SURF-Face: Face Recognition Under Viewpoint Consistency Constraints

Philippe Dreuw
dreuw@cs.rwth-aachen.de
Pascal Steingrube
steingrube@cs.rwth-aachen.de
Harald Hanselmann
hanselmann@cs.rwth-aachen.de
Hermann Ney
ney@cs.rwth-aachen.de

Human Language Technology and Pattern Recognition
RWTH Aachen University
Aachen, Germany

Abstract

We analyze the usage of Speeded Up Robust Features (SURF) as local descriptors for face recognition. The effect of different feature extraction and viewpoint consistency constrained matching approaches are analyzed. Furthermore, a RANSAC based outlier removal for system combination is proposed. The proposed approach allows to match faces under partial occlusions, and even if they are not perfectly aligned or illuminated.

Current approaches are sensitive to registration errors and usually rely on a very good initial alignment and illumination of the faces to be recognized.

A grid-based and dense extraction of local features in combination with a block-based matching accounting for different viewpoint constraints is proposed, as interest-point based feature extraction approaches for face recognition often fail.

The proposed SURF descriptors are compared to SIFT descriptors. Experimental results on the AR-Face and CMU-PIE database using manually aligned faces, unaligned faces, and partially occluded faces show that the proposed approach is robust and can outperform current generic approaches.

1 Introduction

In the past, a large number of approaches that tackled the problem of face recognition were based on generic face recognition algorithms such as PCA/LDA/ICA subspace analysis [20], or local binary pattern histograms (LBP) [19] and its extensions. The discrete cosine transform (DCT) has been used as a feature extraction step in various studies on face recognition, where the proposed local appearance-based face recognition approach in [4] outperformed e.g. the holistic approaches. Nowadays, illumination invariance, facial expressions, and partial occlusions are one of the most challenging problems in face recognition [5, 12, 24], where face images are usually analyzed locally to cope with the corresponding transformations.

Local feature descriptors describe a pixel in an image through its local neighborhood content. They should be distinctive and at the same time robust to changes in viewing conditions. Many different descriptors and interest-point detectors have been proposed in the literature, and the descriptor performance often depends on the interest point detector [15]. Recently, a comparative study in [3] has shown the superior performance of local features for face recognition in unconstrained environments.

The SIFT descriptor [10] seems to be the most widely used descriptor nowadays, as it is distinctive and relatively fast to compute. SIFT has been used successfully for face authentication [2], SIFT face-specific features have been proposed in [11], and extensions towards 3D Face recognition have been presented e.g. in [14].

Due to the global integration of Speeded Up Robust Features (SURF) [1], the authors claim that it stays more robust to various image perturbations than the more locally operating SIFT descriptor. SURF descriptors have been used in combination with a SVM for face components [9] only. However, no detailed analysis for a SURF based face recognition has been presented so far.

We provide a detailed analysis of the SURF descriptors for face recognition, and investigate whether rotation invariant descriptors are helpful for face recognition. The SURF descriptors are compared to SIFT descriptors, and different matching and viewpoint consistency constraints are benchmarked on the AR-Face and CMU-PIE databases. Additionally, a RANSAC based outlier removal and system combination approach is presented.

2 System Overview

Local features such as SIFT or SURF are usually extracted in a sparse way around interest points. First, we propose our dense and grid-based feature extraction for face recognition. Second, different matching approaches are presented.

2.1 Feature Extraction

Interest Point Based Feature Extraction. Interest points need to be found at different scales, where scale spaces are usually implemented as an image pyramid. The pyramid levels are obtained by Gaussian smoothing and sub-sampling. By iteratively reducing the image size, SIFT [10] uses a Difference of Gaussians (DoG) and Hessian detector by subtracting these pyramid layers. Instead, in SURF [1] the scale space is rather analyzed by up-scaling the integral image [23] based filter sizes in combination with a fast Hessian matrix based approach. As the processing time of the filters used in SURF is size invariant, it allows for simultaneous processing and negates the need to subsample the image hence providing performance increase.

