

Part-based Face Recognition Using Near Infrared Images

Ke Pan[†] Shengcai Liao[‡] Zhijian Zhang[†] Stan Z. Li[‡] Peiren Zhang[†]

[†] University of Science and Technology of China
Hefei 230026, China

[‡] Center for Biometrics and Security Research & National Laboratory of Pattern Recognition
Institute of Automation, Chinese Academy of Sciences, Beijing 100080, China

<http://www.cbsr.ia.ac.cn>

Abstract

Recently, we developed NIR based face recognition for highly accurate face recognition under illumination variations [10]. In this paper, we present a part-based method for improving its robustness with respect to pose variations. An NIR face is decomposed into parts. A part classifier is built for each part, using the most discriminative LBP histogram features selected by AdaBoost learning. The outputs of part classifiers are fused to give the final score. Experiments show that the present method outperforms the whole face-based method [10] by 4.53%.

1. Introduction

Face recognition is one of the most prominent areas in computer vision for several decades. It has attracted much attention due to its potential application values as well as theoretical challenges. Numerous face recognition methods have been proposed, such as PCA [17] and LDA [3]. Current systems can do fairly accurate recognition under constrained scenarios using these face recognition methods. However, when the scenarios changes, such as head rotations, illumination variations, and facial expressions, the recognition becomes more difficult.

Recently, Li *et al.* [16, 10] have developed a method and system for illumination invariant face recognition using near infrared images. They build an active near infrared imaging system that is able to produce face images of good condition regardless of visible lights in the environment. They further show that the resulting face images encode intrinsic information of the face, subject only to a monotonic transform in the gray tone, thus combining with Local Binary Pattern (LBP) features to compensate for the monotonic transform, they derive an illumination invariant

face representation. Using this system, high accurate face recognition can be achieved, with the only difficulty of pose variations and facial expressions.

To deal with the above scenarios changes, several component-based face recognition methods have been proposed. Heisele *et al.* [7] introduce one component-based method. Facial components are extracted and combined into a single feature vector. Then the feature vector is classified by a Support Vector Machine. The component-based method is compared with two comparable global systems. The two global systems recognize faces by classifying a single feature vector consisting of the gray values of the whole face image. A variation of that system consists of a set of viewpoint-specific SVM classifiers and involves clustering during training. The component-based method is shown more robust than the two global approach with respect to face rotation up to 40°. In a later paper [8], Ivanov *et al.* improve their previous component-based method [7] and obtain a better result by using use Support Vector Machine as the classifier for each component. Then, the output of the classifiers are fused using vote, sum rule and product rule.

Weyrauch *et al.* [18] propose a component-based face recognition method using 3D morphable models for pose and illumination invariant face recognition. 3D morphable model is used to compute 3D face models from three input images in database. The 3D models are rendered under varying pose and illumination conditions to build a large set of synthetic images. Then, the component-based face recognizer is trained using these images. A component-based face detector is used to detect the face and extract components from the face. This system is much robust than the holistic face recognition system.

Timo Ahonen *et al.* [2] extract the LBP texture features from each face region (part) to form a enhanced facial feature vector. A classifier is built from LBP histogram fea-

tures for each region. Weights are obtained manually to fuse the part-based classifiers.

In this paper, we present a part-based face recognition method for near infrared (NIR) based face recognition. Following our previous approach [10], Subwindow Local Binary Pattern Histograms (SLBPHs) are extracted from a near infrared face image to describe the face therein. Static AdaBoost learning is applied to analyze the discrimination power of the SLBPHs at all facial locations. A face is then decomposed into several facial parts, such as eyes, nose, mouth, and their combinations, according to the analysis result. Then a part classifier is built for each part using boosted SLBPH features, and the outputs of part classifiers are fused to give the final score. We apply max rule, sum rule, and LDA-based sum rule for score fusion, and analyze their effects on performance. Experiments show that the part-based method outperforms the state-of-the-art NIR face recognition method [10] by 4.53%.

