A New Method For Combined Face Detection And Identification Using Interest Point Descriptors

Sebastian Stein
School Of Computing
University Of Dundee
sstein@computing.dundee.ac.uk

Gernot A. Fink
Department Of Computer Science
TU Dortmund
gernot.fink@tu-dortmund.de

Abstract— **Although face recognition has been an active research area for decades, the general problem is still unsolved. While many methods addressing either face detection or face identification emerged, little attention has been paid to the combination or integration of detection and identification.**

In this paper we propose a combined method for face detection and identification using SIFT descriptors. This combined method includes an existing detection model and a new identification method based on object class invariants (OCIs), which is invariant to translation, scale, in-plane rotation and small 3D viewpoint changes. These models are combined using a bounding box around the OCI to filter face features for identification. We show the highly competitive performance of the newly developed identification method and the effectiveness of the proposed combination scheme.

I. Introduction

In computer vision, face recognition attracts the attention of researchers since more than 35 years and is still an unsolved problem [20]. This history reflects not only the complexity involved in automating this task, but also the strong interest in machine recognition systems, which is mainly driven by the wide range of potential application areas.

For law enforcement, surveillance and authentication systems, face recognition provides an intuitive and rather userfriendly identification method. It does not rely on the participants cooperation or even his or her knowledge, as opposed to identification based on fingerprints and iris scans. Furthermore, face recognition systems may take advantage of the numerous surveillance cameras already present in public and private areas today. In the context of social networking platforms, on which vast amounts of face images are accumulated, automatic photo tagging functionality based on face recognition greatly improves the user experience [5]. In many other areas such as virtual reality, ubiquitious computing, human-machine-interaction and entertainment, numerous useful and creative applications are imaginable.

The major challenge for face recognition systems today is to cope with the substantial face appearance variations caused by changes in viewpoint and illumination as well as by partial occlusion.

The face recognition problem consists of two distinct tasks, which are generally addressed separately in the literature:

This research has been performed when Sebastian Stein was working towards his M.Sc. degree at TU Dortmund.

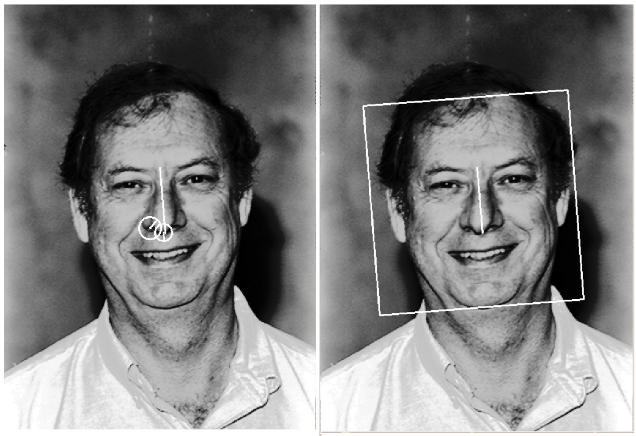

Fig. 1. Recognition process: Prediction of a vector from the base of the nose to the forehead (OCI) based on SIFT features (left) 2) Construction of a bounding box around the OCI (right) 3) Identification based on features inside the bounding box (not illustrated).

1) *Detect* and localize faces in a scene image
2) *Identify* faces of known individuals and discard unknown faces

Although many sophisticated approaches to solving either of these subproblems emerged during the past decades (see e.g. [7], [8], [19], [20]), to the knowledge of the authors no previous research has been focussing on the combination of face detection and identification.

In this paper we propose a system combination scheme for face detection and identification based on an object class invariant (OCI) model [14] and interest point descriptors (see Fig. 1).

The remainder of this paper is organized as follows: Section II reviews related work on face recognition based on interest point methods. Section III describes the face detection model, our face identification model and our system combination scheme. In Section IV we presents our evaluation results.

II. Related Work

In this section we will mainly focus on interest point methods and refer the interested reader to [7], [8], [19], [20] for a more general introduction to face recognition.

Interest points, also known as keypoints, are mathematically well-defined image landmarks, which are likely to be found in different images of the same object. Interest point descriptors, which are the *features* extracted at these locations, describe the local image content and are designed to be stable under certain image transformations such as scaling or rotation. Prominent representatives are the *Scale Invariant Feature Transform* (SIFT), proposed in [9], and *Speeded Up Robust Features* (SURF), introduced in [2].

