A Multiview Face Identification Model With No Geometric Constraints

Jerry Jun Yokono † Tomaso Poggio ‡

† Sony Intelligence Dynamics Laboratories, Inc.
3-14-13-4F Higashigotanda, Shinagawa-ku, Tokyo, 141-0022, JAPAN
‡ Center for Biological and Computational Learning, M.I.T.
Bldg. 46-5155, 43 Vassar St., Cambridge, MA 02139, USA

jerry@idl.sony.co.jp tp@ai.mit.edu

Abstract

Face identification systems relying on local descriptors are increasingly used because of their perceived robustness with respect to occlusions and to global geometrical deformations. Descriptors of this type – based on a set of oriented Gaussian derivative filters – are used in our identification system. In this paper, we explore a pose-invariant multiview face identification system that does not use explicit geometrical information. The basic idea of the approach is to find discriminant features to describe a face across different views. A boosting procedure is used to select features out of a large feature pool of local features collected from the positive training examples. We describe experiments on well-known, though small, face databases with excellent recognition rate.

1. Introduction

After the seminal work of Kanade [28], there have been many face identification algorithms proposed in the literature. However, the task of face identification still remains a challenging problem because of its fundamental difficulties regarding various factors in the real-world such as illumination changes, rotation in both in-plane and out-plane, facial expressions, and clutter backgrounds. One of the most popular approaches is to use the eigenspace approach originally proposed by Turk and Pentland [29] and later extended in a Bayesian framework by Moghaddam and Pentland [30]. Although these approaches show good performances on frontal face identification, they have limited performance for both in-plane and in depth face rotations. As Heisele [16] pointed out, face rotations lead to considerable position changes of facial components, making feature localization difficult and thus lead to poor recognition performance. Therefore, it is necessary to use flexible geometrical models even in the frontal face identification system. Heisele [16] proposed component-based face detection/recognition system. Facial components are selected automatically using a statistical error bound; a SVM classifier is then generated for each component, followed by a higher-level combination classifier. Wiskottt and Malsburg [31] used another type of flexible geometrical model called "elastic graph matching" with Gabor local features. These systems – which have component/local features and their flexible geometrical constraints – have been shown to have better recognition performance than the global/holistic approaches. Those component or local feature based algorithms have also been proposed for the task of object identification. Lowe [7] developed an object identification system that works well in cluttered scenes and achieves rotational and scale invariance by using a unique local descriptor called "SIFT". Inspired by his work, the use of local features for object recognition has become popular. Several recognition systems combined with statistical learning have been since proposed [13][14]. Although they are not face identification systems, they are related to our work. Let us clarify the main differences relative to our work. Wallraven [13] proposed local kernels for SVM to learn the object model from multiple image samples. Their system yields good performance though their formulation has theoretical problems (the kernel is not positive definite). Csurka [14] introduced the notion of a "bags of keypoints" for object recognition. Their system computes SIFT local features; a clustering technique was used to build a "visual vocabulary" and a SVM classifier was trained to classify each category such as faces and cars. The input features to SVMs are the number of occurrences of each key feature. The occurrences are represented by a binned histogram where features are thresholded using Euclidian distance. The main problem of this approach is that all the features have the same threshold, that is the same probability to describe the object. Our framework is more general and we do not use any heuristic such as a threshold ; those thresholds are learned for each feature during the learning. Additionally, we are exploring the relative peformance of full-multiview object/face identification. Therefore, we do not perform clustering of the local features [14] since we found that clustering blurs the representation power of each local feature . We show in this paper that our framework can achieve the following three properties in a unified framework: (1) tuning a threshold for each local feature matching, (2) selecting discriminant local features out of a large feature pool, (3) constructing a strong classifier by combining the selected features.

In the literature, multiview (pose invariant) face identification approaches can be categorized into either multiview-based [1][2][3][21][22][23][26] or single-view based [20] approach. In the multiview-based approach, the training is done using multiview face images and a test image is assumed to be matched to one of the existing head pose model whereas the single-view based approach uses a canonical head pose for recognition. Normally, with the multiview-based approach, one might have "view specific models" [21] which makes the recognition process more complicated and even more time consuming. On the other hand, our framework can be applied without any explicit view-tuned models; we do not exploit strong geometric constraints. Our approach is significantly differs from other approaches in the literature (except [26]) in that just one model describes the whole multiview face space.