Grid-Based Feature Extraction. Usually, a main drawback of an interest point based feature extraction is the large number of false positive detections. This drawback can be overcome by the use of hypothesis rejection methods, such as RANSAC [7] (c.f. subsection 2.4).

However, in face recognition an interest point detection based feature extraction often fails due to missing texture or ill illuminated faces, so that only a few descriptors per face are extracted. Instead of extracting descriptors around interest points only, local feature descriptors are extracted at regular image grid points who give us a dense description of the image content.

2.2 Local Feature Descriptors

In general, local feature descriptors describe a pixel (or a position) in an image through its local content. They are supposed to be robust to small deformations or localization errors, and give us the possibility to find the corresponding pixel locations in images which capture the same amount of information about the spatial intensity patterns under different conditions.

In the following we briefly explain the SIFT [11] and SURF [1] descriptors which offer scale and rotation invariant properties.

2.2.1 Scale Invariant Feature Transform (SIFT)

The SIFT descriptor is a 128-dimensional vector which stores the gradients of 4×4 locations around a pixel in a histogram of 8 main orientations [11]. The gradients are aligned to the main direction resulting in a rotation invariant descriptor. Through computation of the vector in different Gaussian scale spaces (of a specific position) it becomes also scale invariant. There exists a closed source implementation of the inventor Lowe[1] and an open source implementation[2] which is used in our experiments.

In certain applications such as face recognition, rotation invariant descriptors can lead to false matching correspondences. If invariance w.r.t. rotation is not necessary, the gradients of the descriptor can be aligned to a fixed direction. The impact of using an upright version of the SIFT descriptor (i.e. U-SIFT) is investigated in section 3.

2.2.2 Speeded Up Robust Features (SURF)

Conceptually similar to the SIFT descriptor, the 64-dimensional SURF descriptor [1] also focusses on the spatial distribution of gradient information within the interest point neighborhood, where the interest points itself can be localized as described in subsection 2.1 by interest point detection approaches or in a regular grid. The SURF descriptor is invariant to rotation, scale, brightness and, after reduction to unit length, contrast.

For the application of face recognition, invariance w.r.t. rotation is often not necessary. Therefore, we analyze in section 3 the upright version of the SURF descriptor (i.e. U-SURF). The upright versions are faster to compute and can increase distinctivity [1], while maintaining a robustness to rotation of about $\pm$ 15 $^\circ$, which is typical for most face recognition tasks.

Due to the global integration of SURF descriptors, the authors [1] claim that it stays more robust to various image perturbations than the more locally operating SIFT descriptor. In section 3 we analyze if this effect can also be observed if we use SURF descriptors for face recognition under various illuminations. For the extraction of SURF-64, SURF-128, U-SURF-64, and U-SURF-128 descriptors, we use the reference implementation[3] described in [1]. Additionally, another detailed overview and implementation[4] is provided in [6].

2.3 Recognition by Matching

The matching is carried out by a nearest neighbor matching strategy $m(X,Y)$: the descriptor vectors $X := \{x_1, \cdots, x_I\}$ extracted at keypoints $\{1, \cdots, I\}$ in a test image X are compared to all descriptor vectors $Y := \{y_1, \cdots, y_J\}$ extracted at keypoints $\{1, \cdots, J\}$ from the reference

[1] http://www.cs.ubc.ca/~lowe/keypoints/
[2] http://www.vlfeat.org/~vedaldi/code/siftpp.html
[3] http://www.vision.ee.ethz.ch/~surf/
[4] http://code.google.com/p/opensurf1/

images $Y_n, n = 1, \cdots, N$ by the Euclidean distance. Additionally, a ratio constraint is applied: only if the distance from the nearest neighbor descriptor is less than α times the distance from the second nearest neighbor descriptor, a matching pair is detected. Finally, the classification is carried out by assigning the class $c = 1, \cdots, C$ of the nearest neighbor image $Y_{n,c}$ which achieves the highest number of matching correspondences to the test image X. This is described by the following decision rule:

$$X \rightarrow r(X) = \arg\max_c \left\{ \max_n \left\{ m(X, Y_{n,c}) \right\} \right\} = \arg\max_c \left\{ \max_n \left\{ \sum_{x_i \in X} \delta(x_i, Y_{n,c}) \right\} \right\} \quad (1)$$

with

$$\delta(x_i, Y) = \begin{cases} 1 & \min_{y_j \in Y} \{d(x_i, y_j)\} < \alpha \cdot \min_{y'_j \in Y \setminus y_j} \{d(x_i, y_{j'})\} \\ 0 \end{cases} \quad (2)$$

