Appeared in Proc. of International Conference on Image Analysis and Recognition (ICIAR), (Burnaby, Canada), June 2011

Can Gender Be Predicted from Near-Infrared Face Images?*

Arun Ross and Cunjian Chen

Lane Department of Computer Science and Electrical Engineering
West Virginia University, USA
arun.ross@mail.wvu.edu, cchen10@mix.wvu.edu

Abstract. Gender classification based on facial images has received increased attention in the computer vision literature. Previous work on this topic has focused on images acquired in the visible spectrum (VIS). We explore the possibility of predicting gender from face images acquired in the near-infrared spectrum (NIR). In this regard, we address the following two questions: (a) Can gender be predicted from NIR face images; and (b) Can a gender predictor learned using VIS images operate successfully on NIR images and vice-versa? Our experimental results suggest that NIR face images do have some discriminatory information pertaining to gender, although the degree of discrimination is noticeably lower than that of VIS images. Further, the use of an illumination normalization routine may be essential for facilitating cross-spectral gender prediction.

Keywords: Biometrics, Faces, Gender, Near-Infrared, Cross-Spectral.

1 Introduction

Automated gender identification plays an important role in Human-Computer Interaction (HCI), upon which more complex visual systems are built [13]. Recognizing a person's gender will enhance the HCI's ability to respond in a user-friendly and socially acceptable manner. In the realm of biometrics, gender is viewed as a soft biometric trait that can be used to index databases or enhance the recognition accuracy of primary traits such as face [6]. Predicting gender based on human faces has been extensively studied in the literature [14,1]. Two popular methods are due to Moghaddam et al. [14] who utilize a support vector machine (SVM) for gender classification of thumbnail face images, and Baluja et al. [1] who present the use of Adaboost for predicting gender. Gutta et al. [5] consider the use of hybrid classifiers consisting of an ensemble of RBF networks and decision trees. Other approaches utilize gender-specific information, such as hair, to enhance gender prediction [11], or genetic algorithms to select features that encode gender information [15]. A systematic overview on the topic of gender classification from face images can be found in [13].

* Funding from the Office of Naval Research is gratefully acknowledged.

M. Kamel and A. Campilho (Eds.): ICIAR 2011, Part II, LNCS 6754, pp. 120–129, 2011.

Appeared in Proc. of International Conference on Image Analysis and Recognition (ICIAR), (Burnaby, Canada), June 2011

Though gender classification has received much attention from the research community, previous work has focused on face images obtained in the visible spectrum (VIS). The aim of this study is to explore gender classification in the near-infrared spectrum (NIR) using learning-based algorithms. The use of NIR images for face recognition has become necessary especially in the context of a night-time environment where VIS face images cannot be easily discerned [9]. Further, NIR images are less susceptible to changes in ambient illumination. Thus, cross-spectral matching has become an important topic of research [8,12,2]. To the best of our knowledge, this is the first work that explores gender recognition in NIR face images. In this regard, we address the following questions:

- **Q1.** Can gender be predicted from NIR face images?
- **Q2.** Can a gender predictor learned using VIS images operate successfully on NIR images, and vice-versa?

To answer **Q1**, we use an existing gender prediction mechanism based on SVM [14]. In order to address **Q2**, we hypothesize that an illumination normalization scheme may be necessary prior to invoking the gender classifier.

In the next section, we describe the design of the gender classifier (predictor), with special emphasis on illumination normalization approaches for cross-spectral gender prediction. Then, we report experimental results that demonstrate the possibility of assessing gender from NIR face images. Finally, we discuss the difficulties in cross-spectral gender classification and indicate future directions.

2 Proposed Methods

In order to address the questions raised above, we utilize a gender prediction scheme based on SVMs. Such a scheme has been shown to be efficient in the VIS domain [14,4]. The SVM-based classification scheme is described below. Given a facial image x_i, in either the VIS or NIR domains, the feature extractor is applied to obtain a discriminant feature set s_i. The gender classifier, G, is then invoked to predict the gender from the image (Figure 1):

$$G(\alpha(x_i)) = \begin{cases} 1 & \text{if } x_i \text{ is male} \\ -1 & \text{if } x_i \text{ is female.} \end{cases} \quad (1)$$

Here $s_i = \alpha(x_i)$ represents the transformation of the raw image into a feature vector.