A. Face Detection

Many face detection methods such as [17] are single viewpoint in nature and yield no pose annotation of the face in a detected region, although estimation of the face pose may be beneficial for further processing detected faces (e.g. identification). In [11], LeCun et al. integrate face detection and pose estimation using convolutional neural networks requiring large amounts of annotated data for training. Toews and Arbel propose in [16] a probabilistic face detection model requiring a smaller number of training images and predicting the vector from the nose to the forehead, which may be seen as a simple form of pose estimation. This vector is used as a common reference geometry (OCI) to model geometrical feature densities with respect to the face pose. Model training involves determining clusters of SIFT features in appearance space which are distinctive for their respective positions relative to the OCI. By matching features from a test image to these model features, OCI geometries are projected onto the test image and considered to be true face hypotheses if they pass a validation based on a probabilistic face model.

This model serves as the detection module for our recognition system and will be described in more detail in Section III-A, as we altered the model formulation for better understandability and used a different clustering method for model training.

B. Face Identification

Face identification models based on interest point descriptors generally create *face templates* by storing the sets of features extracted from face training images in a database. A *matching strategy* defines the subset of template features which is matched with a feature extracted from a test image in order to avoid false feature correspondences. Finally, a *similarity measure* defines a scalar similarity between the set of test features and a face template.

In [4], Bicego et al. compare three different matching schemes using SIFT descriptors, namely minimum pair distance matching, matching features around the eyes and the mouth, and matching on a regular grid with overlapping subregions. Only test and training features of corresponding subregions are matched. These matching strategies have been evaluated on the BANCA[1] database, where regular grid matching performed best (henceforth refered to as SIFT_Grid).

Luo et al. refine the approach of matching corresponding face regions using SIFT in [10]. They divide the face image into five subregions applying K-means clustering to the image positions of interest points from face training images. In the identification stage, each feature is assigned to the cluster with the closest center location and only features belonging to the same cluster are matched (henceforth refered to as SIFT_Cluster). Their comparative evaluation carried out on the FERET database [13], [12] shows a superior performance of this approach using five clusters over SIFT_Grid (evaluation results are presented in Section IV in the context of our evaluation).

Rather than dividing the image into fixed subregions, Du et al. propose in [6] a sliding window approach in combination with SURF descriptors. Only template features lying within a specified search window around the image position of the currently inspected test feature are considered for matching. Their unfortunately very scarce evaluation results indicate a lower identification accuracy compared to SIFT_Cluster using the standard 64-dimensional SURF descriptor and a comparable accuracy when using the augmented 128-dimensional SURF descriptor.

In this paper we introduce a different region based matching strategy using the OCI. The major contributions of this work may be summarized as follows: We develop a new face identification model using SIFT features and the OCI, which does not rely on conventional face alignment. With this identification model using the pose predicted by the face detection method introduced above, we develop a combination scheme based on a bounding box around the OCI, which is invariant to scale, location and in-plane rotation.

III. Combined Face Detection And Identification

A. Face Detection

In this Section we briefly describe the formulation of the face detection model invented by Toews and Arbel [14], [15], [16], the model training procedure and the geometrical relationships between the OCI and model features.

1) Model Formulation: A model feature represents a cluster of interest point descriptors in appearance space and image space (i.e. feature geometry). A cluster in appearance space is described by its mean appearance a_i and a radius T_i^a. In terms of geometry, a feature cluster is described by its mean geometry g_i and a set of global thresholds T^g : $(T^{pos}, T^{scale}, T^{angle})$, delimiting a cluster in its 2D position, scale and angle, respectively. A model feature may thus be described by the tuple f_i : $(a_i, T_i^a, g_i, p^{face}, p^{bg})$, where p^{face} and p^{bg} denote the feature occurrence probabilities with respect to the classes $C := \{face, bg\}$, corresponding to faces and the background.

The common reference frame (OCI) is described in terms of its geometry g_o : $(x_o, y_o, \sigma_o, \theta_o)$, that is its 2D image position, length and angle.

