Heterogeneous Face Recognition from Local Structures of Normalized Appearance

Shengcai Liao, Dong Yi, Zhen Lei, Rui Qin, and Stan Z. Li*

Center for Biometrics and Security Research,
Institute of Automation, Chinese Academy of Sciences,
95 Zhongguancun Donglu, Beijing 100190, China
{scliao,dyi,zlei,rqin,szli}@nlpr.ia.ac.cn

Abstract. Heterogeneous face images come from different lighting conditions or different imaging devices, such as visible light (VIS) and near infrared (NIR) based. Because heterogeneous face images can have different skin spectra-optical properties, direct appearance based matching is no longer appropriate for solving the problem. Hence we need to find facial features common in heterogeneous images. For this, first we use Difference-of-Gaussian filtering to obtain a normalized appearance for all heterogeneous faces. We then apply MB-LBP, an extension of LBP operator, to encode the local image structures in the transformed domain, and further learn the most discriminant local features for recognition. Experiments show that the proposed method significantly outperforms existing ones in matching between VIS and NIR face images.

Keywords: Face Recognition, Heterogeneous, MB-LBP, DoG.

1 Introduction

In recent years, many face recognition systems are developed for practical use. Applications from the real world now faces an emergency problem that face images are always collected in quite different environments or even captured with different equipments. We call these images heterogeneous face images, and the corresponding new arisen problem as heterogeneous face recognition problem. Some examples of this problem appeared in real-world application are shown in [9,21]. A typical problem is that matching between near infrared (NIR) and visible light (VIS) face images in the situation that enrolment is finished with controlled indoor VIS face images, while authentication would be done using NIR face images to avoid the influence of the variational environment illumination.

Lin *et al.* proposed an algorithm called Common Discriminant Feature Extraction (CDFE) [9]. In the algorithm, two transforms were simultaneously learned to project the samples in both modalities respectively to a common feature space. However, though the authors claimed that they used a smooth factor to avoid overfitting, the algorithm still lacked generalization capability that they only got impressive result on an close-set experimental setting, while had poor performances on non-overlapping test set (*cf.* Section 4).

* Corresponding author.

M. Tistarelli and M.S. Nixon (Eds.): ICB 2009, LNCS 5558, pp. 209–218, 2009.

Yi *et al.* [21] proposed another algorithm for heterogeneous face matching. There they also projected samples of two modalities onto a common subspace respectively, but instead they utilized the well-known CCA algorithm. Yang *et al.* [19] also proposed a CCA-based method for matching between 2D and 3D faces using sub-patches. Like CDFE, both methods failed to learn the essential property of heterogeneous faces that would tend to overfit on the training data (*cf.* Section 4).

In this paper, we present a novel algorithm and analyze the heterogeneous problems in NIR-VIS face recognition. It is obvious from the Lambertian model that the albedo term no longer serves as an ideal component for heterogeneous face recognition. Consequently, traditional approaches which directly match two kinds of face images would cause problems. Then what other representations could serve as an invariant property between heterogeneous faces? Finally we discover that the local image structures could be learned for a robust representation for the heterogeneous problem. The reason is that, no matter how the light source changes (homogeneous or heterogeneous), the local relationships of a face would not change too much. Therefore, it would be a robust representation if we learned the most intrinsic local image structures.

To achieve the goal, two steps are adopted in this work. First, we use Difference-of-Gaussian filtering to normalize the appearance of heterogeneous face images, so that they look similar with each other. This contributes further recognition. Second, after appearance normalization, we apply the Multi-scale Block LBP (MB-LBP) for feature representing, which is proved to be effective for encoding local image structures [8]. We call the resulting feature as Local Structure of Normalized Appearance (LSNA). Based on the representation we further learn the most discriminant features for recognition.

Through the above two stages, a robust local structure representation is learned for heterogeneous face recognition that generalizes well on unseen persons. Furthermore, the final model is universal that it is suitable for both homogeneous and heterogeneous faces.

The rest of this paper is organized as follows: In Section 2, we analyze problems in heterogeneous face recognition, present an appearance normalization method and give the explanation that why it is effective for heterogeneous face recognition problem. In Section 3 we apply MB-LBP to encode the intrinsic local structures between both kinds of source images, and learn a discriminant subspace based on it. The experiment results are given in Section 4. Finally, we summarize this paper in Section 5.