2.1 Feature Extractor

Previous work on gender classification in the visible domain utilized features extracted via Principal Component Analysis (PCA) [14,4,15] or Haar-like features [1]. Features based on local binary patterns (LBP) have also been used [10].

Appeared in Proc. of International Conference on Image Analysis and Recognition (ICIAR), (Burnaby, Canada), June 2011

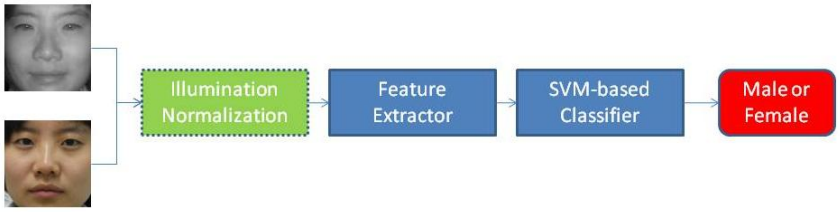

Fig. 1. Cross-spectral gender classifier

In this work, we use features derived from PCA since they have been successfully tested in previous literature. Consider a labeled set of N training samples $\{(x_i, y_i)\}_{i=1}^{N}$, where x_i is the facial image and y_i is the associated class label. Here, $y_i \in \{-1, 1\}$, where a -1 (+1) indicates a female (male). The PCA is performed on the covariance matrix of vectorized images,

$$\Sigma_g = \frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(x_i - \bar{x})^T, \tag{2}$$

where x_i is the sample image after vectorization and $\bar{x}$ is the mean vector of the training set. The eigenvectors can be obtained through the decomposition,

$$\Sigma_g \Phi_g = \Phi_g \Lambda_g, \tag{3}$$

where Φ_g are the eigenvectors of Σ_g and Λ_g are the corresponding eigenvalues. The number of eigenvectors used in our experiment is 60. The gender features can be extracted by projecting the sample image into the subspace expanded by eigenvectors:

$$s_i = \Phi_g^T(x_i - \bar{x}). \tag{4}$$

Here, s_i is the feature vector to represent the gender information of sample x_i. The feature vectors corresponding to the training set and their label information $\{s_i, y_i\}$ are stored in the database. In the test (i.e., evaluation) stage, when an unknown face image is presented, the same feature extractor is invoked to obtain the feature vector, which is input to the classifier G to predict the gender.

2.2 Gender Classifier

To build the gender classifier G using SVM, we need a labeled set of N training samples $\{(s_i, y_i)\}_{i=1}^{N}$. The gender classifier seeks to find an optimal hyperplane defined as

$$f(s) = \sum_{i=1}^{M} y_i \alpha_i \cdot k(s, s_i) + b, \tag{5}$$

where $k(s, s_i)$ represents the kernel function and the sign of $f(s)$ determines the class label of s (gender). The linear kernel is the simplest function, and it is

Appeared in Proc. of International Conference on Image Analysis and Recognition (ICIAR), (Burnaby, Canada), June 2011

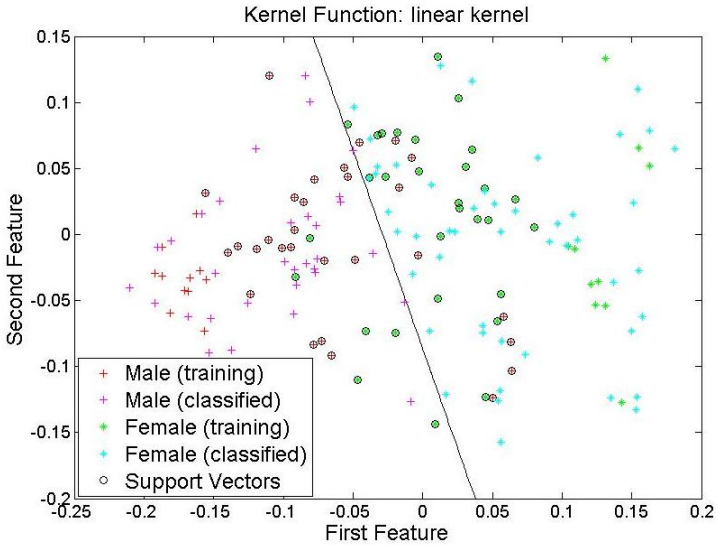

Fig. 2. Illustration of a SVM-based gender classifier with linear kernel on the HFB database [9]

computed by the dot product $< s, s_i >$ plus an optional constant c. Any vector s_i that has a non-zero α_i is called a support vector (SV) of the optimal hyperplane that separates the two classes. The common kernels used are the radial basis function (RBF) kernel and the linear kernel. In this work, the linear kernel was used. An example of gender classification using SVM is shown in Figure 2. Here, the dimension of the extracted feature vector is reduced to two by PCA. The classifier was trained using images from the VIS spectrum and tested on images in the NIR spectrum.