The rest of the paper is organized as follows: In Section 2, we introduce an illumination invariant face representation using LBP filtering on NIR images. Feature selection and classifier learning is presented in Section 3, and facial part decomposition and classifier fusion is shown in Section 4. In Section 5, we demonstrate and analysis the experimental results. Finally, we summarize this paper in Section 6.

2. Illumination Invariant Face Representation

Li *et al.* [10] have recently developed a method and system for illumination invariant face representation using near infrared images. They build an active near infrared imaging system that is able to produce front-lighted face images regardless of visible lights in the environment. It is further shown that the active NIR imaging is subject mainly 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. They use local binary pattern (LBP) features [13, 1, 6] to compensate for the monotonic transform in the NIR images because the ordering relationship between pixels and hence LBP code are not changed by any monotonic transform. Therefore, LBP code generated from an active NIR face image gives rise to an illumination invariant representation of the face.

2.1. Active NIR Imaging

In [16, 10], an active NIR imaging system is designed to overcome the problem arising from uncontrolled environmental lights so as to produce face images of a good illumination condition for face recognition. The good illumination condition means that the lighting is from the frontal direction and the image has suitable pixel intensities, *i.e.* having good contrast and not saturated.

In the imaging system, active lights in the near infrared

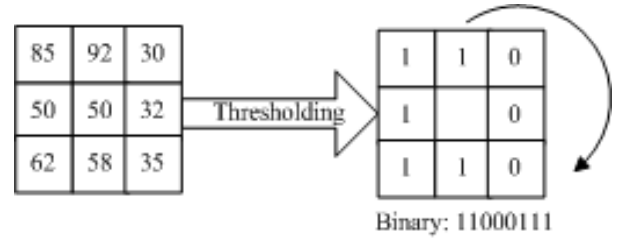


Figure 1. The Basic LBP Operator.

(NIR) spectrum between 780-1100nm are mounted on the camera. NIR light-emitting diodes (LEDs) are used as active radiation sources, which are strong enough for indoor use and are power-effective. A convenient wavelength is 850nm. Such NIR lights are almost invisible to human eyes, yet most CCD and CMOS sensors have sufficient response at this spectrum point. The strength of the total LED lighting should be as strong as possible, at least stronger than expected environmental illumination, yet does not cause sensor saturation. A concern is the safety to human eyes. When the sensor working in the normal mode is not saturated, the safety is guaranteed.

Furthermore, the system uses a long pass optical filter to minimize the environmental lighting. In this case, visible light is cut off while having minimum reduction of the intended active lighting.

2.2. Local Binary Pattern

As shown in [10], the above imaging system produces NIR images that contain most relevant, intrinsic information about a face, subject only to multiplying constant or a monotonic transform due to lighting intensity changes. Then LBP operator can be further applied to compensate the monotonic transform to achieve an illumination invariant representation of faces for indoor face recognition applications.

LBP is originally introduced by Ojala [13]. It's a powerful descriptor for texture information. The operator labels the pixels of an image by thresholding the 3×3 neighborhood of each pixel with the center value and considering the result as a binary number. An illustration of LBP operator is shown in Fig. 1.

Uniform LBP is an extension of the basic local binary patterns [14]. A local binary pattern is called uniform if the binary pattern contains at most two bitwise transitions from 0 to 1 or vice versa when the bit pattern is considered circular. We use $LBP_{8,1}^{u_2}$ operator in this paper, which is the basic uniform LBP. Base on this operator, Subwindow LBP Histogram (SLBPH) is used to provide dissimilarity measures in [10]. In their method, a histogram of the base LBP codes is computed over a local region centered at each pixel, with each histogram bin being the number of occur-

rences of the corresponding LBP code in the local region. There are 59 bins for $LBP_{8,1}^{u2}$ operator, thus a uniform LBP histogram is considered as a set of 59 individual features. We adopt SLBPH as feature representation in this paper.