2 Appearance Normalization by Difference-of-Gaussian

2.1 Problem Analysis

According to the Lambertian model, an image $I(x, y)$ under a point light source is formed as follows.

$$I(x, y) = \rho_\omega(x, y)\mathbf{n}(\mathbf{x}, \mathbf{y})\mathbf{s} \tag{1}$$

where $\rho_\omega(x, y)$ is the albedo of the facial surface material at point (x, y) with the light source of wavelength ω, $\mathbf{n} = (\mathbf{n_x}, \mathbf{n_y}, \mathbf{n_z})$ is the surface normal in the 3D space, and $\mathbf{s} = (\mathbf{s_x}, \mathbf{s_y}, \mathbf{s_z})$ is the lighting direction. Here, albedo $\rho_\omega(x, y)$ reflects the spectral properties of facial skin, and $\mathbf{n}(\mathbf{x}, \mathbf{y})$ is the geometric shape of the face.

Existing 2D face recognition researches mainly focus on visual face images, there the top-most factor that affects the face recognition performance is the direction of the incidence lighting relative to the face surface normal. With homogeneous lightings, $\rho_\omega(x, y)$ is the intrinsic property of the face , and is the ideal part for face recognition, while illumination variations are the extrinsic factor that should be removed from the product $\mathbf{n}(\mathbf{x}, \mathbf{y})\mathbf{s}$. In fact, over the past years much effort has been made to model illumination on visual faces and correct illumination directions (*cf.* [13], [4], [16], [3], [18]). Researches of this kind are all toward an aim to discover the intrinsic property: invariant $\rho_\omega(x, y)$ at point (x, y).

However, when dealing with heterogeneous images, such as matching between NIR and VIS images, the term $\rho_\omega(x, y)$ will not remain to be invariant any more. That is because the reflectance of the same face surface under heterogeneous light sources is different. Actually it is a function of the wavelength ω of the light source. Fig. 2 shows some examples of VIS vs. NIR face images (column (a) and (b)). The comparison demonstrates that the appearance of both kinds looks so different, particularly in the eyes part and the mouth part. It indicates that the reflectance of the same face surface under VIS and NIR light source is different. As a result, the above methods modelling illumination would fail to achieve their original goal under heterogeneous lighting condition, because the assumption of the invariant $\rho_\omega(x, y)$ does not hold any more.

Therefore, heterogeneous face recognition is confronted with more difficulties. From the Lambertian model we know that even with the same pose and lighting direction, one can not directly match two heterogeneous face images because of the distinct $\rho_\omega(x, y)$. As a result, traditional face recognition methods might encounter troubles when directly applied on heterogeneous face images.

From the above analysis, we know that $\rho_\omega(x, y)$ no longer serves as an ideal component for heterogeneous face recognition. Hence the problem could not be done directly upon image itself, even if we got the $\rho_\omega(x, y)$ part. The reason why previous researches [9,21,19] get poor performance on unseen data (*cf.* Section 4) is possibly because the learned model directly based on images lacks generalization.

Then what other representations are really essential for both homogeneous and heterogeneous face recognition? A good many of previous researches prove that local features can be used for robust representations, though they have not been examined on heterogeneous face recognition problem. It suggest us that local image structures might be the really intrinsic property of faces in despite of light sources. To testify this, our work adopts a two-stage procedure to deal with heterogeneous face recognition problem as described in the following materials.

2.2 Local Filtering

The first stage is to normalize the appearance of the input face images under both light sources, since they look quite different. In other words, we try to find a common image space where the appearance of both kinds of face images looks similar, so that it would be more easier for further recognition.

Unlike previous researches [9,21,19], in which they all learn a common subspace that lacks visual interpretation and the ability of generalization, here we try to first get a regularized appearance for heterogeneous face images using filtering without learning.

To achieve this purpose, we adopt the Difference-of-Gaussian (DoG) filtering to normalize the appearance, for its efficiency and effectiveness. It is successfully used in the well-known SIFT feature [11] for robustly identifying keypoint locations and scales. In [17], DoG filtering is also used in a preprocessing chain for illumination normalization.

