Learning to Fuse 3D+2D Based Face Recognition at Both Feature and Decision Levels

Stan Z. Li, ChunShui Zhao, Meng Ao, and Zhen Lei

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. 2D intensity images and 3D shape models are both useful for face recognition, but in different ways. While algorithms have long been developed using 2D or 3D data, recently has seen work on combining both into multi-modal face biometrics to achieve higher performance. However, the fusion of the two modalities has mostly been at the decision level, based on scores obtained from independent 2D and 3D matchers.

In this paper, we propose a systematic framework for fusing 2D and 3D face recognition at both feature and decision levels, by exploring synergies of the two modalities at these levels. The novelties are the following. First, we propose to use Local Binary Pattern (LBP) features to represent 3D faces and present a statistical learning procedure for feature selection and classifier learning. This leads to a matching engine for 3D face recognition. Second, we propose a statistical learning approach for fusing 2D and 3D based face recognition at both feature and decision levels. Experiments show that the fusion at both levels yields significantly better performance than fusion at the decision level.

1 Introduction

Face recognition has attracted much attention due to its potential values for applications as well as theoretical challenges. Many representation approaches have been introduced. Principal Component Analysis (PCA) [1] computes a reduced set of orthogonal basis vector or eigenfaces of training face images. A new face image can be approximated by weighted sum of these eigenfaces. PCA provides an optimal linear transformation from the original image space to an orthogonal eigenspace with reduced dimensionality in the sense of the least mean square reconstruction error. , Linear Discriminant Analysis (LDA) [2] seeks to find a linear transformation by maximizing the between-class variance and minimizing the within-class variance. Independent component analysis(ICA) [3] uses high-order statistics to generate image bases. Elastic bunch graph matching (EBGM) [4,5] uses Gabor wavelets to capture the local structure corresponding to spatial frequency (scale), spatial localization, and orientation selectivity.

Local Binary Pattern (LBP), originally proposed as a descriptor for textures [6], provides a simple yet effective way to represent faces [7,8]. There, the face image is equally divided into small blocks and LBP features are extracted for each blocks to represent the texture of a face locally and globally. Weighted Chi square distance of

W. Zhao, S. Gong, and X. Tang (Eds.): AMFG 2005, LNCS 3723, pp. 43–53, 2005.

these LBP histograms is used as a dissimilarity measure for comparing the two images. The above works have shown that LBP based methods can produces good results for face recognition in 2D images.

Boosting learning with local features have recently been proposed as a promising approach. Jones and Viola [9] propose a general idea of boosting local features and training a classifier on difference between two face image feature vectors (Haar wavelets). Zhang *et al.* present an LBP-based boosting learning algorithm [10]. Such works are for 2D face recognition.

While using 2D intensity images to recognize a face has long history of research [11], recent advances in 3D range sensor has made it possible to overcome some limitations in 2D based face recognition methods such as illumination and pose changes. Early work on 3D face recognition was based on curvature features [12], following this type of work in 3D range image understanding starting from mid-1980's [13]. Later developments in 2D face recognition have influenced 3D face recognition [14].

It may be advantageous to combine information contained in both 2D and 3D data to overcome limitations in 2D or 3D based methods while 2D and 3D images encodes different information. Methods have been proposed to combine information in both modalities into multi-model face biometrics to achieve higher performance [14]. For example, in [15,16], the weighted sum rule is applied to combine the two matching scores. A recent performance evaluation on the 2D and 3D modalities and their fusion has shown that multi-modal 3D+2D face recognition performs significantly better than using either 3D or 2D alone [17].

So far, the fusion of 3D+2D modalities has been at the decision level, using scores from 2D and 3D matchers. The 3D recognition result and the 2D recognition result are each produced without reference to the other modality. It is desirable to explore synergies of the two modalities at the feature level as well [14]. The work presented here explores such synergies in the proposed framework of AdaBoost learning (with LBP feature). This is new for solving the problem of 3D+2D face fusion.

