

Chapter 15

Face Recognition Using Near Infrared Images

Stan Z. Li and Dong Yi

15.1 Introduction

Face recognition should be based on intrinsic factors of the face, such as, the 3D shape and the albedo of the facial surface. Extrinsic factors that include illumination, eyeglasses, and hairstyle, are irrelevant to biometric identity, and hence their influence should be minimized. Out of all these factors, variation in illumination is a major challenge and needs to be tackled first.

Conventional visual (VIS) image based face recognition systems, academic and commercial, are compromised in accuracy by changes in environmental illumination, even for cooperative user applications in an indoor environment. In an in-depth study on influence of illumination changes on face recognition [1], Adini et al. examined several distance measures and local image operators, including Gabor filters, local directive filters, and edge maps, which were considered to be relatively insensitive to illumination changes for face recognition. Several conclusions were made: (i) lighting conditions, and especially light angle, drastically change the appearance of a face; (ii) when comparing unprocessed images, the changes between images of a person under different illumination conditions are larger than those between images of two persons under the same illumination; (iii) all the local filters under study, are not capable overcoming variations due to changes in illumination direction. The influence of illumination is also shown in evaluations such as Face Recognition Vendor Test [17].

Near infrared (NIR) based face recognition [11–14], as opposed to the conventional visible light (VIS) based methods, is an effective approach for overcoming the impact of illumination changes on face recognition. It uses a special purpose

S.Z. Li (✉) · D. Yi

Center for Biometrics and Security Research & National Laboratory of Pattern Recognition,
Institute of Automation, Chinese Academy of Sciences, Beijing, China
e-mail: szli@cbsr.ia.ac.cn

D. Yi

e-mail: d yi@cbsr.ia.ac.cn

imaging device to capture front-lighted NIR face images [3–5], normalizing the illumination direction. Using a proper face feature representation, such as Local Binary Pattern (LBP) [2, 9, 18], variation in the illumination strength is also overcome. These lead to a complete illumination-invariant face representation. Problems caused by uncontrolled environmental illumination are minimized thereby, and difficulties in building the face matching engine are alleviated. The NIR approach usually achieves significantly higher performance than the VIS approach for cooperative user application scenarios in uncontrolled illumination environment. NIR face recognition products and integrated systems have been in the market and are used in many applications (refer to Chap. 1).

In this chapter, we introduce the NIR face recognition approach, describe the design of active NIR face imaging system, illustrate how to derive from NIR face image an illumination invariant face representation, and provide a learning based method for face feature selection and classification. Experiments are presented.

15.2 Active NIR Imaging System

The key aspect in the NIR face recognition approach is a special purpose NIR image capture hardware system [14]. Its goal is to overcome the problem arising from uncontrolled environmental light and produce face images of a good illumination condition for face recognition. Good illumination means (i) that the lighting to the face is from the frontal direction and (ii) that the face image has suitable pixel intensities.

To achieve illumination from the frontal direction, active NIR illuminators, for example, space light-emitting diodes (LEDs) around the camera lens, are used to illuminate the face from the front such that front-lighted NIR face images are acquired. This is similar to a camera with a flash light but the NIR lights work in the invisible spectrum of NIR, being nonintrusive to human eyes.

The following are the main requirements for the NIR imaging system:

1. The active NIR lights should be nonintrusive to human eyes.
2. The direction of the NIR lighting to the face should be fixed.
3. The active NIR light signals arriving at the camera sensor should override the signals from other light sources in the environment.

Here, the NIR lights mean the active NIR lights from the NIR imaging system, excluding NIR components in the environment such as sunlight and light bulbs.

This selective capture (of NIR light from the imaging system) can be achieved by the following methods:

1. Choose illuminators such as LEDs in an invisible spectrum. While a 850 nm LED light looks a dim dark red, 940 nm is entirely invisible.
2. Mount the NIR LEDs around the camera lens so as to illuminate from the frontal direction.

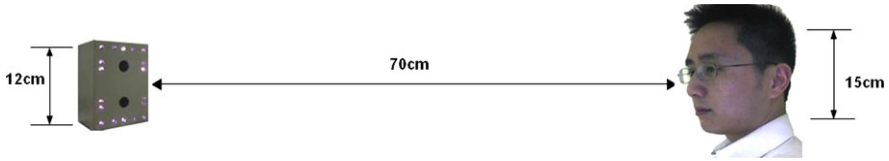


Fig. 15.1 An active NIR imaging device and the face

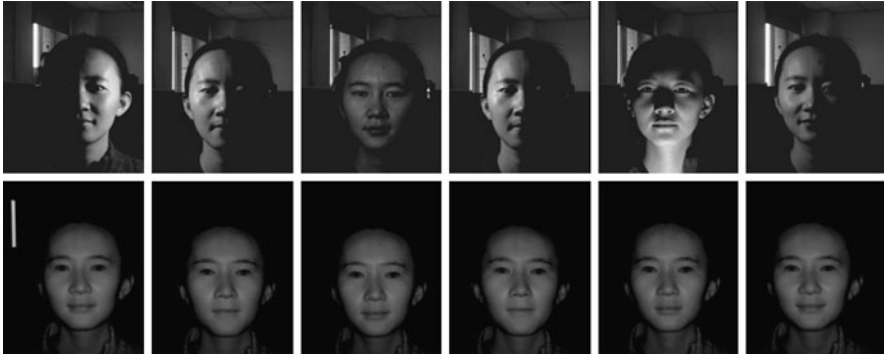


Fig. 15.2 Color images (*top*) captured by a color camera versus NIR images (*bottom*) captured by an NIR imaging system. While unfavorable lighting is obvious in the color face images, it is almost unseen in the NIR face images

3. Choose the NIR LEDs to be powerful enough to override environmental light sources that may affect the system. The most favorable or least challenging environment is when there is total darkness and the least favorable or most challenging is sunlight in summer. A short camera exposure should be used to avoid over-exposure when the LED power is high.
4. Use an optical filter to minimize lights from the environment. One option is to use a long pass filter that filters out visible components of environmental lights. A better but more expensive option, is to use a narrow band pass filter that matches the wavelength of the chosen active LEDs.

