A Near-infrared Image Based Face Recognition System*

Stan Z. Li, Lun Zhang, ShengCai Liao, XiangXin Zhu, RuFeng Chu, Meng Ao, Ran He
Center for Biometrics and Security Research & National Laboratory of Pattern Recognition
Institute of Automation, Chinese Academy of Sciences, Beijing, China
http://www.cbsr.ia.ac.cn

Abstract

In this paper, we present a near infrared (NIR) image based face recognition system. Firstly, we describe a design of NIR image capture device which minimizes influence of environmental lighting on face images. Both face and facial feature localization and face recognition are performed using local features with AdaBoost learning. An evaluation in real-world user scenario shows that the system achieves excellent accuracy, speed and usability.

1. Introduction

Face recognition has received significant attention during the past years[1, 2, 3, 4]. This is partly due to recent technology advances initially made by work on eigenfaces[5, 6], and partly due to its wide potential applications.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. Variations brought about by extrinsic factors make individual face manifolds highly complex [7, 8, 9]; It is difficult for conventional methods to achieve high accuracy,even in cooperative-user conditions, such as access control, machine readable travelling documents, computer login and ATM.

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 are two ways to validate this assumption: by processing the face image or 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 is 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.

To avoid the problem caused by illumination changes(and other changes), several solutions have been investigated into. One is to use 3D technique data obtained from a laser scanner or 3D vision method[10, 11]. For real-world applications, the mainly disadvantage are intensive computation and high capture device price. The other is to use invisible imagery. Recognition of faces using infrared imaging sensors has become an area of growing interest[12, 13]in recent years. Thermal infrared technique have been used in face recognition system, which have advantages in face detection, detection of disguised faces, and face recognition under poor lighting conditions[14]. However, thermal infrared is not desirable because of the higher cost of thermal sensors and its instability in different temperature. Whereas near infrared (NIR) has attracted more and more attention due to its preferable attribute and low cost[15, 16, 17], which is also adopted in this paper. In [15], 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 [16], facial features are detected by analyzing the horizontal and vertical projections of the face area, following a homomorphic-filtering pre-processing. However, in these methods, the physical properties of facial features are not taken into account. As a matter of fact, thick spectacle frames or high specular reflection would make the accuracy decrease drastically. 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 paper, we present an NIR image based face recognition system. The main contributions are in the following:

*This work was supported by the following funding: Chinese National Natural Science Foundation Project 60518002, and Chinese National 863 Program Projects 2004AA1Z2290 & 2004AA119050.

First, we employ a novel design of image capture device to obtain filtered NIR images containing most relevant, intrinsic information for face detection and recognition, with extrinsic factors minimized. Then we construct a facial feature detection system based on local feature representation and tree-like boosting classifier that can precisely localize face and eye in NIR images with very high speed, even specular reflectance on eyeglasses. Finally, the recognition engine is built upon a learning-based procedure, with Local binary patterns [18, 19] features and an AdaBoost classifier. An evaluation in real-world user scenario shows that the system can achieve excellent accuracy, speed and usability.

The rest of the paper is organized as follows: Section 2 describes the design of the imaging hardware. Sections 3 describes the LBP representation and AdaBoost learning method. Section 4 presents the system, including facial feature localization and face recognition modules. Section 2 describes the result of system evaluation.

2. Hardware Design

In order to avoid the problems arising from variable illumination, we mount some near infrared (NIR) light-emitting diodes(LEDs)on the hardware device to provide frontal, active lighting. Then we use a long pass optical filter on the camera lens to cut off visible light while allowing NIR light to pass. The wavelength of the NIR LEDs is 850nm, most of CCD or CMOS sensors have sufficient response to this spectrum. After clearly adjust the strength and positon of the NIR active lighting, we can achieve nearly "idealized" face images for recognition. Figure 1shows example images of a face illuminated by both frontal NIR and side environmental light. We can see that the quality of face images is barely influenced by the environmental light. This will be of great benefit to face recognition.

