Illumination Invariant Face Recognition Using Near-Infrared Images

Stan Z. Li, *Senior Member*, *IEEE*, RuFeng Chu, ShengCai Liao, and Lun Zhang

Abstract—Most current face recognition systems are designed for indoor, cooperative-user applications. However, even in thus-constrained applications, most existing systems, academic and commercial, are compromised in accuracy by changes in environmental illumination. In this paper, we present a novel solution for illumination invariant face recognition for indoor, cooperative-user applications. First, we present an active near infrared (NIR) imaging system that is able to produce face images of good condition regardless of visible lights in the environment. Second, we show that the resulting face images encode intrinsic information of the face, subject only to a monotonic transform in the gray tone; based on this, we use local binary pattern (LBP) features to compensate for the monotonic transform, thus deriving an illumination invariant face representation. Then, we present methods for face recognition using NIR images; statistical learning algorithms are used to extract most discriminative features from a large pool of invariant LBP features and construct a highly accurate face matching engine. Finally, we present a system that is able to achieve accurate and fast face recognition in practice, in which a method is provided to deal with specular reflections of active NIR lights on eyeglasses, a critical issue in active NIR image-based face recognition. Extensive, comparative results are provided to evaluate the imaging hardware, the face and eye detection algorithms, and the face recognition algorithms and systems, with respect to various factors, including illumination, eyeglasses, time lapse, and ethnic groups.

Index Terms—Biometrics, face recognition, near infrared (NIR), illumination invariant, local binary pattern (LBP), statistical learning.

1 Introduction

1.1 Motivations

Face recognition as a primary modality for biometric authentication has received increasing interest in the recent years. This is partly due to recent technology advances [1], [2] initially made by the work on eigenfaces [3], [4] and partly due to increased concerns in security.

Existing biometric systems are developed for *cooperative user applications*, such as access control, machine readable traveling document (MRTD), computer logon, and ATM. In such applications, a user is required to cooperate with the camera to have his/her face image captured properly in order to be permitted some access; this is in contrast to more general scenarios, such as face recognition under surveillance, where a person should be recognized without intentional, cooperative effort. Another aspect is that most of the current systems are designed for indoor use.

To achieve reliable results, face recognition should be performed based on intrinsic factors of the face only, mainly related to 3D shape and reflectance of the facial surface. Extrinsic factors, including eyeglasses, hairstyle, expression, posture, and environmental lighting, which make distributions of face data highly complex [5], [6], [7], should be minimized for reliable face recognition. Among several extrinsic factors, problems with uncontrolled environmental lighting is the topmost issue to solve for reliable face-based biometric applications in practice. From the end-user point of view, a biometric system should adapt to the environment, not vice versa.

However, most current face recognition systems, academic and commercial, are based on face images captured in the *visible light* (VL) spectrum; they are compromised in accuracy by changes in environmental illumination, even for cooperative user applications indoors. In an in-depth study on the influence of illumination changes on face recognition [8], Adini et al. examined several distance measures and several 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 are made there: 1) Lighting conditions, and especially light angle, drastically change the appearance of a face. 2) When comparing unprocessed images, the changes between the images of a person under different illumination conditions are larger than those between the images of two people under the same illumination. 3) All of the local filters under study are insufficient by themselves to overcome variations due to changes in illumination direction. The influence of illumination is also shown in the recent Face Recognition Vendor Test [9].

1.2 Contributions of This Work

In this paper, we present a novel solution for achieving illumination invariant face recognition for indoor, cooperative-user applications, using *active near infrared* (active NIR) imaging techniques, and for building accurate and fast face recognition systems. The solution consists of NIR imaging hardware [10], [11], [12] and NIR-based face recognition algorithms [13], [14], [15]. The main contributions are summarized in the following.

• *The authors are with Center for Biometrics and Security Research and National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, 95 Zhongguancun East Road, Haidian District, Beijing 100080, China.*
E-mail: {szli, rfchu, scliao, lzhang}@cbsr.ia.ac.cn.

Manuscript received 1 Jan. 2006; revised 3 July 2006; accepted 6 Oct. 2006; published online 13 Feb. 2007.
Recommended for acceptance by S. Prabhakar, J. Kittler, D. Maltoni, L. O'Gorman, and T. Tan.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org and reference IEEECS Log Number TPAMISI-0051-0106.
Digital Object Identifier no. 10.1109/TPAMI.2007.1014.

The first contribution is an imaging hardware system for producing face images of a good illumination condition (Section 2). Active NIR lights are used to illuminate the face from the frontal direction during image acquisition. We use powerful enough active NIR lighting to prevail over VL in the indoor environment, and further reduce VL by using an optical filter. As such, the face in the image is illuminated from the frontal direction, with the unwanted effect of uncontrollable VL minimized. Such an imaging system provides an excellent basis for subsequent face recognition.

There could be two approaches to illumination invariant face recognition: by a highly nonlinear face matching engine with an illumination *variant* representation or by an illumination *invariant* face representation with a less complicated face matching engine. We adopt the latter approach by taking advantage of active NIR imaging.

As the second contribution, we derive an illumination invariant face representation on the basis of active NIR images in Section 3. We show that the active NIR imaging is mainly subject to an approximately monotonic transform in the gray tone due to variation in the distance between the face and the NIR lights and camera lens. Noting that the ordering relationship between pixels is not changed by any monotonic transform, we use local binary pattern (LBP) features [16], [17], [18] to compensate for the monotonic transform in the NIR images. Therefore, LBP filtering of an active NIR image generates an illumination invariant representation of faces. This provides a different approach from those of existing works (cf. Section 1.4) for illumination invariant face recognition.

The third contribution is methods for building a highly accurate face recognition engine using LBP features from NIR images (Section 4). While there are a large number of LBP features, not all are useful or equally useful for face recognition. We use AdaBoost learning to learn from training examples to select a small subset of the most discriminative LBP features and thereby construct a strong classifier. We also present another method that uses LDA, building a discriminative classifier from LBP features. Both perform better than the state-of-the-art LBP method of [17], [18].

The fourth contribution is a method for reliable eye detection in active NIR images (Section 5). The frontal active NIR lighting can cause unwanted specular reflections on eyeglasses. This makes accurate eye localization more difficult in active NIR than normal VL images, which cannot be easily tackled by a simple eye detector. We provide a simple-to-complex architecture to give a satisfactory solution to overcome this critical issue of face recognition using active NIR images.

The fifth contribution is a highly accurate and fast face recognition system for cooperative face recognition applications indoor (Section 5). The system has been employed in machine readable traveling documents (MRTD) systems at the China-Hong Kong border, the largest border crossing point in the world, since June 2005. It was also demonstrated at the Second Summer School for Advanced Studies on Biometrics for Secure Authentication [19] and ICCV 2005 [13].

1.3 Summary of Main Results

Extensive experimental results are provided in and before Section 6 to demonstrate the reliability of the present solution and contrast to existing methods. First, we demonstrate by case studies advantages of the active NIR imaging system over conventional visible light (VL) imaging systems in face recognition under different VL directions. Whereas VL images of the same face under different lighting directions are negatively correlated, NIR imaging produces closely correlated images of faces of the same individual. However, it is also shown that, even with the good basis offered by the NIR imaging system, a straightforward matching engine, such as correlation or PCA-based or LDA-based, is insufficient to achieve high accuracy; more advanced techniques are required to deal with the complexity in the pattern recognition problems therein, which also motivated the work of using AdaBoost learning with LBP features for constructing a nonlinear classifier for face recognition.

Second, we provide technology evaluation to evaluate different methods, including the present AdaBoost method, and PCA and LDA methods performed on the raw NIR image data and on LBP feature data. It is shown that the LBP+AdaBoost method produces better results than the other methods: At the false alarm rate (FAR) of 10^{-3}, it achieves the recognition rate of 91.8 percent, as opposed to that of 32.0 percent for the Image+PCA method, 62.4 percent for Image+LDA method, and 69.9 percent for the LBP+LDA method. The comparisons justify that the learning-based method indeed offers a good solution for achieving highly accurate face recognition. Moreover, results are also provided to demonstrate the reliability of the LBP+AdaBoost method with respect to factors of illumination, eyeglasses, time lapse, and ethnic groups.

Third, we provide scenario evaluation results to evaluate an active NIR-based face recognition system for time attendance and access control (a scenario evaluation meant for evaluating face recognition systems is considered harder and more comprehensive than a technology evaluation of algorithms [20]). The results show that the system has excellent accuracy, speed, and usability.

1.4 Review of Related Works

Much effort has been made to model illumination on faces and correct illumination directions in order to achieve illumination invariant face recognition. Georghiades et al. [21] proved that face images with the same pose under different illumination conditions form a convex cone, called the illumination cone. Ramamoorthi [22] and Basri and Jacobs [23] independently used the spherical harmonic representation to explain the low dimensionality of face images under different illumination. Nayar and Bolle [24] and Jacobs et al. [25] proposed algorithms for face image intrinsic property extraction by Lambertian model without shadow. Shashua and Raviv [26] proposed a simple yet practical algorithm, called the quotient image, for extracting illumination invariant representation. Gross and Brajovic [27] and Wang et al. [28] developed reflectance estimation methods by using the idea of the ratio of the original image and its smooth version from Retinex [29] and center-surround filters [30]. These works are shown to improve recognition performance, but have not led to a face recognition method that is illumination invariant.