In this paper, we propose a systematic framework for fusing 2D and 3D information at both feature and decision levels. The main contributions are the following: First, we propose to use LBP features as a representation of faces in 3D data. An AdaBoost learning procedure [18,19,20] is then applied for feature selection and classifier learning. Second, with LBP as a unified representation of faces in both 2D and 3D images, we propose to use AdaBoost learning to fuse 2D and 3D information at both feature and decision levels. The same AdaBoost learning procedure as used for 3D face recognition is used for 3D+2D fusion. 3D and 2D LBP histograms are computed, respectively, and then combined into a 3D+2D feature set. AdaBoost is applied to select effective feature from a 3D+2D feature pool, construct weak classifiers based on the selected features, and then combine the weak classifiers into a strong one. Thus, the learning procedure fuses the 3D and 2D modalities at both feature and decision levels. Experiments show that the AdaBoost learning method produces significantly better results than the baseline PCA method. AdaBoost learning based fusion performs significantly better than fusion of PCA based scores. Experimental results clearly demonstrate the advantages of the two level fusion over the exiting decision level fusion such as presented in a recent PAMI paper [17].

The rest of this paper is organized as follows: In section 2, the LBP representation is described. In section 3, we propose an AdaBoost learning method for 3D face recognition. In section 4, we propose the boosting based fusion of 3D+2D modalities. Experimental results are presented in section 5.

2 Feature Representation

Face images are preprocessed so that they are aligned in a predefined way. For 2D data, the alignment and cropping is done according to the eye centers. For 3D data, the face is rotated about the vertical axis so that the nose tip becomes the closest point and then cropped; after that, a median filter is applied to remove high noise; this is followed by hole-filling. Fig.1 shows some examples. LBP features are then extracted from the cropped and preprocessed images.

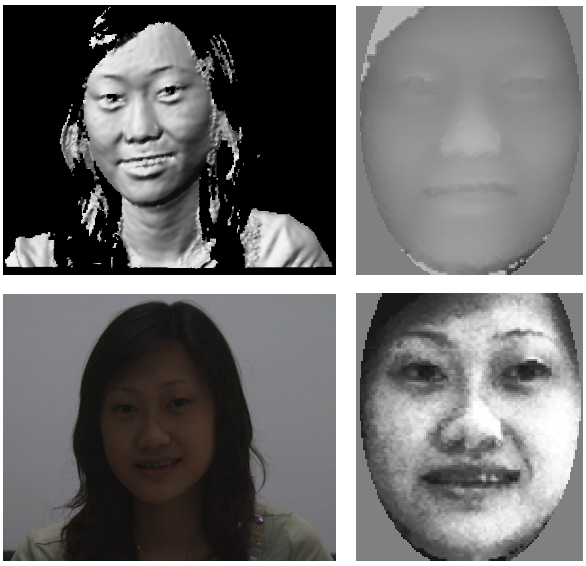

Fig. 1. 3D (top) and 2D (bottom) face images of a person before (left) and after (right) alignment and cropping

2.1 Local Binary Pattern

The LBP operator was originally introduced by Ojala [6] as texture description. LBP features have performed very well in various applications, including texture classification and segmentation. The basic form of an LBP 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 the basic LBP operator is shown in Fig.2. Note that the binary LBP code is circular.

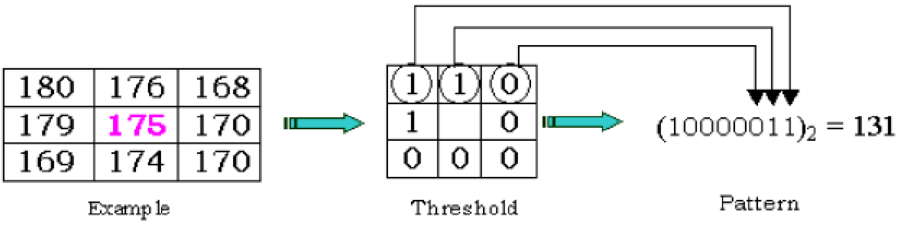

Fig. 2. Calculation of LBP code from 3x3 subwindow (from [8])

The major limitation of the basic LBP operator is its small spatial support area. Features calculated in a local 3×3 neighborhood cannot capture large scale structure that may be the dominant features of some textures. The LBP operator can be extended to use neighborhoods of different size [6]. Another extension to the original operator is to use so called uniform patterns [6]. An LBP is called uniform if it contains at most two bitwise 0-1 or 1-0 transitions. There are 58 uniform LBP code patterns for 8-bits LBP code, and 256-58=198 non-uniform LBP patterns.