In [10], the nearest neighbor ratio scaling parameter was set to $\alpha = 0.7$, in [1] it was set to $\alpha = 0.5$. In some preliminary experiments we found out that α is not a sensitive parameter in the system, with $\alpha = 0.5$ performing slightly better.

The same matching method can be used for all descriptors. However, the SURF descriptor has an additional improvement as it includes the sign of the Laplacian, i.e. the trace of the Hessian matrix, which can be used for fast indexing. Different viewpoint consistency constraints can be considered during matching, accounting for different transformation and registration errors, and resulting in different matching time complexities (c.f. Figure 1).

Maximum Matching. No viewpoint consistency constraints are considered during the matching, i.e. each keypoint in an image is compared to all keypoints in the target image. Here, the unconstrained matching allows for large transformation and registration errors, with the cost of the highest matching time complexity.

Grid-Based Matching. Outliers are removed by enforcing viewpoint consistency constraints: due to an overlaid regular grid and a block-wise comparison in Equation 2, only keypoints from the same grid-cell are compared to corresponding keypoints from the target image. Here, non-overlapping grid-blocks allow for small transformation and registration errors only. Furthermore, the matching time complexity is drastically reduced.

Grid-Based Best Matching. Similar to the Grid-Based Matching, we additionally allow for overlapping blocks. As a keypoint can match now with multiple keypoints from neighboring grid-cells, only the local best match is considered. Here, on the one hand overlapping grid-blocks allow for larger transformation and registration errors, on the other hand stronger and maybe unwanted deformations between faces are possible.

2.4 Outlier Removal

The use of the spatial information about matching points can help to reduce the amount of falsely matched correspondences, i.e. outliers. With the assumption that many parts of a face nearly lie on a plane, with only small viewpoint changes for frontal faces, a given homography (transformation) between the test and train images can reject outlier matches which lie outside a specified inlier radius.

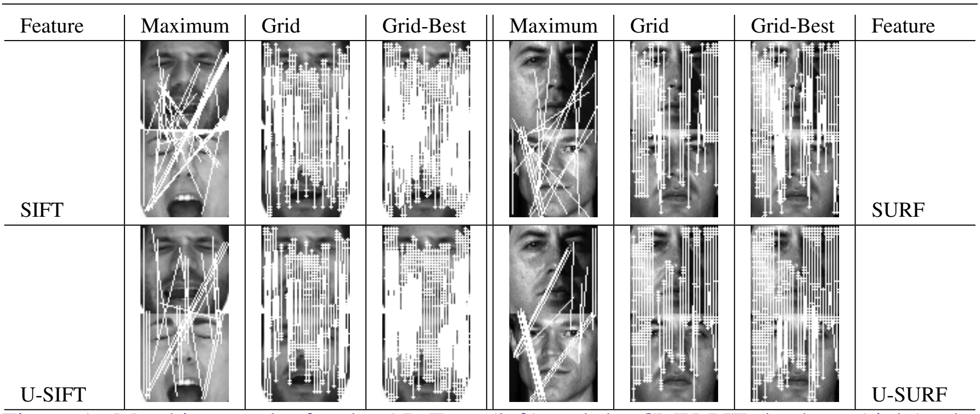

Figure 1: Matching results for the AR-Face (left) and the CMU-PIE database (right): the columns with Maximum matching show false classification examples, the columns with Grid matchings show correct classification examples. In all cases, the corresponding upright descriptor versions reduce the number of false matches.