The posterior probability of a class $c \in C$ given a set of features $\{f_i\}$ extracted from a test image is described as

$$P(c|\{f_i\}) = \frac{P(\{f_i\}|c)P(c)}{P(\{f_i\})} = \frac{P(c)\prod_i P(f_i|c)}{\prod_i P(f_i)}, \qquad (1)$$

[1] Available at http://www.ee.surrey.ac.uk/CVSSP/banca/

assuming conditional feature independance. A face hypothesis $H : (g_o, \{f_i\})$ is evaluated using Bayes' decision ratio

$$\gamma(H) = \frac{P(face|\{f_i\})}{P(bg|\{f_i\})} = \frac{P(face)}{P(bg)} \prod_i \frac{P(f_i|face)}{P(f_i|bg)}. \quad (2)$$

Following the notation introduced above, the individual feature probabilities $P(f_i|c)$ expand to

$$P(f_i|face) = P(a_i|face)P(g_i|face, g_o)p_i^{face} \quad (3)$$

$$\text{and}\, P(f_i|bg) = P(a_i|bg)P(g_i|bg)p_i^{bg}, \quad (4)$$

for features representing a face and the background. For further information see [16].

2) Geometrical Relationships: In order to compare features derived from different images in terms of their geometries, their *normalized* geometries are calculated with respect to the normalized OCI geometry $g_o : (x_o = 0, y_o = 0, \sigma_o = 1, \theta_o = 0°)$, which involves shifting, scaling and rotating the feature geometry according to the labeled OCI geometry.

A feature f_j is considered to *agree geometrically* with a feature f_i, if their normalized geometries differ by less than the global geometrical thresholds $T^g : (T^{pos}, T^{scale}, T^{angle})$:

$$\begin{aligned} GeoAgg(f_i, f_j) =& (\,|\, x_i - x_j \,| < T^{pos} \cdot \sigma_i \,) \wedge \\ & (\,|\, y_i - y_j \,| < T^{pos} \cdot \sigma_i \,) \wedge \\ & (\,|\, log\, \sigma_i - log\, \sigma_j \,| < T^{scale} \,) \wedge \\ & (\,|\, \theta_i - \theta_j \,| < T^{angle} \,) \end{aligned} \quad (5)$$

Using the normalized geometrical relationship of a model feature to its OCI and the absolute geometry of a corresponding feature extracted from a test image, we can *project* an OCI geometry into absolute image dimensions. At first, we need to calculate the OCI geometry with respect to a normalized feature geometry $g_o : (x_o = 0, y_o = 0, \sigma_o = 1, \theta_o = 0°)$. In a second step, this OCI geometry is shifted, scaled and rotated according to the absolute geometry of the test feature.

3) Feature Clustering: Rather than applying a mean-shift on feature appearance and geometry for feature clustering as in [16], we propose to use a two step clustering approach:

At first, features are clustered in terms of their appearance using agglomerative clustering: All face training features are considered as potential model feature clusters. In each iteration, the pair of remaining clusters with smallest Euclidean distance is merged. The resulting mean appearance a' is calculated as the weighted mean appearance of the original clusters:

$$a' = \frac{n_1 a_1 + n_2 a_2}{n_1 + n_2}, \quad (6)$$

where n_1 and n_2 denote the number of training features corresponding to the respective clusters. This process halts when the minimum distance between any pair of remaining clusters exceeds a predefined threshold T^d.

After appearance clustering, all clusters with less than two support features are discarded. The remaining clusters are inspected individually. Given an appearance cluster i with supporting training features f_j^i, we seek to determine whether for any geometry g_j^i more than half of the support features agree geometrically. Otherwise the appearance cluster i is considered to be not distinctive for any geometry and is discarded. Thus, we count for each geometry g_j^i the number of geometrically agreeing features $f_k^i; k \neq j$. If the condition defined above holds for any geometry g_j^i, the final cluster geometry is calculated by averaging the geometries of all support features, which agree geometrically with g_j^i.

Given the mean appearance, the mean geometry and the supporting training features for a particular cluster, all remaining model parameters may be calculated. The appearance cluster radius T_i^a is set to the maximum distance of all support features to the cluster mean appearance. The appearance probabilities $P(a_i|c)$ and the geometrical probabilities $P(g_i|face, g_o)$ are assumed to be Gaussian, while occurrence probabilities p_i^{face} and p_i^{bg} are binomial. The geometrical distribution $P(g_i|bg)$ is assumed to be uniform, as background features may appear anywhere in the image equally likely.