In this report, we show how a full-multiview (invariant to rotations both in-plane and in depth) face identification system can be designed and how viewpoint-robust local features can be selected during the learning procedure. Although there are several related approaches in the literature which use local features and feature selection for the face identification [17] [18], their performance is limited because of the strong geometry constraints on the face (without actually exploiting the multiview face approach). Comparisons with two other statistical classifiers, SVM and boosting, are also described in this paper. The main contribution of this paper is (1) to propose a framework that can effectively select viewpoint-specific distinctive features for full-multiview face/object recognition, (2) to propose a simple (actually simpler than conventional approaches) yet powerful framework for integrating full face poses into just one model (3) to show promising experimental results on face identification (though they should be validated on larger databases).

In section 2, we review a simple-cell type local descriptor based on the Gaussian derivatives, implemented using "steerable filters" [6]. Section 3 overviews the system. We describe the main experiments in section 4 and 5. Section 6 concludes the paper.

2. A simple-cell type local descriptor

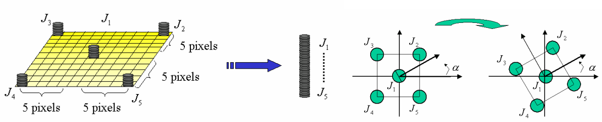

Figure 1. Gaussian derivatives up to the third order with four orientations and three scales are computed at corner-like points and four neighboring pixel locations.

Gaussian derivatives are filters with spatial orientation selectivity as well as frequency selectivity. The steerable filter [6] response for the *n*th order Gaussian derivative $G_n(\theta)$ to an arbitrary orientation θ is, given the Gaussian distribution *G:*

$$G = \exp(-\frac{x^2 + y^2}{2\sigma^2})$$

$$G_1(\theta) = \cos(\theta)G_1(0^o) + \sin(\theta)G_1(90^o)$$

$$G_2(\theta) = k_{21}(\theta)G_2(0^o) + k_{22}(\theta)G_2(60^o) + k_{23}(\theta)G_2(120^o)$$

$$k_{2i}(\theta) = \frac{1}{3}\{1 + 2\cos(2(\theta - \theta_i))\}$$

$$G_3(\theta) = k_{31}(\theta)G_3(0^o) + k_{32}(\theta)G_3(45^o) + k_{33}(\theta)G_3(90^o) + k_{34}(\theta)G_3(135^o)$$

$$k_{3i}(\theta) = \frac{1}{4}\{2\cos(\theta - \theta_i) + 2\cos(3(\theta - \theta_i))\}$$

where $k_{in}(\theta)$ is the coefficient of the basis. Gaussian derivatives can be seen as the asymptotic form of Gabor filters. Yokono and Poggio [19] evaluated their invariance and selectivity concluding that Gaussian derivatives with higher order derivatives make the descriptor more powerful than Gabor descriptor for the specific object identification task. We use Gaussian derivatives up to the third order, with four orientations and three widths: 1,2, and 4. The vector length of the jet descriptor associated with a location in the image is 3x3x4=36. The descriptor can be made more powerful by combining the neighboring four jets, which are five pixels away from the center pixel. In this case the length of the local descriptor is 36x5=180. The local descriptor used in the experiments is shown in Figure 1.
Design details of the descriptor can be found in [19][34][35].

3. Boosting local features with no geometric constraints

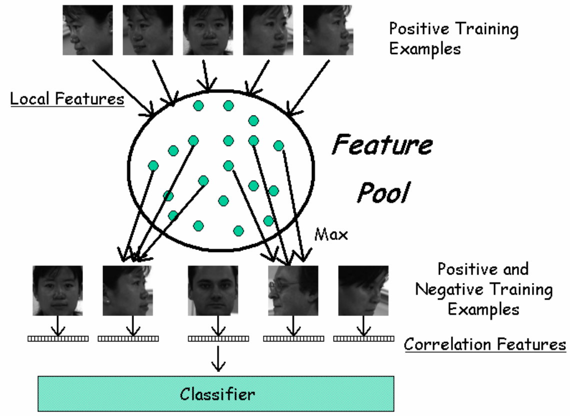

Figure 2. System Overview: Local features are collected from positive training examples. Then maximum correlations over all the local features in a sample are computed and used as inputs to the classifier. Note that when considering multiclass recognition, a different set of features is used for each class.