Given an input image $I(x, y)$, DoG filtering is computed as

$$D(x, y|\sigma_0, \sigma_1) = (G(x, y, \sigma_0) - G(x, y, \sigma_1)) * I(x, y), \tag{2}$$

where $*$ is the convolution operator, and

$$G(x, y, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x^2+y^2)/2\sigma^2}. \tag{3}$$

Here $\sigma_0 < \sigma_1$, so that it constructs a bandpass filter. This simple filtering is actually a subtraction of two Gaussian filtering, thus is efficient to compute. Fig. 1 shows an example of DoG filter.

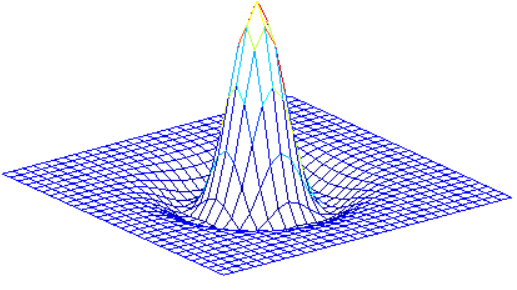

Fig. 1. An example of DoG filter

There are a number of reasons why we choose DoG filtering for normalizing appearance of heterogeneous face images, besides its computation efficiency.

Though the appearance of heterogeneous face images looks different (see Fig. 2 (a) and (b)), their local structures would be similar because they are all faces, particularly for the same person. DoG is a typical differential filter with excitatory and inhibitory lobes, like LoG (Laplacian-of-Gaussian) operator and Gabor

filter. Differential filters are often used for representing image structures. They encode local relationships via comparing neighboring regions. We know that both DoG and LoG are scale-space operators with Gaussian kernel. Lowe [11] shows that the difference-of-Gaussian function provides a close approximation to the scale-normalized Laplacian of Gaussian, $\sigma^2\nabla^2 G$, which is studied by Lindeberg for representing visual structures [10]. Lindeberg also shows an interesting similarity with biological vision that these scale-space operators closely resemble receptive field profiles registered in neurophysiological studies of the mammalian retina and visual cortex [10].

Furthermore, DoG filtering also helps to reduce both illumination variation, image noise and aliasing, while preserving enough details for recognition. As we know, both image shading and highlight area caused by illumination lie in low frequency domain, while noise and aliasing are high frequency information. Hence suppressing both the lowest and highest spatial frequencies by bandpass filtering contributes to further recognition.

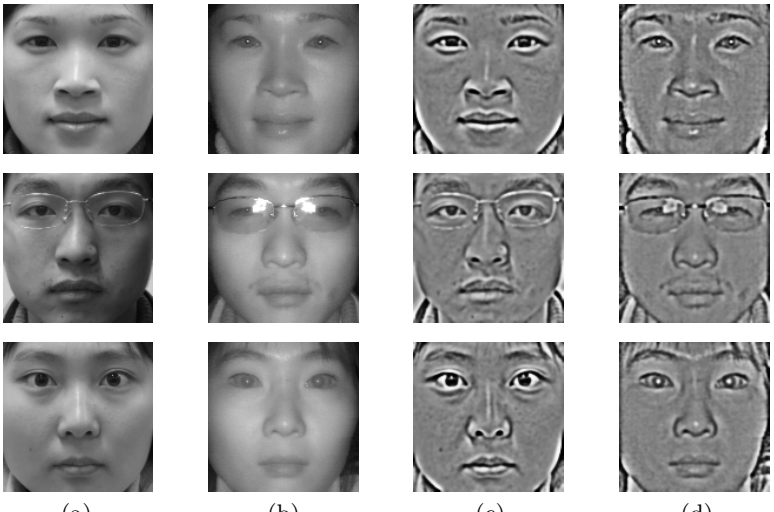

Fig. 2. (a) VIS face images of three people. (b) The corresponding NIR images of the same person as (a) in each row. (c) The corresponding DoG filtered images of (a). (d) The corresponding DoG filtered images of (b).

Fig. 2 shows some preprocessed VIS and NIR face images (column (c) and (d)) using DoG filtering. The parameters are set to $\sigma_0 = 1.0$ and $\sigma_1 = 2.0$ by default, as suggested in [17]. Compared with (a) and (b), we could clearly see that the distinct appearances of VIS and NIR images have been normalized to be similar after preprocessing. Besides, the image shadows and highlights are reduced after DoG filtering. The mouth part is normalized to be more similar, but the eyes part still looks different, so we still need to learn discriminant local structures for further recognition.

