Highly Accurate and Fast Face Recognition Using Near Infrared Images*

Stan Z. Li, RuFeng Chu, Meng Ao, Lun Zhang, and Ran He

Center for Biometrics and Security Research & National Laboratory of Pattern Recognition,
Institute of Automation, Chinese Academy of Sciences,
95 Zhongguancun Donglu Beijing 100080, China
`http://www.cbsr.ia.ac.cn`

Abstract. In this paper, we present a highly accurate, realtime face recognition system for cooperative user applications. The novelties are: (1) a novel design of camera hardware, and (2) a learning based procedure for effective face and eye detection and recognition with the resulting imagery. The hardware minimizes environmental lighting and delivers face images with frontal lighting. This avoids many problems in subsequent face processing to a great extent. The face detection and recognition algorithms are based on a local feature representation. Statistical learning is applied to learn most effective features and classifiers for building face detection and recognition engines. The novel imaging system and the detection and recognition engines are integrated into a powerful face recognition system. Evaluated in real-world user scenario, a condition that is harder than a technology evaluation such as Face Recognition Vendor Tests (FRVT), the system has demonstrated excellent accuracy, speed and usability.

1   Introduction

Face recognition has a wide range of applications such as face-based video indexing and browsing engines, multimedia management, human-computer interaction, biometric identity authentication, and surveillance. Interest and research activities in face recognition have increased significantly in the past years [16, 17, 5, 20].

In *cooperative user* scenarios, a user is required cooperate with the face camera to have his/her face image captured properly, in order to be grated for the access; this is in contrast to more general scenarios, such as face recognition under surveillance. There are many cooperative user applications, such as access control, machine readable traveling documents (MRTD), ATM, computer login, e-commerce and e-government. In fact, many face recognition systems have been developed for such applications.

However, even in such a favorable condition, most existing face recognition systems, academic and commercial, are confounded by even moderate illumination changes. When the lighting differs from that for the enrollment, the system would either fail to recognize (false rejection) or make mistaken matches (false acceptance).

To avoid the problem caused by illumination changes (and other changes), several solutions have been investigated into. One technique is to use 3D (in many case, 2.5D) data obtained from a laser scanner or 3D vision method (*cf.* papers [3, 21]). Because 3D data captures geometric shapes of face, such systems are affected less by environmental lighting and it can cope with rotated faces because of the availability of 3D (2.5D) information for visible points. The disadvantages are the increased cost and slowed speed as well as the artifacts due to speculation. Recognition performances obtained using a single 2D image or a single 3D image are similar. [4].

Invisible imagery has recently received increased attention in the computer vision community, as seen from the IEEE workshop series [6, 13]. Thermal or far infrared imagery has been used for face recognition (*cf.* and a survey paper [10]). While thermal based face recognition systems are advantages for detecting disguised faces or when there is no control over illumination, they are subject to environmental temperature, emotional and health conditions, and generally do not perform as well as 2D based systems for the cooperative scenario.

The use of near infrared (NIR) imagery brings a new dimension for applications of invisible lights for face detection and recognition [7, 11, 14]. In [7], face detection is performed by analyzing horizontal projections of the face area using the fact that eyes and eyebrows regions have different responses in the lower and upper bands of NIR. In [11], a method of homomorphic-filtering is used as a pre-processing before extracting facial features. In [14], face recognition is done using hyperspectral images captured in 31 bands over an NIR range of 0.7μm-1.0μm; invariant features are extracted from such images.

In this paper, we present a highly accurate, real-time system for face recognition in cooperative user applications. The contributions are the following: First, we present a novel design of camera hardware. The camera delivers filtered NIR images containing mostly relevant, intrinsic information for face detection and recognition, with extrinsic factors minimized. This alleviates much difficulty in subsequent processing. Second, we present learning based algorithms, using a local feature representation, for effective face/eye detection and face recognition in filtered NIR images. The algorithms can achieve high accuracies with high speed. The most important contribution is the methodology learned from the building of this successful system for how to make face recognition really work.

The present system has been tested for a real application of access control and time attendance. This is a scenario evaluation[15], an evaluation condition that is harder than a technology evaluation such as FRVT tests. The working conditions are under varying indoor locations and illumination conditions, with cooperative users. After a period of one month, the system has demonstrated excellent accuracy, speed, usability and stability under varying indoor illumination, even in the complete darkness. It has achieved an equal error rate below 0.3%.