3.1. Correlation features

An overview of the system is shown in Figure 2. The basic idea of our approach -- motivated by Morgenstern and Heisele[8] -- is to collect local features from the positive training images and thus create a feature pool. The pool features represent a "dictionary" of features that describes the face. When considering multi-class recognition, a feature pool is created for each class with a different set of local features. Local features based on the Gaussian derivatives are computed on corner-like points detected by Harris measure. The advantage of using corner-like points is that those points usually have high information. Once the feature pool is created, all the positive and negative training images are used to compute "*correlation features*": for each feature x_i in the pool X, the maximum of the (normalized) correlation over all the local features in a sample is computed for each training sample. Therefore, if N features are in the pool, every training image has a N-dimensional feature vector. Let *V* be the set of features in the image, a *correlation feature* C has elements $C(i),\ i \in [\boldsymbol{1,...,N}]$, such that

$$C(i) = \max_{v \in V}(\frac{(x_i - \overline{x_i})(v - \overline{v})}{\sqrt{(x_i - \overline{x_i})^2 (v - \overline{v})^2}})$$

We call this vector – which is the input to the classifier -- a "*correlation feature*". Taking the max over all the features in a sample means that we do not use geometric information. Our expectation was that even without geometric information, the descriptors are sufficiently discriminant. The *correlation features* might be common features across the viewpoint of the face or distinctive features for a specific viewpoint. At run time local features on the corner-like points are used to compute a *correlation feature* which is fed into the previously trained classifier.

3.2. Discrete AdaBoost and Gentle AdaBoost

We may use any kind of classifiers such as Support Vector Machine and boosting. AdaBoost was originally proposed by Freund and Schapire [32] and is a successful classifier in a host of real-world application [4]. Later, Friedman et. al. [9] proposed a modified version of AdaBoost that uses additive regression as a weak learner and adaptive Newton steps for the optimization. They called the original AdaBoost Discrete AdaBoost and claimed that their new Gentle AdaBoost often outperforms Discrete AdaBoost [9][15]. Both algorithms are listed in the Algorithm Box. As can be seen, Gentle AdaBoost uses real-valued regression rather than the *{-1, +1}* of Discrete AdaBoost.

3.3. Boosting local features

In terms of performance, SVM and boosting (and as a matter of fact also square-loss regularization) are usually quite similar. An advantage of using boosting is that it effectively performs feature selection during the learning. For instance, if the system holds initially 6000 features in the feature pool, it is necessary when using SVM to compute at run time all the correlation values for all the features. On the other hand, boosting may selects fewer features, say 200, thereby considerably speeding up computation at run time. Since we are effectively using a decision stump (binary split decision tree) as weak classifiers [32], the learning procedure tunes the threshold for each feature. Good features can be selected by taking the minimum error of the features. In the experiments, we use both SVM and boosting and compare the results.

4. Frontal face identification

4.1. MIT CBCL datasets

Figure 3. Sample images from the MIT face database. The database contains 10 people with approximately 200 images per person. It has various changes in illumination, scale, pose, and facial expression.

4.1. 1 Experimental setup

In the first experiment, we performed face identification on MIT face database (*1). The face database contains faces of 10 people with approximately 200 images per person. It has both male and female face images collected from various ethnic subjects. Sample images are shown in Figure 3. As shown in these images, there are variations in illuminations, face positions (not aligned to center), slight scale changes, and pose changes up to about 45 degrees of depth rotation and up to 30 degrees of in-plane rotation. Images are 70x70 gray values. For each run, N images are randomly chosen as training images and the remaining images are used for testing. Gaussian derivative based local features are computed on the corner-like points detected by the Harris corner detector for every training image. Approximately 50 points per image are detected in our experiments. These local features from positive examples are collected to build a feature pool. For instance, when we use 30 training images, approximately 50x30 = 1500 features are in the pool. Then, all the positive and negative images are used to make the *correlation feature* vectors where each image represents approximately 1500-dimensional correlation feature vector. All the results are averages of 10 runs. We report here the performance of the binary classifier and multiclass face identification.

