in S$$

$$d_A(x_i,x_j) \ge l,\quad (i,j) \in D \qquad (7)$$

According to the equivalence of differential relative entropy between two multivariate Gaussians and the LogDet divergence between the covariance matrices [9], a slack variable is introduced to handle the non-linear problem as:

$$min_{A>0,\xi}\ D_{ld}(A,A_0) + \gamma * D_{ld}(diag(\xi),diag(\xi_0))$$

$$\text{s.t}\ \ tr(A(x_i - x_j)\,(x_i - x_j)^T \le \xi_{c(i,j)}\ (i,j) \in S$$

$$tr(A(x_i - x_j)\,(x_i - x_j)^T \ge \xi_{c(i,j)}\ (i,j) \in D$$

In order to solve the convex optimization problem, the sequential optimization is applied by repeatedly computing the projections of current solution onto a single constraint.

The comparison results with metric learning are shown in Figure 7. We choose DSIFT feature to illustrate the importance of metric learning. As can be observed from this figure, the performance without metric learning is lower which suggests that the genuine/impostor pairs cannot be effectively separately. However, after the supervised metric learning, the weighted DSIFT feature vector can achieve greater separation between genuine and impostor pairs resulting in superior performance.

The Cholesky decomposition is applied to matrix A as $A = R^T\ R$. And then, the features are transformed by $\frac{Rx}{\|Rx\|}$. Instead of using cosine similarity in [11], we trained a binary SVM using the element-wise multiplication of features pair to classify the genuine and impostor pairs. The SVM decision value is normalized as the genuine and impostor score pairs.

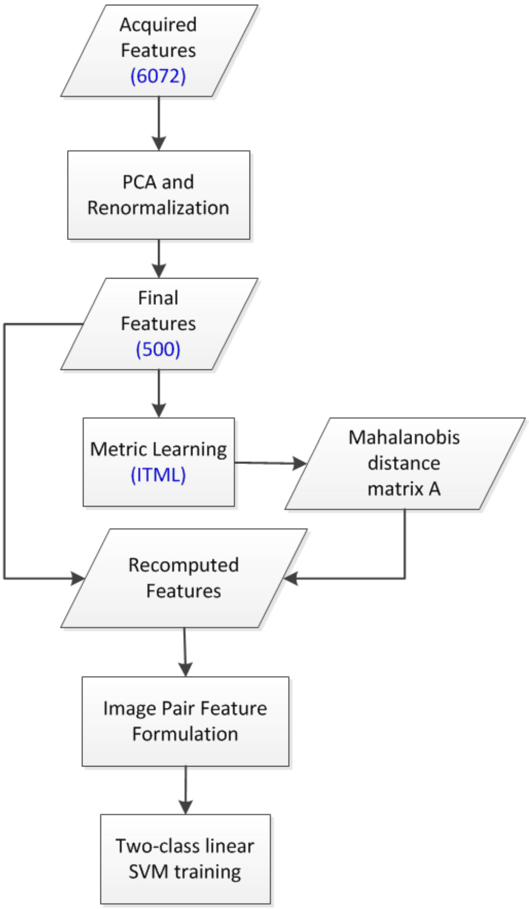


Figure 6: Key steps during match score generation.

III. EXPERIMENTs AND RESULTS

We utilized recently released UBIPr dataset from reference [1] as this dataset is more realistic and has largest number of subject's images than other periocular datasets available in the literature in the best of our knowledge. This dataset has images with complete periocular region (unlike UBIRIS v1 or v2) and higher pose variations. The dataset has 10,252 images acquired from 344 subjects. These images are acquired from varying distances: Five different distances (3m, 4m, 5m, 6m, 7m, 8m) and with different poses. The eye corner position has been made available in this dataset and the periocular region is cropped by using the center of the left and right eye corners. In our experiments, the periocular region is set to 1.2w × 0.9w where w is the eye width.

A. Experiment Protocol

The first 34 subjects in the dataset are used as training subject, and the remaining 300 subjects are used as testing subject. There are total 1,000 genuine and 20,000 impostor pairs are randomly selected from the training subjects for ITML metric learning and SVM training. In order to reduce the computation requirements, the first 10 images from each of the test subjects are used. We have a total of 13,860 genuine pairs, and 4,727,800 impostor pairs for the test phase.

B. Experiment Results

In order to compare with existing hand-crafted features, like DSIFT, LBP and HOG, we follow the parameters from [1] to divide the normalized image into 8 × 6 blocks and then compute the histogram of HOG, DSIFT, and LBP features. We use the VLfeat [2] toolbox to generate the hand-crafted features.