The rest of the paper is organized as follows: Section 2 describes the design of the imaging hardware and presents an analysis of characteristics and amicability of resulting images for subsequent face processing. Section 3 describes the software part, including the feature representation, and the learning based methods for face/eye detection and face recognition. Section describes the system evaluation (Section 4).

2   Imaging Hardware

The goal of making the special-purpose hardware is to avoid the problems arising from environmental lighting, towards producing nearly idealized face images for face recognition. By the word "idealized", we mean that the lighting is frontal and of suitable strength. Environmental lighting is generally existing but from un-controlled directions and it is difficult to normalize it well by using an illumination normalization method. This is in fact a major obstacle in traditional face recognition.

To overcome the problem, we decide to use some active lights mounted on the camera to provide frontal lighting and to use further means to reduce environmental lighting to minimum. We propose two principles for the active lighting: (1) the lights should be strong enough to produce clear frontal-lighted face image but not cause disturbance to human eyes, and (2) the resulting face image should be affected as little as possible after minimizing the environmental lighting.

Our solution for (1) is to mount near infrared (NIR) light-emitting diodes (LEDs) on the hardware device to provide active lighting. When mounted on the camera, the LEDs provide the best possible straight frontal lighting, better than mounted anywhere else.

For (2), we use a long pass optical filter on the camera lens to cut off visible light while allowing NIR light to pass. The long pass filter is such that the wavelength points for 0%, 50%, 88%, and 99% passing rates are 720, 800, 850, and 880nm, respectively. The filter cuts off visible environmental lights (< 700nm) while allowing the NIR light (850nm) to pass.

As a result, this imaging hardware device not only provides appropriate active frontal lighting but also minimizes lightings from other sources. Figure 1 shows example images of a face illuminated by both frontal NIR and a side environmental light. We can see that the lighting conditions are likely to cause problems for face recognition with the conventional color (and black and white) images, the NIR images are mostly frontal-lighted by the NIR lights only, with minimum influence from the environmental light, and are very suitable for face recognition. The effect of remaining NIR component of environmental lights in the NIR image (such as due to the lamp light for making the example images) is much weak than that of the NIR LED lights.
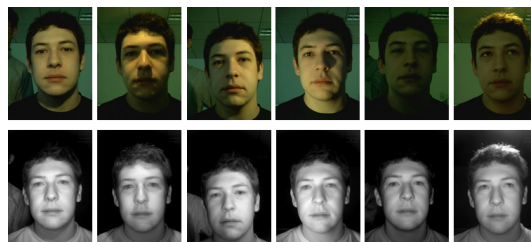

Fig. 1. Upper-row: 5 color images of a face. Lower-row: The corresponding NIR-filtered images.

3   Learning-Based Algorithms

Both detection and matching are posed as a two-class problem of classifying the input into the positive or negative class. The central problem in face/eye detection is to classify each scanned sub-window into either face/eye or non-face/eye; the positive sub-windows are post-processed by merging multiple detects in nearby locations. For face matching, the central problem is to develop a matching engine or a similarity/distance function for the comparison of two cropped face images. In this regard, we adopt the intrapersonal and extrapersonal dichotomy proposed in [12], and train a classifier for the two-class classification. The trained classifier outputs a similarity value, based on which the classification can be done with a confidence threshold.

3.1   Learning for Face/Eye Detection

A cascade of classifiers are learned from face/eye and non-face/eye training data. For face detection, an example is a 21x21 image, containing a face or nonface pattern. For eye detection, an example is a 21x15 image, containing an eye or noneye pattern. Sub-regions of varying sizes from 5×5 to 11×11 with step size 3 in both directions are used for computing the LBP histogram features for the local regions, which generates all possible features composed of all the 59 scalar features at all the locations.

Figure 2 show statistics on the training results. On the left shows the face and nonface distributions as functions of number of weak classifiers. We can see that the two classes are well separated, and a large number (more than 95% in the data) of nonface examples are rejected at the first two stages. The ROC indicates that the overall detection rate is 96.8% given the false alarm rate of 10^{-7}. On the right compares the ROC curves with that of the baseline algorithm of [18].
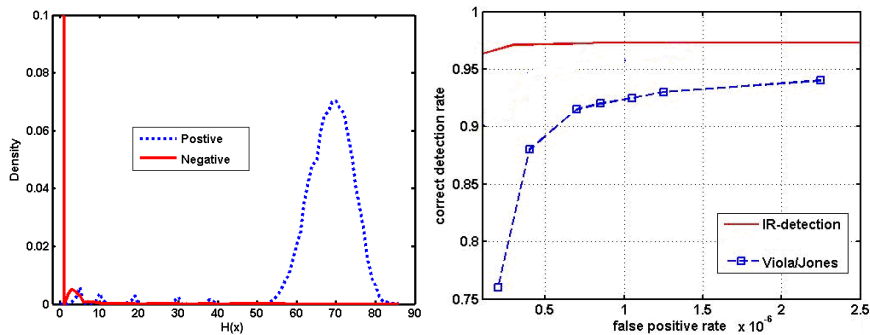