Figure 15.1 illustrates a hardware device and its positional relationship with the face. The device consists of 18 NIR LEDs, an NIR camera, a VIS color camera, and the casing. The NIR LEDs and the NIR camera are for NIR face image acquisition. The hardware and the face are relatively positioned in such a way that the lighting is frontal and the NIR rays provide nearly homogeneous illumination on the face, which is an excellent illumination condition for face recognition. The VIS image may be used for visual feedback and human computer interaction (HCI). The imaging device works at a rate of 30 frames per second with the USB 2.0 protocol for 640×480 images.

Figure 15.2 depicts images of a face illuminated by NIR LED lights from the front, a lamp aside and environmental lights. We can see the following: (i) the lighting conditions are likely to cause problems for face recognition with the color im-

ages; (ii) the NIR images, with the visible light composition cut off by the filter, are mostly frontal-lighted (by the NIR lights), with minimum influence from the side lighting. Based on the two set of images that were obtained the following observations can be made.

In outdoor environments, the sunlight contains very a strong NIR component, much stronger than what can be overridden by the NIR imaging system described above. The NIR imaging hardware must be enhanced to overcome the influence from the sunlight. The key factor is the ratio between the power of the controlled active NIR illumination and the power from other light sources (recorded by the image sensor).

One solution for such an enhanced NIR imaging system is provided in [21]. The system uses a powerful NIR illumination flash and synchronizes it with image sensor exposure. The main points of the enhanced imaging system are summarized as follows:

1. Use a powerful NIR illuminator, such as a narrow band NIR laser generator.
2. Set very short imaging exposure time, for example, 50 μ s.
3. Synchronize the time of NIR flash output to the exposure window of sensor.
4. Use a narrow band-pass optical filter that matches the wavelength of the chosen active NIR flash.

The enhanced NIR imaging (ENIR) system may include an additional processing step to further reduce the influence from strong and uncontrolled lights in the environment. This is done by taking difference of two successive frames. The first frame is captured by illuminating the subject's face with an active light source and the second is captured after turning this light source off. The second frame is used to represent the face subject under ambient static lights. The difference operation is performed to reduce or eliminate the strong NIR component in the ambient light or sunlight [10], and thus output an image of the face illuminated by the active NIR lighting in the frontal direction. As a result, this ENIR imaging system not only provides appropriate active frontal lighting but also minimizes ambient light as well as outdoor sunlight.

Figure 15.3 shows some images captured when the active light source on/off and the finally output of the ENIR camera under the sunlight. From Fig. 15.3(c), we can see that ENIR imaging hardware can work properly under the sunlight. Figure 15.10 shows face images captured by the ENIR imaging system in the sunlight. It demonstrates the effect of reducing NIR component from sunlight.

15.3 Illumination Invariant Face Representation

In this section, we first provide an analysis using a Lambertian surface imaging model to show that the NIR images contain the most relevant, intrinsic information about a face, subject only to a multiplying constant or a monotonic transform due to lighting intensity changes. We then present an Local Binary Pattern (LBP) based representation to amend the degree of freedom of the monotonic transform to achieve an illumination invariant representation for face recognition applications.

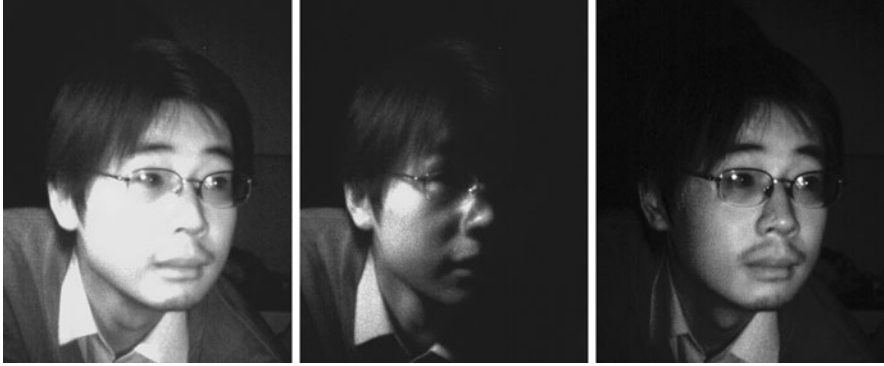


Fig. 15.3 From left to right: **a** Face image in mixture of sunlight and active NIR light; **b** face image in sunlight only, **c** image difference (**a**)–(**b**)

15.3.1 Modeling of Active NIR Images

According to the Lambertian model, an image $I(x, y)$ under a point light source is formed according to the following criterion:

$$I(x, y) = \rho(x, y) \mathbf{n}(x, y) \mathbf{s} \quad (15.1)$$

where, $\rho(x, y)$ is the albedo of the facial surface material at point (x, y) , $\mathbf{n} = (n_x, n_y, n_z)$ is the surface normal (a unit row vector) in 3D space, and $\mathbf{s} = (s_x, s_y, s_z)$ is the lighting direction (a column vector, with magnitude). Albedo $\rho(x, y)$ reflects the photometric properties of facial skin and hairs. $\mathbf{n}(x, y)$ is the geometric shape of the face. The most important factor that affects the face recognition performance is the direction of the incident lighting relative to the face surface normal. The product of $\rho(x, y)$ and $\mathbf{n}(x, y)$ is the intrinsic property of the face at a fixed pose and is the only factor needed for face detection and recognition. Therefore, \mathbf{s} is the extrinsic property that should be removed. Assume $\mathbf{s} = \kappa \mathbf{s}^0$, where κ is a multiplying constant that is introduced to account for possible changes in the strength of the lighting caused by changes in the distance between the face and the LED lights, and $\mathbf{s}^0 = (s_x^0, s_y^0, s_z^0)$ is a unit column vector of the lighting direction. Let $\theta(x, y)$ be the incident angle between the lighting and the face surface normal at point (x, y) , then $\cos \theta(x, y) = \mathbf{n}(x, y) \mathbf{s}^0$. Equation (15.1) can be expressed as

$$I(x, y) = \kappa \rho(x, y) \cos \theta(x, y). \quad (15.2)$$

A less restrictive modeling of constant κ would be to use a monotonic transform instead of a constant. It's evident we see that the face image $\rho(x, y) \cos \theta(x, y)$ changes as the lighting direction changes, given albedo $\rho(x, y)$ and 3D shape $\mathbf{n}(x, y)$ fixed. The present hardware design is aimed at preserving the intrinsic property while minimizing variation due to the extrinsic factor of environmental lights. When the active NIR lighting is from the (nearly) frontal direction (see Fig. 15.1),