2.2 Local Histograms of LBP Code

LBP histograms over local regions provides a more reliable description when the pattern is subject to alignment errors. Considering the uniform LBP scheme, and denoting all the non-uniform LBP patterns with a single bin, then there are a set of $L + 1 = 59$ possible LBP code types for the 8-bit LBP code. Let us denote this set by $\mathcal{L} = \{0, 1, \ldots, L\}$ such that $LBP(x, y) \in \mathcal{L}$, and the local LBP histogram over a block $S_{(x,y)}$ centered at (x, y) by $H_{(x,y)} = (H_{(x,y)}(0), H_{(x,y)}(1), \ldots, H_{(x,y)}(L))$. The histgram can be defined as

$$H_{(x,y)}(\ell) = \sum_{(x',y') \in S(x,y)} I\{LBP(x', y') = \ell\}, \qquad \ell \in \mathcal{L} \tag{1}$$

where $I(\cdot) \in \{0, 1\}$ is an indication function of a boolean condition, and $S(x, y)$ is a local region centered at (x, y) which in our case is a 20x15 block.

The histogram $H_{(x,y)}$ contains information about the distribution of the local micro-patterns, such as edges, spots and flat areas, over the block $S_{(x,y)}$. It effectively gives a description of the face at two different levels of locality: individual LBP labels contain information about the patterns at the pixel-level, whereas the frequencies of the labels in the histogram produce information on regional level [7]. The collection of the histograms at all possible pixels $\{H_{(x,y)} \mid \forall(x, y)\}$, called the global LBP histogram, provides the global level description.

In [7], the face image is partitioned into a number (49) of blocks and a weight is empirically assigned to each block. Denote the corresponding histograms between the probe and a gallery by $H^P_{(x,y)}$ and $H^G_{(x,y)}$, respectively. Several possible dissimilarity measures are available to compare local two histograms. The following Chi square distance is reported to work better for small sample size [7]:

$$\chi^2(H^P_{(x,y)}, H^G_{(x,y)}) = \sum_{\ell \in \mathcal{L}} \frac{(H^P_{(x,y)}(\ell) - H^G_{(x,y)}(\ell))^2}{(H^P_{(x,y)}(\ell) + H^G_{(x,y)}(\ell))} \tag{2}$$

A possible scheme for matching between two images is based on a weighted sum of χ^2 distances [7].

3 Learning for 3D Face Recognition

In this section, we describe a method which uses LBP features and AdaBoost learning for 3D face recognition with the LBP features. While in [7], a face image is partitioned into blocks, We consider every block centered at each pixel location. This yields a large number of possible blocks, and hence a large number of local histograms $H_{(x,y)}$. Instead of assigning a weight to each block, we derive the weights using an AdaBoost learning method. As a result of the learning, those blocks which are more discriminative for classification will be assigned larger weights and those which are useless or give conflict information will be assigned near-zero weights. The learning also produces the final classifier.

Face recognition is a multi-class problem. To dispense the need for a training process for faces of a newly added person, we use a large training set describing intra-personal or extra-personal variations [21], and train a "universal" two-class classifier. An ideal intra-personal difference should be an image with all pixel values being zero, whereas an extra-personal difference image should generally have much larger pixel values. However, instead of deriving the intra-personal or extra-personal variations using difference images as in [21], the training examples to our learning algorithm is the set of differences between each pair of local histograms $H_{(x,y)}$ at the corresponding locations. The positive examples are derived from pairs of intra-personal differences and the negative from pairs of inter-personal differences.