3 Learning Discriminant Local Structures

As mentioned before, the local structures of heterogeneous face images remain similar for the same person, which provides potential information for discriminating different persons. Therefore, we further apply MB-LBP, an extension of LBP operator, for multi-scale analysis of the local image structures after the appearance normalization.

Over the past years, Local Binary Patterns (LBP) is proved to be a powerful local descriptor for texture analysis and image structure representation [14,15,1]. The original LBP operator labels the pixels of an image by thresholding the 3×3-neighborhood of each pixel with the center value and considering the result as a binary string or a decimal number. Then the histogram of the labels can be used as a texture descriptor. An illustration of the basic LBP operator is shown in Fig. 3(a).

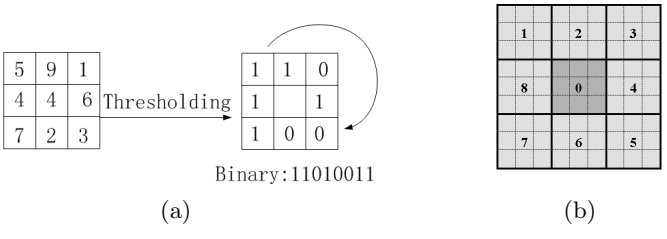

Fig. 3. (a) The basic LBP operator. (b) The 9×9 MB-LBP operator. In each sub-region, average sum of image intensity is computed. These average sums are then thresholded by that of the center block. MB-LBP is then obtained.

Recently, an extended LBP called MB-LBP [8,22] has been proposed for encoding multi-scale image structures. There, the computation is done based on average values of block subregions, instead of individual pixels (*cf.* Fig. 3(b)), and the size of the block indicates the scale. This way, MB-LBP only incurs a little more computation cost considering integral images, yet yields a more robust representation. Besides, the multi-scale representations encode not only microstructures but also macrostructures of image patterns, and hence provides a more complete image representation than the basic LBP operator.

In this work, we adopt the MB-LBP operator to encode local image structures on the normalized appearance, the resulting feature is denoted as Local Structure of Normalized Appearance (LSNA). we also use the histogram of statistically effective MB-LBP to represent the final feature set, as described in [8].

Considering the whole set of multi-scale features, MB-LBP would generate an over-complete representation. However, those excessive measures contain much redundant information, thus a further processing is needed to remove the redundancy and build effective classifiers. In this paper we use Gentle AdaBoost algorithm [2] to select the most effective MB-LBP features, as in [8].

Unlike in [9] and [21] where two kinds of heterogeneous face images are considered in two different spaces, we treat all the appearance normalized images as lying in the same space. Thus the intra-/extra-personal notation [12] is still suitable here, and the feature selection is then directly applied on both kinds of heterogeneous face images in the training set after appearance normalization. Finally, using the learned discriminant features, we apply R-LDA [20] on the whole training set to construct a universal subspace for distinguishing different persons, whether they are homogeneous or heterogeneous. Therefore, the learned model is suitable for both kinds of face recognition, in despite of the light sources.

In the test phase, for identifying two given images (homogeneous or heterogeneous), the test procedure of the proposed approach is as follows. First, both images are preprocessed using DoG filtering for appearance normalization. And then, the selected MB-LBP features are extracted on each output correspondingly. After that, the extracted feature vectors are both projected onto the learned subspace. Finally, the cosine distance is measured on the two projected vectors, and the output value is compared with a certain threshold to decide the identify result.

4 Experiments

4.1 Data Preparation

In order to evaluate the performance of the heterogeneous face recognition, we focus on the NIR vs. VIS problem and collect a corresponding NIR-VIS database for experiment. The whole database contains 4582 images from 202 subjects, including 2095 VIS and 3002 NIR images. All of them are frontal face images.

The database is divided into two distinct sets for training and testing.The training set contains 3464 images, including 1580 VIS and 1884 NIR images from the former 150 persons, and the left 52 persons constitute the test set. There is no intersection for both face image and person between training set and test set, for the purpose of constructing an open-set test protocol. In the test set, 515 VIS images compose the target set and 1118 NIR ones consist the probe set.

