AuthenMetric F1:
A Highly Accurate and Fast Face Recognition System

Stan Z. Li and His Face Team

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

AuthenMetric F1 *is a face-based biometric authentication system. Like Formula One (F1), the system is fast and accurate. Better than that F1, this F1 makes no disasters, hence can save money and human lives. As a demo at ICCV, it must have been made of scientific ingredients of computer vision, such as innovative imaging hardware, local feature extraction and statistical learning; it is further enhanced by engineering inspirations. The system is high in both speed and accuracy. A users can be identified quickly once he/she sees him/herself in a mirror or LCD, at a distance of an arm away from the camera. The performance is regardless of illumination changes. Interested? Come to user-experience the magic.*

Technically, we present a highly accurate, realtime face recognition system for co-operative 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 [19, 24, 6, 27]. This is partly due to recent technology advances initially made by work on eigenfaces [21, 23], and partly due to increased concerns in security. However, the problem of face recognition remains a great challenge after several decades of research.

Whereas the shape and reflectance are the intrinsic properties, the appearance of a face is affected by extrinsic factors, including illumination, pose and expression, as well as inaccuracies made in pre-processing stages such as face alignment. Variations brought about by extrinsic factors make individual face manifolds highly complex [3, 20, 22]; it is difficult for conventional methods to achieve high accuracy. Speed is also an important issue for many applications. Although progress has been encouraging since 1990's, the task remains a difficult endeavor, as clearly shown in recent Face Recognition Vendor Test [9], specially for un-cooperative user scenarios such as face recognition in visual surveillance.

To achieve high accuracy, the recognition should be performed based on intrinsic properties, and the algorithms should be able to deal with unfavorable influences due to extrinsic factors and mis-alignment. There two approaches to validate this assumption: by processing the face image and by minimizing extrinsic factors before the image is processed. The former is the approach that has been adopted by most of current research and has not been very successful. The latter is what we adopt and are presenting in this paper. Our assumptions are that the user is cooperative and that an application is in a moderate environment such as in door. These are valid for many useful applications.

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 [4, 28]). 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. [5].

Invisible imagery has recently received increased attention in the computer vision community, as seen from the IEEE workshop series [7, 16]. Thermal or far infrared imagery has been used for face recognition (*cf.* and a survey paper [11]). 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 [8, 12, 17]. In [8], 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 [12], facial features are detected by analyzing the horizontal and vertical projections of the face area, following a homomorphic-filtering pre-processing. In [17], face recognition is done using hyperspectral images captured in 31 bands over an NIR range of 0.7μm-1.0μm; invariant features are extracted using the fact that multi-band spectral measurements of facial skin differ significantly from person to person; and recognition is performed based on the invariant features.

In this demo, 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 camera hardware device consists of a CCD/CMOS USB camera, NIR light-emitting diode (LED) lights, and an optical filter. The NIR LEDs, mounted on the device, illuminate the face with frontal lights, with a strength that is stronger than expected environmental illumination. The system is further enhanced by highly accurate and efficient algorithms for face detection and recognition [26]. Local binary pattern (LBP) [15, 1] is used as the base features. AdaBoost learning [10, 25] is applied to learn most effective features and classifiers. The algorithms perform very well in both accuracy and speed with the filtered NIR images. Technical details is presented in [13].

The resulting **AuthenMetric F1** system delivers excellent results running on a P4 3GHz PC: The face and eye detection sub-system runs at 25ms per frame. The face recognition sub-system takes 40ms per input probe image for a database of 1000 persons, 5 images per person. It has so far performed nearly perfectly. For experienced users, The overall average speed is with 1 second per recognition session.

2 Imaging Hardware

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. 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.

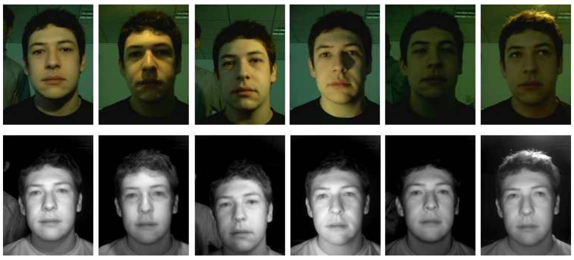

Fig. 1. Upper-row: 5 color images of a face. The lights are from different lamp lights as well as the NIR LED lights mounted on the camera. Lower-row: The corresponding NIR-filtered images. While unfavorable lightings are obvious in the color images, they are much reduced in the NIR-filtered face images. Only the NIR images were used for the recognition.

This imaging hardware device not only provides appropriate active frontal lighting but also minimizes lightings from other sources. Figure 6 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.

3 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.