Other directions have also been explored to overcome problems caused by illumination changes. One direction is to use 3D (in many case, 2.5D) data obtained from a laser range scanner or 3D vision method (cf. papers [31], [1]). Because such data captures geometric shapes of face, such systems are less affected by environmental lighting. Moreover, it can cope with rotated faces because of the availability of 3D (2.5D) information for visible surfaces. The disadvantages are the

increased cost and slowed speed as well as specular reflections. More importantly, it is shown that the 3D method may not necessarily produce better recognition results: recognition performances achieved by using a single 2D image and by a single 3D image are similar [32]. A commercial development is A4Vision [33]. It is basically a 3D (or 2.5D) face recognition system, but it creates 3D mesh of the face by means of triangulation based on an NIR light pattern projected onto the face. While not affected by lighting conditions, background colors, facial hair, or make-up, it has problems in working under conditions when the user is wearing glasses or opening the mouth, due to limitations of its 3D reconstruction algorithm.

Imaging and vision beyond the visible spectrum has recently received much attention in the computer vision community, as seen from the IEEE workshop series [34], [35]. Thermal or far infrared imagery has been used for face recognition (cf. a survey paper [36]). This class of techniques is advantageous for identifying faces under uncontrolled illumination or for detecting disguised faces. Its performance can be enhanced by the fusion of coregistered visible and thermal IR images using a fusion of experts' methodologies [37], [38]. Their disadvantages include instability due to environmental temperature, emotional and health conditions, and poor eye localization accuracy [39], [40]. A large-scale study [39] showed that they did not perform as well as visible light image-based systems, in a scenario involving time lapse between gallery and probe and with relatively controlled lighting.

The use of near infrared (NIR) imaging brings a new dimension for face detection and recognition [41], [42], [43]. Dowdall et al. [41] presented an NIR-based face detection method; faces are detected by analyzing horizontal projections of the face area and by using the fact that eyes and eyebrows regions have different responses in the lower and upper bands of NIR. Li and Liao [42] presented a homomorphic-filtering preprocessing to reduce inhomogeneous NIR lighting and a facial feature detection method by analyzing the horizontal and vertical projections of the face area. Pan et al. [43] presented an NIR-based face recognition method in which hyperspectral images are captured in 31 bands over a wavelength range of 0.7μm-1.0μm; multiband spectral measurements of facial skin sampled at some facial points are used for face recognition; they are shown to differ significantly from person to person.

Further investigations of using NIR images for face localization and recognition are found in [10], [44], [45]. All of those works use the "bright pupil" effect, namely, specular reflection of active NIR lights on pupils to detect eyes in NIR images. In addition, Zhao and Grigat's system [44] uses DCT coefficients as features and an SVM as the classifier. Zou et al.'s work [45] derives their matching methods based on an LDA transform and shows that the NIR illuminated faces are better separable than faces under varying ambient illumination.

The method of using the "bright pupil" effect to detect eyes has a serious drawback that limits its applications. It assumes that the "bright pupils" are present in the eyes and can be detected using an algorithm. However, the assumption can be invalidated, for example, when there are specular reflections of NIR lights on the eyeglasses, with "narrow eyes" where eyelids may occlude "bright pupils," when an eye is closed, or when the eyes are looking aside (see examples in Fig. 7). These happen for most of the face images with eyeglasses and also for a significant percentage of face images without eyeglasses. Therefore, we have considered eye detection a main problem to be solved in face recognition using active NIR images.

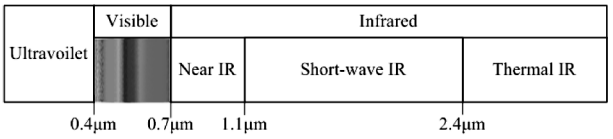


Fig. 1. Radiation spectrum ranges.

2 Active NIR Imaging System

The goal of making this special-purpose hardware is to overcome the problem arising from uncontrolled environmental lights so as to produce face images of a good illumination condition for face recognition. By "a good illumination condition," we mean that the lighting is from the frontal direction and the image has suitable pixel intensities, i.e., having good contrast and not saturated.

2.1 Hardware Design

We propose two strategies to control the light direction: 1) Mount active lights on the camera to provide frontal lighting and 2) minimize environmental lighting. We set two requirements on the active lighting: 1) The lights should be strong enough to produce a clear frontal-lighted face image without causing disturbance to human eyes and 2) the minimization of the environmental lighting should have minimum reduction of the intended active lighting.

Radiation spectrum ranges are shown in Fig. 1. While far (thermal) infrared imaging reflects heat radiation, NIR imaging is more like normal visible light imaging, though NIR is invisible to the naked eyes. Ultraviolet radiation is harmful to the human body and cannot be used for face recognition applications.

Our solution for requirement 1, is to choose the active lights in the near infrared (NIR) spectrum between 780-1,100 nm and mount them on the camera. We use NIR light-emitting diodes (LEDs) as active radiation sources, which are strong enough for indoor use and are power-effective. A convenient wavelength is 850 nm. Such NIR lights are almost invisible to the human eye, yet most CCD and CMOS sensors have sufficient response at this spectrum point.

When mounted on the camera, the LEDs are approximately coaxial to the camera direction and, thus, provide the best possible straight frontal lighting, better than mounting anywhere else; moreover, when the LEDs and camera are together, control of the lights can be easier using a circuit in the box. The geometric layout of the LEDs on the camera panel may be carefully designed such that the illumination on the face is as homogeneous as possible.

The strength of the total LED lighting should be such that it results in the NIR face images with good S/N ratio when the camera-face distance is between 50-100 cm, a convenient range for the user. A guideline is that it should be as strong as possible, at least stronger than expected environmental illumination, yet not cause sensor saturation. A concern is the safety of human eyes. When the sensor working in the normal mode is not saturated, the safety is guaranteed.

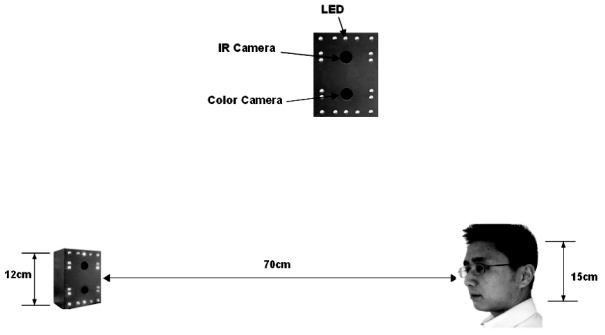


Fig. 2. Active NIR imaging system (upper) and its geometric relationship with the face (lower).

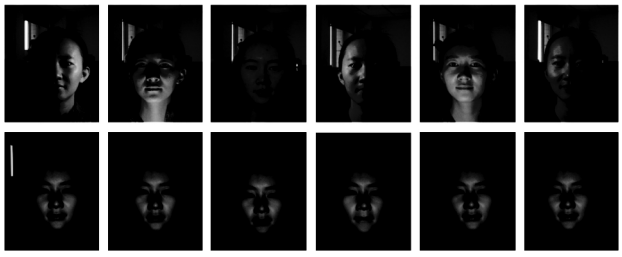


Fig. 3. Color images taken by a color camera versus NIR images taken by the present NIR imaging system. While unfavorable lighting is obvious in the color face images, it is almost unseen in the NIR face images.

Our solution for requirement 2, above is to use a long pass optical filter to cut off visible light while allowing NIR light to pass. We choose a filter such that ray passing rates are 0, 50, 88, and 99 percent at the wavelength points of 720, 800, 850, and 880 nm, respectively. The filter cuts off visible environmental lights (< 700 nm) while allowing most of the 850nm NIR light to pass.

Fig. 2 illustrates a design of the hardware device and its relationship with the face. The device consists of 18 NIR LEDs, an NIR camera, a color camera, and the box. The NIR LEDs and camera are for NIR face image acquisition. The color camera capture color face images may be used for fusion with the NIR images or for other purposes. The hardware and the face are relatively positioned in such a way that the lighting is frontal and NIR rays provide nearly homogenous illumination on face. The imaging hardware works at a rate of 30 frames per second with the USB 2.0 protocol for 640×480 images and costs less than 20 US dollars.

Fig. 3 shows example images of a face illuminated by NIR LED lights from the front, a lamp aside and environmental lights. We can see the following: 1) The lighting conditions are likely to cause problems for face recognition with the color images. 2) 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, and provide a good basis for face recognition.