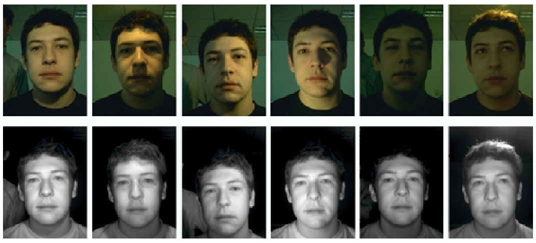

Figure 1. Upper-row:5 color images of a face. Lower-row:The corresponding NIR-filtered images.

An NIR face image captured from our device is amicable for face detection and recognition because it delivers intrinsic, most relevant information about a face for person identification, without the need for complicated preprocessing. 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 at surface point (x, y), $\mathbf{n} = (n_x, n_y, n_z)$ is the surface normal (a column vector) in the 3D space, and $\mathbf{s} = (s_x, s_y, s_z)$ is the lighting direction (with magnitude). Multiple point lights may be approximately linearly combined into a single $\mathbf{s}$. Here, albedo $\rho(x, y)$ is the photometric property facial skin and hairs, $\mathbf{n}(x, y)$ does the geometric shape of the face; the product of $\rho(x, y)\mathbf{n}(x, y)$ well encodes the intrinsic information about the face at a fixed pose.

When the lighting of the face is in the frontal direction as a result of the hardware design, the image is approximated as

$$I(x, y) \propto \rho(x, y)n_z(x, y) \tag{2}$$

It is subject to a multiplying constant due to a change in the absolute intensity of the lighting due to a variation in the distance between the face and the lights (however, such variations can be easily coped with by using the LBP representation, as will be detailed later). This much simplifies the subsequent processing tasks such as face detection, facial feature detection and thereby face recognition.

3. Face Processing Algorithms

3.1. Local Feature Representation

We adopt local binary patterns to encode the local micro-structures of images. The original LBP operator, introduced by Ojala [18], is a simple yet powerful method for texture analysis. The basic form is illustrated in Fig.2. The binary bits describing a local 3x3 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 anti-clockwise. 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.

Obviously, the LBP code representation is invariant to any monotonic transformation on pixel values because the code does not change the ordering relationships between pixel values. Therefore, it is very suitable for representing faces which are illuminated from a fixed direction but with the lighting intensity changes. Several extensions to the basic form of LBP can be made, including using multi-scale neighborhoods, and using a "uniform" subset of the LBP string [20]. The LBP operator at a scale can be denoted as LBP$_{(P,R)}$ where R is the radius of the circle surrounding the

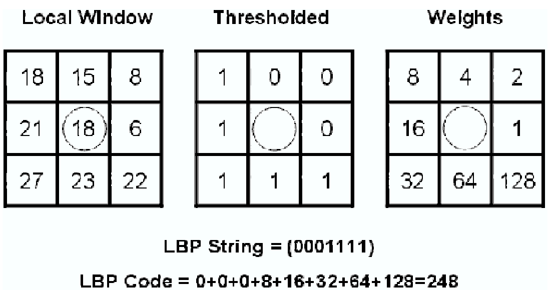

Figure 2. The LBP code for a 3x3 window.

center, and P is the number of pixels on the circle. Varying R and P gives rise to multi-scale LBP operators.

The uniform subset is denoted by $\text{LBP}^{u2}_{(P,R)}$. Considering neighboring relations between bits in the circular sense, an LBP string is called uniform if it contains at most 2 bitwise transitions from 0 to 1 or vice versa (see [20] for details). For example, (0001111) and (11111111) are uniform LBP patterns whereas (0001101) is not. There are a set of 58 possible uniform LBP patterns. All the non-uniform ones can be represented by a wildcard code which we denote as code 0. This way, the use of the $\text{LBP}^{u2}_{(8,1)}$ operator results in a set of 59 LBP patterns, denoted by $= \{0, 1, 2, \ldots, 58\}$. It is noted that the uniform subset accounts for about 90% of all configurations for the (8,1) neighborhood and for 70% for the (16,2) neighborhood. The extensions have shown to produce excellent results for texture related classification in terms of accuracy and computational complexity in many empirical studies [20].