With the two-class scheme, the face matching procedure will work in the following way: It takes the probe face image and a gallery face image as the input; computes a difference-based feature vector from the two images; and then calculated a similarity score for the feature vector using some matching function. A decision is made based on the score, to classify the feature vector into the positive class (coming from the same person) or the negative class (different persons). The following presents an AdaBoost learning algorithm for training such a two-class classifier using the positive and negative examples of the 2D or 3D face data.

In AdaBoost learning, we are given a training set of N labeled examples from two classes, $\mathbf{S} = (x_1, y_1), \ldots, (x_N, y_N)$, where x_i is the data $y_i \in \{+1, -1\}$ is the class label. Associated with the training examples is a distribution $w_t = (w_{t,1}, \ldots, w_{t,N})$ which is updated after each learning iteration t. An AdaBoost procedure adjust the distribution in such a way that more difficult examples will receive higher weights. It 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

$$H(x) = \operatorname{sign}\left(\sum_{t=1}^{T} \alpha_t h_t(x)\right) \tag{3}$$

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, shown in Fig. 3, is aimed to derive α_t and $h_t(x)$. The AdaBoost learning procedure in effect solves the following three fundamental problems: (1) learning effective features from the candidate feature set (step 3), (2) constructing weak classifiers each of which is based on one of the selected features (step 1-3), and (3) combining the learned weak classifiers into a stronger classifier (the output step).

```
Input: Given labeled examples $S$;
Set the initial $w_1$ to the uniform distribution;
For $t = 1, \dots, T$:
  1. Train a weak classifier $h_j : x \rightarrow \{-1, +1\}$;
  2. Calculate $w_t$-weighted error
       $e_j = P[h_j(x_i) \neq y_i \mid w_t]$;
  3. Choose $h_k(x)$, such that $e_k < e_j, \forall j \neq k$;
  4. Let $e_t = e_k$.
  5. Choose $\alpha_t = \frac{1}{2} \log \left( \frac{1-e_t}{e_t} \right)$;
  6. Update $w_{t+1,i} \leftarrow w_{t,i} \exp(-\alpha_t y_i h_i(x_i))$;
  7. Normalize $w_{t+1}$ to $\sum_i w_{t+1,i} = 1$;
Output $H(x)$ as in Equ.(3).
```

Fig. 3. The AdaBoost learning procedure

In our system, a weak classifier is defined based on a single feature (*i.e.* an LBP histogram bin value). A weak classifier gives an output of +1 or -1, by thresholding the feature, at an appropriate threshold value learned with a weak learner procedure. This is unlike the Chi square distance based weak classifiers used in [10]. We find that the bin based weak classifiers can do a better job in both training and testing.

4 Learning to Fuse 2D and 3D

Now we present a method for fusing 2D and 3D information at both feature and decision levels. In the fusion of the 2D and 3D information, we do not make assumptions on how the information is correlated between 2D and 3D nor do we require that there are correspondences between 2D and 3D images. The only requirement is that faces in 2D and 3D images are properly aligned and normalized, respectively, as a result of pre-processing. Then, everything is learned automatically. We use the same AdaBoost learning procedure as above for the 3D+2D fusion as follows:

For every pixel location in an image (2D or 3D), an LBP code is computed. There are $L + 1 = 59$ possible LBP code types. A histogram of 59 bins is calculated, over a local sub-window centered at the pixel, to account for the distributions of the 59 types of features in the sub-window. For each intra-pair or inter-pair of 2D or 3D images, the Chi square distance is computed, according to Eq.(2), to account for the differences of the two corresponding local LBP histograms, and will be used as the feature to measure the dissimilarity between the two local image patches. The distributions of that Chi distances for the positive and negative examples at the local patch are then analyzed by considering all the intra-pairs or inter-pairs. Such statistics are computed over all the image locations and for both 2D and 3D images.