Fig. 15.4 LBP code for 3×3 window

Local Window	Thresholded	Weights																											
<table border="1"> <tr><td>18</td><td>15</td><td>8</td></tr> <tr><td>21</td><td>18</td><td>6</td></tr> <tr><td>27</td><td>23</td><td>22</td></tr> </table>	18	15	8	21	18	6	27	23	22	<table border="1"> <tr><td>1</td><td>0</td><td>0</td></tr> <tr><td>1</td><td>0</td><td>0</td></tr> <tr><td>1</td><td>1</td><td>1</td></tr> </table>	1	0	0	1	0	0	1	1	1	<table border="1"> <tr><td>8</td><td>4</td><td>2</td></tr> <tr><td>16</td><td>0</td><td>1</td></tr> <tr><td>32</td><td>64</td><td>128</td></tr> </table>	8	4	2	16	0	1	32	64	128
18	15	8																											
21	18	6																											
27	23	22																											
1	0	0																											
1	0	0																											
1	1	1																											
8	4	2																											
16	0	1																											
32	64	128																											

LBP String = (0001111)

LBP Code = $0+0+0+8+16+32+64+128=248$

that is, $\mathbf{s}^0 = (0, 0, 1)$, the image can be approximated by

$$I(x, y) = \kappa \rho(x, y) n_z(x, y), \quad (15.3)$$

where, $n_z(x, y)$ is the z component of the surface normal that can be acquired by a range imaging system. An active NIR image $I(x, y)$ combines information about both surface normal component $n_z(x, y)$ and albedo map $\rho(x, y)$ and, therefore, provides the required intrinsic property about a face for face recognition.

15.3.2 Compensation for Monotonic Transform

Given a face, the constant κ is the only factor affecting the intensity of the face image. This monotonic transform in intensity could be calibrated by histogram equalization, histogram specification, or some monotonic transform invariant features.

The degree of freedom in κ or in a monotonic transform can be compensated by using LBP features to achieve an illumination invariant representation of faces. The basic form of the LBP operator is illustrated in Fig. 15.4. The binary bits describing a local 3×3 sub-window are generated by thresholding the 8 pixels in the surrounding locations by the gray value of its center; the feature vector is formed by concatenating the thresholded binary bits in an anticlockwise manner. There are a total of 256 possible values and, hence, 256 LBP patterns denoted by such an LBP code; each value represents a type of LBP local pattern. Such a basic form of LBP can be extended to multi-scale LBP, $\text{LBP}_{(P,R)}$, where R is the radius of the circle surrounding the center and P is the number of pixels on the circle. An LBP (P, R) string is called uniform, denoted by $\text{LBP}_{(P,R)}^{u2}$, if the neighboring bits (the circular sense) contain at most two bitwise transitions from 0 to 1 or vice versa (see [18] for details).

From the analysis, we see that the NIR imaging and LBP features together lead to an illumination invariant representation of faces. In other words, applying the LBP operator to an active NIR image generates illumination invariant features for faces. The illumination invariant face representation provides great advantages for face recognition in varying illumination.

An LBP-based face matching method is described in [2, 9]. In this method, the image is divided into $7 \times 7 = 49$ blocks. An LBP histogram is calculated for

each block. A χ^2 distance is calculated between two histograms for matching and a weighted sum of the χ^2 distance is then used for matching between two face images. The method is shown to achieve very good results on the FERET database. However, such a method still lacks optimality in terms of the block division and the weights.

Recently, several LBP variants are proposed for face recognition, such as Multi-scale Block LBP (MB-LBP) [15], Local Ternary Pattern (LTP) [19] and so on. These features are also robust to intensity monotonic changes and can be used in NIR face recognition.

The following describes a procedure for extracting LBP histogram features:

1. Computing Base LBP Features

- Computing $\text{LBP}_{8,1}^{u2}$ codes for every pixel location in the image.

2. LBP Code Histogramming

- A histogram of the base LBP codes is computed over a local region centered at each pixel, each histogram bin being the number of occurrences of the corresponding LBP code in the local region. There are 59 bins for $\text{LBP}_{8,1}^{u2}$.
- An LBP histogram is considered as a set of 59 individual features.

3. Gathering LBP Histograms

- For a face image of size $W \times H$, with the interior area of size $W' \times H'$, the total number of LBP histogram features is $D = W' \times H' \times 59$ (number of valid pixel locations times the number of LBP histogram bins).

For example, if $W \times H = 120 \times 142$ and a local region for histogramming is a rectangle of size 16×20 , the interior area is of size $W' \times H' = 104 \times 122$ pixels. Then there are a total of $104 \times 122 \times 59 = 748\,592$ elements in the LBP histogram feature pool.

15.4 NIR Face Classification

Of the large number of LBP histogram features present in a feature pool. Some are useful for face recognition, some are not so useful, and some may be contradictory. They must be selected or weighted to achieve the best performance. In this section, we present an AdaBoost [7, 20] based learning method for selecting best LBP features and constructing a face classifier.

Given a training set of LBP features of faces subject to image noise, slight pose changes, and alignment errors, the learning method finds a good set of discriminative features among a large number of candidates and then build a strong classifier based on the selected features. Once trained, the classifier is able to recognize faces without having to be retrained when a new individual client is added.

15.4.1 AdaBoost Based Feature Selection

As an AdaBoost procedure essentially learns a two-class classifier, we convert the multi-class problem into a two-class problem using the idea of intra- and extra-class differences [16]. Two face examples are considered as intra-class if they are of the same person, or extra-class, otherwise. However, in terms of facial features, the difference data are derived between samples in the features rather than in the original image space, that is, a difference is taken between the facial features (for example, LBP histograms) of two face examples. In this manner, a training set of positive (intra-class) and negative (extra-class) training examples can be obtained for the learning procedure.