Recently, the LBP representation has been used for face detection and recognition [19, 21]. For face detection [21], a 19x19 face image is divided into 9 overlapping regions of 10x10 pixels. From each region, a 16-bin histogram is computed using the $\text{LBP}_{4,1}$ operator, results into a single 144-bin histogram. The $\text{LBP}^{u2}_{8,1}$ operator is applied to the whole 19x19 face image, resulting in a 59-bin histogram. The two histograms are concatenated to form a (59+144=203)-bin histogram as a face representation. An SVM classifier is then trained with face and nonface examples represented in the 203-Dim feature vectors.

For face recognition [19, 21], 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)

$$\chi^2(S, M) = \sum_{b=1}^{B} \frac{(S_b - M_b)^2}{(S_b + M_b)} \qquad (3)$$

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.

3.2. AdaBoost Learning

We extract a feature vector for each interior location in a face/eye image. The $\text{LBP}^{u2}_{8,1}$ is used as the base operator. For each interior pixel location, a histogram is computed over a local sub-region of radius R centered at that pixel. For the interior area of size $W' \times H'$ for a face image of size $W \times H$, the feature vector is of $D = W' \times H' \times 59$ dimensions (determined by both pixel locations and the LBP pattern types), each entry is a number of counts for the occurrences of the corresponding LBP pattern in the local region.

We use the LBP histograms at all interior pixels to describe a face (and nonface/noneye in the case of face/eye detection). There are $D = W' \times H' \times 59$ dimensions in this initial feature set. However, not all the local regions are useful or equally useful, nor are so for the all the LBP patterns (histogram bins). Also, manually assigning weights to different blocks or locations is not optimized as mentioned in the above; especially when there are a large number of possible regions which we start with, it is intractable to manually assign weights to them. Therefore, instead of using a Chi-square distance [19, 21] and weighted sum of block matches for matching between two faces, we adopt a statistical learning approach. 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 [22] is used for the selection of most effective features and construction of classifiers for face/eye detection and face matching. The learning procedure performs two tasks: (1) learning to select the most effective features and (2) thereby construct a best classifier in some sense.

4. Building the System

4.1. Face and Eye Detection

For face detection, an example is a 21x21 image, containing a face or nonface pattern. The training set consisted of $43,000$ frontal upright NIR faces images, about 40% of them with glasses and 15% with eye closed. 10^7 nonface examples are used. 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. The candidate weak classifier at a pixels location is made by thresholding on a scalar feature, and a cascade of strong classifiers is trained using AdaBoost [23]. Given a detection rate of 99%, the false alarm rate is reduced to below 10^{-7} with a trained cascade of 7 strong classifiers, composed of 87 weak classifiers.

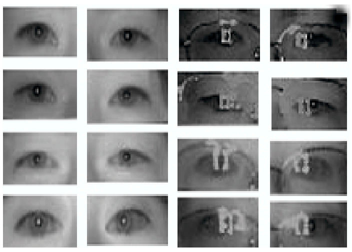

Figure 3. Eye examples can be grouped into four subclasses

Similarly for eye detection, about $86,000$ eye and $3,000,000$ noneye examples are used where the noneye examples are collected mainly from the upper-part of face images. The examples are 21x15 images. Eye localization with glasses is more difficult than face detection when specular reflection is present on glasses. A single eye detector for all cases, with and without glasses, is often overstrained. We therefore group the eye patterns into four categories, left and right eyes, with and without glass, shown in Figure 3. Then we train multiple specialized eye detectors using corresponding group of examples.

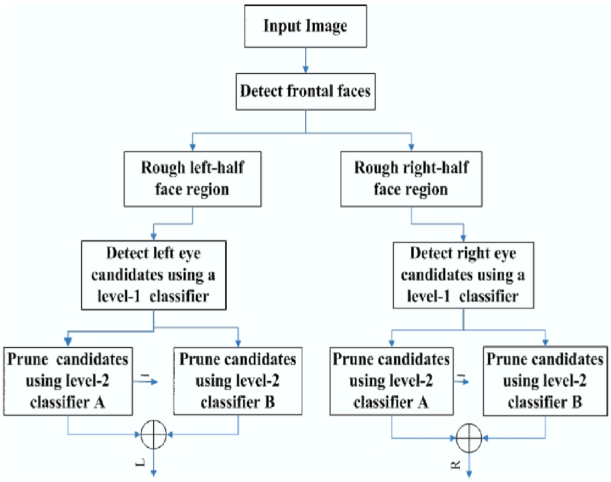
Input Image
Detect frontal faces
Rough left-half face region
Rough right-half face region
Detect left eye candidates using a level-1 classifier
Detect right eye candidates using a level-1 classifier
Prune candidates using level-2 classifier A
Prune candidates using level-2 classifier B
Prune candidates using level-2 classifier A
Prune candidates using level-2 classifier B
L
R