In the training phase, a candidate weak classifier at a pixels location is made by thresholding on a scalar feature. In each iteration of AdaBoost learning, an optimal threshold is learned for each candidate feature to minimize the weighted training error with that feature, giving a weak classifier; the optimal weak classifier is learned by picking up the one with minimum overall weighted training error. After a strong classifier is learned, the subsequent strong classifier in the cascade is then learned with nonface examples booststrapped using the already learned cascade of classifiers.

For face, about $22,000$ NIR face and over 10^7 nonface examples are used. Given a target detection rate of 99% for each stage, the false alarm rate is reduced to below 10^{-7} with a trained cascade of 7 strong classifiers, composed a total of 87 weak classifiers. Similarly for eye detection, about $44,000$ NIR eye and $300,000$ noneye examples are used where the noneye examples are collected mainly from the upper-part of face images. The eye detector also has excellent separability between the eye and noneye

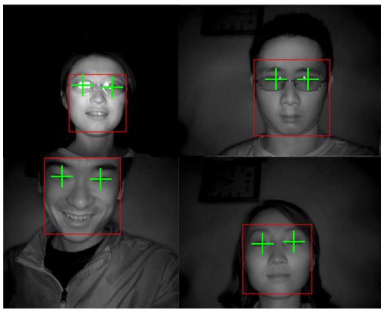

Fig. 2. Examples of face and eye detection in NIR images.

classes. Fig.2 shows some examples of face and eye detection. Figure 3 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 [25].

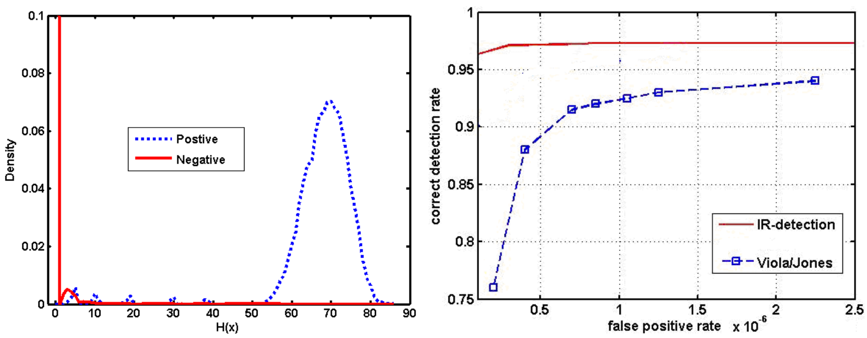

Fig. 3. On the left are the face (blue, dashed) and nonface (red, solid) distributions, and on the right compares the ROC curves of the IR face detection and visible light face detection of [25].

In the testing phase, the search for face detection is performed over the entire image, whereas that for eye detection is limited to a sub-region in the upper-part of the face rectangle. The detectors run very fast, at a speed of 32.3ms per frame on a P4 3.0GHz PC, for detecting the face and eyes in a 640x480 image.

4 Face Recognition

Face matching is a multi-class problem. One possibility is to train a classifier using positive training examples of each client and negative examples from all other clients. This would require a training process when a new client is added, which is an inconvenience because the training would takes time. Therefore, we adopt the intrapersonal and extrapersonal dichotomy [14] to convert the multi-class problem into one of two-class. The idea is to train a two-class classifier in the training phase, with intrapersonal and extrapersonal training examples. In this work, the differences are derived between LBP feature vectors, instead of doing differences between images as in [14]. This is important because doing image differences between images would lose important information in the LBP encoding. In the testing phase, the matcher compares two learned LBP feature vectors, calculate the similarity score, and make a decision whether they come from the same individual or not.

For face matching, a face image is of size $W \times H = 110 \times 120$, cropped based on the locations of the two eyes. The interior area is of size $W' \times H' = 94 \times 100$ pixels for sub-regions composed of pixels within an ellipse of "horizontal radius" of $R_W = 8$ and "vertical radius" of $R_H = 10$. An original LBP feature vector is a histogram of LBP codes computed for a sub-region. The $\text{LBP}^{u2}_{(8,1)}$ operator is used, and so the original feature vector for a face is of $94 \times 100 \times 59 = 53100$ dimensions.

An intrapersonal LBP feature vector is then derived as the difference between two original LBP feature vectors computed from a pair of two images of the same person, whereas an extrapersonal LBP feature vector is the difference between two vectors computed from a pair of images of two different persons, both being 53100 dimensional. In the AdaBoost training phase, the intrapersonal and extrapersonal training data are generated from all intrapersonal pairs and extrapersonal pairs, respectively. In the testing phase, the input is two face images, or two feature vectors derived therefrom; the LBP-difference is computed from the two vectors and sent to the trained classifier. The classifier outputs a similarity score and makes a decision to answer the question of whether or not the two images (or two original LBP feature vectors) belong to the same person.