The Random Sample Consensus (RANSAC) [7] algorithm samples randomly from a small set of matching candidates and estimates a homography between these points by minimizing the least squared error. The amount of sample correspondences can vary but have to be greater or equal than four. In the case of four correspondences the estimation has zero error, as a homography is well-defined by four correspondences. In practical tasks the use of five or more sample points smoothes the transformation and often yields to better performance. In our experiments in section 3 we empirically optimized it to six sample points.

Once a transformation has been estimated, all points of a test image will be projected to the train image. If a projected point lies in a given radius to its corresponding point it is classified as an *inlier* for that particular homography, otherwise it is declared as an *outlier*. After a given number of iterations the maximum amount of inliers of all estimated homographies will be used as a measurement to determine the likelihood of the similarity between the test and the train image.

RANSAC-Based System Combination. Opposed to a combination of the descriptors on a feature-level (i.e. early combination), we introduce a novel RANSAC-based system combination and outlier removal: by merging all matching candidates (i.e. late fusion) of different descriptor-based face recognition systems, the homography is estimated now on the merged matching candidates accounting for different transformations.

3 Experimental Results

In this section, we study whether SURF features are suitable and robust enough for face recognition. Comparisons to a SIFT based approach are provided for the AR-Face and the CMU-PIE databases under various conditions. If available, we provide comparative results from the literature.

3.1 Databases

AR-Face. The AR-Face database has been created by Aleix Martinez and Robert Benavente at the Computer Vision Center of the University of Barcelona [13]. It consists of frontal view face images of 126 individuals, 70 men and 56 women. The database was generated in two sessions with a two week time delay. In each session 13 images per individual were taken with differences in facial expression, illumination and partial face occlusion. Similar to the work presented in [4, 5], we only use a subset of 110 individuals for the AR-Face database.

CMU-PIE. The CMU-PIE database contains 41,368 images of 68 people. Each person is imaged under 13 different poses, 43 different illumination conditions, and with 4 different expressions [21]. We only use the frontal images from the illumination subset of the CMU-PIE database.

3.2 Conditions and Results

Manually Aligned Faces. In this experimental setup, the original face images from both databases have been manually aligned by the eye-center locations [8]. The images are rotated such that the eye-center locations are in the same row. After that the faces are cropped and scaled to a 64×64 resolution. An example can be seen in Figure 2.

For the AR-Face database, seven images from the first session are used for training, the remaining seven images from the second session are used for testing. The results of these experiments on the AR-Face database are listed in Table 1.

For the CMU-PIE database, we define for each of the 68 individuals from the illumination subset set a "one-shot" training condition, where we use a single frontally illuminated image for training, and the remaining 20 images, which are taken under varying illumination conditions, for testing. This scenario simulates situations where only one reference image, e.g. a passport image, is available. Since for one individual only 18 images are present in the database, the total number of training images is 68, and the total number of test images is 1357. The results for the CMU-PIE database are listed in Table 2.

It can directly be seen from Table 1 and Table 2 that the grid-based extraction as well as the grid-based matching methods can achieve the best error rates (c.f. also Figure 1). We empirically optimized the feature extraction grid to 64×64 with a step size of 2 (named "64x64-2 grid" in the following), resulting in 1024 grid points. For the grid-based matchings we used 8×8 block sizes, and a 50% block overlap for the Grid-Best matching method. Both tables show that an interest-point based extraction of the local feature descriptors is insufficient for face recognition as in average only a few interest-points (IPs) are detected per image. Even if the SIFT detector can find more IPs than the Hessian Matrix based SURF detector, the results are worse in both cases.

SURF and SIFT descriptors perform equally well in practically all cases in Table 1, whereas the SURF cannot outperform the SIFT descriptors in Table 2. Interestingly, the upright and rotation dependent descriptor versions U-SURF-64, U-SURF-128, and U-SIFT can improve the results in almost all cases. The visual inspection in Figure 1 furthermore shows that our approach achieves its recognition performance by robustly solving the problem, instead of e.g. exploiting accidental low-level regularities present in the test [17].