All the faces are cropped to 128×128 according to the automatically detected eye positions [5]. Fig. 2 has already shown some examples for VIS and NIR face images from this database.

4.2 Performance Evaluation

To verify whether an algorithm learns the essential property of the heterogeneous faces, the performance evaluation runs on an open-set test protocol as described in the previous subsection. Fig. 4 illustrates the performance evaluation result on the test set in ROC curve. Our proposed method (denoted as "LSNA") is compared with other existing methods. "CDFE" result is obtained using the approach of [9]. "PCA-CCA" and "LDA-CCA" are algorithms proposed by [21]. Besides, we also show the result of [8], denoted as "MB-LBP". Note that all parameters are carefully turned to give the best result for the compared methods, such as the smooth factor in CDFE.

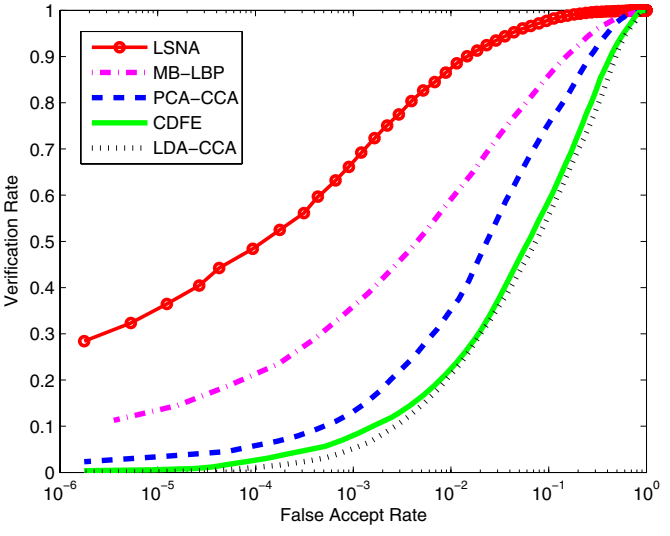

Fig. 4. Performance comparison with ROC curve

From the comparison it can be clearly seen that our approach significantly outperforms all other methods. Existing subspace projection based methods obtain poor results in the non-overlapping test set. As analyzed before, they are all applied directly on heterogeneous face images without appearance normalization and the intrinsic local-structure learning, as a result, they lack generalization ability and tend to overfit in the training data.

On the other hand, the proposed method learns the intrinsic local structure from the appearance-normalized heterogeneous face images, the resulting uniform model depends little on the training data, thus it generalizes well and gets a much better result on the test set. The verification rate of the proposed method at 0.1% false acceptance rate is 67.5%, and is 87.5% at false acceptance rate of 1%.

It can also be seen that the proposed approach outperforms MB-LBP method, which indicates that LSNA benefits a lot from the appearance normalization stage. The results also support our previous analysis that DoG filtering helps to normalize the heterogenous face appearance and contribute to recognition.

5 Summary and Conclusions

In this paper, we have presented an algorithm for heterogeneous face recognition and analyzed the particular problem of NIR to VIS face recognition. Because different types of face images have different skin spectra-optical properties, we first normalize the appearance of all heterogeneous face images using DoG filtering, and then we utilize MB-LBP to encode local image structures in the transformed space, and finally learn an universal model for both modalities of face

recognition. Experiments have shown the effectiveness of the proposed method in matching between NIR and VIS faces.

Since heterogeneous face recognition is a difficult problem, which features are more suitable for this problem is still unknown till now. Future works would be investigating ordinal filters [7,6] for more flexible encodings to tackle the problem, and applying our method for other heterogeneous face recognition problems.

Acknowledgements. This work was supported by the following fundings: National Natural Science Foundation Project #60518002, National Science and Technology Support Program Project #2006BAK08B06, National Hi-Tech (863) Program Projects #2006AA01Z192, #2006AA01Z193, and #2008AA01Z124, Chinese Academy of Sciences 100 People Project, and AuthenMetric R&D Funds.