2.2 Active NIR versus Visible Light Images

In visible light images, an intrapersonal change due to different lighting directions could be larger than an extrapersonal change under similar lighting conditions. This can be illustrated by an analysis on correlation coefficients and matching scores, shown in Fig. 4. There are seven pairs of face images; each pair is taken of the same person's face but

Visible Light Face Images Taken with Left and Right Lamp Lights

Correlation Coefficients

	Right 1	Right 2	Right 3	Right 4	Right 5	Right 6	Right 7	
Left 1	**-0.6542**	-0.5765	-0.5346	0.7595	-0.4883	0.7560	0.7090	Right 1
Left 2	-0.5672	**-0.5160**	0.6385	-0.5262	0.6991	-0.4554	-0.4256	Right 2
Left 3	-0.5319	0.6548	**-0.4408**	-0.4856	0.6128	-0.3953	-0.4555	Right 3
Left 4	0.7211	-0.4917	-0.4971	**-0.5199**	-0.4479	0.6361	0.7769	Right 4
Left 5	-0.5148	0.6402	0.8881	-0.4807	**-0.3730**	-0.3812	-0.3537	Right 5
Left 6	0.7101	-0.4242	-0.3948	0.4931	-0.3877	**-0.3565**	0.7064	Right 6
Left 7	0.7943	-0.5213	-0.4986	0.8237	-0.4675	0.6570	**-0.4564**	Right 7
	Left 1	Left 2	Left 3	Left 4	Left 5	Left 6	Left 7	

Matching Scores

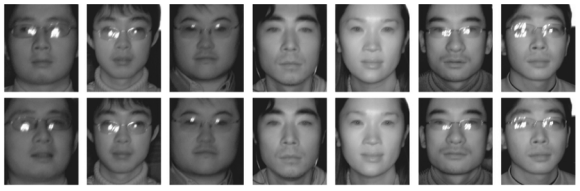
	Right 1	Right 2	Right 3	Right 4	Right 5	Right 6	Right 7	
Left 1	**0.4794**	0.5781	0.5884	0.5660	0.5975	0.5975	0.5761	Right 1
Left 2	0.5368	**0.4787**	0.6093	0.6036	0.6107	0.5955	0.6338	Right 2
Left 3	0.5398	0.5486	**0.4744**	0.5754	0.6146	0.6177	0.5887	Right 3
Left 4	0.5154	0.5258	0.5068	**0.4517**	0.6032	0.5762	0.6138	Right 4
Left 5	0.5335	0.5587	0.5227	0.5104	**0.4810**	0.6192	0.6067	Right 5
Left 6	0.5518	0.5435	0.5288	0.5178	0.5665	**0.5298**	0.6142	Right 6
Left 7	0.5785	0.5816	0.5334	0.5259	0.5604	0.5414	**0.4708**	Right 7
	Left 1	Left 2	Left 3	Left 4	Left 5	Left 6	Left 7	

Fig. 4. Correlation coefficients and matching scores between visible light face images with lamp lights from the right (row 1) and the left (row 2) directions. Top: Face images, each column belonging to the same person. Middle: Correlation coefficients. Bottom: LBP+AdaBoost match-

illuminated by a visible light lamp from left and right, respectively. The correlation table shows intrapersonal and extrapersonal correlation coefficients. There, the diagonal entries (in bold font) are for the intrapersonal pairs (e.g., entry (i, i) is the correlation between the two images in column i); the upper triangles are for the extrapersonal right-right pairs (e.g., entry (i, j) is the correlation between the two images in column i and j in the first row); the lower triangle entries are for the extrapersonal left-left pairs (e.g., entry (i, j) is the correlation between the two images in column i and j in the second row). We see that the correlation coefficients between two images of faces illuminated from left and right are all negative numbers regardless of the face identity; those between two images under similar lighting directions can be either positive or negative. The mean and variance are —0.4738 and 0.1015 for the intrapersonal pairs and 0.0327 and 0.5932 for the extrapersonal pairs. Therefore, it is not a surprise that a PCA matching engine has no chance of making correct matches in this case.

The score table is produced by an advanced matching engine trained using AdaBoost with LBP features on a visible light face image training set (the AdaBoost-trained matching engine produces better results than trained using LDA). The mean and variance of the scores are 0.4808 and 0.0238 for intrapersonal pairs and 0.5694 and 0.0359 for extrapersonal pairs. We see that the scores for intrapersonal pairs under different lighting directions are generally lower than those for extrapersonal pairs under similar lighting directions. This means that reliable recognition cannot be achieved with visible light images, even using the advanced matching engine.

Active NIR Light Face Images Taken with Left and Right Lamp Lights



Correlation Coefficients

	Right 1	Right 2	Right 3	Right 4	Right 5	Right 6	Right 7	
Left 1	**0.5467**	0.5407	0.6775	0.6290	0.6503	0.6062	0.5600	Right 1
Left 2	0.5269	**0.9190**	0.5938	0.7091	0.6389	0.6694	0.6257	Right 2
Left 3	0.6628	0.6288	**0.8992**	0.7207	0.7812	0.7044	0.5688	Right 3
Left 4	0.6936	0.6569	0.7321	**0.9684**	0.8040	0.7757	0.6697	Right 4
Left 5	0.7004	0.6426	0.8269	0.7944	**0.9918**	0.8321	0.6864	Right 5
Left 6	0.7610	0.6079	0.7631	0.7991	0.8100	**0.9154**	0.6264	Right 6
Left 7	0.5240	0.6740	0.6042	0.7145	0.7384	0.6536	**0.9090**	Right 7
	Left 1	Left 2	Left 3	Left 4	Left 5	Left 6	Left 7	

Matching Scores

	Right 1	Right 2	Right 3	Right 4	Right 5	Right 6	Right 7	
Left 1	**0.6253**	0.2925	0.3435	0.3268	0.3270	0.3494	0.3158	Right 1
Left 2	0.3035	**0.6709**	0.2201	0.3096	0.3167	0.3081	0.3442	Right 2
Left 3	0.3288	0.2127	**0.6592**	0.2845	0.3389	0.3146	0.2742	Right 3
Left 4	0.3292	0.2929	0.2861	**0.7013**	0.2909	0.3300	0.3845	Right 4
Left 5	0.3618	0.2975	0.3158	0.3003	**0.6914**	0.3620	0.2651	Right 5
Left 6	0.3562	0.3169	0.2981	0.3148	0.3490	**0.6640**	0.2778	Right 6
Left 7	0.3032	0.3224	0.2777	0.3705	0.2711	0.2905	**0.7128**	Right 7
	Left 1	Left 2	Left 3	Left 4	Left 5	Left 6	Left 7	

Fig. 5. Analysis on properties of active NIR light face images with lamp lights from the right (row 1) and the left (row 2) directions. Top: The images, each column belonging to the same person. Middle: Correlation coefficients. Bottom: LBP+AdaBoost Matching scores.

The impact of environmental lighting is much reduced by the present NIR imaging system, as shown by the correlations and scores in Fig. 5. There, the corresponding NIR face images are taken in the same visible light conditions as in Fig. 4 and the explanations of the two tables are similar to those for Fig. 4. The correlation coefficients between NIR images are all positive, regardless of the visible light conditions and person identity. They have mean and variance of 0.8785 and 0.1502 for intrapersonal pairs and 0.6806 and 0.0830 for extrapersonal pairs. However, the intrapersonal correlation coefficients may not necessarily be higher than the extrapersonal ones, meaning possible recognition errors, even with the NIR images. Therefore, a better matching engine than correlation or PCA is still needed for highly accurate face recognition, even with NIR face images.

The score table shows the matching score produced by an LBP+AdaBoost classifier trained on active NIR images. The mean and variance of the scores are 0.6750 and 0.0296 for intrapersonal pairs and 0.3113 and 0.0358 for extrapersonal pairs. By examining all of the entries, we can see that the intrapersonal scores are consistently much higher than the extrapersonal ones. The above case studies suggest that the proposed active NIR imaging system with an advanced LBP+AdaBoost recognition engine can yield the highest recognition performance of all of the schemes.

3 Illumination Invariant Face Representation

In this section, we first provide an analysis from the Lambertian imaging model to show that such 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 LBP-based representation to amend the degree of freedom of the monotonic transform to achieve an illumination invariant representation of faces for indoor face recognition applications.

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:

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

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 the 3D space, and $\mathbf{s} = (s_x, s_y, s_z)$ is the lighting direction (a column vector, with magnitude). Here, 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 topmost factor that affects the face recognition performance is the direction of the incidence lighting relative to the face surface normal. The product of $\rho(x, y)\mathbf{n}(x, y)$ is the intrinsic property of the face at a fixed pose and is the only thing needed for face detection and recognition and, therefore, $\mathbf{s}$ is the extrinsic property that should be removed. Assume $\mathbf{s} = \kappa\mathbf{s}^0$, where κ is the strength of the lighting and $\mathbf{s}^0 = (s^0_x, s^0_y, s^0_z)$ is a unit column vector of the lighting direction, and let $\theta(x, y)$ be the incidence angle between the lighting and the face surface vector at point (x, y), $\cos\theta(x, y) = \mathbf{n}(x, y)\mathbf{s}^0$. Equation (1) can be expressed as

$$I(x, y) \propto \kappa\rho(x, y)\cos\theta(x, y). \tag{2}$$

A less restrictive modeling of constant κ would be a monotonic transform instead of a constant. 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 (cf. Fig. 2), i.e., $\mathbf{s}^0 = (0, 0, 1)$, the image can be approximated by

$$I(x, y) = \kappa\rho(x, y)n_z(x, y), \tag{3}$$

where κ is a multiplying constant due to possible changes in the strength of the lighting caused by changes in the distance between the face and the LED lights and $n_z(x, y)$ is exactly the depth information (2.5 map) that can be acquired by a range imaging system. An active NIR image $I(x, y)$ combines information about both depth map $n_z(x, y)$ and albedo map $\rho(x, y)$ and, therefore, provides the wanted intrinsic property about a face for face recognition.

3.2 Compensation for Monotonic Transform

We use an LBP representation to compensate for the degree of freedom in κ or in a monotonic transform to achieve an illumination invariant representation of faces for indoor face recognition applications. The basic form of the LBP operator is illustrated in Fig. 6. The binary bits describing a local 3×3 subwindow 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 anticlockwise. 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

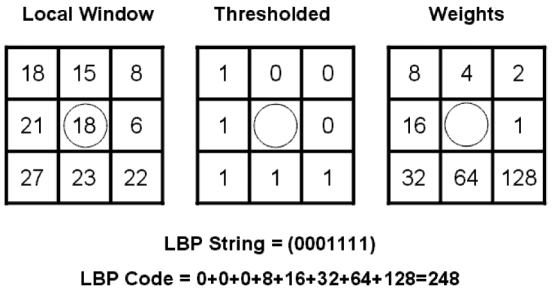


Fig. 6. LBP code for 3×3 window.

multiscale LBP, $\mathrm{LBP}_{(P,R)}$, where R is the radius of the circle surrounding the center and P is the number of pixels on the circle. An $\mathrm{LBP}(P,R)$ string is called uniform, denoted by $\mathrm{LBP}^{u2}_{(P,R)}$, if the neighboring bits (the circular sense) contain at most two bitwise transitions from 0 to 1 or vice versa (see [46] for details).