Assume that a training set of N examples is given as positive and negative classes, $\mathbf{S} = (x_1, y_1), (x_2, y_2), \dots, (x_N, y_N)$ where x_i is a training example (the difference between two feature vectors) and $y_i \in \{+1, -1\}$ is the class label. The AdaBoost procedure can be used to learn a set of best T features stagewise and thereby constructs a sequence of T weak classifiers, $h_t(x) \in \{+1, -1\}$, and linearly combine the weak classifiers in an optimal way into a stronger classifier,

$$H(x) = \text{sign} \left(\sum_{t=1}^T \alpha_t h_t(x) \right), \quad (15.4)$$

where $\alpha_t \in \mathbf{R}$ are the combining weights. The AdaBoost learning procedure is originally aimed at deriving α_t and $h_t(x)$ so that an upper error bound is minimized [7]. The reader is referred to [7, 20] for AdaBoost learning.

AdaBoost assumes that a procedure is available for learning a weak classifier $h_t(x)$ from the training examples weighted by the current distribution w_t . We use a weak classifier based on a single scalar feature, that is, an LBP histogram bin value. Therefore, when AdaBoost constructs a $h_t(x)$, it need to select a good feature for it. In this way, AdaBoost can provide a good subset of features.

In the test phase, the learned $H(x)$ can be used to classify face images. The difference is calculated between the selected features of the two face images. A weak decision h_t can be made in terms of each selected feature. The weak decisions are linearly combined with the weights α_t to give the predict value $H(x)$. The final decision can be made by comparing $H(x)$ with a threshold value. If greater, the two face images are considered as belonging to the same person (intra-class), otherwise, they are belonging to different persons (inter-class). Moreover, a cascade of AdaBoost classifiers [20] can be constructed to cope with complex distributions of two classes.

15.4.2 LDA Classifier

LDA reduces the dimensionality by linearly projecting the original feature vector in high dimensional space to a lower dimensional subspace such that the ratio of

within-class scatter over between-class scatter is minimized. This minimized the classification error when the distributions of the class data are Gaussian [8]. The high dimensional data may be preprocessed using the PCA transform to make the within-class scatter matrix nonsingular, before LDA is applied. The basis images of such a combined projection \mathbf{P} of PCA and LDA is called Fisherfaces [6]. More advanced forms of LDA, such as direct LDA or regularized LDA, could be used to obtain \mathbf{P} . The input of LDA, in our context, is the space of selected features.

Given two input vectors \mathbf{x}_1 and \mathbf{x}_2 in the space of selected features, their LDA projections are calculated as $\mathbf{v}_1 = \mathbf{P}\mathbf{x}_1$ and $\mathbf{v}_2 = \mathbf{P}\mathbf{x}_2$ and the following cosine score (or called “cosine distance” in some of the literature) is used for the matching:

$$H(\mathbf{v}_1, \mathbf{v}_2) = (\mathbf{v}_1 \cdot \mathbf{v}_2) / \|\mathbf{v}_1\| \|\mathbf{v}_2\|. \quad (15.5)$$

In the test phase, the projections \mathbf{v}_1 and \mathbf{v}_2 are computed from two input vectors \mathbf{x}_1 and \mathbf{x}_2 , the other for the input face image and one for an enrolled face image. By comparing the score $H(\mathbf{v}_1, \mathbf{v}_2)$ with a threshold, a decision can be made whether \mathbf{x}_1 and \mathbf{x}_2 belong to the same person.

15.5 Experiments

In this section, results with the NIR face recognition system of [14] are presented to illustrate the advantages of the NIR face recognition method. Performances of LBP + AdaBoost and LBP + LDA NIR face matching engines are compared with several existing baseline and face matching engines. Case studies regarding effects of eyeglasses, time lapse, and weak illumination are reported. Finally, results on a data set collected under sunlight in an outdoor environment are presented to demonstrate the performance of ENIR system [21].

15.5.1 Basic Evaluation

In the training phase, the training set of positive examples were derived from intra-class pairs of LBP histogram features, the negative set from extra-class pairs, each example being a 748 592 dimensional vector. There were 10^4 face images of about 1000 persons, 10 images each person, all Chinese. A training set of about 45×10^3 positive and 5×10^7 negative examples were collected from the training images. A cascade of 5 strong classifiers were trained, with about 1500 weak classifiers. The ROC curves for the training set are shown on the top of Fig. 15.5, where the FAR is reduced to below 10^{-7} with an accuracy of 94.4%.

A technology evaluation was done with a test set of 3237 images. The test set contained 35 persons, with 80 to 100 images per person. None of the test images were in the training set. This generated 149 217 intra-class (positive) and 5 088 249

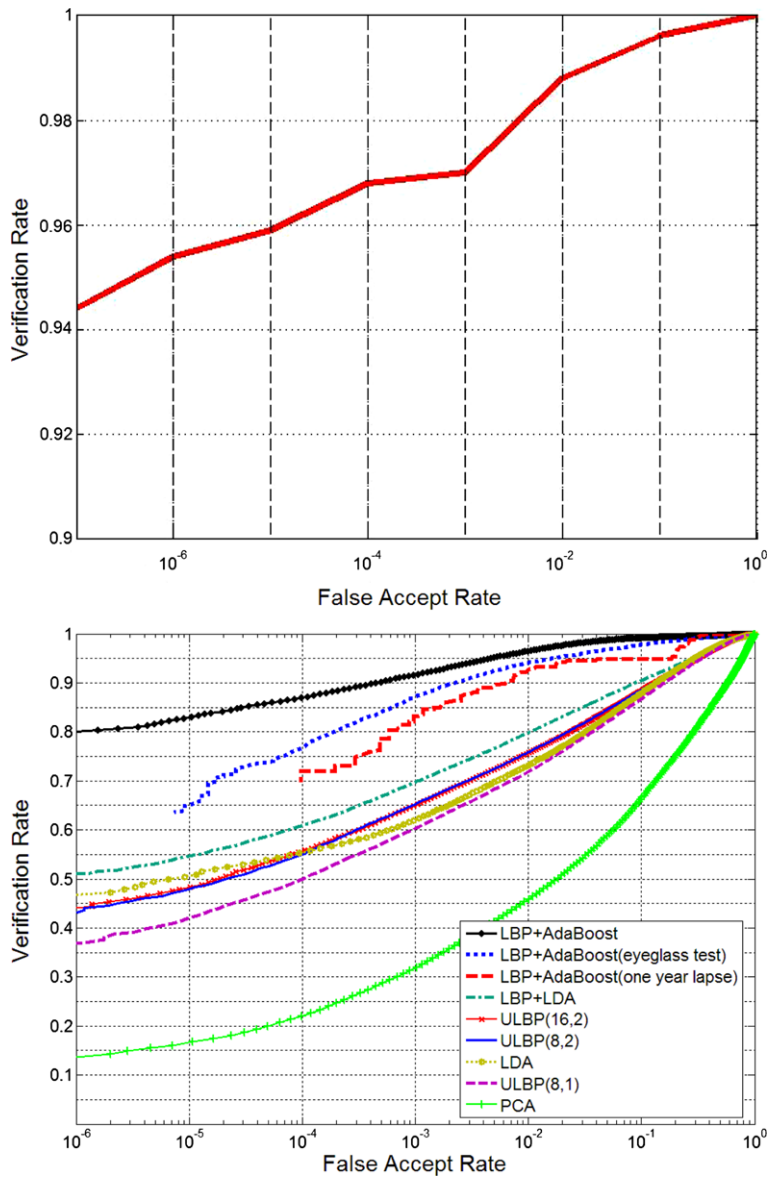
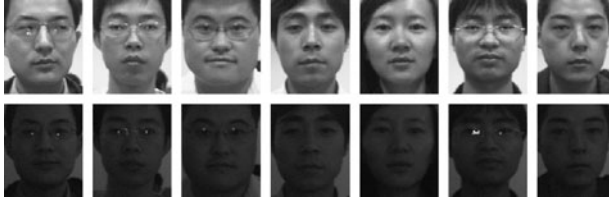