3. Learning Optimal Features and Classifier

Though SLBPH features construct a powerful representation, considering all scales and locations, the complete feature set is very huge, which contains much redundancy. Thus it is important how to select efficient SLBPH features to construct face classifiers. In this paper, we apply AdaBoost to select the most significant features and learn a strong classifier for each facial part.

The basic AdaBoost [4] is for two class problems. A training set of N labeled examples is given for two classes, $\mathbf{S} = (\mathbf{x}_1, \mathbf{y}_1), \dots, (\mathbf{x}_N, \mathbf{y}_N)$, where x_i is a training example 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

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

where $\alpha_t \in \mathbb{R}$ 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 to derive α_t and $h_t(x)$ so that an upper error bound is minimized [5]. The procedure keeps a distribution $w_t = (w_{t,1}, \dots, w_{t,N})$ 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.

While an AdaBoost procedure essentially learns a two-class classifier, we convert the multi-class problem into a two-class one using the idea of intra- and extra-class difference [12]. However, here the difference data are derived from the SLBPH features rather than from the images. A difference is taken between two SLBPH feature sets, which is intra-class if the two face images are of the same person, or extra-class if not. to construct weak classifier for the above AdaBoost learning, we generate samples using histogram bin difference as dissimilarity measure between SLBPH features of two faces.

4. Fusion of Part-based Classifiers

To achieve a high accurate face recognition system, more effort should be made to overcome some scenario changes, such as pose and expression variations. In this paper, we propose a part-based solution to this aim. Firstly, according to a statistical analysis, we decompose the NIR face image



Figure 2. The first 100 subwindows of SLBPH feature learned by AdaBoost. Only the highlighted facial parts are contributed in face recognition, and the brightest parts indicate the most discriminative power.

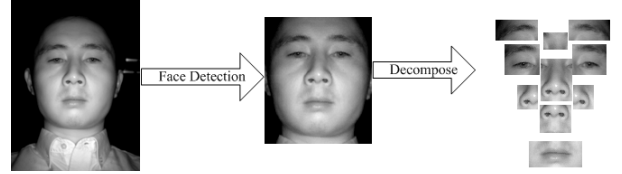


Figure 3. Face detection, normalization and decomposition.

into several parts, such as eyes, noses, and mouth. Then, AdaBoost learning is applied to each part to build a part-based classifier, and finally the output of each part classifier is fused to give the final score.

4.1. Facial Part Decomposition

Human face consists of several different parts, such as eyes, nose, and mouth. Earlier researches show that each facial part has a different contribution in face recognition [8, 2]. In this paper, we propose a statistical learning method to analyze the discriminant power of each facial part. Using SLBPH as feature representation, AdaBoost [4] can be further applied to learn the most discriminative features for face recognition. Since each learned SLBPH feature covers a subregion of the whole face, areas covered by more SLBPH features have more importance in face recognition.

We select the first 100 SLBPH features learned by AdaBoost to demonstrate how facial parts contribute in face recognition. Given each subwindow of SLBPH feature with equal gray intensity, we further overlap them all in the same face according to their locations. Then the brightest parts indicate the most discriminative power for face recognition. The result is shown in Fig. 2.

According to the above analysis, we decompose the whole face into several facial parts. When a new image comes into our system, we first detect the face, and normalize it to a fixed size according to the eye coordinates. Then, we decompose the face into several different parts. The process is shown in Fig. 3.

When we decompose the face into parts, the size of subwindows are restricted by the size of each part. The maximum size of the subwindows is the size of the biggest part. It can be seen that, in whole face, there are some big subwindows that are larger than any decomposed part. These

subwindows are definitely lost if we only consider the subwindows in each part. We can also find out that some subwindows covers more than one part. If we only consider the subwindows of each part, these subwindows will also be lost. To make it up, we consider the whole face as one big facial part.