AdaBoost is applied to select most effective features from the complete 3D+2D difference feature set. At each iteration, the best LBP feature is selected, among all the locations for the 2D and 3D images, according to the distributions of the Chi square distances of the LBP histograms, such that the feature provides the best discriminative power. A weak classifier is then constructed by thresholding the Chi square distance. The weak classifiers are then combined into a strong one. This way, the AdaBoost based procedure provides a systematic approach for 3D+2D fusion at both feature and decision levels.

5 Experimental Results

The purpose of the experiments presented below is to compare the proposed boosting learning methods with the baseline PCA methods in their performance for 3D, 2D and 3D+2D face recognition.

5.1 Data Description

A large 3D+2D database is created for the experiments using a Minolta 3D digitizer, which produces a range image and the corresponding color image. The images are taken near-frontal but with varying pose, expression, and lighting changes. The database is composed of 2305 images of persons. It is divided into three sets. The composition of the data for the training, gallery and probe sets is summarized in Table 1. The images are preprocessed and cropped into 138x118 pixels. Figure 4 gives some examples of the preprocessed imaged.

Table 1. Data Composition

3D Data	Num. of Images	Num. of Persons
Train	945	246
Gallery	252	252
Probe	1108	252

2D Data	Num. of Images	Num. of Persons
Train	945	246
Gallery	252	252
Probe	1108	252

Before PCA the pixel vectors are first scaled such that the mean value of the vectors is zero and the standard deviation is one. We choose the top 99 percent of the energy and distance metric is L2. By computing the distance between the images in 2D and 3D set, respectively, we can get two similarity scores matrix. But the performance of the PCA on 2D or 3D is not good enough. Therefore, we fuse the scores to improve the

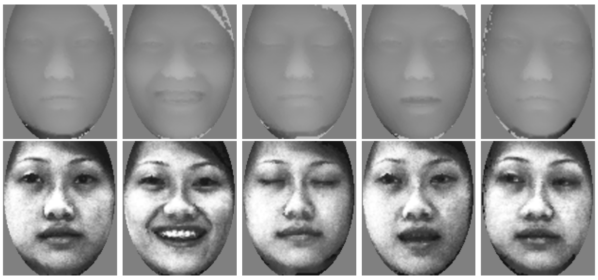

Fig. 4. Examples of 3D images and the corresponding 2D images of a person

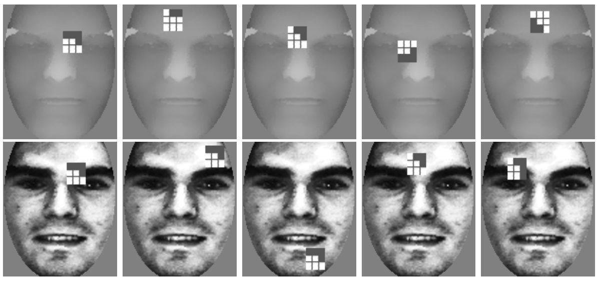

Fig. 5. The first 5 features for 3D (top) and for 2D (bottom) learned by AdaBoost

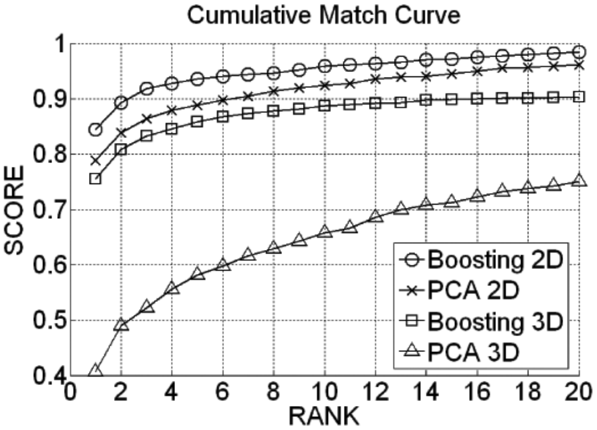

Fig. 6. Cumulative Match Curves for 3D and 2D

classifying performance. Before fusing the scores from each modality, the scores are normalized to [0, 100] and then fused by the sum rule. The weight is computed according to the method being mentioned in [6]. By fusing at decision levels, the performances are improved significantly.