Fig. 15.5 *Top:* ROC curve of LBP + AdaBoost method for face verification on the training set. *Bottom:* ROC curves for various compared methods

extra-class (negative) pairs. Several other methods were included in the evaluation (for comparison), using the same set of training and test images. They were: (i) PCA on the NIR images (with Mahalanobis distance), (ii) LDA on the NIR images (with cosine distance), (iii) the LBP + LDA method, (iv) the original LBP

Visible Light Face Images under Controlled and Weak Light Condition



LBP+AdaBoost Matching Scores for intra- and extra-class pairs

	Ctrl 1	Ctrl 2	Ctrl 3	Ctrl 4	Ctrl 5	Ctrl 6	Ctrl 7
Weak 1	0.4888	0.4831	0.4751	0.4689	0.4764	0.4745	0.4658
Weak 2	0.4831	0.5131	0.4578	0.5202	0.4926	0.5014	0.4709
Weak 3	0.4808	0.4588	0.4804	0.4724	0.4814	0.4653	0.4619
Weak 4	0.4531	0.4563	0.4415	0.4688	0.4570	0.4582	0.4570
Weak 5	0.4892	0.4904	0.4757	0.4893	0.5275	0.4780	0.4824
Weak 6	0.4648	0.4835	0.4768	0.5085	0.4864	0.4935	0.4691
Weak 7	0.4686	0.4683	0.4636	0.4708	0.4887	0.4687	0.5068

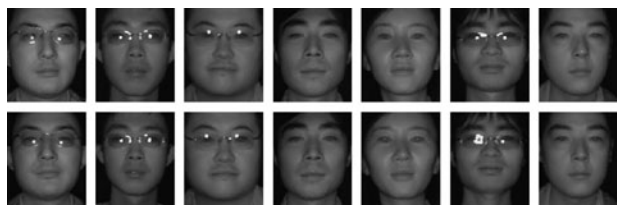
Fig. 15.6 *Top:* Images captured under controlled illumination (*row 1*) and under weak light conditions (*row 2*), each column belonging to the same person. *Bottom:* LBP + AdaBoost matching scores

method developed by Ahonen et al. [2] and Hadid et al. [9] (χ^2 distances between LBP histograms in 7×7 image blocks) with three operators: $\text{LBP}_{(8,1)}^{u2}$, $\text{LBP}_{(8,2)}^{u2}$ and $\text{LBP}_{(16,2)}^{u2}$. On the bottom of Fig. 15.5 shows the ROC curves derived from the scores for the intra- and extra-class pairs. By the VR values at FAR = 0.1%, the compared methods can be ranked in order of decreasing VR as: LBP + AdaBoost (VR = 91.8%), LBP + LDA (69.9%), $\text{LBP}_{(8,2)}^{u2}$ (65.29%), $\text{LBP}_{(16,2)}^{u2}$ (65.29%), Image + LDA (62.4%), $\text{LBP}_{(8,1)}^{u2}$ (60.7%), and Image + PCA (32.0%). See later for explanations of the “LBP + AdaBoost (eyeglass test)” and “LBP + AdaBoost (one year lapse)” curves.

15.5.2 Weak Illumination

Figures 15.6 and 15.7 present case studies to compare performance of visible light (VIS) and NIR image based face matching methods under weak illumination. The LBP + AdaBoost classifier for VIS images was trained using VIS images, whereas the one for NIR images was the one using for other tests. In the tables, the diagonal entries (in bold font) are for the intra-class pairs between controlled and weak illumination. For the VIS case, the mean and variance are 0.4970 and 0.0201 for intra-class pairs, and 0.4747 and 0.0154 for extra-class pairs. There are several cases of mismatch because the intra-class scores are not necessarily higher than the extra-class ones. In contrast, the NIR solution well separates the two classes, with

Active NIR Light Face Images under Controlled and Weak Light Conditions



LBP+AdaBoost Matching Scores for intra- and extra-class pairs

	Ctrl 1	Ctrl 2	Ctrl 3	Ctrl 4	Ctrl 5	Ctrl 6	Ctrl 7
Weak 1	0.6202	0.3262	0.3963	0.3354	0.3492	0.3747	0.3558
Weak 2	0.3551	0.6573	0.3226	0.3213	0.3637	0.4054	0.3442
Weak 3	0.3886	0.3489	0.7144	0.3244	0.3759	0.3249	0.3139
Weak 4	0.3902	0.3062	0.3545	0.6812	0.4123	0.3069	0.3144
Weak 5	0.4135	0.3289	0.4507	0.3851	0.6882	0.3780	0.3510
Weak 6	0.3459	0.4163	0.3520	0.3292	0.2996	0.6948	0.2849
Weak 7	0.3697	0.2847	0.2831	0.3374	0.3656	0.2778	0.6162

Fig. 15.7 *Top*: Images captured under controlled illumination (*row 1*) and under weak light conditions (*row 2*), each column belonging to the same person. *Bottom*: LBP + AdaBoost matching scores

the mean and variance of 0.6675 and 0.0377 for intra-class pairs, and 0.3492 and 0.0403 for extra-class pairs, and correctly matches all the pairs.