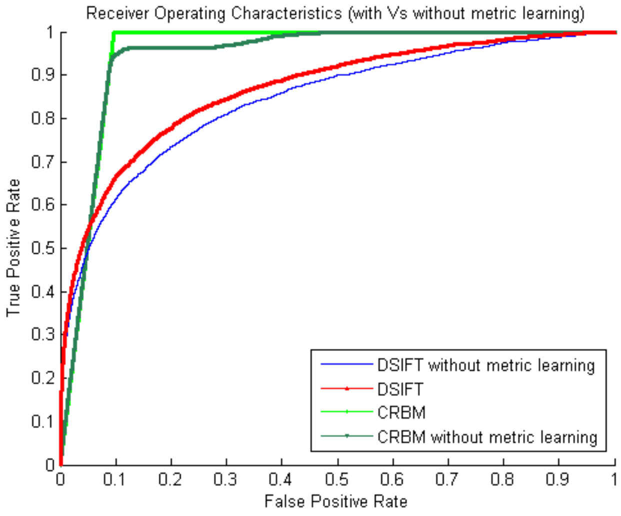


Figure 7: Performance **c**omparison with the usage of metric learning using ROC (300 subjects test images).

In the convolutional RBM model, the hyper-parameter is finely tuned using cross-validation. The input image is resized to 80 × 80, the learned filter size is set to 11 × 11, the filter number is set to 16, and the pooling size is set to 3 × 3.

In order to extract HOG features, we used 8-orientation bin. For each of these features, the 2 × 2 normalization is performed for each pixel. In the LBP feature extraction, uniform LBP feature is used to generate 58 binary codes, and finally 2,784 features are generated. In the DSIFT feature, the bin size for SIFT descriptor is set to 6 pixels, and the step is set to 25 pixels which consistent with the block size. These hand-crafted features are processed with the same PCA dimension reduction, ITML metric learning and binary SVM.

We can observe that the ROC curve surpasses the hand-crafted features in most regions except for the low-acceptance region. We concatenate the decision values of four SVM (CRBM, HOG, LBP, DSIFT) to train a second SVM. The performance comparison of hand-crafted features and CRBM feature are shown in the Figure 8. This figure also provides performance from the combination of CRBM and hand-crafted features. The combination of features further improves the performance (low false-acceptance region in ROC). The EER has been greatly reduced from the 20% to 6.4%.

C. Score Combination

We also explored the score-level combination of the best performing hand-crafted features, *i.e.*, DSIFT, with the CRBM features. Several score-level fusion methods were experimented, including the SVM fusion [3], weighted-minimum fusion [7] and the weighted product [8]. The results from the DSIFT and the hand-crafted SVM fusion are also included in Figure 8 for the comparison. The best results with the EER of 11% are achieved from the weighted-minimum score combination. Performance in the low FRR region is also greatly improved over those from the CRBM or DSIFT.

The weighted minimum and weighted product score combination [8] is computed as follows:

$$s_{\min} = \min(w * \mathrm{sa}, \mathrm{sb})\,, 1 \le w \le 2$$
$$s_{\mathrm{prod}} = \mathrm{sa} * \mathrm{sb}^{w-1}\,, 1 \le w \le 2 \qquad (8)$$

The scores were firstly normalized in the range 0-1 and then the least square error of the trained data is used to generate the optimized w and the new fused scores. For SVM fusion, we use the trained label and trained scores to generate a linear SVM. Then the SVM score is generated. In addition to identification or verification experiments, we also performed experiments for periocular recognition. The weighted-product combination with handcraft scores, can achieve average rank-one accuracy of 50.1% which is higher than those using DSIFT 33.8% and using hand-crafted features 40.4%.

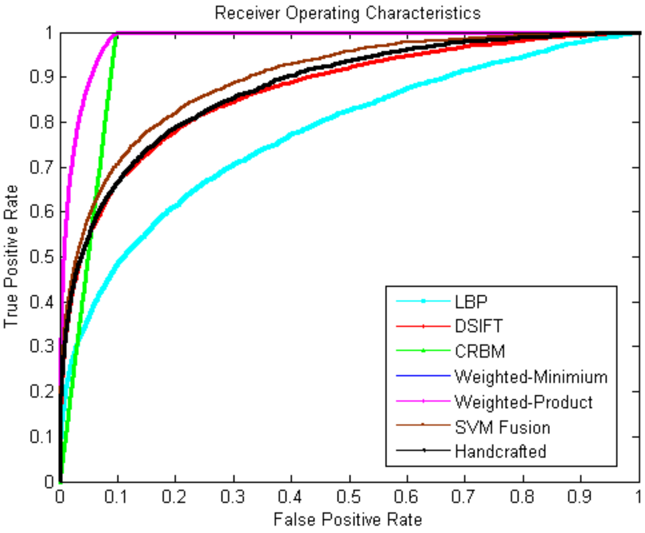


Figure 8: Performance **c**omparison between CRBM and hand-crafted features using ROC (300 subjects test images).

IV. Conclusions and Further Work

In this paper, a new approach for the periocular recognition using unsupervised feature learning has been successfully investigated. Based upon the framework of CRBM, we show that significantly superior results can be achieved with the score level combination of hand-crafted features. We also evaluated the effectiveness of supervised metric learning for the periocular biometrics matching. Our experimental results suggest that the supervised Manhalanobis distance outperforms in comparison with traditional metric space. In the context of periocular matching, the experimental results presented in this paper also suggests that the nonlinear score level combination can better separate the genuine pairs than the traditional weighted sum and linear SVM fusion approaches.