References

1. Ahonen, T., Hadid, A., Pietikainen, M.: Face recognition with local binary patterns. In: Pajdla, T., Matas, J. (eds.) ECCV 2004. LNCS, vol. 3021, pp. 469–481. Springer, Heidelberg (2004)
2. Friedman, J., Hastie, T., Tibshirani, R.: Additive logistic regression: a statistical view of boosting. Technical report, Department of Statistics, Sequoia Hall, Stanford Univerity (July 1998)
3. Gross, R., Brajovic, V.: An image preprocessing algorithm for illumination invariant face recognition. In: Proc. 4th International Conference on Audio- and Video-Based Biometric Person Authentication, Guildford, UK, June 9-11 (2003)
4. Jacobs, D., Belhumeur, P., Basri, R.: Comparing images under variable illumination. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 610–617 (1998)
5. Li, S.Z., Chu, R., Liao, S., Zhang, L.: Illumination invariant face recognition using near-infrared images. IEEE Transactions on Pattern Analysis and Machine Intelligence 29(4), 627–639 (2007)
6. Liao, S., Lei, Z., Li, S.Z., Yuan, X., He, R.: Structured ordinal features for appearance-based object representation. In: Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures, pp. 183–192 (2007)
7. Liao, S., Lei, Z., Zhu, X., Sun, Z., Li, S.Z., Tan, T.: Face recognition using ordinal features. In: Zhang, D., Jain, A.K. (eds.) ICB 2006. LNCS, vol. 3832, pp. 40–46. Springer, Heidelberg (2006)
8. Liao, S., Zhu, X., Lei, Z., Zhang, L., Li, S.Z.: Learning multi-scale block local binary patterns for face recognition. In: Lee, S.-W., Li, S.Z. (eds.) ICB 2007. LNCS, vol. 4642, pp. 828–837. Springer, Heidelberg (2007)
9. Lin, D., Tang, X.: Inter-modality face recognition. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3954, pp. 13–26. Springer, Heidelberg (2006)
10. Lindeberg, T.: Scale-space theory: A basic tool for analysing structures at different scales. Journal of Applied Statistics 21, 225–270 (1994)
11. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60, 91–110 (2004)

12. Moghaddam, B., Nastar, C., Pentland, A.: A Bayesain similarity measure for direct image matching. Media Lab Tech. Report No. 393, MIT (August 1996)
13. Nayar, S.K., Bolle, R.M.: Reflectance based object recognition. International Journal of Computer Vision 17(3), 219–240 (1996)
14. Ojala, T., Pietikainen, M., Harwood, D.: A comparative study of texture measures with classification based on feature distributions. Pattern Recognition 29(1), 51–59 (1996)
15. Ojala, T., Pietikainen, M., Maenpaa, M.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence 24(7), 971–987 (2002)
16. Shashua, A., Raviv, T.R.: The quotient image: Class based re-rendering and recognition with varying illuminations. IEEE Transactions on Pattern Analysis and Machine Intelligence 23(2), 129–139 (2001)
17. Tan, X., Triggs, B.: Enhanced local texture feature sets for face recognition under difficult lighting conditions. In: Proceedings of the IEEE International Workshop on Analysis and Modeling of Faces and Gestures (2007)
18. Wang, H.T., Li, S.Z., Wang, Y.S.: Generalized quotient image. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 498–505 (2004)
19. Yang, W., Yi, D., Lei, Z., Sang, J., Li, S.Z.: 2D-3D face matching using CCA. In: Proc. IEEE International Conference on Automatic Face and Gesture Recognition, Amsterdam, The Netherlands, September 17-19 (2008)
20. Ye, J., Xiong, T., Li, Q., Janardan, R., Bi, J., Cherkassky, V., Kambhamettu, C.: Efficient model selection for regularized linear discriminant analysis. In: Proceedings of the ACM Conference on Information and Knowledge Management, pp. 532–539 (2006)
21. Yi, D., Liu, R., Chu, R., Lei, Z., Li, S.Z.: Face matching from near infrared to visual images. In: Proceedings of the 2nd IAPR/IEEE International Conference on Biometrics, Seoul, Korea (August 2007)
22. Zhang, L., Chu, R., Xiang, S., Liao, S., Li, S.Z.: Face detection based on multi-block lbp representation. In: Lee, S.-W., Li, S.Z. (eds.) ICB 2007. LNCS, vol. 4642, pp. 11–18. Springer, Heidelberg (2007)