In the training phase, the weak classifiers, strong classifiers and cascade are learned in a similar way to the training phase for face/eye detection. A cascade of classifiers are learned from intre- and extra-class training data. There are 10^4 face images of 1000 persons, 10 images each. A training set of about 45×10^3 intrapersonal and 5×10^7 extrapersonal examples are generated. The target false rejection rate is 1% for training a strong classifier. The resulting cascade consists of 10 strong classifiers with about 1800 weak classifiers. The false acceptance rate is reduced to below 10^{-7} with an accuracy of 94.4% on the training set. Figure 4 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].)

In the testing phase, the face and then eyes are detected; the face region is cropped according to the eye centers; the learned features are computed over the crop region and

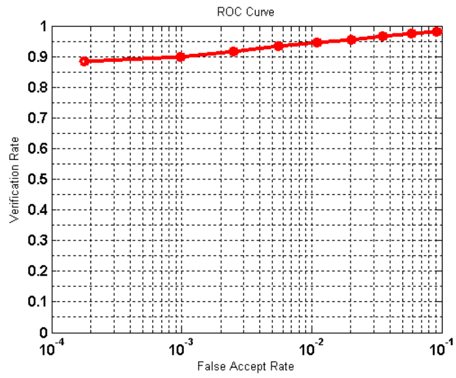

Fig. 4. ROC Curves for verification on a test data set.

then matched to each gallery feature vector; and finally the matching score is delivered. The matching engine runs at a speed of 56 ms per frame on the P4 3.0GHz PC, for a database of 1000 persons, 5 images per person.

5 Building the System

The total processing time for face/eye detection and face matching is less than 90ms on the P4 3.0GHz PC, excluding times used for getting images from USB cameras. With the very fast and powerful engines, we are able to build a highly accurate, realtime face recognition system, even if the detection rate and correct acceptance rate are not perfect. This is because we can afford to keep processing the incoming images until a match is found with a sufficient confidence. Using this simple strategy, we are able to make a successful identification for an enrolled person within 1 second for experienced users who know how to present their faces so as to be identified quickly. For first-time users, the speed is less than 3 second per person. The system uses a time-out strategy, say 5 seconds, to reject a user for a recognition session. A rejection is mostly due to unavailability of the personal information in the database (not enrolled). Few are due to insufficient cooperation of the user.

The present system has been tested for a real application of access control and time attendance. This is a scenario evaluation[18], 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%.

6 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 the novel imaging hardware device and the software algorithms derived based on local features and statical learning.

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. M. Bichsel and A. P. Pentland. “Human face recognition and the face image set’s topology”. *CVGIP: Image Understanding*, 59:254–261, 1994.
4. 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.
5. 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.
6. R. Chellappa, C. Wilson, and S. Sirohey. “Human and machine recognition of faces: A survey”. *Proceedings of the IEEE*, 83:705–740, 1995.
7. CVBVS. In *IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications*, 1999-2003.
8. J. Dowdall, I. Pavlidis, and G. Bebis. “Face detection in the near-IR spectrum”. *Image and Vision Computing*, 21:565–578, July 2003.
9. Face Recognition Vendor Tests (FRVT). *http://www.frvt.org*.
10. 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.
11. 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.
12. 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.
13. S. Z. Li and *et. al.* “Highly accurate and fast face recognition using near infrared images”. In *IAPR International Conference on Biometric*, Hong Kong, Jan 2006.
14. B. Moghaddam, C. Nastar, and A. Pentland. “A Bayesain similarity measure for direct image matching”. *Media Lab Tech Report* No.393, MIT, August 1996.
15. T. Ojala, M. Pietikainen, and D. Harwood. “A comparative study of texture measures with classification based on feature distributions”. *Pattern Recognition*, 29(1):51–59, January 1996.

16. OTCBVS. In *IEEE International Workshop on Object Tracking and Classification in and Beyond the Visible Spectrum*, 2004-2005.
17. 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.
18. 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.
19. A. Samal and P. A.Iyengar. "Automatic recognition and analysis of human faces and facial expressions: A survey". *Pattern Recognition*, 25:65–77, 1992.
20. P. Y. Simard, Y. A. L. Cun, J. S. Denker, and B. Victorri. "Transformation invariance in pattern recognition - tangent distance and tangent propagation". In G. B. Orr and K.-R. Muller, editors, *Neural Networks: Tricks of the Trade*. Springer, 1998.
21. L. Sirovich and M. Kirby. "Low-dimensional procedure for the characterization of human faces". *Journal of the Optical Society of America A*, 4(3):519–524, March 1987.
22. M. Turk. "A random walk through eigenspace". *IEICE Trans. Inf. & Syst.*, E84-D(12):1586–1695, December 2001.
23. M. A. Turk and A. P. Pentland. "Eigenfaces for recognition". *Journal of Cognitive Neuroscience*, 3(1):71–86, March 1991.
24. 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.
25. 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.
26. 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.
27. 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.
28. W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld. Face recognition: A literature survey. *ACM Computing Surveys*, pages 399–458, 2003.

The Team