Unaligned Faces. As most generic face recognition approaches do not explicitly model local deformations, face registration errors can have a large impact on the system performance [18]. In contrast to the manually aligned database setup, a second database setup has been generated where the faces have been automatically detected using the OpenCV implementa-

Table 1: Error rates for the AR-Face database with manually aligned faces using different descriptors, extractors, and matching constraints.

Descriptor	Extraction	# Features	Error Rates [%]		
			Maximum	Grid	Grid-Best
SURF-64	IPs	5.6 (avg.) $\times$ 64	80.64	84.15	84.15
	64x64-2 grid	1024 $\times$ 64	0.90	0.51	0.90
U-SURF-64	64x64-2 grid	1024 $\times$ 64	0.90	1.03	0.64
SURF-128	64x64-2 grid	1024 $\times$ 128	0.90	0.51	0.38
U-SURF-128	64x64-2 grid	1024 $\times$ 128	1.55	1.29	1.03
SIFT	IPs	633.78 (avg.) $\times$ 128	1.03	95.84	95.84
	64x64-2 grid	1024 $\times$ 128	11.03	0.90	0.64
U-SIFT	64x64-2 grid	1024 $\times$ 128	**0.25**	**0.25**	**0.25**
Modular PCA [22]			14.14		
Adaptively weighted Sub-Pattern PCA [22]			6.43		
DCT [4]			4.70		

Table 2: Error rates for the CMU-PIE database with manually aligned faces using different descriptors, extractors, and matching constraints.

Descriptor	Extraction	# Features	Error Rates [%]		
			Maximum	Grid	Grid-Best
SURF-64	IPs	6.80 (avg.) $\times$ 128	93.95	95.21	95.21
	64x64-2 grid	1024 $\times$ 64	13.41	4.12	7.82
U-SURF-64	64x64-2 grid	1024 $\times$ 64	**3.83**	**0.51**	**0.66**
SURF-128	64x64-2 grid	1024 $\times$ 128	12.45	3.68	3.24
U-SURF-128	64x64-2 grid	1024 $\times$ 128	5.67	0.95	0.88
SIFT	IPs	723.17 (avg.) $\times$ 128	43.47	99.33	99.33
	64x64-2 grid	1024 $\times$ 128	27.92	7.00	9.80
U-SIFT	64x64-2 grid	1024 $\times$ 128	16.28	1.40	6.41
Spherical Harmonics [25]			1.80		
Modeling Phase Spectra using GMM [16]			12.83		

Figure 2: Example of a manually aligned image and an unaligned image.

Table 3: Error rates for the AR-Face database with automatically detected and unaligned faces using different descriptors, grid-based extractors, and grid matching constraints.

Descriptor	Error Rates [%]	
	AR-Face	CMU-PIE
SURF-64	5.97	15.32
U-SURF-64	5.32	5.52
SURF-128	5.71	11.42
U-SURF-128	5.71	**4.86**
SIFT	5.45	8.32
U-SIFT	**4.15**	8.99

tion of the Viola & Jones [23] face detector. The first-best face detection is used to crop and scale the images to a common size of 64×64 pixels (c.f. Figure 2).

Here we focus on the robustness of the descriptors w.r.t. face registration errors. Similar to the aligned databases, we use a dense and grid-based descriptor extraction with a grid-matching. The results of these experiments on the AR-Face and the CMU-PIE database are presented in Table 3. For the AR-Face, almost all faces are detected correctly resulting in similar error rates (c.f. Table 1), whereas due to the extreme variations in illumination in the CMU-PIE database, the face detector does not always succeed in finding the face in the image.

Similar to the results presented in Table 1 and Table 2 the results in Table 3 show that the upright descriptor versions perform better than their corresponding rotation invariant descriptors. Furthermore, it can be observed that the upright SURF descriptors seem to be more robust to illumination than the upright SIFT descriptors, as their average error rate is lower.