Fig. 2. On the left are the face (blue, dashed) and nonface (red, solid) distributions, and on the right compares the ROC curves of the IR face detection and visible light face detection of [18].

3.2 Learning for Face Recognition

Recently, the LPB representation has been used for face detection and recognition. In [1, 9], an input face image is divided into 42 blocks of size w by h pixels. Instead of using the LBP patterns for individual pixels, the histogram of 59 bins over each block in the image is computed to make a more stable representation of the block. The Chi-square distance is used for the comparison of the two histograms (feature vectors)

$$\chi^2(S, M) = \sum_{b=1}^{B} \frac{(S_b - M_b)^2}{(S_b + M_b)} \tag{1}$$

where S_b and M_b are to the probabilities of bin b for the corresponding histograms in the gallery and probe images and B is the number of bins in the distributions. The final matching is based on the weighted chi-square distance over all blocks.

We believe that the above scheme lacks optimality. First, a partition into blocks is not optimized in any sense and ideally all possible pixel locations should be considered. Second, manually assigning a weight to a block is not optimized. Third, there should be better matching schemes than using the block comparison with the Chi-distance.

Therefore, we adopt a statistical learning approach [19], instead of using a Chi-square distance [1, 9] and weighted sum of block matches for matching between two faces. The need for a learning is also due to the complexity of the classification. The classification here is inherently a nonlinear problem. An AdaBoost learning procedure [8] is used for these purposes, where we adopt the intrapersonal and extrapersonal dichotomy [12] to convert the multi-class problem into one of two-class. See [19] for more details of the methods.

Figure 3 shows the ROC curve for the present method obtained on a test data set, which shows a verification rate (VR) of 90% at FAR=0.001 and 95% at FAR=0.01. In comparison, the corresponding VR's for the PCA (with Mahalanobis distance) and LDA on the same data set are 42% and 31%, respectively, FAR=0.001; and 62% and 59% at FAR=0.01. (Note that it is not unusual that LDA performs worse than PCA [2].)
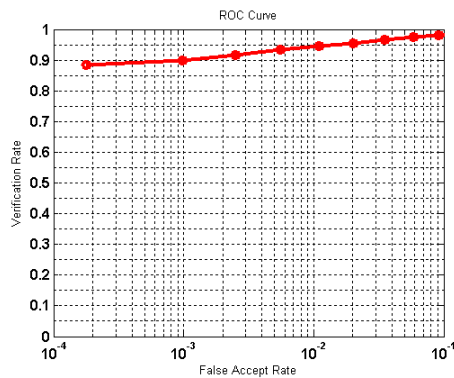

Fig. 3. ROC Curves for verification on a test data set

4   System Evaluation

Our tests are in the form of scenario evaluation [15], for 1-N identification in an access control and time attendance application in an office building. The participation protocol was the following: 1470 persons were enrolled under environmental conditions different from those of the client sites, with 5 templates per enrolled person recorded. Of these persons, 100 were workers in the building and most others were collected from other sources unrelated to the building environment. The 100 workers were used as the genuine clients while the others were used as the background individuals. On the other hand, additional 10 workers were used as the regular imposters, and some visitors were required to participate as irregular imposters. This provided statistics for calculating correct rejection rate and false acceptance rate.

The 100 clients and 10 imposters were required to report to the system 4 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. Some did more than 4 times a day. Some clients deliberately challenged the system by doing strange face or occluding the face with a hand, so that the system did not recognize them. We counted these as visitor imposter 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 results show that the system achieved an equal error rate below 0.3%. Hence, we conclude that the system has achieved high performance for cooperative face recognition.