4.2. Classifier Fusion

In this system, a strong AdaBoost classifier is learned for each facial part, so that each part classifier gives out a score. The final result is reached by fusions of all part classifiers. Before fusion, we apply Z score normalization to each output of part classifiers, so that they have zero mean and unit standard variance individually.

Kittler et al. [9] have developed a theoretical framework for consolidating the evidence obtained from multiple classifiers using schemes like the sum rule, product rule, max rule, min rule, median rule and majority voting. In this paper, we use sum rule and max rule to do score fusion and compare the performances.

The max rule and sum rule are used for score fusion in this paper. The max rule approximates the mean of the posterior probabilities by the maximum value, so the final score is given by

$$s = \max_i F_i(\vec{x}) \quad (2)$$

where F_i is the i th part classifier.

The sum rule assumes that the posterior probabilities computed by the individual classifiers do not deviate much from the prior probabilities. In this case, the final score is calculate as

$$s = \sum_{i=1}^N F_i(\vec{x}) \quad (3)$$

where F_i is the i th part classifier, and N is the number of facial parts.

We also apply a weighted sum fusion, which is given out by

$$s = \sum_{i=1}^N w_i * F_i(\vec{x}) \quad (4)$$

where w_i is the weight for the i th part classifier.

To learn optimal weights w_i for fusion, we apply Linear Discriminant Analysis (LDA) [11, 15] on training data. LDA learns an optimal linear subspace, where each class is best separated. Since we treat face recognition as a two-class problem, the resulting projection matrix is actually a vector, with coefficients corresponding to optimal weights of each part classifier for fusion.

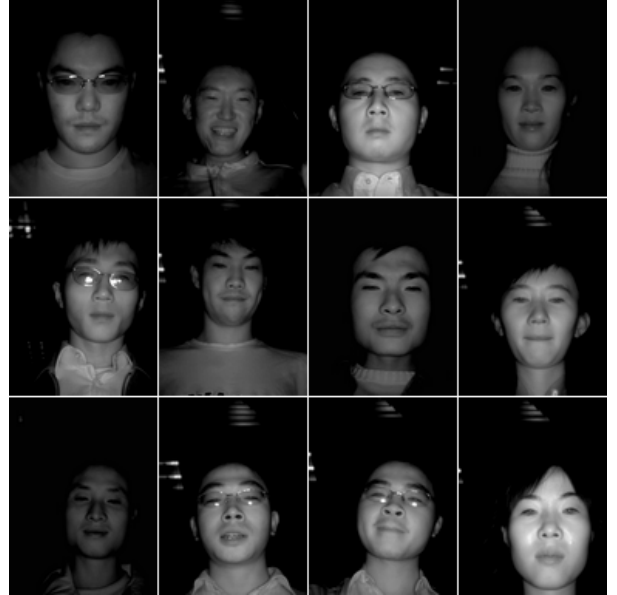


Figure 4. Examples of near infrared images.

5. Experiments

5.1. Database

The near infrared image database used in our experiments are captured by AuthenMetric F1 which is described in [16, 10]. The captured images are 480×640 pixels. Fig 4 shows some examples of the database. In our experiments, all face images are cropped into 142 pixels height and 120 pixels wide, according to their eye coordinates.

There are 10^4 face images of about 1000 persons, 10 images each person in the database, all Chinese. We select a subset of the database for experiments. The selected images are split into two sets, training set and test set, in the same way as in [10]. In the training phase, the training set of positive examples were derived from intra-personal pairs of SLBPH features, the negative set from extra-personal pairs. A training set of about 45×10^3 positive and 5×10^7 negative examples were collected from the training images. A technology evaluation was done with a test set of 3237 images. The test set contained 35 persons, with 80 to 100 images per person. None of the test images were in the training set. This generated 149,217 intra-class (positive) and 5,088,249 extra-class (negative) pairs.