(*1) http://cbcl.mit.edu/software-datasets/heisele/facerecognition-database.html

Discrete AdaBoost Algorithm (Freund & Schapire [32])

1.Initialize weights $w_i = 1/N, \quad i = 1,2,...,N$, where N is the number of samples.

2. Repeat for $m = 1,2,...,M$:

(a) Train the weak classifier (ex. stump) $f_m(x) \in \{-1,+1\}$ using weights w_i .

(b) Compute error $err_m = E_w[1_{(y \neq f_m(x))}], c_m = \log((1 - err_m)/err_m)$.

(c) Update weights by $w_i \leftarrow w_i \exp[c_m \cdot 1_{(y_i \neq f_m(x_i))}], i = 1,2,...,N$ and normalize

3. Final strong classifier output is

$$sign[\sum_{m=1}^{M} c_m f_m(x)]$$

Gentle AdaBoost Algorithm (Friedman, Hastie & Tibshirani [9])

1.Initialize weights $w_i = 1/N, \quad i = 1,2,...,N$, where N is the number of samples.

2.Repeat for $m = 1,2,...,M$:

(a) Fit the regression stump $f_m(x)$ by minimizing a weighted squared error

(b) Update the function by $F(x) \leftarrow F(x) + f_m(x)$

(c) Update weights by $w_i \leftarrow w_i e^{-y_i f_m(x_i)}$

3.Final strong classifier output is

$$sign[F(x)] = sign[\sum_{m=1}^{M} f_m(x)]$$

4.1.2 Number of training images

We conduct experiments by changing the number of training images. ROC curve indicating the classifier performance is shown in Figure 4. Due to the difficulty of the database, 10 images are not enough for the high performance. If we use 50 images for training, recognition is almost perfect.

4.1.3 Number of weak classifiers

We also conduct experiment how number of weak classifiers of the boosting affects the performance. As expected, if we use more weak classifiers, performance increases significantly. When 30 images are used for the training, a classifier using 1000 stumps is almost perfect as shown in ROC curve (Figure 5).

4.1.4 Multiclass Recognition

We also perform multiclass face recognition. The system has to identify the specific person out of 10 subjects. We trained 10 one-vs-all binary classifiers and take the maximum output. Results are shown in Table1. As shown in the table, if we have more images available for training, there is no significant difference in performance between SVMs and Boostings. However, the boosting performance is worse than SVMs when less than 10 images are used for training. Figure 6 shows recognition rate of each class. Person 5 and person 6 are difficult due to the much variation in illumination and depth rotation.

# of training images / Methods	5 images	10 images	20 images	30 images
Linear SVM	82.1 %	90.3 %	95.0 %	97.2 %
RBF SVM	**84.3 %**	**92.4 %**	**95.9 %**	**97.2 %**
AdaBoost (200 stumps)	60.8 %	84.9 %	93.8 %	96.7 %
GentleBoost (200 stumps)	68.5 %	87.7 %	93.8 %	96.7 %

Table 1. Recognition performance on MIT face database.

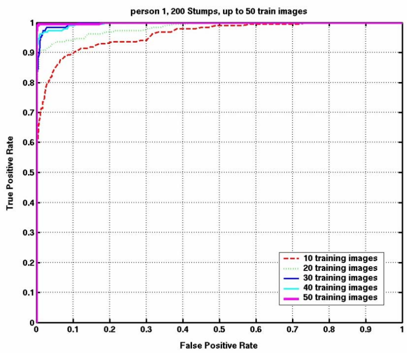

Figure 4. ROC curve of a boosted face classifier using 200 stumps. The performance increases, as more training images are available.