As discussed in the above, the pixel intensities in an NIR image are subject to a multiplying constant κ due to changes in the distance between the face and the LED lights. This degree of freedom can be fixed by using an LBP-based representation. To be less restrictive and more realistic, let us relax the effect of the multiplying constant to a *monotonic transform*, $\mathcal{T}$. Then, active NIR images can be modeled by

$$I(x,y) = \mathcal{T}(\rho(x,y)n_z(x,y)). \tag{4}$$

Let us be given an image $I'(x,y) = \rho(x,y)n_z(x,y)$ and a transformed image $I''(x,y) = \mathcal{T}(I'(x,y)) = \mathcal{T}(\rho(x,y)n_z(x,y))$. The ordering relationship between pixels in an image is not changed by any monotonic transform, namely, if $I'(x_1,y_1) > I'(x_2,y_2)$, then $I''(x_1,y_1) > I''(x_2,y_2)$. Therefore, the LBP codes generated from I'' are exactly the same as the ones generated from I'.

From the analysis, we see that the NIR imaging and LBP features together lead to an illumination invariant representation of faces for indoor face recognition applications. 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 convenience for face recognition.

The vision group at Oulu University developed a method of LBP-based matching [17], [18]. There 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.

4 Statistical Learning of Best Features and Classifiers

In this section, we present two statistical learning methods, one based on LDA and one on AdaBoost, for building face recognition classifiers. Given a training set of LBP features of faces subject to image noise, slight pose changes, and alignment errors, such a learning method performs a transform to the find most discriminative features and thereby build a strong classifier. We assume and ensure that a large set of training examples is available to sufficiently represent differences between individual faces such that, once trained, the classifier is able to recognize faces without need to be retrained when a new individual client is added.

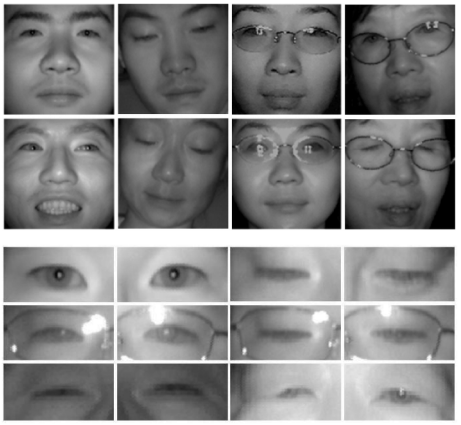


Fig. 7. Face examples (upper) and eye examples (lower) with and without eyeglasses, with eyes open and closed, with "bright pupils" present and absent.

The LBP features are derived from LBP histogram statistics as follows:

1. Base LBP Operator.
   a. $\mathrm{LBP}^{u2}_{8,1}$ is used. The LBP codes are computed for every pixel location over the image.
2. Histogramming of LBP Codes.
   a. 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 $\mathrm{LBP}^{u2}_{8,1}$.
   b. An LBP histogram is considered as a set of 59 individual features.
3. Feature Set of LBP Histograms.
   a. Assuming 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).
   b. In our case, $W \times H = 120 \times 142$, a local region for histogramming is a rectangle of size 16×20, and, therefore, the interior area is of size $W' \times H' = 104 \times 122$ pixels, and there are a total of $104 \times 122 \times 59 = 748,592$ elements in the LBP histogram feature set of a face image.

While the initial LBP histogram feature set is of high cardinality, 748,592 in this work, the intrinsic dimension of the face pattern may not be so high. The present LBP + LDA and LBP+AdaBoost methods are developed for reducing the dimensionality and construct classifiers in the reduces feature space, in an optimal sense.

4.1 LBP+LDA Classifier

LDA reduces the dimensionality by linearly projecting the original feature vector in high dimensional space to a lower dimensional subspace. The projection matrix is calculated from intraclass and extraclass scatter matrices. For memory reasons, the 748,592-dimensional LBP histogram features are down-sampled to 10,000 by uniformly sampling. The

10,000-dimensional data is preprocessed using the PCA transform to make the within-class scatter matrix nonsingular. The LDA projection matrix $\mathbf{P}$, composed of so-called Fisherfaces [47], are then computed. Of course, other more advanced forms of LDA, such as direct LDA or regularized LDA, could be used to obtain $\mathbf{P}$.

Given two input vectors $\mathbf{x}_1$ and $\mathbf{x}_2$, 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:

$$H(\mathbf{v}_1, \mathbf{v}_2) = (\mathbf{v}_1 \cdot \mathbf{v}_2)/\|\mathbf{v}_1\| \, \|\mathbf{v}_2\|. \tag{5}$$

In the test phase, the projections $\mathbf{v}_1$ and $\mathbf{v}_2$ are computed from two input vectors $\mathbf{x}_1$ and $\mathbf{x}_2$, one 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 $\mathbf{x}_1$ belong to the same person.

4.2 LBP+AdaBoost Classifier

While an AdaBoost procedure essentially learns a two-class classifier, we convert the multiclass problem into a two-class one using the idea of intra and extraclass difference [48]. However, here the difference data are derived from the LBP Histogram features rather than from the images. A difference is taken between two LBP histogram feature sets (748,592-dimensional), which is intraclass if the two face images are of the same person, or extraclass if not.

A training set of N labeled examples is given for two classes, $\mathbf{S} = (x_1, y_1), \ldots, (x_N, y_N)$, where x_i is a training example (which is the difference between two LBP histogram feature sets in this case) and $y_i \in \{+1, -1\}$ is the class label. The procedure learns a sequence of T weak classifiers, $h_t(x) \in \{-1, +1\}$, and linearly combines it in an optimal way into a stronger classifier,

$$H(x) = \text{sign}\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right), \tag{6}$$

where $\alpha_t \in \mathbf{R}$ are the combining weights. We can consider the real-valued number $\sum_{t=1}^{T} \alpha_t h_t(x)$ as the score and make a decision by comparing the score with a threshold.

An AdaBoost learning procedure is aimed at deriving α_t and $h_t(x)$ so that an upper error bound is minimized [49]. The procedure keeps a distribution $w_t = (w_{t,1}, \ldots, w_{t,N})$ for the training examples. The distribution is updated after each learning iteration t. The AdaBoost procedure adjusts the distribution in such a way that more difficult examples will receive higher weights. The subsequent weak classifier is designed according to the weighted distribution of the training examples. We train a cascade of strong classifiers [50] to handle complicated data distributions. The cascade strategy is also used in face/eye detection.

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. In our system, a weak classifier is based on a single scalar feature, i.e., an LBP histogram bin value; a weak classification decision, represented by +1 or -1, is made comparing the scalar feature with an appropriate threshold, the threshold being computed to minimize the weighted error on the training set.

In the test phase, two sets of selected LBP features are compared, one for the input face image and one for an enrolled face image. The difference is taken between the two sets. Each element in the difference set is compared with a threshold to give a weak decision h_t. The weak decisions for a strong classifier are linearly combined with the weights α_t to give the predict value $H(x)$. When a cascade is used, a cascade of strong classifiers is evaluated and a final decision is made.

5 Building the System

Highly accurate and fast face recognition systems can be built using the present hardware and recognition engine. Such a system includes three main software modules, as usual: face detection, eye localization, and face matching. Each of these performs a two-class classification, i.e., classifying the input into the positive or negative class:

- **Face Detection** classifies each scanned subwindow into either face or nonface.
- **Eye Localization** classifies each scanned subwindow into either eye or noneye.
- **Face Recognition** evaluates the similarity between the input face and each enrolled face and compares the similarity to decide whether the two faces belong to the same person.

The fundamental learning problem here is to learn a generally nonlinear classifier to classify between face and nonface classes for face detection, between eye and noneye classes for eye localization, and between intrapersonal and extrapersonal classes for face recognition. These are done in a unified framework of local features with AdaBoost learning. A learned classifier is a nonlinear mapping from the input subwindow to a confidence score. The final classification decision can be done by comparing the score with a confidence threshold. While how to build the face matching engine has been described in the previous section, here we include how to build the face detector and eye localizer and how to integrate the modules into a system.

In this paper, AdaBoost classifiers [50], with the extended Haar features proposed in [51], are used for face and eye detection. This design meets the requirement on the accuracy and is faster than LBP-based detectors because of the simpler form of Haar filters. Our face detection module is a rather standard AdaBoost face detector, trained using 50,000 NIR face examples, including difficult ones shown in Fig. 7. However, accurate eye localization in active NIR images is a difficult task when the active NIR lighting causes specular reflections on eyeglasses. This critical issue will be solved in the following.

5.1 Eye Detection

Detection of eyes in active NIR images is more challenging than in normal visible light images due to likely specular reflections on eyeglasses. On the right of Fig. 7 shows some examples of eyes. At the upper-left corner is an ideal example of eyes which can be easily detected, with a "bright pupil" detector [10], [44], [45] or an appearance-based detector. The other examples are difficult ones. Specular reflection on eyeglasses is the most serious problem. Eyelid occlusion, which happens among people in some ethnic groups and senior people and eye closing due to blinking are among other problems. Eye detection in these situations cannot be done by using a simple eye detector.