2.3 Illumination Normalization

As stated earlier, we hypothesize that the use of an illumination normalization scheme may be necessary to accommodate cross-spectral gender prediction where the training and test sets have images pertaining to different spectral bands.

Self Quotient Image (SQI): According to the Lambertian model, the image formation process is described as follows:

$$I(x, y) = \rho_w(x, y)n(x, y)s, \tag{6}$$

where $\rho_w(x, y)$ is the albedo of the facial surface, n is the surface normal and s is the lighting reflection. To reduce the impact of illumination, we need to separate out the extrinsic factor s from ρ and n. The self-quotient image, Q, of I is defined as [18],

$$Q = \frac{I(x,y)}{I(\hat{x},y)} = \frac{\rho_w(x,y)n(x,y)s}{G * [\rho_w(x,y)n(x,y)s]}, \tag{7}$$

where $\hat{I}$ is the smoothed version of I and G is the smoothing kernel.

Retinex Model: The retinex approach is based on the reflectance illumination model instead of the Lambertian model. It is an image enhancement algorithm [7] proposed to account for the lightness and color constancy of the dynamic range compression properties of the human vision system. It tries to compute the invariant property of reflectance ratio under varying illumination conditions [3,18]. The retinex model is described as follows:

$$I(x,y) = R(x,y)L(x,y), \tag{8}$$

where $I(x,y)$ is the image, $R(x,y)$ is the reflectance of the scene and $L(x,y)$ is the lighting. The lighting is considered to be the low-frequency component of the image $I(x,y)$, and is thus approximated as,

$$L(x,y) = G(x,y) * I(x,y), \tag{9}$$

where $G(x,y)$ is a Gaussian filter and $*$ denotes the convolution operator. The output of the retinex approach is the image $R(x,y)$ that is computed as,

$$R(x,y) = \frac{I(x,y)}{L(x,y)} = \frac{I(x,y)}{G(x,y) * I(x,y)}. \tag{10}$$

Discrete Cosine Transform (DCT) Model: Since illumination variations typically manifest in the low-frequency domain, it is reasonable to normalize the illumination by removing the low-frequency components of an image. DCT can be first applied to transform an image from the spatial domain to the frequency domain, and then estimate the illumination of the image via low-frequency DCT coefficients which appear in the upper-left corner of the DCT [3]. By setting the low-frequency components to zero and reconstructing the image, variations due to illumination can be reduced. The 2D $M \times N$ DCT can be computed as,

$$C(u,v) = \alpha(u)\alpha(v) \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} I(x,y) \times cos\left[\frac{\pi(2x+1)u}{2M}\right] cos\left[\frac{\pi(2y+1)v}{2N}\right]. \tag{11}$$

Here $\alpha(u)$ and $\alpha(v)$ are the normalization factors.

CLAHE Normalization: The CLAHE (Contrast Limited Adaptive Histogram Equalization) [19] method applies contrast normalization to local blocks in the image such that the histogram of pixel intensities in each block approximately matches a pre-specified histogram distribution. This scheme is applied to blocks of size 16×16. CLAHE is effective at improving local contrast without inducing much noise. It utilizes the normalized cumulative distribution of each gray level, x, in the block [2]:

$$f(x) = \frac{N-1}{M} \times \sum_{k=0}^{x} h(k). \tag{12}$$

Here, M is the total number of pixels in the block, N is the number of gray levels in the block, and h is the histogram of the block. To improve the contrast, the CLAHE technique transforms the histogram of the block such that the histogram height falls below a pre-specified threshold. Gray level counts beyond the threshold are uniformly redistributed among the gray levels below the threshold. The blocks are then blended across their boundaries using bilinear interpolation.