15.5.3 Eyeglasses

This section presents a case analysis of influence of eye glasses on face matching, as shown by the images, and the score table in Fig. 15.8. In the tables, the diagonal entries (in bold font) are for the intra-class pairs without and with glasses (that is, between the two images in the same column); the lower triangle entries for the extra-class no-glass pairs (that is, between two different images in the first row); and the upper triangle for the extra-class glass pairs (that is, between two different images in the second row). The mean and variance of correlations (not shown here due to page limit) are 0.9306 and 0.0419 for intra-class pairs, and 0.7985 and 0.0761 for extra-class pairs of either wearing no glasses or wearing glasses. Compared with correlation, LBP + AdaBoost matching engine can well separate between the two classes—the intra-class scores are consistently higher than those of extra-class scores.

Statistics were also obtained using 1500 images of 30 subjects, 50 images per subject of which 25 are with glasses and 25 without. The no-eyeglass images were used as the gallery set and the eyeglass images as the probe set. The ROC curve is labeled “LBP + AdaBoost (eyeglass test)” in the bottom of Fig. 15.5 (the portion for FAR smaller than 10^{−5} is unavailable because of the limited data points). At FAR =

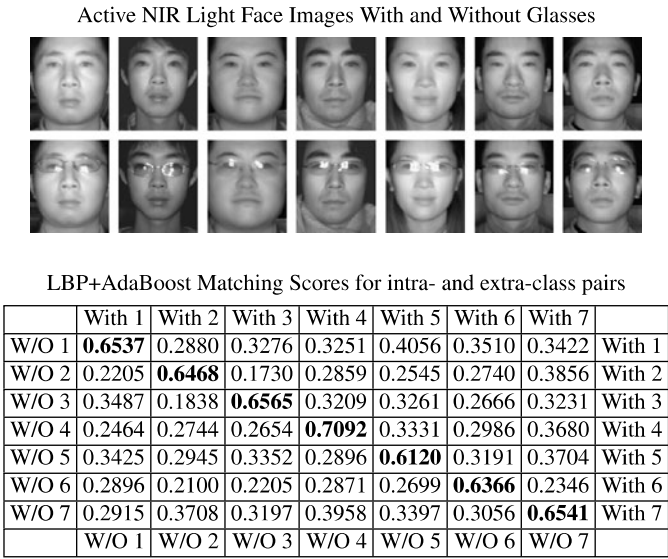


Fig. 15.8 Analysis on effects of glasses. *Top*: Images without glasses (*row 1*) and with glasses (*row 2*), each column belonging to the same person. *Bottom*: LBP + AdaBoost matching scores

0.1%, the VR was 87.1%, as opposed to 91.8% of the “LBP + AdaBoost” curve for the no-eyeglasses vs. no-eyeglasses and eyeglasses vs. eyeglasses comparisons.

15.5.4 Time Lapse

Tests were performed to evaluate effect of time lapse on NIR face recognition. Figure 15.9 presents a case analysis of time lapse effect on the matching scores. The NIR images of 7 individuals were acquired in Spring 2005 and Spring 2006, respectively. The table shows the matching scores produced by the LBP + AdaBoost classifier trained on active NIR images. The mean and variance of the scores are 0.6421 and 0.0198 for intra-class pairs (in bold font), and 0.3045 and 0.0462 for extra-class pairs. The LBP + AdaBoost matching engine well separates between the two classes, the intra-class scores are consistently higher than the extra-class ones.

Statistics were also obtained using 750 images of 30 persons, 25 images per person; of the 25 images, 10 were captured one year ago and used as the gallery set, and 15 were current images used as the probe set. The ROC curve is labeled “LBP + AdaBoost (one year lapse)” in the bottom of Fig. 15.5 (the portion for FAR smaller than 10^{-4} is unavailable because of the limited data points). At FAR = 0.1%, the VR was 83.24%, as opposed to 91.8% for images of no significant time lapse (the “LBP + AdaBoost” curve).

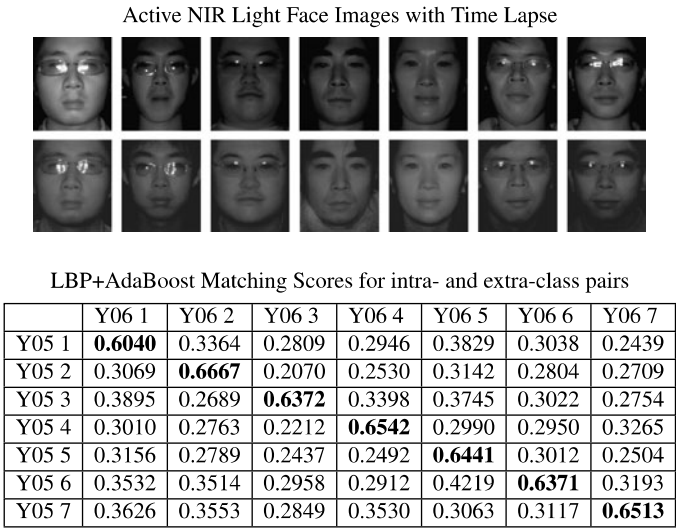


Fig. 15.9 Analysis on effects of time lapse. *Top*: NIR face images of spring 2005 (*row 1*) and spring 2006 (*row 2*), each column belonging to the same person. *Bottom*: LBP + AdaBoost matching scores where Y05 and Y06 denote Spring 2005 and Spring 2006, respectively

15.5.5 Outdoor Environment

These following case studies were performed to evaluate the robustness of ENIR [21] system in an outdoor environment and compare its performance to VIS and NIR systems. To this end, a independent test set containing 20 persons is collected under the following lighting conditions:

- 1. Normal indoor frontal lighting
- 2. Strong indoor frontal lighting
- 3. Strong indoor non-frontal lighting
- 4. Outdoor frontal sunlight
- 5. Outdoor side sunlight
- 6. Outdoor back sunlight

where the strong indoor lighting is provided by a 1000 W tungsten-halogen lamp with a color temperature of 2500–3000 K. While capturing face images, the lamp is placed at 3.5–4 m away from the subject. The lamp contains a wide range of NIR compositions and can be use to simulate sunlight in an indoor environment.