To tackle the difficulties, we propose a coarse-to-fine, easy-to-complex architecture to overcome the problem, as in Fig. 8. The motivation for designing such an architecture is the following: By an observation on characteristics of NIR images

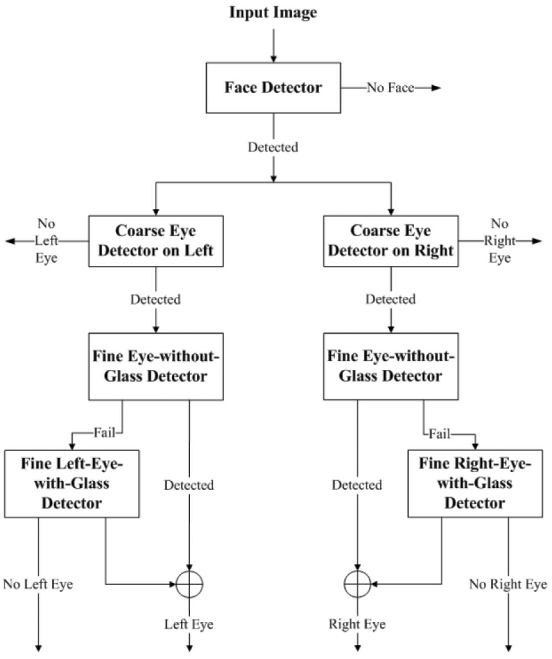


Fig. 8. Architecture for face and eye detection in NIR Images, with or without glasses.

of eyes with glasses, we see 1) that the images of eyes with and without glasses appear quite different and 2) that an image of an eye with glasses is not left-right symmetric because the specular reflections tend to be more in the inner part of the face rather than around the centers of the eyes. A mixture of left and right eye images, with and without glasses, would contain a large variation. However, our detector for eyes without glasses can perform very accurately and fast and have a priori knowledge of the positions of the left and right eyes in the face area for more effective applications of the left and right eye detectors. Based on these, we manually group the eye images into three subsets, 1) eyes without glasses, 2) left eyes with glasses, and 3) right eyes with glasses, and train an effective eye detector for each case. The resulting detectors can be very accurate because the variation in each subset is reduced considerably.

The architecture includes a sequence of increasingly complex eye detectors, trained with specific eye data. After a face is detected, the left and right eyes are detected, respectively, by applying a coarse eye detector, trained using all eye examples, including left and right eyes, with and without glasses. This detects all possible eye subwindows, and rejects more than 95 percent noneye subwindows. Then, a fine eye-without-glass detector is applied to each of the coarse eye subwindows to verify if there is an eye-without-glass pattern. If successful, an eye is detected; otherwise, the subwindow is passed to a fine eye-with-glass detector to verify if there is an eye-with-glass pattern. A final decision is made after merging multiple detects. This design enables accurate and fast detection of faces and eyes, with and without glasses.

5.2 Implementation

The active NIR imaging hardware and the trained face detector, eye detectors, and face matching engine are integrated onto a P4 3.0 GHz PC. The camera interface uses the USB 2.0 protocol and runs at 30 frames per second. The total processing time for face/eye detection and face matching in a 640×480 image is, on average, about 78 ms, in which the face and eye detection consumes 43 ms, and the matching engine does 35 ms for a database of 1,000 people.

Fig. 9. Application systems. (a) An access control and time attendance system in operation. (b) A prototype machine readable travel document (MRTD) system.

With such a speed, we are able to build a highly accurate and fast face recognition system, even if the detection and recognition engines do not achieve 100 percent accuracies. The system keeps processing incoming images from the video camera until a match is found with a sufficient confidence. Usually, an enrolled person can be successfully identified within 1-2 seconds. When a person cannot be identified after a timeout period, say 5 seconds from the first attempt, the system gives a "sorry" message; this usually happens when the user is not in the database or is not cooperative enough to look into the camera.

Fig. 9 shows two application systems that use the present NIR face recognition technology for biometric authentication. On the left is a system for access control and time attendance,[1] where the NIR face biometric is used for one to many identification. The access control of office and lab rooms is done via intranet. On the right is an MRTD system in which the NIR face biometric is used for one-to-one verification. The NIR face biometric module has been deployed at ShenZhen (China)-Hong Kong border, the world largest border crossing point, as a biometric option for Hong Kong passengers and vehicle drivers since June 2005 and at ZhuHai (China)-Macau border since March 2006.

6 Performance Evaluations

A biometric system may be evaluated in three major steps: a technology evaluation, a scenario evaluation, and an operational evaluation [20], [52]. In a technology evaluation, the testing is performed in laboratories using a prepared set of data. A scenario evaluation aims to evaluate the overall capabilities of the entire system in a specific scenario. An operational evaluation is very similar to a scenario evaluation except that it is performed at the actual site using the actual subjects/areas. So far, most available reports are on technology evaluation and few reports are available on scenario or operational evaluations.

In the following, we evaluate the performance of our NIR face and eye detection modules, and LBP+AdaBoost and LBP+LDA face matching engines in comparison with several

1. See a video demo at http://www.cbsr.ia.ac.cn/demos/index-en.htm.

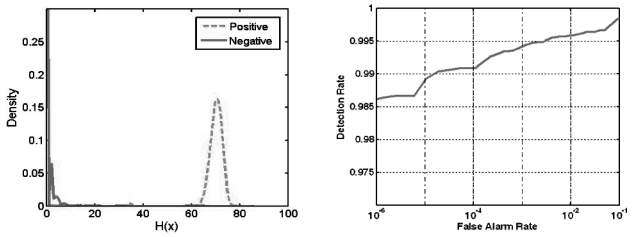


Fig. 10. Distributions of faces and nonfaces (left) and ROC curve (right) for face detection, on the test set.

existing baseline and state-of-the-art face matching engines. We then present case studies regarding the effects of eyeglasses, time lapse, and weak illumination, and report on the ability of the LBP+AdaBoost method to generalize to unseen ethnic faces. Finally, we present scenario evaluation of the NIR face biometric system for access control and time attendance.

6.1 Face and Eye Detection

The face detection cascade was trained using 178,000 positive examples of all types of faces (Fig. 7), of size 20×20 pixels. A cascade of seven strong classifiers was trained for face detection, composed of a total of only 436 weak classifiers. With NIR images, the Haar+AdaBoost detector could easily detect faces with or without glasses, and eyes *without* glasses, open or closed, and with slight pose variation. Fig. 10 shows the face and nonface distributions and the ROC curve for a test set. The test set is composed of 8,600 positive examples and about 20,000,000 negative examples (subwindows) derived from 8,600 NIR images, none of which were in the training set. We can see the two classes were well separated for NIR images. At the FAR of 10^{-6}, the detection rate was 98.7 percent; at the FAR of 10^{-7}, it was 98 percent. The speed of the face detector was about 18 ms per frame.

For training the eye detectors, the eye training set (of size 24×12 pixels) was composed of original eye images and slightly rotated versions of the original ones and their mirrors. Among this, 60 percent were without glasses and 40 percent were with glasses and 15 percent wwere of closed eyes. The noneye examples were collected mainly from the upperpart of faces. The training method is the following: First, the coarse eye detector was trained using 170,000 positive examples of all types of eyes and about 30,000,000 negative examples of noneyes. The resulting detector consisted of a cascade of four strong classifiers, with a total of only 56 weak classifiers. It detected all possible eye subwindows, and rejected more than 95 percent of the noneye subwindows (i.e., with FAR of about 5 percent).

Next, the eye-without-glass detector was trained using 271,000 of positive examples of eye-without-glasses and about 20,000,000 negative examples of noneyes. A cascade of seven strong classifiers were trained, with a total of 326 weak classifiers. This achieved a detection rate of nearly 98 percent with an FAR of 10^{-2} (see Fig. 11). The test set is composed of 10,000 eye-without-glasses positive examples and 10,000,000 negative examples derived from 5,000 images. Finally, the eye-with-glass detector was trained using 91,000 positive examples of eye-with-glasses and about 20,000,000 negative examples of noneyes. A cascade of four strong classifiers was trained, with a total of 225 weak classifiers. This achieved a detection rate of 87 percent, with the FAR of 10^{-2} (see Fig. 11). The test set is composed of 7,200 eye-with-glasses positive examples and 10,000,000 negative examples derived from 3,600 images, none of which

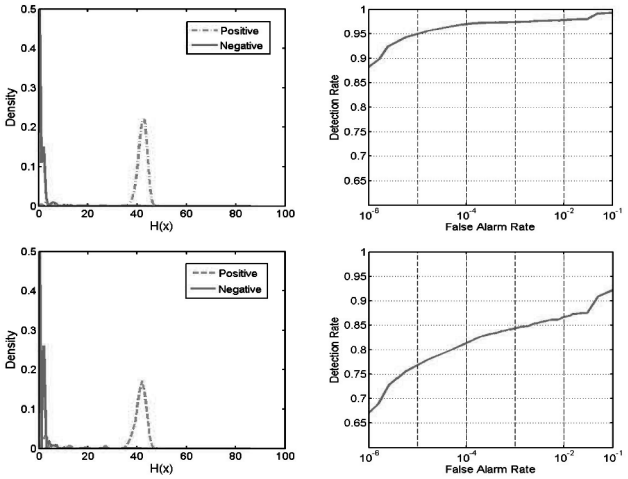


Fig. 11. Training performances for detection of eyes without glasses (first row) and with glasses (second row) on the test set. Shown are the distributions of positive and negative examples (left) and the ROC curves (right).

were in the training set. The speed of eye detectors was about 25 ms per frame.