Partial Occlusions. The experimental setup with partial occlusions is available only for face images from the AR-Face database. Similar to the manually aligned condition, we use the same subset of 110 individuals but with partial occlusions in the train and test set. We follow the same "one-shot" training experiment protocol as proposed in [5]: one neutral image per individual is used from the first session for training, five images per individual are used for testing. Overall, five separate test cases are conducted on this data set with 110 test images each: one neutral test case *ARneutral*, two test cases *AR1sun* and *AR2sun* related to upper face occlusions, and two test cases *AR1scarf* and *AR2scarf* related to lower face occlusions.

The results of these experiments are listed in Table 4. Again, it can be observed that the upright versions perform better than their corresponding rotation invariant versions. The best average result is achieved using the U-SIFT descriptor. The fact that most part of the face information is located around the eyes and the mouth can especially be observed for the upper face occlusions. However, the performance drop for partial occlusions is not as significant as for the baseline system reported in [5], proofing the robustness of our proposed approach.

Table 4: Error rates for the AR-Face database with partially occluded faces using different descriptors, 1024 grid-based extractors and grid-based matching constraints.

Descriptor	Error Rates [%]					
	AR1scarf	*AR1sun*	*ARneutral*	*AR2scarf*	*AR2sun*	Avg.
SURF-64	2.72	30.00	0.00	4.54	47.27	16.90
U-SURF-64	4.54	23.63	0.00	4.54	47.27	15.99
SURF-128	1.81	23.63	0.00	3.63	40.90	13.99
U-SURF-128	1.81	**20.00**	0.00	3.63	41.81	13.45
SIFT	1.81	24.54	0.00	2.72	44.54	14.72
U-SIFT	**1.81**	20.90	**0.00**	**1.81**	**38.18**	**12.54**
U-SURF-128+R	1.81	19.09	0.00	3.63	43.63	13.63
U-SIFT+R	2.72	**14.54**	0.00	**0.90**	35.45	10.72
U-SURF-128+U-SIFT+R	**0.90**	16.36	**0.00**	2.72	**32.72**	**10.54**
DCT [5], baseline	8.2	61.8	7.3	16.4	62.7	31.28
DCT [5], realigned	2.7	1.8	0.0	6.4	4.5	3.08

The performance for the upper face occlusions (*AR1sun* and *AR2sun*) of our system might be further improved by extracting the local features at larger scales to reduce the number of false positive matches of the sun glasses with e.g. mustache hairs.

The results in Table 4 furthermore show that the RANSAC post-processing step improves the system performance even for the already spatially restricted Grid-Based matching (see subsection 2.3). An outlier removal with an empirically optimized removal-radius set to 3 pixel of the best system (i.e. U-SIFT+R) can further descrease the error rate, and by combining the best two systems (i.e. U-SURF-128+U-SIFT+R), the best average error rate of 10.54% is achieved. A combination of all descriptors did not lead to further improvements. This is in accordance to experiments in other domains where the combination of different systems can lead to an improvement over the individual ones.

4 Conclusions

In this work, we investigated the usage of SURF descriptors in comparison to SIFT descriptors for face recognition. We showed that using our proposed grid-based local feature extraction instead of an interest point detection based extraction, SURF descriptors as well as SIFT descriptors can be used for face recognition, especially in combination with a grid-based matching enforcing viewpoint consistency constraints.

In most cases, upright descriptor versions achieved better results than their corresponding rotation invariant versions, and the SURF-128 descriptor achieved better results than the SURF-64 descriptor. Furthermore, the experiments on the CMU-PIE database showed that SURF descriptors are more robust to illumination, whereas the results on the AR-Face database showed that the SIFT descriptors are more robust to changes in viewing conditions as well as to errors of the detector due to facial expressions. Especially for partially occluded faces, the proposed RANSAC-based system combination and outlier removal could combine the advantages of both descriptors.

The different database conditions with manually aligned faces, unaligned faces, and partially occluded faces showed the robustness of our proposed grid-based approach. Additionally, and to the best of our knowledge, we could outperform many generic approaches known from the literature on the same benchmark sets. Interesting for future work will be the evaluation of the proposed approach on the Labeled Faces in the Wild dataset [24].

Acknowledgements. We would like to thank Hazim Ekenel and Mika Fischer (University of Karlsruhe, Germany), Mauricio Villegas (UPV, Spain), and Thomas Deselaers (ETH Zürich, Switzerland).