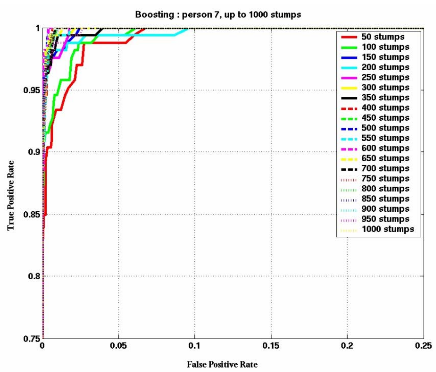

Figure 5. ROC curve as a function of number of weak classifiers. When 1000 stumps are used, performance is almost perfect.

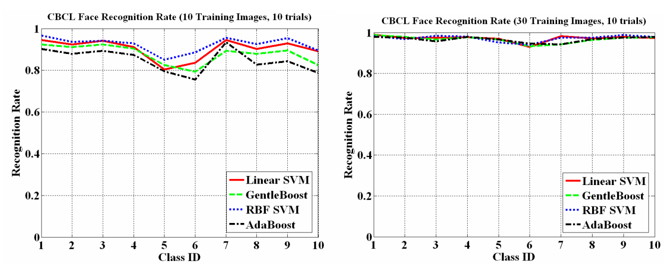

Figure 6. Performance of each person. Figures show average recognition rate on 10 runs when 10 images are used for training (left) and 30 images are used for training (right). Some people are more difficult than the others to recognize. More the training images are available, performance of SVM and boosting getting closer.

4.2. ORL datasets

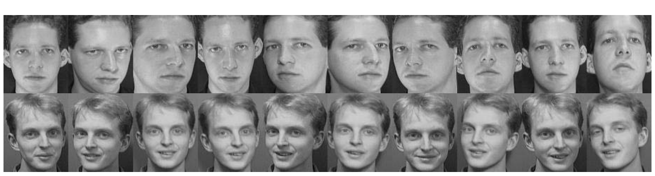

Figure 7. Sample images from ORL face database. The database contains 40 subjects with minor variation of illumination, scale, and pose. Faces are aligned to the center of the image with a uniform background. There are 10 face images per subject.

We also performed multiclass face recognition on the Cambridge ORL face database. This database is an often used benchmark test set for face identification. The images are 112x92 pixels and have minor variation in facial expression and scale and pose. Example images are shown in Figure 7. It contains 40 different subjects with 10 images each. Although this database is relatively easy, compared to the MIT database (since the faces are aligned in the center of the image and the subjects are placed on the uniform background), it is interesting to compare the results with other techniques reported in the literature. When 5 images are used for training and testing, our implementation of one-shot-learning system[19][34][35] achieved 97.5% recognition rate which is state-of-the-art performance for this (quite easy) database. We tested our new approach to this data set. For each subject, N randomly chosen images are used for training and the remaining 10-N images are used for testing. Thus a total of Nx40 images are used for training and (10-N)x40 images are used for the testing. The number of local features depends on the image but approximately 70 features are found in an image. For instance, when 5 images are used for training, there are 70x5=350 features in the feature pool. Maximum correlations are computed for all the features. We trained the one-vs-all SVM classifiers and boosting classifiers for all the 40 people. In the testing, all the test data are classified into one of the 40 categories. The input feature vector is classified by all the 40 classifiers. The classifier with maximum value provides the final decision. We ran the experiment 30 times since the result is slightly different when different images are chosen as training and testing and considered the average result. Table 2 shows the results with other techniques such as Eigenface[10], SOM+CN[10], and ARENA[11]. As we can see from the table, the result is excellent: even with only 1 training image, the recognition rate of the SVM classifier is over 80%. When we use 3 images for training, performance is almost perfect. When 5 images are used for the training, 12 out of 30 runs show100% recognition rate. We should note here that boosting performs poorly compare to SVM. However, when the number of training images is more than 10, the performance of boosting increases significantly. We already showed the result on MIT database containing more number of examples.