The experimental results in this paper were presented on larger database of 344 subjects using recently introduced Ubir periocular database [1]. This is plausible explanation for the lower overall accuracy shown in this paper as compared to prior publications in periocular biometrics which used database with relatively smaller number of subjects. However the nonlinear combination of simultaneously extracted DSIFT features can significantly help to achieve superior results. The promising experimental results illustrated in this paper using larger database suggest great potential for the unsupervised feature learning approaches in the periocular biometrics recognition.

Further work is required to evaluate the performance of the proposed periocular identification approach on other databases, like UBIRIS V2 or those acquired under the near infrared imaging. Combination of periocular match scores with the face match scores can augment the accuracy for the face recognition and worth investigating in further extension of this work. Similar combination with the simultaneously acquired iris match scores, and/or from the larger areas [20] surrounding eye with the eyebrow and skin texture, can also be effective in improving the performance for the iris recognition and is suggested for further work.

References

[1] C. N. Padole, H. Proenca, "Periocular recognition: analysis of performance degradation factors," *Proc. ICB 2012*, New Delhi, India, pp. 439 – 445, March-April 2012.

[2] A. Vedaldi and B. Fulkerson, *VLFeat: An Open and Portable Library of Computer Vision Algorithms*, http://www.vlfeat.org/, 2008

[3] G. B. Huang, H. Lee, and E. Learned-Miller, "Learning hierarchical representations for face verification with convolutional deep belief networks," *Proc. CVPR 2012*, Providence, Rhode Island, pp. 2518-2525, Jun. 2012.

[4] S. Bharadwaj, H. Bhatt, M. Vatsa, and R. Singh. "Periocular biometrics: When iris recognition fails," *Proc. 4th IEEE Conf. on Biometrics: Theory Applications and Systems*, BTAS 2010, Washington DC, pp. 1–6, Sept. 2010.

[5] G. E. Hinton, S. Osindero, and Y.-W. The, "A fast learning algorithm for deep belief networks," *Neural Computation*, vol. 18, no. 7, pp. 1527–1554, 2006.

[6] K. Sohn, D. Y. Jung, H. Lee, and A. O. Hero, "Efficient learning of sparse distributed convolutional feature representations for object recognition," *Proc. ICCV 2011*, pp. 643-2650, 2011.

[7] A. Kumar and A. Passi, "Comparison and combination of iris matchers for reliable personal authentication," *Pattern Recognition*, vol. 43, no. 3, pp. 1016-1026, Mar. 2010.

[8] A. Kumar and C. Wu, "Automated human identification using ear imaging," *Pattern Recognition*, vol. 41, no. 5, March 2012.

[9] J. V . Davis, B. Kulis, P . Jain, S. Sra, and I. S. Dhillon, "Information-theoretic metric learning," *Proc. ICML*, 2007.

[10] Z. Cao, Q.Yin, X. Tang, and J. Sun, "Face recognition with learning-based descriptor," *Proc. CVPR 2010*. Jun. 2010.

[11] H. V. Nguyen and L. Bai, "Cosine similarity metric learning for face verification," *Proc. ACCV 2010*, LNCS 6493, pp. 709-720, 2010.

[12] Y. Bengio, P. Lamblin, D. Popovici, and H. Larochelle.," Greedy layer-wise training of deep networks," *Proc. NIPS*, 2007.

[13] H. Lee, A. Battle, R. Raina, and A. Y. Ng, "Efficient sparse coding algorithms," *Proc. NIPS 2007*, pp. 801-808, 2007.

[14] B. A. Olshausen and D. J. Field, "Emergence of simple-cell receptive field properties by learning a sparse code for natural images," *Nature*, 381:607–609, 1996.

[15] M. Ranzato and G. E. Hinton, "Modeling pixel means and covariances using factorized third-order Boltzmann machines," *Proc. CVPR, 2010*. pp. 2551-2558, Jun. 2010.

[16] M. Ranzato, F.-J. Huang, Y .-L. Boureau, and Y . LeCun, "Unsupervised learning of invariant feature hierarchies with applications to object recognition," *Proc. CVPR, 2007*, pp. 1-8, Minneapolis, Jun. 2007.

[17] H. Lee, R. Grosse, R. Ranganath, and A. Y. Ng," Unsupervised learning of hierarchical representations with convolutional deep belief networks," *Communications of the ACM*, vol. 54, pp. 95–103, 2011.

[18] A Krizhevsky, I Sutskever, G Hinton, "ImageNet classification with deep convolutional neural networks," *NIPS 2012*, pp. 1106-1114, 2012.

[19] T. Mensink, J. Verbeek, F. Perronnin, and G. Csurka, "Metric learning for large scale image classification: generalizing to new classes at near-zero cost," *Proc. ECCV 2012*.

[20] J. M. Smereka, B. V. K. Vijaya Kumar, "What is a 'good' periocular region for recognition?," *Proc. CVPR 2013*, Portland, Oregon, CVPR'W 2012, pp. 117-124, Jun. 2012.