Figure 4. Flowchart of tree-like classifier

This suggests a classifier tree, shown in Figure 4, using a coarse-to-fine strategy, where every detector is a cascade of strong classifiers. First, a coarse (level-1) eye detector is applied. This is a cascade of 4 strong classifiers, trained with all types of eye examples. It consists of only 25 weak classifiers, but can detect all possible eye candidates, while rejecting more than 95% non-eye examples. Then eye candidates are then passed to a specialized no-glass eye detector (classifier-A) at level-2. It works accurately with a cascade of 6 strong classifiers. If failed, the subwindow is then passed to a specialized eye-with-glass detector (classifier-B) trained using examples of eyes with glasses. The result of the latter two eye detectors is finally fused to make the final decision for the eye locations. Figure 5 shows some examples of face and eye detection.

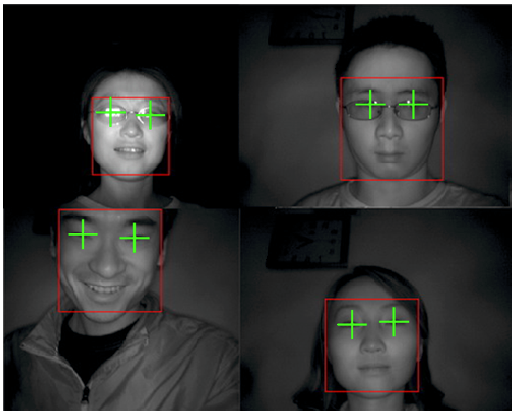

Figure 5. Examples of face and eye detection in NIR images.

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.

4.2. 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 intra-personal and extra-personal dichotomy [24] 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 intra-personal and extra-personal training examples. In this work, the differences are derived between LBP feature vectors, instead of doing differences between images as in [24]. 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 $\text{LBP}_{(8,1)}^{u2}$ operator is used, and so the original feature vector for a face is of $94 \times 100 \times 59 = 554600$ dimensions.

An intra-personal 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 extra-personal LBP feature vector is the difference between two vectors computed from a pair of images of two different persons, both being 554600 dimensional. In the AdaBoost training phase, the intra-personal and extra-personal training data are generated from all intra-personal pairs and extra-personal 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.

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 intra-personal and 5×10^7 extra-personal 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 6 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.

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 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.

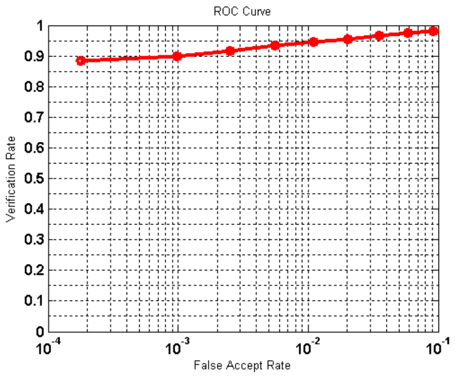

Figure 6. ROC Curves for verification on a test data set.

5. System Evaluation

The system evaluation was in the form of scenario evaluation [25], for 1-N identification in an access control and time attendance application in the CASIA office building. The participation protocol was the following: 870 persons were enrolled, with 5 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 persons while the other 840 were used as background persons. 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 persons 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 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, whereas some did more than 4 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.

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 summarized as follows:

- 1725 client sessions were performed, with 1721 successful acceptances and 4 false rejections, meaning a success rate of 99.77%;
- 1096 imposters (visitors included) were performed, with 3 false acceptances and 1093 rejections, meaning a success rate of 99.72%;
- There were 840 background persons who did not participate in the tests, none of those were matched to.