# of training images	1 image	3 images	5 images
Eigenface [10]	61.4 %	81.8 %	89.5 %
SOM+CN [10]	70.0 %	88.2 %	96.5 %
ARENA [11]	74.7 %	92.2 %	97.1 %
Linear SVM	84.4 %	96.5 %	99.3 %
RBF SVM	**84.5 %**	**96.8 %**	**99.3 %**
Gentle AdaBoost	74.7 %	74.8 %	78.6 %
One Shot System [19]	50.3 %	82.9 %	97.5 %

Table 2. Recognition performance on the ORL database. Our system outperforms other previously proposed techniques in the literature on this database. In particular, when 1 image is used for training, SVM classifiers show excellent results.

5. Multiview face identification

5.1. Integrating multiview to one model

In the previous section, we applied our system to the frontal face recognition task and showed high recognition performance on the ORL and MIT face databases. The next challenge was to apply the system to multiview face recognition. One way to accomplish multiview recognition from collection of 2D images is to use reference view [21]. In that method, a set of images separated by N degrees in depth rotation are chosen as references. For a planar object such as a painting, even if the viewpoint is changed 45 degrees, the images are still similar enough for detection. On the other hand, for a rigid object like faces, 30 degrees interval might not be too much and we have to tune this interval heuristically. In our approach, this problem is dealt with same way as framework described in the previous section. Local features are extracted from face images taken from the various viewpoint and good features are selected during the learning procedure. The selected features are common features across the view or distinctive features of a certain viewpoint. As noted in the previous section, local features from the positive training examples are used to create a "feature pool" and "correlation features" are computed for all the positive and negative examples.

5.2. Results on CMU PIE face database

Figure. 8 Sample images from CMU PIE database. We use 48 subjects from the database since those subjects have images under two different conditions: "lights" and "facial expression".

CMU PIE is a commonly used dataset [24][26] for evaluation of multiview face identification performance. We use 48 subjects from this dataset, which have face images taken under different light conditions and different facial expressions. Sample images from this database are shown in Fig.8. We cropped face regions and resized to 100x100 pixels in the experiments.

The experiment is conducted by making train/test subsets for each person using one of the conditions (lights or expression) and evaluate the performance by changing number of training samples. For each subject, N randomly chosen images are used for training and the remaining images in the same condition are used for testing. Following the same procedure of the previous section, we first construct the feature pool and train the one-vs-all SVM classifiers and boosting classifiers for all the 48 people. In the testing, the classifier with maximum value provides the final decision. Note that when only one image is used for training, the classifier might capture only one view of the face. Testing is done using all the viewpoints (including frontal and profile and other views). Results are shown in Table 3 and Table 4,. As we can see from the tables, the results are excellent even when the number of training samples is only one, with better than 40% of identification rate when tested on light changes. This means that the one view model allows ± 36 degrees in-depth head pose changes. Since the training images are chosen randomly from all the multiview images, sometimes the classifier cannot model the full view even when 3 images are used for training. This leads to no performance increase despite of the increase in the number of training images; however, when 9 images are used for training, the RBF SVM classifier achieves almost 90% identification correctness, which is a state-of-the-art performance on this database. Fig. 9 shows some of the images from one of the subjects under lighting changes. Fig.10 shows initially detected corner points and selected local features during the boosting procedure. It is interesting to point out that the features in the half-profile views are more likely to be selected rather than the frontal views. This observation implies that near-frontal views can be recognized by a small set of common features across the viewpoints while near-profile views require more view-specific features. Another implication is that local features from half-profile views are more distinctive than frontal view features for the face identification.

Table 4 shows the performance on the "expression" subset. Due to the variation of facial expressions, the result shows lower identification performance than the "light" experiment. We also show the sample images of facial expression variations in Fig. 11.

It is clear that the boosting algorithm performs poorly in most of our experiments when the training samples are few in number. However, recent advances in boosting algorithm [5] shows almost

Figure. 9 Sample images from "lights" subset. This subset has approximately 73 images per person. Even though the lighting condition is changed, normalized correlation of the two local features is still high, thus yielding good identification performance (see Table 3).

Figure. 10 Initially detected corner points (top row) and selected local features (bottom row). Red circle regions indicate the support regions of the local descriptors.

Figure. 11 Sample images from "facial expression" subset. This subset contains approximately 50 images per person with various facial expressions. There are images of some subjects with and without glasses. These factors lead to failure in local feature matching.