The VIS and NIR cameras are saturated easily by sunlight, and hence the test set does not include VIS and NIR images outdoor. The number of VIS and NIR face images in the set are $4 \text{ (indoor conditions)} \times 20 \text{ (person)} \times 10 \text{ (images/person)} = 800$. The number of ENIR face images is $800 + 3 \text{ (outdoor conditions)} \times 20 \text{ (person)} \times 10 \text{ (images/person)} = 1400$. The resolution of VIS, NIR and ENIR images are all 640×480 .

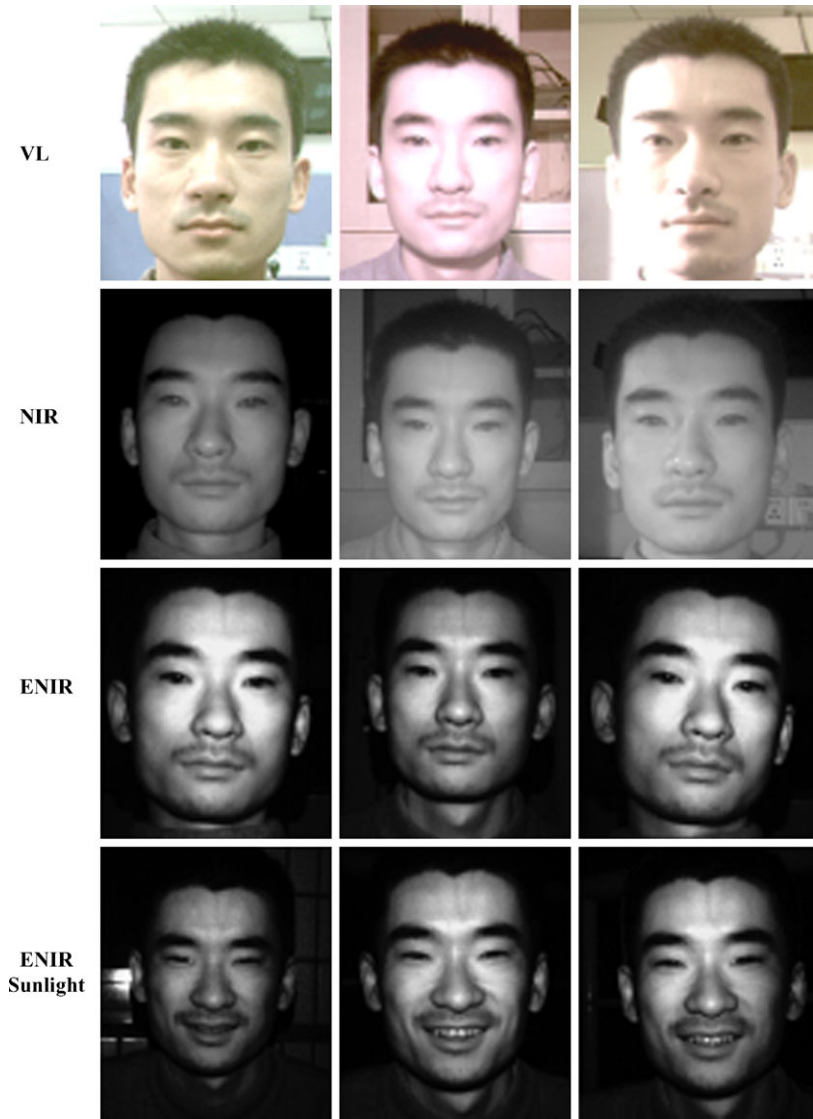


Fig. 15.10 Face images in the outdoor test set. Rows 1–3: VIS, NIR and ENIR images under 3 indoor lighting conditions. Row 4: ENIR images captured outdoors

Figure 15.10 shows some face images in the test set. We can see that VIS face images are unstable to the direction and strength of the strong lighting, NIR images are less unstable, and ENIR images are the most stable ones under various lighting conditions.

In the experiments, the face images under normal indoor frontal lighting are used as gallery and the other images are used as probe. Two protocols are used to compare the three imaging systems:

1. For indoor images: comparing ROCs of all the 3 systems.
2. For outdoor images: evaluating ROC of the ENIR system only, while the other systems could not function to a satisfactory extent.

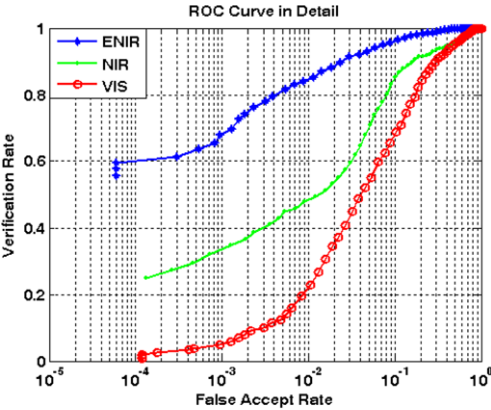
Figure 15.11 shows the ROC curves for all experiments, in which the ENIR system is the most stable one under all lighting conditions. In Fig. 15.11(a), the performance of VIS and NIR systems dropped significantly under strong frontal halogen lamp light, whereas the ENIR system still has relatively high verification rate ($VR = 69\%$ when $FAR = 0.001$ and $VR = 85\%$ when $FAR = 0.01$). Similar trend is also shown in Fig. 15.11(b). Figure 15.11(c) shows the results of the ENIR system under outdoor sunlight in three different directions. The highest verification rate is $VR = 50\% @ FAR = 0.001$ and $VR = 69\% @ FAR = 0.01$ when the sunlight is from the subjects' back direction. From the above figures, we can see the performance under frontal lighting is always worse when compared to nonfrontal lighting. A possible explanation is that human eyes are easily disturbed by the frontal lighting, and this can cause a significant change in facial expression, which is another challenging problem in face recognition.