B. Face Identification Using The OCI

In this section we present our face identification model using the OCI.

A person specific face template consists of a set of features extracted from a face training image $\{f_j\}_i^{temp} = \{f_j\}^{training}$. In order to use the OCI geometry as a common reference frame, feature geometries are replaced by their normalized counterparts with respect to the labeled OCI geometries in the training images.

Similarly, all features extracted from a test image are geometrically normalized to their respective OCI geometries.

As all feature geometries - those derived from the test image and those stored in the database - are described with respect to a common reference frame, we define a region based feature selection method as follows: Each test feature f_k^{test} is matched with the subset of features of a template f_j^{temp}, whose normalized geometries g_j^{temp} agree with the normalized geometry g_k^{test} of the test feature.

This region based matching strategy differs in several ways from the ones proposed by other authors: Rather than fixing the number of regions (i.e. the number of grid cells or the number of geometrical clusters), our approach keeps the *size* of a region fixed while the number of distinct regions is undefined. Furthermore, regions derived from geometrical agreement are not only bounded in 2D pixel coordinates but also in angle and scale dimensions. These additional restrictions may further reduce the number of false correspondences (depending on the choice of the thresholds T^g for geometrical agreement). Assuming that the OCI geometry is known (either labeled by hand or located by the previously described face detector), this region based matching strategy is applicable in uncontrolled environments without further alignment or registration.

For measuring the similarity between a set of test features and a face template, we apply a simple voting scheme similar to [9]: All test features f_k^{test} are matched with all geometrically agreeing features f_j^i of all templates i in the database. For each test feature f_k^{test}, we determine the template feature f_k^{min} with minimum Euclidean distance $d(f_k^{test}, f_j^i)$. We define a function $temp(f_k^{min})$, which maps the feature f_k^{min} to the template i that includes f_k^{min}. A test feature f_k^{test} *votes* for a particular subject i, if f_k^{min} corresponds to i:

$$vote(f_k^{test}, i) = 1\left[temp(f_k^{min}) = i\right], \tag{7}$$

where $1\left[expression\right]$ is one if the expression is true and zero otherwise. Although Lowe proposes in [9] to discard ambiguous feature matches, our experimental evaluation showed that this verification procedure reduces identification performance and is therefore not incorporated into our voting scheme. The final similarity between a set of test features and a face template i is defined as the number of test features voting for subject i:

$$S(\{f_k\}^{test}, i) = \sum_k vote(f_k^{test}, i) \tag{8}$$

In a closed identification scenario, where all faces in test images are assumed to be known, the subject corresponding to the template with the highest similarity measure

$$argmax_i S(\{f_k\}^{test}, i) \tag{9}$$

is considered to be present in the image.

C. System Combination

In order to combine the face detection and face identification models, we need to select a subset of features extracted from a scene image which shall be used for identification. An optimal system combination scheme would discard all background features and keep only the features extracted from the face we intend to identify. We approach this objective by defining a bounding box about the predicted OCI geometry, balancing the number of falsely rejected face features and the number of falsely kept background features. Recall that feature selection for identification is performed using *geometrical agreement* as defined in Section III-A. With the geometrical thresholds $T^g : (T^{pos}, T^{scale}, T^{angle})$ and the normalized geometry of a template feature f_i, we can determine the normalized bounding box $BB_i : (x_i^{min}, y_i^{min}, x_i^{max}, y_i^{max})$ in horizontal and vertical direction about the normalized feature position g_i as follows:

$$\begin{aligned} x_i^{min} &= x_i - T^{pos} \cdot \sigma_i \qquad & y_i^{min} &= y_i - T^{pos} \cdot \sigma_i \\ x_i^{max} &= x_i + T^{pos} \cdot \sigma_i \qquad & y_i^{max} &= y_i + T^{pos} \cdot \sigma_i \end{aligned} \tag{10}$$

Based on the set of feature specific bounding boxes $\{BB_i\}$ corresponding to all template features, we can determine the bounding box BB_Max including all positions normalized to the OCI, which are relevant for identification.

In several cases this bounding box is too wide and tends to include a fairly large number of unrelated features. Consider for example scene images with overlapping faces or training images with inprecisely defined face regions. Therefore, we determine a bounding box BB_Max_i for each template $i = 1, \cdots, N$ in the database and derive in a second step the average bounding box BB_Avg of all BB_Max_i.