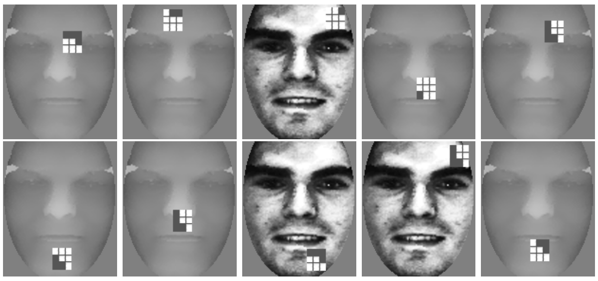

Fig. 7. The first 10 LBP features learned by boosted fusion of 3D+2D, ranked 1 to 10 from left to right, from top to bottom. Among these top 10, 7 features are from 3D data and 3 from 2D.

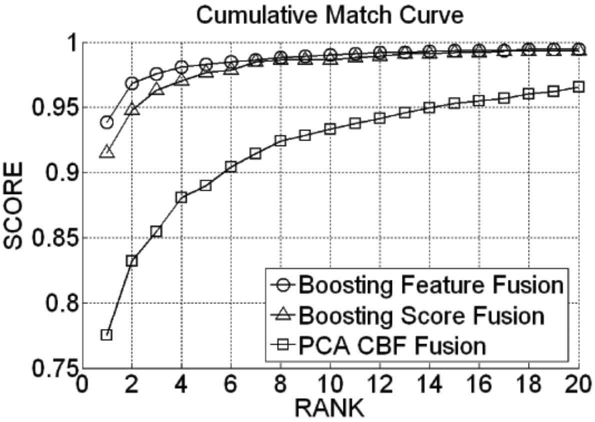

Fig. 8. Cumulative Match Curves for 3D+2D fusion

5.2 Boosted 3D and 2D Face Recognition

An AdaBoost classifier is trained for 3D faces and another trained for 2D faces recognition, separately. The 3D model 83 weak classifiers whereas the 2D model has 170 weak classifiers. Fig.5 shows the first 5 features for 3D and the first 5 for 2D. The comparative results are shown in Fig.6 in terms of cumulative match curves (CMC). From the CMC curves we conclude that the boosting learning method is superior to the PCA method.

5.3 Boosted Fusion of 3D+2D Face Recognition

For 3D+2D fusion, we trained a boosted model selected 97 most significant features. Of the 97 features, 59 are from 3D and 38 from 2D. Fig.7 shows the first 10 features for the 3D+2D fusion. We notice that the first 2 features in the AdaBoost 3D+2D fusion model (Fig.7) correspond to the first 2 features of 3D only model (Fig.5); and that there are more 3D features than 2D ones.

To contrast with the proposed AdaBoost learning fusion scheme, two non-boosting fusion schemes are included: The first is the PCA-based 3D+2D fusion (called "CBF' score fusion, described at the end of Section 3 of [17]), which is the baseline fusion performance. The second uses a sum rule to fuse the two AdaBoost classification scores. The comparative results are shown in Fig.8 in terms of cumulative match curves (CMC). From the CMC curves we conclude that fusing AdaBoost scores performs better than fusing PCA scores; and that fusion at both feature and decision levels by the proposed AdaBoost learning achieves the best performance of the three compared schemes.

6 Conclusion

In this paper, we explore synergies of 3D and 2D modalities by proposing a systematic framework for fusing 2D and 3D face recognition at both feature and decision levels. To our knowledge, this is the first work of this kind and is the main contribution of the paper. Another contribution is the novel LBP+AdaBoost learning method for 3D face recognition. We have demonstrated by experiments the effectiveness of the two contributions in 3D face recognition and in 3D+2D fusion. The successful fusion of 3D+2D at both feature and decision level has verified a conjecture made in [14] that "it is at least potentially more powerful to exploit possible synergies between the the two modalities in the interpretation of the data."