Note that the FARs for the eye detectors were conditioned on the face detection. Considering that the FAR is 10^{-6} for the face detection (and 5 percent for the coarse eye detection), the chance that a noneye subwindow would be classified as eyes would be below 10^{-7}, which is low enough for applications. On the other hand, given the FAR = 1 percent, the detection rate for eyes without glasses was 98 percent, a very good number. However, the detection of eye-with-glasses was more difficult: The detection rate was 87 percent. Nonetheless, the tradeoff was that it took a bit longer time (at most 20 percent of increase) for a successful detection of eyes with glasses.

6.2 Face Recognition—Basic Evaluation

In the training phase, the training set of positive examples was derived from intrapersonal pairs of LBP histogram features, the negative set from extrapersonal pairs, each example being a 748,592-dimensional vector. There were 10^4 face images of about 1,000 people, 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 five strong classifiers were trained, with about 1,500 weak classifiers. The ROC curves for the training set is shown on the left of Fig. 12, where the FAR is reduced to below 10^{-7} with an accuracy of 94.4 percent.

A technology evaluation was done with a test set of 3,237 images. The test set contained 35 people, with 80 to 100 images per person. None of the test images were in the training set. This generated 149,217 intraclass (positive) and 5,088,249 extraclass (negative) pairs. Several other methods were included in the evaluation for comparison, using the same set of training and test images. They were:

1. PCA on the NIR images (with Mahalanobis distance),
2. LDA on the NIR images (with cosine distance),
3. the LBP + LDA method,
4. the original LBP method developed by Ahonen et al. [17] and Hadid et al. [18] (χ^2 distances between LBP histograms in 7×7 image blocks) with three operators: $\mathrm{LBP}^{u2}_{(8,1)}$, $\mathrm{LBP}^{u2}_{(8,2)}$, and $\mathrm{LBP}^{u2}_{(16,2)}$.

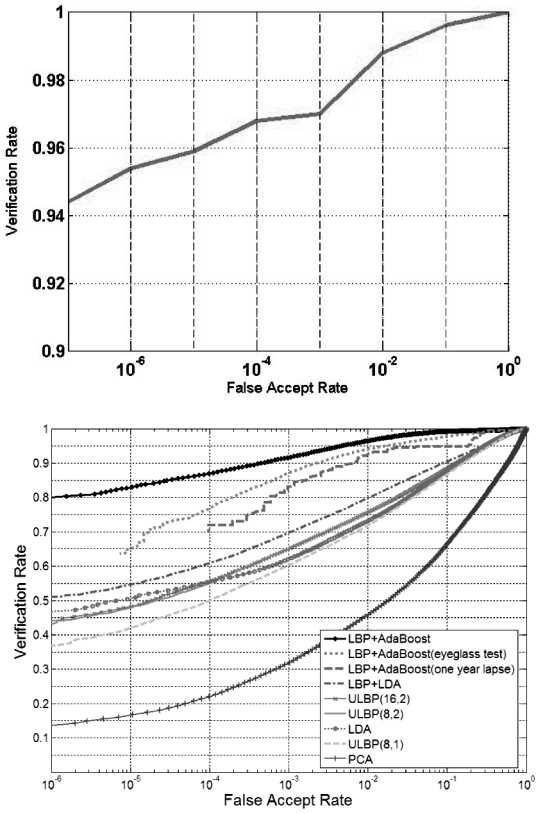


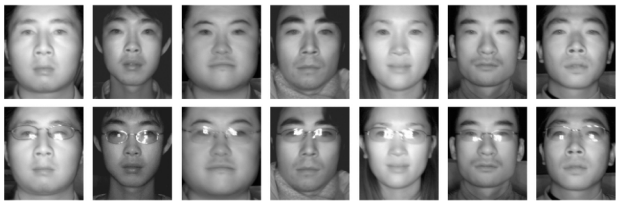
Fig. 12. Top: ROC Curve of LBP+AdaBoost method for face verification on the training set. Bottom: ROC curves for various compared methods, an ROC curve for testing the eyeglass effect, and an ROC for testing the time lapse effect.

The bottom of Fig. 12 shows the ROC curves (1-1 verification rate (VR) versus FAR) derived from the scores for the intra and extraclass pairs. By the VR values at FAR = 0.1 percent, the compared methods can be ranked in order of decreasing VR as: LBP+AdaBoost (VR = 91.8 percent), LBP+LDA (69.9 percent), $LBP^{u2}_{(8,2)}$ (65.29 percent), $LBP^{u2}_{(16,2)}$ (65.29 percent) Image+LDA (62.4 percent), $LBP^{u2}_{(8,1)}$ (60.7 percent), and Image+PCA (32.0 percent). See later for explanations of the "LBP+AdaBoost (eyeglass test)" and "LBP+AdaBoost (one year lapse)" curves.

6.3 Face Recognition—Eyeglasses versus No-Eyeglasses

The following presents a case analysis of influence of eye glasses on face matching, as shown by the images and the score table in Fig. 13. In the tables, the diagonal entries (in bold font) are for the intrapersonal pairs without (W/O) and with glasses (i.e., between the two images in the same column), the lower triangle entries for the extrapersonal no-glass pairs (i.e., between two different images in the first row), and the upper triangle for the extrapersonal glass pairs (i.e., 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 intrapersonal pairs and 0.7985 and 0.0761 for extrapersonal pairs of either wearing no glasses or wearing glasses. However, the intrapersonal correlations may not necessarily be higher than the extrapersonal ones, which means recognition errors. However, the LBP+AdaBoost matching engine can separate between the two classes

Active NIR Light Face Images With and Without Glasses

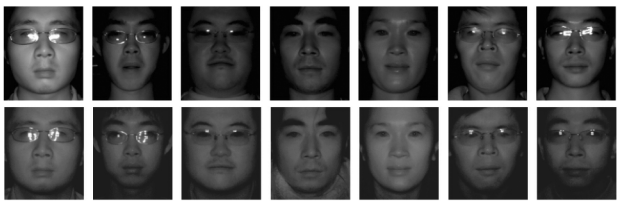


LBP+AdaBoost Matching Scores for intra- and extra-personal pairs

	With 1	With 2	With 3	With 4	With 5	With 6	With 7	
W/O 1	**0.6537**	0.2880	0.3276	0.3251	0.4056	0.3510	0.3422	With 1
W/O 2	0.2205	**0.6468**	0.1730	0.2859	0.2545	0.2740	0.3856	With 2
W/O 3	0.3487	0.1838	**0.6565**	0.3209	0.3261	0.2666	0.3231	With 3
W/O 4	0.2464	0.2744	0.2654	**0.7092**	0.3331	0.2986	0.3680	With 4
W/O 5	0.3425	0.2945	0.3352	0.2896	**0.6120**	0.3191	0.3704	With 5
W/O 6	0.2896	0.2100	0.2205	0.2871	0.2699	**0.6366**	0.2346	With 6
W/O 7	0.2915	0.3708	0.3197	0.3958	0.3397	0.3056	**0.6541**	With 7
	W/O 1	W/O 2	W/O 3	W/O 4	W/O 5	W/O 6	W/O 7	

Fig. 13. 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.

Active NIR Light Face Images with Time Lapse



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

	Y06 1	Y06 2	Y06 3	Y06 4	Y06 5	Y06 6	Y06 7
Y05 1	**0.6040**	0.3364	0.2809	0.2946	0.3829	0.3038	0.2439
Y05 2	0.3069	**0.6667**	0.2070	0.2530	0.3142	0.2804	0.2709
Y05 3	0.3895	0.2689	**0.6372**	0.3398	0.3745	0.3022	0.2754
Y05 4	0.3010	0.2763	0.2212	**0.6542**	0.2990	0.2950	0.3265
Y05 5	0.3156	0.2789	0.2437	0.2492	**0.6441**	0.3012	0.2504
Y05 6	0.3532	0.3514	0.2958	0.2912	0.4219	**0.6371**	0.3193
Y05 7	0.3626	0.3553	0.2849	0.3530	0.3063	0.3117	**0.6513**

Fig. 14. 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.

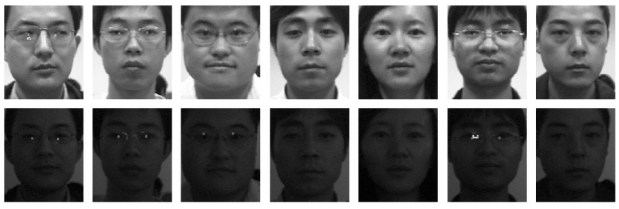
well—the intrapersonal scores are consistently higher than than extrapersonal scores. Also, we can find that the statistics (mean and variance) of LBP+AdaBoost scores for this case study and for the case in Fig. 5 are quite consistent.

Statistics were also obtained using 1,500 images of 30 people, 50 images per person, 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. 12 (the portion for FAR smaller than 10^{-5} is unavailable because of the limited data points). At FAR = 0.1 percent, the VR was 87.1 percent, as opposed to 91.8 percent of the "LBP+AdaBoost" curve for the no-eyeglasses versus no-eyeglasses and eyeglasses versus eyeglasses comparisons.