Now consider a set of test features $\{f_k^{test}\}$ with geometries g_k^{test} in absolute image dimensions. The detector localizes a face with OCI geometry g_o. Given a normalized bounding box $BB : (x^{min}, y^{min}, x^{max}, y^{max})$, how can we determine the subset of features of $\{f_k^{test}\}$ which are inside the bounding box BB around the detected OCI geometry? At first, we need to normalize the whole set of test features with respect to the OCI geometry g_o as described in Section III-A. These locations can easily be compared to the boundaries defined by bounding box BB and features located outside BB may be discarded. In summary, we project test feature geometries into the normalized image space and filter all features with respect to the normalized bounding box.

IV. Evaluation

In this section we present evaluation results of detection, identification and combined detection and identification on the FERET database [13], [12].

A. Data Set

The FERET database is a standard large scale face database widely used to evaluate face recognition algorithms. The Gray FERET database consists of 14051 grayscale images of 1169 subjects under varying viewing conditions. The collection of subjects is very diverse in terms of age and ethnicity. The photographs have been recorded in a total of 15 sessions over the course of three years. The images are neither aligned nor cropped to the face region. Therefore, ground truth OCI geometries highly vary between images! Pictures are taken in front of a neutral background, but show a more or less large part of upper body clothing. Some subjects wear glasses or jewelery and some have facial hair.

The evaluation protocol defines a gallery set of images comprising one image per individual for training and several *partitions* (test sets) corresponding to different face variations and viewing conditions. In this paper we make use of the partition *fb* showing each individual with a different facial expression than in the training set, the partitions *dup1* and *dup2* consisting of frontal face images of subjects collected in later sessions and the partition *fc* showing faces under different illumination conditions.

The Color FERET database contains a subset of Gray FERET images in color format, higher resolution and without lossy compression.

For our experiments we labeled all images used for training and testing with the OCI. Additionally, we manually defined for all images a paraxial rectangular *region of interest* (ROI) enclosing the face region. This ROI is used for detection to separate face features from background features, and for identification to discard features corresponding to

Fig. 2. Example detection: 1) Labeled ground truth OCI (left) 2) predicted OCI based on four SIFT features (right).

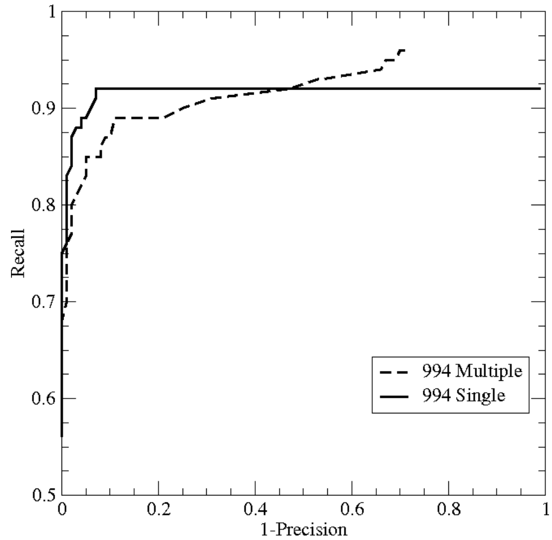

Fig. 3. Detector performance allowing multiple/ single detector responses

clothing. As subjects wear the same clothes in some test images as in the gallery, including these features artificially improves identification performance. All images of the Color FERET database have been converted to gray scale images. SIFT feature extraction is performed using David Lowe's demo software[2].

B. Detection

For detector model training, all features inside the ROI are used as face model features, and all remaining features are used as background features. The appearance clustering threshold T^d was set to 0.5, as this threshold produced in our experiments the maximum number of appearance clusters consisting of at least two face model features. The geometry thresholds were set to $T^g : (T^{pos} = 0.5, T^{scale} = log(1.5), T^{angle} = \pi/2)$ as in [16]. Evaluation results are visualized with Recall vs. 1-Precision plots.

We consider a positive response (i.e. an OCI geometry g_o) as true positive, if g_o agrees geometrically with the ground truth OCI g_o^{gt}. For multiple detector responses, we count only the first true positive response as such and count all other responses regardless their geometrical agreement with the ground truth as false positives. Note that this method represents a *conservative* performance measure.