5.2. Training Part-Classifiers

There are huge number of features and samples in the training set. If we use all the features and samples for training in one time, the classifiers are not actually trainable. To tackle such a problem, we adopt a two level training scheme. Level 1 is designed for feature selection. In this procedure, we randomly split the feature set of each fa-

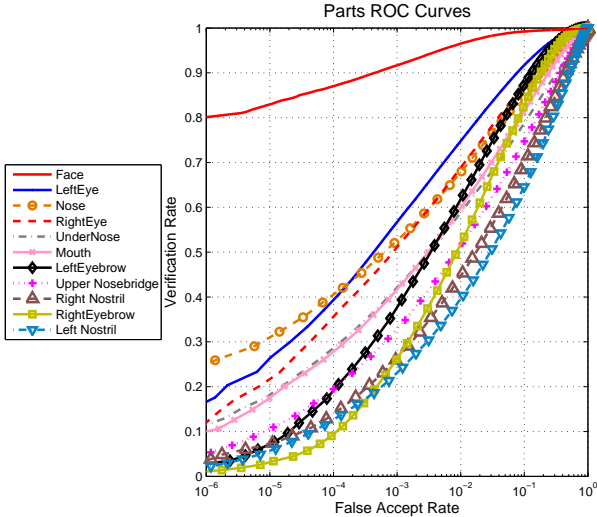


Figure 5. ROC curves of part classifiers

cial parts into 2 or 3 subsets, which contains about 20,000 features each. Then, training samples are randomly bootstrapped, with extra samples 8 to 12 times more than intra samples. In the end, features selected by AdaBoost are collected together for level 2 training. In level 2, since the number of features is dramatically reduced during level 1, so much more samples can be used to train a final strong classifier by AdaBoost.

5.3. Results

The ROC curves for the performances of all facial parts are shown in Fig. 5. From the result we can see that, discarding holistic face, the left eye and the nose part has the best performance in all facial parts. Remember that in Fig. 2, these parts are also the brightest ones. The results shows that our facial part decomposition according to the statistical analysis is reasonable.

We observe an interesting phenomena in Fig. 5: all the left side of the same part has a better performance than the right side except nostril part. Maybe this is due to the non-symmetric distribution of the training face images.

From the figure we can also see that the nostril part and the upper nose bridge part get the worse performance, which is observed similar in Fig. 2. Since these two parts give little contributions for part-based face recognition, for speed consideration, we drop these two parts. Finally we choose the eyes, eyebrows, nose, mouth, under nose part and the holistic face to form a part-based face recognition.

The results of each facial parts are fused to obtain the final score. Three fusion rules are applied in our experiment, as described in section 4. For the weighted sum rule, the weights for fusion are optimally learned by LDA on the training set. Table 1 gives the optimal weights and their corresponding parts. From the table we can see that, for

Table 1. Optimal weights Learned by LDA for part-based fusion.

Facial Parts	Weights
Face	0.7408
Left Eyebrow	0.0072
Right Eyebrow	0.0133
Left Eye	0.2638
Right Eye	0.2092
Nose	0.3629
Mouth	0.2337
Under Nose	0.3575

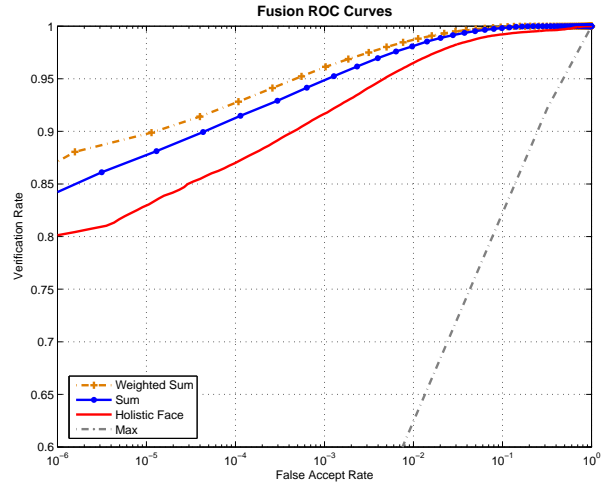


Figure 6. ROC curves of fusion results.

optimal weighted sum fusion, the contribution of each part classifier does not all direct ratio to their individual recognition rates. That is why we use LDA to learn optimal weights instead of taking weights from recognition rates.