6.4 Face Recognition—Time Lapse

Tests were performed to evaluate the effect of time lapse on NIR face recognition. Fig. 14 presents a case analysis of the time lapse effect on the matching scores. The NIR images of seven individuals were acquired in Spring 2005 and Spring

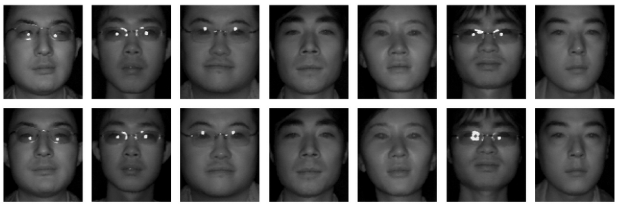
Visible Light Face Images under Controlled and Weak Light Condition



LBP+AdaBoost Matching Scores for intra- and extra-personal 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. 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.



LBP+AdaBoost Matching Scores for intra- and extra-personal 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. 16. 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.

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 intrapersonal pairs (in bold font) and 0.3045 and 0.0462 for extrapersonal pairs. The LBP+AdaBoost matching engine separates between the two classes well; the intrapersonal scores are consistently higher than the extrapersonal ones.

Statistics were also obtained using 750 images of 30 people, 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. 12 (the portion for FAR smaller than 10^{-4} is unavailable because of the limited data points). At FAR=0.1 percent, the VR was 83.24 percent, as opposed to 91.8 percent for images of no significant time lapse (the "LBP+AdaBoost" curve).

6.5 Face Recognition—Weak Illumination

Figs. 15 and 16 present case studies to compare the performance of visible light (VL) and NIR image-based face matching methods under weak illumination. The LBP+AdaBoost classifier for VL images was trained using VL images, whereas the one for NIR images was the one using other tests. In the tables, the diagonal entries (in bold font) are for the intrapersonal pairs between controlled and weak illumination. For the VL case, the mean and variance are 0.4970 and 0.0201 for intrapersonal pairs and 0.4747 and 0.0154 for extrapersonal pairs and there are several mismatches because the intrapersonal scores are not necessarily higher than the extrapersonal ones. In contrast, the NIR solution separates the two classes well, with the mean and variance of 0.6675 and 0.0377 for intrapersonal pairs and 0.3492 and 0.0403 for extrapersonal pairs, producing correct matches for all pairs.

6.6 Scenario Evaluation

The scenario evaluation was associated with an access control and time attendance application in the CASIA office building. The system configured for this purpose consisted of one server and three clients. The server was used for enrollment, template generation, database and log information management, and communication with the clients. A client was used to do online face recognition and communication with the server to log recognition results and times. The tests lasted over a period of a month.

The tests were done in the form of one-to-many matching, with the following protocol: 870 people were enrolled, with five templates per person recorded. Of these, 30 were workers in the building and 840 were collected from other sources unrelated to the building environment. The 30 workers were used as the client people while the other 840 were used as background people. Images of the 870 participating people were not included in the training set. The enrolled population was mostly composed of Chinese people with a few Caucasians. The enrollment was done at sites different from the sites of the client systems. The tests with the client people were aimed at collecting statistics of the correct recognition rate and false rejection rate. On the other hand, 10 workers participated as the regular imposters (not enrolled) and some visitors were requested to participate as irregular imposters. This provided statistics of correct rejection rate and false acceptance rate.

The 30 clients and 10 imposters were required to report to the system four times a day to take time attendance, twice in the morning and twice in the evening when they started working and left the office for lunch and for home. Not all workers followed this rule strictly, whereas some did more than four times a day. Some client people deliberately challenged the system by exaggerated expressions or occluding part of the face with a hand so that the system did not recognize them. We counted these as invalid sessions. Only those client sessions which were reported having problems getting recognized were counted as false rejections. On the other hand, the imposters were encouraged to challenge the system to get false acceptances.

The tests were also extended by moving an additional client on a laptop at various sites in the building. A further extension was to test in complete darkness. These were to test the reliability of the system with changing illumination and environments. Both regular and irregular clients and imposters participated in the extended tests.

After a period of one month evaluation, the system has demonstrated excellent accuracy, speed, usability, and stability under varying indoor illumination, even in the complete darkness. The statistics of the scenario evaluation is

TABLE 1
Scenario Evaluation Statistics

	# Persons	# Sessions	# Sessions Accepted	# Sessions Rejected	Success Rate
Clients	30	1725	1721	4	99.77%,
Imposters	10+visitors	1096	3	1093	99.72%
Background	840	-	-	-	-

summarized in Table 1 where the "Success Rate" means correct acceptance rate for clients and correct rejection rate for imposters. From the statistics, we can see that the equal error rate was below 0.3 percent. Hence, we conclude that the system has achieved high performance for cooperative face recognition.

6.7 Face Recognition—Unseen Ethnic Faces

The present NIR face recognition engine was trained with Chinese faces only. The system has been tested in with faces of other ethnic groups, including Caucasian and African faces. The recognition accuracy appeared not affected by ethnic groups and the system gave perfect performance in face/eye localization, enrollment, and identification in all cases. This suggests that the system is unaffected by different ethic groups even unseen in training data.

7 Summary and Conclusions

We have presented a novel solution, including active NIR imaging hardware, algorithms, and system design, to overcome the problem of illumination variation that every face recognition system has to deal with. An illumination invariant face representation is obtained by extracting LBP features 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 thereby be built. Extensive experiments show the robustness of the present solution in terms of image properties, illumination changes, ethnic groups, and advantages over existing methods.

The current solution is developed for cooperative user applications indoor. It is not yet suitable for uncooperative user applications such as face recognition in video surveillance. Nor is it suitable for outdoor use due to strong NIR component in the sunlight. Future works will be enhancing solutions to overcome these limitations.

Acknowledgments

This work was supported by the following funding sources: Chinese National Natural Science Foundation Project #60518002, Chinese National 863 Program Projects #2004AA1Z2290 and #2004AA119050, Chinese Academy of Sciences 100 people project, and AuthenMetric Co. Ltd.

References

[1] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld, "Face Recognition: A Literature Survey," *ACM Computing Surveys,* pp. 399-458, 2003.

[2] *Handbook of Face Recognition,* S.Z. Li and A.K. Jain, eds. Springer-Verlag, Mar. 2005.

[3] L. Sirovich and M. Kirby, "Low-Dimensional Procedure for the Characterization of Human Faces," *J. Optical Soc. Am. A,* vol. 4, no. 3, pp. 519-524, Mar. 1987.

[4] M.A. Turk and A.P. Pentland, "Eigenfaces for Recognition," *J. Cognitive Neuroscience,* vol. 3, no. 1, pp. 71-86, Mar. 1991.

[5] M. Bichsel and A.P. Pentland, "Human Face Recognition and the Face Image Set's Topology," *CVGIP: Image Understanding,* vol. 59, pp. 254-261, 1994.

[6] P.Y. Simard, Y.A.L. Cun, J.S. Denker, and B. Victorri, "Transformation Invariance in Pattern Recognition—Tangent Distance and Tangent Propagation," *Neural Networks: Tricks of the Trade,* pp. 239-274, 1998.

[7] M. Turk, "A Random Walk through Eigenspace," *IEICE Trans. Information and Systems,* vol. 84, no. 12, pp. 1586-1695, Dec. 2001.

[8] Y. Adini, Y. Moses, and S. Ullman, "Face Recognition: The Problem of Compensating for Changes in Illumination Direction," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 7, pp. 721-732, July 1997.

[9] "Face Recognition Vendor Tests (FRVT)," Nat'l Inst. of Standards and Technology, http://www.frvt.org, 2006.

[10] "A Method for Face Image Acquisition Using Active Lighting," AuthenMetric Co. Ltd., Patent Application No. 200310121340.1, Beijing, China, Dec. 2003.

[11] "A Method for Face Image Acquisition and a Method and System for Face Recognition," AuthenMetric Co. Ltd., Patent Application No. PCT/CN2004/000482, Beijing, China, May 2004.

[12] "An Image Acquisition Apparatus for Face Recognition," AuthenMetric Co. Ltd., Patent Application No. 200520022878.1, Beijing, China, Mar. 2005.

[13] S.Z. Li et al., "AuthenMetric F1: A Highly Accurate and Fast Face Recognition System," *Proc. Int'l Conf. Computer Vision,* Oct. 2005.

[14] S.Z. Li, R.F. Chu, M. Ao, L. Zhang, and R. He, "Highly Accurate and Fast Face Recognition Using Near Infrared Images," *Proc. IAPR Int'l Conf. Biometric,* pp. 151-158, Jan. 2006.

[15] S.Z. Li, L. Zhang, S.C. Liao, X.X. Zhu, R.F. Chu, M. Ao, and R. He, "A Near-Infrared Image Based Face Recognition System," *Proc. Seventh IEEE Int'l Conf. Automatic Face and Gesture Recognition,* pp. 455-460, Apr. 2006.

[16] T. Ojala, M. Pietikainen, and D. Harwood, "A Comparative Study of Texture Measures with Classification Based on Feature Distributions," *Pattern Recognition,* vol. 29, no. 1, pp. 51-59, Jan. 1996.

[17] T. Ahonen, A. Hadid, and M. Pietikainen, "Face Recognition with Local Binary Patterns," *Proc. European Conf. Computer Vision,* pp. 469-481, 2004.

[18] A. Hadid, M. Pietikainen, and T. Ahonen, "A Discriminative Feature Space for Detecting and Recognizing Faces," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* vol. 2, pp. 797-804, 2004.