With respect to a combined detection and identification system which is trained on the same subjects, we train our detection model with the full Color FERET gallery set. For evaluation, we used the test set on expression variations *fb*. Figure 3 shows evaluation results for the cases allowing multiple and only a single detector response (exploiting the fact that there is only a single face in each image). In the case of multiple responses, we achieve 89% recall with a precision of 89%. Allowing just a single detector response, the maximum recall of 92% is obtained with 93% precision. Note that detector evaluation on FERET does not show *true* detection performance but rather indicates localization accuracy. An example detection result is shown in Fig. 2. A correct OCI is detected based on less than ten SIFT features on average.

[2]Available at: http://www.cs.ubc.ca/lowe/keypoints/

TABLE I

IDENTIFICATION PERFORMANCE COMPARISON

Method	**fb**	**fc**	**dup1**	**dup2**
Fisherface [3]	0.94	0.73	0.55	0.31
EBGM [18]	0.90	0.42	0.46	0.24
Local Binary Patterns [1]	0.97	0.79	0.66	0.64
Local Matching Gabor [21]	0.99	0.99	0.85	0.79
SIFT_Grid [4]	0.94	0.35	0.53	0.36
SIFT_Cluster [10]	0.97	0.47	0.61	0.53
SIFT_OCI	0.99	0.92	0.69	0.61

C. Identification

We test the identification model using the labeled ROI and OCI in training and test images. All features of a training image inside the ROI are normalized to the OCI geometry and stored in the database. From each test image, we extract solely features inside the ROI and normalize their geometries to the OCI. Then, we compute the similiarity measures between the set of test features and all templates, and consider the template with the highest similarity as being present in the image. The geometrical thresholds for feature matching are chosen as above. Before feature extraction, all images have been scaled by a factor of two in each dimension and their histograms have been equalized.

We chose to evaluate the performance of the proposed identification model on the Gray FERET database, as many other methods for face identification have been evaluated on this data set. Table I compares the identification accuracy (as true positive rate) of our identification model (SIFT_OCI) with other SIFT based identification methods (SIFT_Grid and SIFT_Cluster), Fisherfaces, Elastic Bunch Graph Matching (EBGM), Local Binary Patterns and Local Matching Gabor. The proposed method outperforms the other SIFT based models on all test sets and shows competitive results to

TABLE II
PERFORMANCE OF COMBINED DETECTION AND IDENTIFICATION

Test Set	Precision	Recall	ID/ Det.	ID/ Exp.
fb	0.9284	0.9153	0.9920	0.9079
fc	0.9057	0.8927	0.9165	0.8181
dup1	0.8979	0.8764	0.6832	0.5987
dup2	0.8407	0.8333	0.5974	0.4978

other established approaches. Note that our preprocessing differs slightly from the others: While we scale face images by a fixed factor, the other considered methods scale face images to occupy a fixed region and align images using the coordinates of the eye centers. While all information we utilize for identification may be acquired by a face detector, accurate localization of eye centers needs to be accomplished in addition to face detection if applied to scene images.

D. Combined Detection and Identification

For the evaluation of the combined model, we use the average template bounding box BB_Avg for feature selection. In our experiments we determine the fraction of correctly detected and correctly identified faces of all face images (ID/Exp.) as well as the fraction of correctly identified faces of all detected faces (ID/Det.). While the former measure corresponds to the performance of the whole recognition system, the latter indicates the portion of the overall system performance achieved by the identification module. The evaluation results are presented in Table II. With respect to detector performance, we see that precision and recall highly depend on the test set. The identification performance (ID/Det.) is comparable to the results achieved with pure face identification (Table I). Thus, the difference in identification accuracy of pure identification and combined detection and identification is almost exclusively due to the detection rate. Therefore, the proposed system combination approach may be considered as being very effective, as the performance decrease is not caused by the system combination but by the rather low detection performance.

V. CONCLUSION

In this article we proposed a new face identification method based on SIFT features and object class invariants and showed its competitive performance on the FERET database. This identification model uses labeled or detected OCI information such that conventional image alignment and rasterization is unnecessary. We further adapted the face detector proposed by Toews and Arbel and developed a system combination scheme based on bounding boxes, which are invariant to scale, location and in-plane rotation. This full recognition model may potentially be used with any type of interest point descriptors, although this still needs to be evaluated.