Difference-of-Gaussian (DoG) Filtering: Another type of normalization is proposed in [16], where the local image structures are enhanced. One of the key components in [16] is the Difference-of-Gaussian (DoG) filtering, which can be computed as,

$$D(x, y|\sigma_0, \sigma_1) = [G(x, y, \sigma_0) - G(x, y, \sigma_1)] * I(x, y). \quad (13)$$

The symbol * is the convolution operator, and the gaussian kernel function based on σ is,

$$G(x, y, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x^2+y^2)/2\sigma^2}. \quad (14)$$

This simple filtering scheme has the effect of subtracting two Gaussian filters.

The output of the various illumination normalization schemes are presented in Figure 3. The goal of illumination normalization is to facilitate cross-spectral gender classification by mitigating the effect of spectral specific features [17].

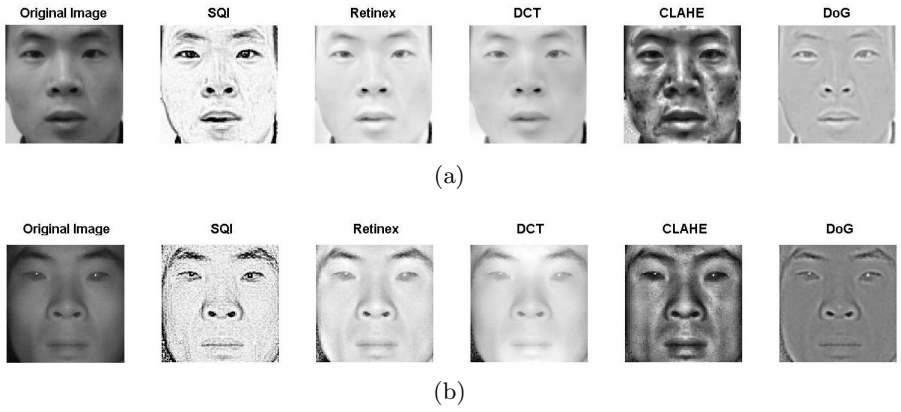

Fig. 3. (a) A VIS image and its corresponding normalized images; (b) A NIR image and its corresponding normalized images

3 Experiments

The HFB face database [9] consists of 100 subjects, including 57 males and 43 females. There are 4 VIS and 4 NIR face images per subject. The following experiments were conducted on this database: (a) Training and testing using VIS images (VIS-VIS); (b) Training and testing using NIR images (NIR-NIR);

Appeared in Proc. of International Conference on Image Analysis and Recognition (ICIAR), (Burnaby, Canada), June 2011

(c) Training using VIS images and testing using NIR images (VIS-NIR); (d) Training using NIR images and testing using VIS images (NIR-VIS). In all cases, the subjects used in the training and test sets were mutually exclusive. 20 male and 20 female subjects were randomly selected for training, with 4 samples for each subject. The remaining subjects were reserved for testing. This random partitioning to generate the training and test sets was done 10 times for each experiment in order to understand the variance in classification accuracy. The image size used in our work was 128×128.

Table 1. Gender classification results on the HFB database when illumination normalization is **not** used for cross-spectral prediction

Scenario	**Classification Rate**	**Best**	**Worst**
VIS-VIS	0.9067 ± 0.0397	0.9708	0.8458
NIR-NIR	0.8442 ± 0.0264	0.9000	0.8042
VIS-NIR	0.5625 ± 0.1289	0.7083	0.3833
NIR-VIS	0.6021 ± 0.0769	0.6667	0.3875

Table 2. Results for cross-spectral gender classification after applying different normalization schemes

	VIS-NIR(N)	NIR-VIS(N)
CLAHE	$\mathbf{0.6617 \pm 0.0724}$	$\mathbf{0.6642 \pm 0.0806}$
DoG	0.6446 ± 0.0331	0.6100 ± 0.0354
SQI	0.4512 ± 0.0693	0.4692 ± 0.0611
Retinex	0.5525 ± 0.0537	0.5921 ± 0.0674
DCT	0.5967 ± 0.0840	0.6392 ± 0.0666

For the VIS-VIS experiments, the average classification rate was 90.67%, with the best performance being 97.08% (Table 1). The performance is comparable to the results reported in previous literature on other datasets [14,1]. This suggests that gender classification can be performed with high accuracy in the VIS domain. For the NIR-NIR experiment, the average performance declined by around 6% compared to VIS-VIS classification resulting in an average accuracy rate of 84.42%. For the VIS-NIR and NIR-VIS experiments, the average classification rates were 56.25% and 60.21%, respectively, suggesting the difficulty in performing cross-spectral gender classification.