15.6 Conclusions

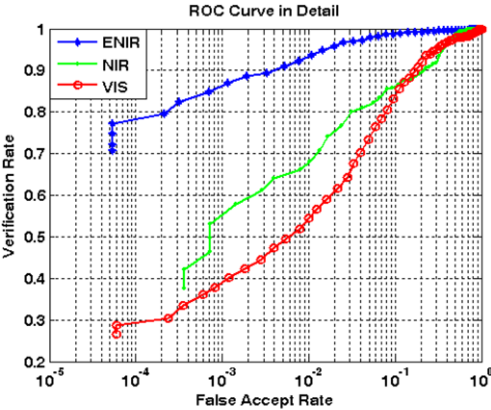
We have presented an effective solution for overcoming the problems caused by illumination variation that severely affects the performance of face recognition systems. The solution consists of active NIR imaging hardware, more efficient algorithms, and a novel system design. An illumination invariant face representation is obtained by extracting LBP features from NIR images. The AdaBoost procedure is used to learn a powerful face recognition engine based on the invariant representation. Highly accurate and fast face recognition systems can be built thereby. Extensive experiments show the robustness of the present solution in terms of image properties, illumination changes, ethnic groups, and advantages over existing methods. An enhanced solution, using a newly developed ENIR imaging device, is presented to deal with strong NIR composition in ambient light such as in the sunlight. The results show that the ENIR system performed significantly better than the VIS and NIR systems in adverse illumination environment. This approach results in face recognition products that perform well for 1-to-many identification for cooperative user applications.

Acknowledgements This work was partially supported by the Chinese National Natural Science Foundation Project #61070146, the National Science and Technology Support Program Project #2009BAK43B26, and the AuthenMetric R&D Funds (2004–2011). The work was also partially supported by the TABULA RASA project (<http://www.tabularasa-euproject.org>) under the Seventh Framework Programme for research and technological development (FP7) of the European Union (EU), grant agreement #257289.

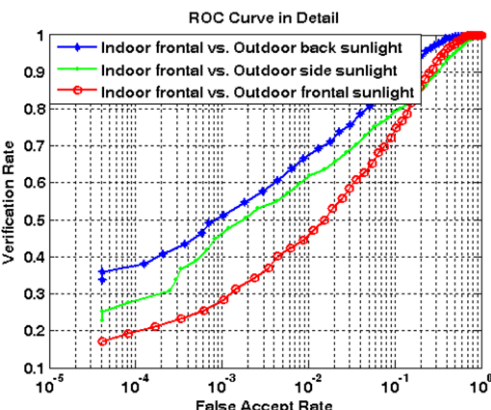
Fig. 15.11 ROC curves of NIR, VIS and ENIR systems on the outdoor test set



(a) Indoor frontal vs. Indoor strong frontal.



(b) Indoor frontal vs. Indoor strong non-frontal.



(c) ENIR system under outdoor sunlight.

References

- Adini, Y., Moses, Y., Ullman, S.: Face recognition: The problem of compensating for changes in illumination direction. *IEEE Trans. Pattern Anal. Mach. Intell.* **19**(7), 721–732 (1997)
- Ahonen, T., Hadid, A., Pietikainen, M.: Face recognition with local binary patterns. In: *Proceedings of the European Conference on Computer Vision*, pp. 469–481, Prague, Czech Republic (2004)
- AuthenMetric Co. Ltd.: A method for face image acquisition using active lighting. Patent Application No. 200310121340.1, 12 December 2003
- AuthenMetric Co. Ltd.: A method for face image acquisition and a method and system for face recognition. Patent Application No. PCT/CN2004/000482, 14 May 2004
- AuthenMetric Co. Ltd.: An image acquisition apparatus for face recognition. Patent Application No. 200520022878.1, 22 March 2005
- Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. In: *Proceedings of the European Conference on Computer Vision*, pp. 45–58 (1996)
- Freund, Y., Schapire, R.E.: A decision-theoretic generalization of online learning and an application to boosting. *J. Comput. Syst. Sci.* **55**(1), 119–139 (1997)
- Fukunaga, K.: *Introduction to Statistical Pattern Recognition*, 2nd edn. Academic Press, Boston (1990)
- Hadid, A., Pietikainen, M., Ahonen, T.: A discriminative feature space for detecting and recognizing faces. In: *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 797–804 (2004)
- Hizem, W., Krichen, E., Ni, Y., Dorizzi, B., Garcia-Salicetti, S.: Specific sensors for face recognition. In: *Proceedings of IAPR International Conference on Biometric*, vol. 3832, pp. 47–54 (2006)
- Li, S.Z., His Face Team: AuthenMetric F1: a highly accurate and fast face recognition system. In: *ICCV2005—Demos*, 15–21 October 2005
- Li, S.Z., Chu, R.F., Ao, M., Zhang, L., He, R.: Highly accurate and fast face recognition using near infrared images. In: *Proceedings of IAPR International Conference on Biometric (ICB-2006)*, pp. 151–158, Hong Kong, January 2006
- Li, S.Z., Zhang, L., Liao, S.C., Zhu, X.X., Chu, R.F., Ao, M., He, R.: A near-infrared image based face recognition system. In: *Proceedings of 7th IEEE International Conference Automatic Face and Gesture Recognition (FG-2006)*, pp. 455–460, Southampton, UK, 10–12 April 2006
- Li, S.Z., Chu, R., Liao, S., Zhang, L.: Illumination invariant face recognition using near-infrared images. *IEEE Trans. Pattern Anal. Mach. Intell.* **26** (2007) (Special issue on Biometrics: Progress and Directions)
- Liao, S., Zhu, X., Lei, Z., Zhang, L., Li, S.Z.: Learning multi-scale block local binary patterns for face recognition. In: *ICB*, pp. 828–837 (2007)
- Moghaddam, B., Nastar, C., Pentland, A.: A Bayesian similarity measure for direct image matching. Media Lab Tech Report No. 393, MIT, August 1996
- NIST. Face Recognition Vendor Tests (FRVT). <http://www.frvt.org>
- Ojala, T., Pietikainen, M., Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.* **24**(7), 971–987 (2002)
- Tan, X., Triggs, B.: Enhanced local texture feature sets for face recognition under difficult lighting conditions. *IEEE Trans. Image Process.* **19**(6), 1635–1650 (2010)
- Viola, P., Jones, M.: Robust real time object detection. In: *IEEE ICCV Workshop on Statistical and Computational Theories of Vision*. Vancouver, Canada, 13 July 2001
- Yi, D., Liu, R., Chu, R., Liu, D., Wang, R., Li, S.Z.: Outdoor face recognition using enhanced near infrared imaging. In: *Proceedings of IAPR International Conference on Biometric*, Seoul, Korea, August 2007