[19] S.Z. Li, "Face Recognition: Challenges and Rresearch Directions," Lecture at the Second Summer School for Advanced Studies on Biometrics for Secure Authentication, http://www.computer-vision.191.it/Biomet-School.htm, June 2005.

[20] P.J. Phillips, A. Martin, C.L. Wilson, and M. Przybocki, "An Introduction to Evaluating Biometric System," *Computer,* pp. 56-63, Feb. 2000.

[21] A.S. Georghiades, P.N. Belhumeur, and D.J. Kriegman, "From Few to Many: Illumination Cone Models for Face Recognition under Variable Lighting and Pose," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 23, no. 6, pp. 643-660, June 2001.

[22] R. Ramamoorthi, "Analytic PCA Construction for Theoretical Analysis of Lighting Variability in Images of a Lambertian Object," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 24, no. 10, pp. 1322-1333, Oct. 2002.

[23] R. Basri and D.W. Jacobs, "Lambertian Reflectance and Linear Subspaces," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 25, no. 2, pp. 218-233, Feb. 2003.

[24] S.K. Nayar and R.M. Bolle, "Reflectance Based Object Recognition," *Int'l J. Computer Vision,* vol. 17, no. 3, pp. 219-240, 1996.

[25] D. Jacobs, P. Belhumeur, and R. Basri, "Comparing Images under Variable Illumination," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* pp. 610-617, 1998.

[26] A. Shashua and T.R. Raviv, "The Quotient Image: Class Based Re-Rendering and Recognition with Varying Illuminations," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 23, no. 2, pp. 129-139, Feb. 2001.

[27] R. Gross and V. Brajovic, "An Image Preprocessing Algorithm for Illumination Invariant Face Recognition," *Proc. Fourth Int'l Conf. Audio and Video-Based Biometric Person Authentication,* pp. 10-18, June 2003.

[28] H.T. Wang, S.Z. Li, and Y.S. Wang, "Generalized Quotient Image," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* pp. 498-505, 2004.

[29] E.H. Land, "An Alternative Technique for the Computation of the Designator in the Retinex Theory of Color Vision," *Proc. Nat'l Academy of Sciences (Physics),* vol. 83, no. 2, pp. 3078-3080, May 1986.

[30] D.J. Jobson, Z. Rahman, and G.A. Woodell, "Properties and Performance of a Center/Surround Retinex," *IEEE Trans. Image Processing,* vol. 6, no. 3, pp. 451-462, Mar. 1997.

[31] K.W. Bowyer, K.I. Chang, and P.J. Flynn, "A Survey of 3D and MultiModal 3D + 2D Face Recognition," *Proc. Int'l Conf. Pattern Recognition,* pp. 358-361, Aug. 2004.

[32] K.I. Chang, K.W. Bowyer, and P.J. Flynn, "An Evaluation of MultiModal 2D+3D Face Biometrics," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 27, pp. 619-624, 2005.

[33] "A4Vision Technology," A4Vision, http://www.a4vision.com/, 2006.

[34] "Computer Vision beyond the Visible Spectrum," *Proc. IEEE Workshop Computer Vision beyond the Visible Spectrum: Methods and Applications,* 1999-2003.

[35] "Object Tracking and Classification in and beyond the Visible Spectrum," *Proc. IEEE Int'l Workshop Object Tracking and Classification in and beyond the Visible Spectrum,* 2004-2005.

[36] S.G. Kong, J. Heo, B. Abidi, J. Paik, and M. Abidi, "Recent Advances in Visual and Infrared Face Recognition—A Review," *Computer Vision and Image Understanding,* vol. 97, no. 1, pp. 103-135, Jan. 2005.

[37] D.A. Socolinsky, L.B. Wolff, J.D. Neuheisel, and C.K. Eveland, "Illumination Invariant Face Recognition Using Thermal Infrared Imagery," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition,* vol. 1, pp. 527-534, Dec. 2001.

[38] D.A. Socolinsky and A. Selinger, "A Comparative Analysis of Face Recognition Performance with Visible and Thermal Infrared Imagery," *Proc. Int'l Conf. Pattern Recognition,* vol. 4, pp. 217-222, Aug. 2002.

[39] X. Chen, P.J. Flynn, and K.W. Bowyer, "Infra-Red and Visible-Light Face Recognition," *Computer Vision and Image Understanding,* vol. 99, pp. 332-358, Sept. 2005.

[40] A. Selinger and D.A. Socolinsky, "Face Recognition in the Dark," *Proc. IEEE CS Conf. Computer Vision and Pattern Recognition Workshops,* pp. 129-129, June 2004.

[41] J. Dowdall, I. Pavlidis, and G. Bebis, "Face Detection in the Near-IR Spectrum," *Image and Vision Computing,* vol. 21, pp. 565-578, July 2003.

[42] D.Y. Li and W.H. Liao, "Facial Feature Detection in Near-Infrared Images," *Proc. Fifth Int'l Conf. Computer Vision, Pattern Recognition and Image Processing,* pp. 26-30, Sept. 2003.

[43] Z.H. Pan, G. Healey, M. Prasad, and B. Tromberg, "Face Recognition in Hyperspectral Images," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 25, no. 12, pp. 1552-1560, Dec. 2003.

[44] S.Y. Zhao and R.R. Grigat, "An Automatic Face Recognition System in the Near Infrared Spectrum," *Proc. Int'l Conf. Machine Learning and Data Mining in Pattern Recognition,* pp. 437-444, July 2005.

[45] X. Zou, J. Kittler, and K. Messer, "Face Recognition Using Active Near-IR Illumination," *Proc. British Machine Vision Conf.,* Sept. 2005.

[46] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution Gray-Scale and Rotation Invariant Texture Classification Width Local Binary Patterns," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 24, no. 7, pp. 971-987, July 2002.

[47] P.N. Belhumeur, J.P. Hespanha, and D.J. Kriegman, "Eigenfaces versus Fisherfaces: Recognition Using Class Specific Linear Projection," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 19, no. 7, pp. 711-720, July 1997.

[48] B. Moghaddam, C. Nastar, and A. Pentland, "A Bayesian Similarity Measure for Direct Image Matching," Media Lab Technical Report No. 393, Massachusetts Inst. of Technology, Aug. 1996.

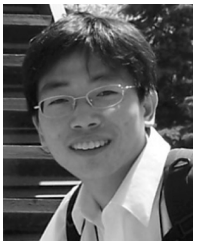
[49] Y. Freund and R. Schapire, "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting," *J. Computer and System Sciences,* vol. 55, no. 1, pp. 119-139, Aug. 1997.

[50] P. Viola and M. Jones, "Robust Real Time Objeect Detection," *Proc. IEE ICCV Workshop Statistical and Computational Theories of Vision,* July 2001.

[51] S.Z. Li and Z. Zhang, "FloatBoost Learning and Statistical Face Detection," *IEEE Trans. Pattern Analysis and Machine Intelligence,* vol. 26, no. 9, pp. 1112-1123, Sept. 2004.

[52] D.M. Blackburn, "Evaluating Technology Properly—Three Easy Steps to Success," *Corrections Today,* vol. 63, no. 1, July 2000.

Stan Z. Li received the PhD degree from Surrey University, United Kingdom. He is currently a professor at the National Laboratory of Pattern Recognition (NLPR), director of the Center for Biometrics and Security Research (CBSR), the Institute of Automation, Chinese Academy of Sciences (CASIA), and director of the Joint Laboratory for Intelligent Surveillance and Identification in Civil Aviation (CASIA-CAUC). He worked at Microsoft Research Asia as a researcher from 2000 to 2004. Prior to that, he was an associate professor at Nanyang Technological University, Singapore. His research interests include face recognition, biometrics, intelligent video surveillance, pattern recognition and machine learning, and image and video processing. He authored the book *Markov Random Field Modeling in Image Analysis* (Springer, first edition, 1995, second edition, 2001), coedited the *Handbook of Face Recognition* (Springer, 2005), and published more than 200 papers in international journals and conferences. He is currently an associate editor of *IEEE Transaction on Pattern Analysis and Machine Intelligence* and the editor-in-chief for the *Encyclopedia of Biometrics* (Springer, 2008). He leads several national and international collaboration projects in biometrics and intelligent video surveillance. He has been actively participating in organizing a number of international conferences and workshops in the fields of computer vision, image processing, pattern recognition, face analysis, and biometrics. He is a senior member of the IEEE and the IEEE Computer Society.



Rufeng Chu received the BS degree in automation from Zhengzhou University, China, in 1997 and the MS degree in electronic engineering from Dalian University of Technology, China, in 2001. He is currently a PhD candidate at the National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences. His research interests include pattern recognition, computer vision, and multimodal human-computer interaction.

Shengcai Liao received the BS degree in mathematics and applied mathematics in 2005 from Sun Yat-sen University. He is currently a graduate student at the Institute of Automation, Chinese Academic of Sciences (CASIA). He is studying pattern recognition and computer vision under the supervision of Stan Z. Li at the Center of Biometrics and Security Research at CASIA. His research interests include machine learning, pattern recognition, and computer vision. His paper "Face Recognition Using Ordinal Features" received the Motorola Best Student Paper at ICB 2006 in Hong Kong.

Lun Zhang received the BE degree in the special class for the gifted young from the University of Science and Technology of China, Hefei, in 2005. He is a master's degree candidate at the Institute of Automation at the Chinese Academy of Science. His research interests include pattern recognition and machine learning.