# of training images / Methods	1 image	3 images	5 images	9 images
Linear SVM	41.9 %	51.4 %	75.4 %	85.2 %
RBF SVM	**44.2 %**	**55.2 %**	**82.7 %**	**89.4 %**
GentleBoost (200 stumps)	42.6 %	41.0 %	65.7 %	79.4 %

Table 3. Recognition performance on light changes.

# of training images / Methods	1 image	3 images	5 images	9 images
Linear SVM	22.2%	37.3%	47.7%	62.1%
RBF SVM	20.6%	42.1%	54.3%	69.7%
GentleBoost (200 stumps)	24.8%	26.3%	42.2%	59.5%

Table 4. Recognition performance on expression changes.

the same classification performance as SVM and our system can be extended using those algorithms when there are a few training samples available.

6. Conclusion

In this paper, we do not exploit any geometric constraints for the face identification. Local features based on sets of oriented Gaussian derivatives are efficiently implemented by "steerable filters". Positive training images are used to extract local features and build a "feature pool". For each feature in the pool, a maximum correlation is computed to make a "correlation feature" to be trained by SVM and boosting. We applied our system to frontal face identification and multiview face identification. The ORL and MIT CBCL and CMU PIE database showed excellent results in relation to other approaches. It should be emphasized that larger databases should be used to obtain absolute levels of performance that are relevant to some real world situations. Even without geometric information, the system achieves state-of-the-art performance. A motivation for applying the system to multiview face identification is that the system effectively integrates automatically multiview face models into one model by collecting a sufficient number of distinctive and subject-specific local features from the training images. The boosting algorithm selects specific features out of a large feature pool while tuning the threshold of matching each feature. Those are distinctive features in a certain viewpoint as well as the common features across viewpoints. Currently our system needs training images collected from various viewpoints. Future work will consider the generation of multiview faces by the morphable models [23] and possibly weak geometry constraints for the model.

Acknowledgements

Priliminary work of this research was done during the visit to CBCL, MIT with the support of Sony Corporation. The authors would like to thank Bernd Heisele and Lior Wolf for very useful suggestions, Pascal Paysan for preliminary work and the Olivetti Research Laboratory and CMU for providing the face databases.

References

[1] D.J. Beymer. "Face recognition under varying pose". MIT AI Lab, Technical Report, 1993.

[2] D.J. Beymer and T. Poggio, "Face recognition from one example view". Int. Conf. On Computer Vision (ICCV), pp. 500-507, 1995

[3] P.N. Belhumeur, D.J. Kriegman, "What is the set of images of an object under all possible lighting condition?". CVPR pp. 52-58, 1997

[4] Paul Viola and Michael J. Jones. "Robust real-time object detection". Technical Report CRL 2001.

[5] L.Wolf. "Robust boosting for learning from few examples". CVPR 2005

[6] W. Freeman and E. Adelson. "The design and use of steerable filters". PAMI, 13(9):891–906, 1991.

[7] D. Lowe. "Distinctive image features from scale-invariant keypoints". IJCV, 2004

[8] C. Morgenstern and B.Heisele, "Component Based Recognition of Objects in an Office Environment", AI Memo, MIT CBCL, 2003

[9] J. Friedman, T. Hastie, R. Tibshirani, "Additive Logistic Regression: a Statistical View of Boosting". Tech Report, Dept. of Statistics, Stanford Univ, 1998

[10] S. Lawrence, C. Giles, A. Tsoi, A, Back, "Face Recognition: A Convolutional Neural Network Approach". Neural Networks, Vol.8, Number 1, pp.98-113, 1997

[11] T. Sim, R. Sukhankar, M. Mullin, S. Baluja, "High-Performance Memory-based Face Recognition for Visitor Identification". Just Research Technical Report, 1999

[12] T. Sim, R. Sukthankar, M. Mullin, S. Baluja, "Memory-based face recognition for visitor identification". IEEE Int. Conf. On Face and Gesture 2000 (FG2000)

[13] C. Wallraven, B. Caputo, A. Graf. "Recognition with Local features: the kernel recipe". ICCV2003

[14] Gabriela Csurka, Cedric Bray, Chris Dance, Lixin Fan. "Visual categorization with bags of keypoints", ECCV 2004

[15] Antonio Torralba, Kevin Murphy and William Freeman. "Sharing features: efficient boosting procedures for multiclass object detection"., Int. Conf. On Computer Vision and Pattern Recognition (CVPR2004)

[16] Heisele, B., P. Ho, J. Wu and T. Poggio. "Face Recognition: Component-based versus Global Approaches". Computer Vision and Image Understanding, Vol. 91, No. 1/2, 6-21, 2003.