The ROC curves of each fusion method are shown in Fig. 6. The result of holistic face is the same as [10] The experiment result shows that the weighted sum rule achieves the best performance, followed by sum rule and holistic face. The corresponding recognition rates are 96.03%, 94.8%, and 91.5% when FAR=0.1%. Clearly, our part-based approach outperformed the holistic approach.

6. Conclusion

This paper presents a part-based method for NIR based face recognition. A face is decomposed into several facial parts according to their discriminative power. Then a part classifier is built for each part using boosted SLBPH features, and the outputs of part classifiers are fused to give the final score. Three fusion rules are applied for classifier fusion, with LDA-weighted sum rule achieving the best result.

Experiments show that the present part-based method outperform the previous whole face-based one [10] by 4.53%. While we have used LDA to learn the optimal weights, we would explore better scheme of weights in the future.

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
- [2] T. Ahonen and M. Pietikainen. "Face description with local binary patterns: Application to face recognition". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2006. 1, 3
- [3] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, July 1997. 1
- [4] Y. Freund and R. Schapire. "Experiments with a new boosting algorithm". In *Proc. 13th International Conference on Machine Learning*, pages 148–146. Morgan Kaufmann, 1996. 3
- [5] 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. 3
- [6] A. Hadid, M. Pietikainen, 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. 2
- [7] B. Heisele, P. Ho, and T. Poggio. "Face recognition with support vector machines: Global versus component-based approach". *IEEE International Conference on Computer Vision*, 2001. 1
- [8] Y. Ivanov, B. Heisele, and T. Serre. "Using component features for face recognition". In *Proceedings of the 6th International Conference on Automatic Face and Gesture Recognition*, 2004. 1, 3
- [9] J. Kittler, M. Hatel, R. P. W. Duin, and J. Matas. "On combining classifiers". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(3):226–239, January 1998. 4
- [10] S. Z. Li, R. Chu, S. Liao, and L. Zhang. "Illumination invariant face recognition using near-infrared images". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(Special issue on Biometrics: Progress and Directions), April 2007. 1, 2, 4, 5, 6
- [11] G. J. McLachlan. *Discriminant Analysis and Statistical Pattern Recognition*. Wiley, New York, 1996. 4
- [12] B. Moghaddam, C. Nastar, and A. Pentland. "A Bayesian similarity measure for direct image matching". *Media Lab Tech Report No.393*, MIT, August 1996. 3
- [13] 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. 2
- [14] T. Ojala, M. Pietikainen, and T. Maenpaa. "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7):971–987, 2002. 2
- [15] A. Ross, A. K. Jain, and J.-Z. Qian. "Information fusion in biometrics". *Lecture Notes in Computer Science (A longer version appears in Pattern Recognition Letters, Vol. 24, Issue 13, pp. 2115-2125, September, 2003., 2091:354–359, 2001.* 4
- [16] Stan Z. Li and His Face Team. "AuthenMetric F1: A Highly Accurate and Fast Face Recognition System". ICCV2005 - Demos, October 15-21 2005. 1, 2, 4
- [17] M. A. Turk and A. P. Pentland. "Eigenfaces for recognition". *Journal of Cognitive Neuroscience*, 3(1):71–86, March 1991. 1
- [18] B. Weyrauch, B. Heisele, J. Huang, and V. Blanz. "Component-based face recognition with 3d morphable models". *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2004. 1