However, upon applying certain illumination normalization schemes (to both the training and test images), we observed an improvement in classification accuracy (Table 2). Two of the most effective normalization schemes were CLAHE and DoG. In our experiment, the CLAHE gave slightly better performance than DoG. Specifically, the CLAHE normalization scheme improved cross-spectral gender classification for the VIS-NIR and NIR-VIS experiments to 66.17% and 66.42%, respectively - this represents an improvement of 18% and 10%, respectively. The SQI scheme decreased the performance after normalization, while the retinex model did not impact the accuracy. The DCT algorithm gave slightly better performance, but not as significant as that of CLAHE and DoG.

Appeared in Proc. of International Conference on Image Analysis and Recognition (ICIAR), (Burnaby, Canada), June 2011

Table 3. Impact of image size on gender classification for the VIS-NIR and NIR-VIS scenarios when the CLAHE normalization method is used

Image Size	VIS-NIR	NIR-VIS
128×128	0.6617 ± 0.0724	0.6642 ± 0.0806
64×64	0.6958 ± 0.0241	0.6596 ± 0.0856
32×32	$\mathbf{0.7179 \pm 0.0208}$	$\mathbf{0.6917 \pm 0.0292}$
16×16	0.6638 ± 0.0362	0.6617 ± 0.0668

4 Discussion

Our experimental results indicate the possibility of performing gender classification using NIR face images although the performance is slightly inferior to that of VIS images. This indicates that the gender information observed in the NIR domain *may* not be as discriminative as in the VIS domain. Cross-spectral gender prediction was observed to be difficult - this suggests that the gender related information available in the NIR and VIS face images are significantly different as assessed by the classifier. The key, therefore, is to reduce the variability between these two type of images by applying an illumination normalization routine. Experiments suggest that certain normalization schemes are better than the others. In particular, the CLAHE scheme proved superior than the other models considered in this work.

Next, we consider the reasons for the inferior performance of the other normalization models. The Lambertian model usually assumes that the term $\rho_w(x, y)$ is constant across different lighting sources. However, since the lighting conditions under NIR and VIS spectra are not homogeneous, estimating an illumination invariant albedo $\rho_w(x, y)$ under the Lambertian model for those two type of images is not possible [12]. Therefore, approaches based on the Lambertian model, such as self-quotient image and its variants, are not useful in our application. Since the reflectance is not a stable characteristic of facial features for images captured under the NIR and VIS spectra, the retinex model also does not result in good performance. The DCT method fails since the illumination in NIR images cannot be simply estimated by the low-frequency coefficients of the image. Only those normalization methods based on local appearance-based features (i.e., CLAHE and DoG) result in better accuracy. This could partly be due to the use of PCA-based features in our experiments. The use of other sophisticated features (such as LBP) for gender classification may be useful when the SQI and retinex models are used for normalization.

When the images (128×128) are downsampled by a factor of 16, the average accuracy of VIS-NIR improved from 66.17% to 71.79% (Table 3). Similarly, the average accuracy of NIR-VIS improved from 66.42% to 69.17%. Another observation has to do with the difference in gender classification of males and females. We ran the VIS-NIR experiments on the HFB database 100 times and observed that the female classification rate was 68% while the male classification rate was 77%.

Appeared in Proc. of International Conference on Image Analysis and Recognition (ICIAR), (Burnaby, Canada), June 2011

Next, we take a look at the histogram distribution of pixel intensities for a VIS image and a NIR image (Figure 4). The VIS image has a dense histogram, while the NIR image has a relatively sparse histogram distribution. This suggests that the VIS image has more intensity values captured than its counterpart. Such a difference suggests a loss in information when forming NIR images. The hypothesis is that histogram normalization can mitigate some of these differences thereby improving cross-spectral gender prediction. We find that by applying the CLAHE normalization approach, it is possible to reduce the difference between the two histograms (Figure 4).