References

1. Turk, M.A., Pentland, A.P.: "Eigenfaces for recognition". Journal of Cognitive Neuroscience **3** (1991) 71–86
2. Belhumeur, P.N., Hespanha, J.P., Kriegman, D.J.: "Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. IEEE Transactions on Pattern Analysis and Machine Intelligence **19** (1997) 711–720
3. Bartlett, M.S., Lades, H.M., Sejnowski, T.J.: "Independent component representations for face recognition". Proceedings of the SPIE, Conference on Human Vision and Electronic Imaging III **3299** (1998) 528–539
4. Lades, M., Vorbruggen, J., Buhmann, J., Lange, J., von der Malsburg, C., Wurtz, R.P., Konen, W.: Distortion invariant object recognition in the dynamic link architecture. IEEE Transactions on Computers **42** (1993) 300–311
5. Wiskott, L., Fellous, J., Kruger, N., v. d. Malsburg, C.: "Face recognition by elastic bunch graph matching". IEEE Transactions on Pattern Analysis and Machine Intelligence **19** (1997) 775–779
6. Ojala, T., Pietikainen, M., Harwood, D.: "A comparative study of texture measures with classification based on feature distributions". Pattern Recognition **29** (1996) 51–59
7. Ahonen, T., Hadid, A., M.Pietikainen: "Face recognition with local binary patterns". In: Proceedings of the European Conference on Computer Vision, Prague, Czech (2004) 469–481
8. Hadid, A., Pietikinen, M., Ahonen, T.: "A discriminative feature space for detecting and recognizing faces". In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Volume 2. (2004) 797–804
9. Jones, M., Viola, P.: "Face recognition using boosted local features". Tech Report TR2003-025, MERL (2003)

10. Zhang, G., Huang, X., Li, S.Z., Wang, Y.: "Boosting local binary pattern (LBP)-based face recognition". In Li, S.Z., Lai, J., Tan, T., Feng, G., Wang, Y., eds.: Advances in Biometric Personal Authentication. Volume LNCS-3338. Springer (2004) 180–187
11. Kanade, T.: Picture Processing by Computer Complex and Recognition of Human Faces. PhD thesis, Kyoto University (1973)
12. Cartoux, J.Y., LaPreste, J.T., Richetin, M.: "Face authentication or recognition by profile extraction from range images". In: Proceedings of the Workshop on Interpretation of 3D Scenes. (1989)
13. Besl, P.J., Jain, R.C.: "Intrinsic and extrinsic surface characteristics". In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, California (1985) 226–233
14. Bowyer, K.W., Chang, Flynn, P.J.: "A survey of 3D and multi-modal 3d+2d face recognition". In: Proceedings of International Conference Pattern Recognition. (2004) 358–361
15. Lu, X., Jain, A.K.: "Integrating range and texture information for 3d face recognition". In: Proc. 7th IEEE Workshop on Applications of Computer Vision (WACV'05), Breckenridge, CO (2005)
16. Tsalakanidou, F., Malassiotis, S., Strintzis, M.G.: "Face localization and authentication using color and depth images". **14** (2005) 152–168
17. Chang, K.I., Bowyer, K.W., Flynn, P.J.: "An evaluation of multi-modal 2D+3D face biometrics". IEEE Transactions on Pattern Analysis and Machine Intelligence (2005) to appear
18. Freund, Y., Schapire, R.: "A decision-theoretic generalization of on-line learning and an application to boosting". Journal of Computer and System Sciences **55** (1997) 119–139
19. Schapire, R.: A brief introduction to boosting. In: Proceedings of the Sixteenth International Joint Conference on Artificial Intelligence. (1999)
20. Viola, P., Jones, M.: "Robust real time object detection". In: IEEE ICCV Workshop on Statistical and Computational Theories of Vision, Vancouver, Canada (2001)
21. Moghaddam, B., Nastar, C., Pentland, A.: "A Bayesain similarity measure for direct image matching". *Media Lab Tech Report* No.393, MIT (1996)