5   Summary and Conclusions

We have presented a highly accurate and fast face recognition system for cooperative user applications. The novel design of the imaging hardware delivers face images amicable for face processing. The statistical learning procedures with local features give to highly accurate and fast classifiers for face/eye detection and face recognition. These, together with engineering inspirations, have made a successful system. Evaluated in real-world user scenario tests, the system has demonstrated excellent accuracy, speed and usability. We believed that this was the best system in the world for cooperative face recognition. The success is ascribed to two reasons: First, the classification tasks herein are made very easy with NIR images captured by the novel hardware device. Second, the learning based methods with the local features by their own are powerful classification engines.

Future work includes the following: The first is to study the performance of the matching engine for face matching after a long time-lapse, while the system has had no problem with faces previously seen about 8 months ago. The second is to improve the imaging hardware and processing software to deal with influence of NIR component in outdoor sunlight.

• *Two patents have been filed for the technology described in this paper.*

References

1. T. Ahonen, A. Hadid, and M.Pietikainen. "Face recognition with local binary patterns". In *Proceedings of the European Conference on Computer Vision*, pages 469–481, Prague, Czech, 2004.

2. J. R. Beveridge, K. She, B. A. Draper, and G. H. Givens. "A nonparametric statistical comparison of principal component and linear discriminant subspaces for face recognition". In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages I:535–542, 2001.

3. K. W. Bowyer, Chang, and P. J. Flynn. "A survey of 3D and multi-modal 3d+2d face recognition". In *Proceedings of International Conference Pattern Recognition*, pages 358–361, August 2004.

4. K. I. Chang, K. W. Bowyer, and P. J. Flynn. "An evaluation of multi-modal 2D+3D face biometrics". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, page to appear, 2005.

5. R. Chellappa, C. Wilson, and S. Sirohey. "Human and machine recognition of faces: A survey". *Proceedings of the IEEE*, 83:705–740, 1995.

6. CVBVS. In *IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications*, 1999-2003.

7. J. Dowdall, I. Pavlidis, and G. Bebis. "Face detection in the near-IR spectrum". *Image and Vision Computing*, 21:565–578, July 2003.

8. Y. Freund and R. Schapire. "A decision-theoretic generalization of on-line learning and an application to boosting". *Journal of Computer and System Sciences*, 55(1):119–139, August 1997.

9. A. Hadid, M. Pietikinen, and T. Ahonen. "A discriminative feature space for detecting and recognizing faces". In *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 797–804, 2004.

10. 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*, 97(1):103–135, January 2005.

11. D.-Y. Li and W.-H. Liao. "Facial feature detection in near-infrared images". In *Proc. of 5th International Conference on Computer Vision, Pattern Recognition and Image Processing*, pages 26–30, Cary, NC, September 2003.

12. B. Moghaddam, C. Nastar, and A. Pentland. "A Bayesain similarity measure for direct image matching". *Media Lab Tech Report* No.393, MIT, August 1996.

13. OTCBVS. In *IEEE International Workshop on Object Tracking and Classification in and Beyond the Visible Spectrum*, 2004-2005.

14. Z. Pan, G. Healey, M. Prasad, and B. Tromberg. "Face recognition in hyperspectral images". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(12):1552–1560, December 2003.

15. P. J. Phillips, A. Martin, C. L. Wilson, and M. Przybocki. "An introduction to evaluating biometric system". *IEEE Computer (Special issue on biometrics)*, pages 56–63, February 2000.

16. A. Samal and P. A.Iyengar. "Automatic recognition and analysis of human faces and facial expressions: A survey". *Pattern Recognition*, 25:65–77, 1992.

17. D. Valentin, H. Abdi, A. J. O'Toole, and G. W. Cottrell. "Connectionist models of face processing: A survey". *Pattern Recognition*, 27(9):1209–1230, 1994.

18. P. Viola and M. Jones. "Robust real time object detection". In *IEEE ICCV Workshop on Statistical and Computational Theories of Vision*, Vancouver, Canada, July 13 2001.
19. G. Zhang, X. Huang, S. Z. Li, Y. Wang, and X. Wu. "Boosting local binary pattern (LBP)-based face recognition". In S. Z. Li, J. Lai, T. Tan, G. Feng, and Y. Wang, editors, *Advances in Biometric Personal Authentication*, volume LNCS-3338, pages 180–187. Springer, December 2004.
20. W. Zhao and R. Chellappa. "Image based face recognition, issues and methods". In B. Javidi, editor, *Image Recognition and Classification*, pages 375–402. Mercel Dekker, 2002.
21. W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, pages 399–458, 2003.