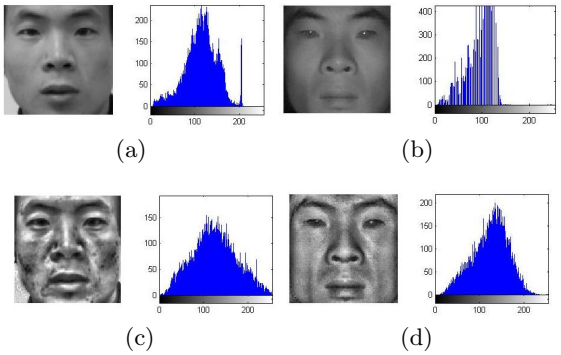

Fig. 4. (a) VIS image before normalization; (b) NIR image before normalization; (c) VIS image after normalization; (d) NIR image after normalization.

5 Summary

This paper presents initial experimental results on gender classification from NIR face images. A classification accuracy of 84.42% was obtained in the NIR-NIR scenario. The paper also discusses cross-spectral gender recognition where training images and test images originate from different spectral bands. The preprocessing operation involving illumination normalization was observed to improve cross-spectral classification accuracy by up to 18%. But this is still lower than the performance obtained for intra-spectral classification (i.e., the VIS-VIS and NIR-NIR scenarios). Currently, we are examining the use of fundamental image formation models to better understand the gender-specific details present in NIR and VIS images.

References

1. Baluja, S., Rowley, H.A.: Boosting sex identification performance. IJCV 71(1), 111–119 (2007)
2. Bourlai, T., Kalka, N.D., Ross, A., Cukic, B., Hornak, L.: Cross-spectral face verification in the short wave infrared (SWIR) band. In: ICPR, pp. 1343–1347 (2010)

3. Chen, W., Er, M.J., Wu, S.: Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain. IEEE SMC-B 36(2), 458–466 (2006)
4. Graf, A.B.A., Wichmann, F.A.: Gender classification of human faces. In: Biologically Motivated Computer Vision, pp. 491–500 (2002)
5. Gutta, S., Wechsler, H., Phillips, P.J.: Gender and ethnic classification of face images. In: FG, pp. 194–199 (1998)
6. Jain, A.K., Dass, S.C., Nandakumar, K.: Can soft biometric traits assist user recognition? In: BTHI, pp. 561–572. SPIE, San Jose (2004)
7. Jobson, D.J., Rahman, Z., Woodell, G.A.: Properties and performance of a center/surround retinex. IEEE TIP 6(3), 451–462 (1997)
8. Klare, B., Jain, A.K.: Heterogeneous face recognition: Matching NIR to visible light images. In: ICPR, pp. 1513–1516 (2010)
9. Li, S.Z., Lei, Z., Ao, M.: The HFB face database for heterogeneous face biometrics research. In: CVPR Workshop, pp. 1–8 (2009)
10. Lian, H.-C., Lu, B.-L.: Multi-view gender classification using local binary patterns and support vector machines. In: Wang, J., Yi, Z., Żurada, J.M., Lu, B.-L., Yin, H. (eds.) ISNN 2006. LNCS, vol. 3972, pp. 202–209. Springer, Heidelberg (2006)
11. Lian, X.-C., Lu, B.-L.: Gender classification by combining facial and hair information. In: ICONIP (2), pp. 647–654 (2008)
12. Liao, S., Yi, D., Lei, Z., Qin, R., Li, S.Z.: Heterogeneous face recognition from local structures of normalized appearance. In: ICB, pp. 209–218 (2009)
13. Makinen, E., Raisamo, R.: Evaluation of gender classification methods with automatically detected and aligned faces. PAMI 30(3), 541–547 (2008)
14. Moghaddam, B., Yang, M.-H.: Learning gender with support faces. PAMI 24(5), 707–711 (2002)
15. Sun, Z., Bebis, G., Yuan, X., Louis, S.J.: Genetic feature subset selection for gender classification: A comparison study. In: WACV, pp. 165–170 (2002)
16. Tan, X., Triggs, B.: Enhanced local texture feature sets for face recognition under difficult lighting conditions. In: AMFG, pp. 168–182 (2007)
17. Štruc, V., Pavešić, N.: Gabor-based kernel partial-least-squares discrimination features for face recognition. Informatica 20, 115–138 (2009)
18. Wang, H., Li, S.Z., Wang, Y., Zhang, J.: Self quotient image for face recognition. In: ICIP, pp. 1397–1400 (2004)
19. Zuiderveld, K.: Contrast limited adaptive histogram equalization, pp. 474–485. Academic Press Professional, San Diego (1994)