This work was partly realised as part of the Quaero Programme, funded by OSEO, French State agency for innovation, and received funding from the European Community's Seventh Framework Programme under grant agreement number 231424 (FP7-ICT-2007-3).

References

[1] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. *Computer Vision and Image Understanding (CVIU)*, 110(3):346–359, 2008. http://www.vision.ee.ethz.ch/~surf/.

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

[3] Javier Ruiz del Solar, Rodrigo Verschae, and Mauricio Correa. Face recognition in unconstrained environments: A comparative study. In *ECCV Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition*, Marseille, France, October 2008.

[4] H.K. Ekenel and R. Stiefelhagen. Analysis of local appearance-based face recognition: Effects of feature selection and feature normalization. In *CVPR Biometrics Workshop*, New York, USA, 2006.

[5] H.K. Ekenel and R. Stiefelhagen. Why is facial occlusion a challenging problem? In *International Conference on Biometrics*, Sassari, Italy, June 2009.

[6] Christopher Evans. Notes on the opensurf library. Technical Report CSTR-09-001, University of Bristol, January 2009. URL http://www.cs.bris.ac.uk/Publications/Papers/2000970.pdf.

[7] M.A. Fischler and R.C. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 25(6):381–395, June 1981.

[8] Ralph Gross. http://ralphgross.com/FaceLabels.

[9] Donghoon Kim and Rozenn Dahyot. Face components detection using surf descriptors and svms. In *International Machine Vision and Image Processing Conference*, pages 51–56, Portrush, Northern Ireland, September 2008.

[10] David G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision (IJCV)*, 60(2):91–110, 2004.

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

[12] A.M. Martinez. Recognizing imprecisely localized, partially occluded and expression variant faces from a single sample per class. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(6):748–763, 2002.

[13] A.M. Martinez and R. Benavente. The AR face database. Technical report, CVC Technical report, 1998.

[14] Ajmal Mian, Mohammed Bennamoun, and Robyn Owens. Face recognition using 2d and 3d multimodal local features. In *International Symposium on Visual Computing (ISVC)*, volume 4291 of *LNCS*, pages 860–870, November 2006.

[15] Krystian Mikolajczyk and Cordelia Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10): 1615–1630, 2005. ISSN 0162-8828. doi: http://dx.doi.org/10.1109/TPAMI.2005.188.

[16] S. Mitra, M. Savvides, and A. Brockwell. Modeling Phase Spectra Using Gaussian Mixture Models for Human Face Identification. *LNCS*, 3687:174, 2005.

[17] Nicolas Pinto, James J. DiCarlo, and David D. Cox. How far can you get with a modern face recognition test set using only simple features? In *CVPR*, Miami, FL, USA, June 2009.

[18] E. Rentzeperis, A. Stergiou, A. Pnevmatikakis, and L. Polymenakos. Impact of face registration errors on recognition. In *Artificial Intelligence Applications and Innovations*, pages 187–194, 2006.

[19] Y. Rodriguez and S. Marcel. Face authentication using adapted local binary pattern histograms. In *European Conference on Computer Vision (ECCV)*, pages 321–332, Graz, Austria, May 2006.

[20] G. Shakhnarovich and B. Moghaddam. Face recognition in subspaces. *Handbook of Face Recognition*, pages 141–168, 2004.

[21] T. Sim, S. Baker, and M. Bsat. The CMU pose, illumination, and expression (PIE) database. In *International Conference on Automatic Face and Gesture Recognition*, 2002.

[22] K. Tan and S. Chen. Adaptively weighted sub-pattern PCA for face recognition. *Neurocomputing*, 64:505–511, 2005.

[23] P. Viola and M.J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.

[24] L. Wolf, T. Hassner, and Y. Taigman. Descriptor based methods in the wild. In *ECCV*, 2008.

[25] L. Zhang and D. Samaras. Face recognition from a single training image under arbitrary unknown lighting using spherical harmonics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(3):351–363, 2006.