We chose to use the FERET database in our experiements for comparison with other identification methods. To the knowledge of the authors, there is currently no database of scene images publicly available showing each subject in multiple images, which would be more appropriate for combined face detection and identification. This aspect will particularly be addressed in future work.

REFERENCES

[1] Timo Ahonen, Abdenour Hadid, and Matti Pietikäinen. Face recognition with local binary patterns. *Lecture Notes in Computer Science: Proceedings of the 8th European Conference on Computer Vision, Prague, Czech Republic*, 3021:469–481, 2004.

[2] Herbert Bay, Tinne Tuytelaars, and Luc Van Gool. SURF: Speeded up robust features. *European Conference On Computer Vision, Graz, Austria*, 2006.

[3] Peter N. Belhumeur, João P. Hespanha, and David J. Kriegman. Eigenfaces vs. Fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:7:711–720, 1997.

[4] Manuele Bicego, Andrea Lagorio, Enrico Grosso, and Massimo Tistarelli. On the use of SIFT features for face authentication. *IEEE International Conference on Computer Vision and Pattern Recognition Workshop, New York City, USA*, 2006.

[5] Jae Young Choi, Wesley De Neve, Yong Man Ro, and Konstantinos N. Plataniotis. Face annotation for personal photos using collaborative face recognition in online social networks. *International Conference on Digital Signal Processing, Santorini, Greece*, pages 240–247, 2009.

[6] Geng Du, Fei Su, and Anni Cai. Face recognition using SURF features. *SPIE The Sixth International Symposium on Multispectral Image Processing and Pattern Recognition,*, 7496, 2009.

[7] Erik Hjelmås. Face detection: A survey. *Computer Vision and Image Understanding*, 83:236–274, 2001.

[8] Anil K. Jain and Stan Z. Li. *Handbook Of Face Recognition*. Springer-Verlag New York, Inc.

[9] David G. Lowe. Distinct image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60:2:91–110, 2004.

[10] Jun Luo, Yong Ma, Erina Takikawa, Shihong Lao, Masato Kawade, and Bao-Liang Lu. Person-specific SIFT features for face recognition. *IEEE International Conference on Acoustics, Speech and Signal Processing, Honolulu, Hawaii, USA*, 2:593–596, 2007.

[11] Margarita Osadchy, Yann LeCun, and Matthew L. Miller. Synergistic face detection and pose estimation with energy-based models. *The Journal of Machine Learning Research*, 8, 2007.

[12] P.J. Philipps, H. Moon, S.A. Rizvi, and P.J. Rauss. The FERET evaluation methodology for face recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22:1090–1104, 2000.

[13] P.J. Phillips, H. Wechsler, J. Huang, and P. Rauss. The FERET database and evaluation procedure for face recognition algorithms. *Image and Vision Computing*, 16:5:295–306, 1998.

[14] Matthew Toews and Tal Arbel. Detection over viewpoint via the object class invariant. *IEEE International Conference on Pattern Recognition, Quebec City, Canada*, 1:765–768, 2006.

[15] Matthew Toews and Tal Arbel. Detecting, localizing and classifying visual traits from arbitrary viewpoints using probabilistic local feature modeling. *IEEE Workshop Analysis and Modeling of Faces and Gestures, Rio de Janeiro, Brazil*, pages 154–167, 2007.

[16] Matthew Toews and Tal Arbel. Detection, localization, and sex classification of faces from arbitrary viewpoints and under occlusion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31:1567–1581, 2009.

[17] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. *IEEE International Conference on Computer Vision and Pattern Recognition, Kauai Marriott, Hawaii, USA*, pages 511–518, 2001.

[18] Laurenz Wiskoktt, Jean-Marc Fellous, Norbert Krüger, and Christoph von der Malsburg. Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19:7:775–779, 1997.

[19] Ming-Hsuan Yang, David J. Kriegman, and Narendra Ahuja. Detecting faces in images: A survey. 24(1):34–58, 2002.

[20] W. Zhao, R. Chellappa, A. Rosenfeld, and P.J. Phillips. Face recognition: A literature survey. *ACM Computing Surveys*, pages 399–458, 2003.

[21] Jie Zou, Quiang Ji, and George Nagy. A comparative study of local matching approach for face recognition. *IEEE Transactions on Image Processing*, 16(10):2617–2628, 2007.