[17] Lei Zhang, Stan Z. Li, ZhiYi Qu, X. Huang. "Boosting Local Feature Based Classifiers for Face Recognition". Workshop on Face Processing in Video 2004.

[18] Peng Yang, Shiguang Shan, Wen Gao, Stan Li, Dong Zhang, "Face Recognition Using Ada-Boosted Gabor Features", IEEE Int. Conf. On Automatic Face and Gesture Recognition (FG2004), pp356-361, 2004

[19] Jerry Jun Yokono and Tomaso Poggio. "Oriented filters for Object Recognition: an empirical study". Proceedings of the IEEE Conference on Face and Gesture Recognition (FG2004). Seoul, Korea. 2004.

[20] T. Maurer and C. Malsburg. "Single-view based recognition of faces rotated in depth". Workshop on Automatic Face and Gesture Recognition, pp.176-181, 1996

[21] A. Pentland and B.Moghaddam and T.Starner. "View-based and modular eigenspaces for face recognition". Int. Conf. On Computer Vision and Pattern Reocognition (CVPR), 1994

[22] T.Vetter and T.Poggio. "Linear object classes and image synthesis from a single example image". IEEE PAMI, 19, pp.733-742, 1997

[23] V.Blanz and T.Vetter. "A morphable model for the synthesis of 3D faces". SIGGRAPH'99, pp.187-194

[24] V.Blanz and S.Romdhani and T.Vetter. "Face identification across different poses and illuminations with a 3D morphable model".

[25] W.Zhao and R.Chellappa. "SFS based view synthesis for robust face recognition". Int. Conf. On Face and Gesture Recognition (FG), 2000

[26] R.Gross, I.Matthews, and S.Baker. "Eigen Light-Fields and Face Recognition Across Pose". Int. Conf. on Automatic Face and Gesture Recognition (FG), 2002

[27] T.F.Cootes, G.J. Edwards and C.J.Taylor. "Active Appearance Models", IEEE PAMI, Vol.23, No.6, pp.681-685, 2001

[28] T. Kanade. "Computer recognition of human faces". Interdisciplinary Systems Research, Vol.47, 1977

[29] M.Turk and A.Pentland. "Eigenfaces for recognition". Journal of Cognitive Neuroscience, 3, pp.72-86, 1991

[30] B.Moghaddam and A.Pentland. "Probabilistic visual learning for object representation". PAMI, 19, pp.696-710

[31] L.Wiskott, J.Fellous, and Von Der Malsburg. "Face recognition by elastic bunch graph matching". IEEE Trans. PAMI 19, pp.775-779

[32] Y. Freund, R. Schapire, "Experiments with a new boosting algorithm". IEEE Int. Conf. On Machine Learning, pp.148-156, 1996

[33] Yokono, J.J. and T. Poggio. Boosting a Biologically Inspired Local Descriptor for Geometry-free Face and Full Multi-view 3D Object Recognition, *CBCL Paper #254/AI Memo #2005-023*, Massachusetts Institute of Technology, Cambridge, MA, July 2005.

[34] Yokono, J.J. and T. Poggio. Rotation Invariant Object Recognition from One Training Example, *CBCL Paper #238/AI Memo #2004-010*, Massachusetts Institute of Technology, Cambridge, MA, April, 2004.

[35] Yokono, J.J. and T. Poggio. Evaluation of Sets of Oriented and Non-oriented Receptive Fields as Local Descriptors, *CBCL Paper #237/AI Memo #2004-007*, Massachusetts Institute of Technology, Cambridge, MA, March, 2004.