

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

<http://www-i6.informatik.rwth-aachen.de/~dreuw/>

Introduction. Recently, a comparative study in [2] has shown the superior performance of local features for face recognition in unconstrained environments. 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. 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.

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 [4] 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 based filter sizes in combination with a fast Hessian matrix based approach.

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. 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.

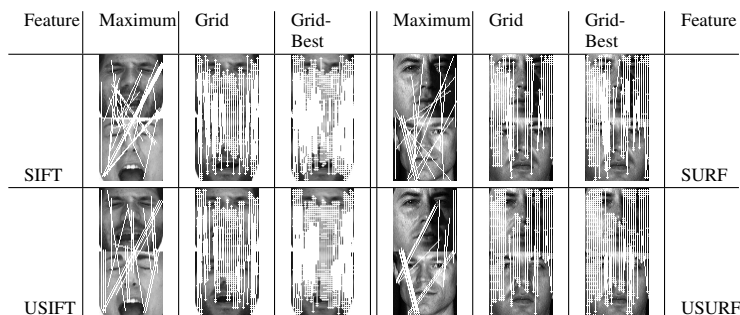
Local Feature Descriptors. The SIFT descriptor is a 128-dimensional vector which stores the gradients of 4×4 locations around a pixel in a histogram of 8 main orientations [4]. Conceptually similar to the SIFT descriptor, the 64-dimensional SURF descriptor [1] also focusses on the spatial distribution of gradient information within the interest point neighborhood. The SURF descriptor is invariant to rotation, scale, brightness and, after reduction to unit length, contrast. In certain applications such as face recognition, rotation invariant descriptors can lead to false matching correspondences. The impact of using an upright version of the SURF and SIFT descriptors (i.e. USURF, USIFT) is investigated.

Recognition by Matching. The matching is carried out by a nearest neighbor matching strategy. Additionally, a ratio constraint is applied: only if the distance from the nearest neighbor descriptor is less than 0.5 times the distance from the second nearest neighbor descriptor, a matching pair is detected. Finally, the classification is carried out by assigning the class of the nearest neighbor image which achieves the highest number of matching correspondences to the test image.

Different viewpoint consistency constraints can be considered during matching, accounting for different transformation and registration errors, and resulting in different matching time complexities:

- **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.
- **Grid-Based Matching:** Due to an overlaid regular grid and a block-wise comparison, outliers are removed by enforcing viewpoint consistency constraints.
- **Grid-Based Best Matching:** Similar to the Grid-Based Matching,

we additionally allow for overlapping blocks.



Experimental Results. 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 several conditions:

- **Manually Aligned Faces:** In this experimental setup, the original face images from both databases have been manually aligned by the eye-center locations [3].
- **Unaligned Faces:** Here we focus on the robustness of the descriptors w.r.t. face registration errors; faces have been detected and cropped using the OpenCV Viola&Jones face detector.
- **Partial Occlusions:** The experimental setup with partial occlusions is available only for face images from the AR-Face database (e.g. faces with sunglasses or scarfs)

Descriptor	Error Rates [%]			
	AR-Face		CMU-PIE	
	Aligned	Unaligned	Aligned	Unaligned
SURF-64	0.51	5.97	4.12	15.32
U-SURF-64	1.03	5.32	0.51	5.52
SURF-128	0.51	5.71	3.68	11.42
U-SURF-128	1.29	5.71	0.95	4.86
SIFT	0.90	5.45	7.00	8.32
U-SIFT	0.25	4.15	1.40	8.99

Conclusions. 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.

- [1] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Surf: Speeded up robust features. *CVIU*, 110(3):346–359, 2008.
- [2] 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.
- [3] Ralph Gross. <http://ralphgross.com/FaceLabels>.
- [4] David G. Lowe. Distinctive image features from scale-invariant keypoints. *IJCV*, 60(2):91–110, 2004.