From the statistics, we can see that the equal error rate was below 0.3%. Hence, we conclude that the system has achieved high performance for cooperative face recognition.

6. Conclusion and Future Work

This paper has described a highly accurate face recognition system using near infrared images. The novel NIR image hardware provides face images invariant to illumination, a good basis for face recognition. The coarse-to-find face and eye detector tree can detect face and eyes accurately in high speed, even with specular reflectance on eyeglasses due to active NIR illumination. The face recognition system achieves high accuracy for cooperative face recognition in indoor environments. The evaluation in real-world user scenario demonstrated excellent accuracy, speed and usability of the system.

Future work includes the following: The first is to improve the imaging hardware and recognition engine to deal with influence of NIR component in outdoor sunlight. The second is to improve tree-like classifier to construct the tree automatically. The third is to study the performance of the matching engine in mobile devices.

References

[1] A. Samal and P. A. Iyengar. Automatic recognition and analysis of human faces and facial expressions: A survey. *Pattern Recognition*, 25:65C77, 1992.

[2] D. Valentin, H. Abdi, A. J. OToole, and G. W. Cottrell. Connectionist models of face processing: A survey. *Pattern Recognition*, 27(9):1209C1230, 1994.

[3] R. Chellappa, C. Wilson, and S. Sirohey. Human and machine recognition of faces: A survey. *Proceedings of the IEEE*, 83:705C740, 1995.

[4] W. Zhao and R. Chellappa. Image based face recognition, issues and methods. In B. Javidi,editor, *Image Recognition and Classification*, pages 375C402. Mercel Dekker, 2002.

[5] L. Sirovich and M. Kirby. Low-dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America A*, 4(3):519C524, March 1987.

[6] M. A. Turk and A. P. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, 3(1):71C86, March 1991.

[7] M. Bichsel and A. P. Pentland. Human face recognition and the face image sets topology. CVGIP: Image Understanding, 59:254C261, 1994.

[8] 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.

[9] M. Turk. A random walk through eigenspace. *IEICE Trans. Inf. & Syst.*, E84-D(12):1586C 1695, December 2001.

[10] 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 358C361, August 2004.

[11] W. Zhao, R. Chellappa, P. Phillips, and A. Rosenfeld. "Face recognition: A literature survey". *ACM Computing Surveys*, pages 399C458, 2003.

[12] J. Wilder, P.J. Phillips, C. Jiang, and S. Wiener "Comparison of visible and infra-red imagery for face recognition". *Proceedings of the Second International Conference on Digital Object Identifier*, Pages:182-187, 1996

[13] D. Socolinsky, A. Selinger, J. Neuheisel. "Face recognition with visible and thermal infrared imagery". *Computer Vision and Image Understanding*, pages:72C114, 2003

[14] 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):103C135, January 2005.

[15] J. Dowdall, I. Pavlidis, and G. Bebis. Face detection in the near-IR spectrum. *Image and Vision Computing*, 21:565C578, July 2003.

[16] 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 26C30, Cary, NC, September 2003.

[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):1552C1560, December 2003.

[18] T. Ojala, M. Pietikainen, and D. Harwood. A comparative study of texture measures with classification based on feature distributions. *Pattern Recognition*, 29(1):51C59, January 1996.

[19] T. Ahonen, A. Hadid, and M. Pietikainen. Face recognition with local binary patterns. *In Proceedings of the European Conference on Computer Vision*, pages 469C481, Prague, Czech, 2004.

[20] T. Ojala, M. Pietikainen, and M. Maenpaa. Multiresolution gray-scale and rotation invariant texture classification width local binary patterns. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971C987, 2002.

[21] 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 797C804, 2004.

[22] 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):119C139, August 1997.

[23] P. Viola and M. Jones. Robust real time object detection. newblock *In IEEE ICCV Workshop on Statistical and Computational Theories of Vision, Vancouver*, Canada, July 13 2001.

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

[25] P. Jonathon Phillips, A. Martin, C. L. Wilson, and M. Przybocki, "An introduction to evaluating biometric system",' *IEEE Computer (Special issue on biometrics)*, pp. 56–63, February 2000.