

# SURF-Face: Face Recognition Under Viewpoint Consistency Constraints

Philippe Dreuw, Pascal Steingrube, Harald Hanselmann and Hermann Ney

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

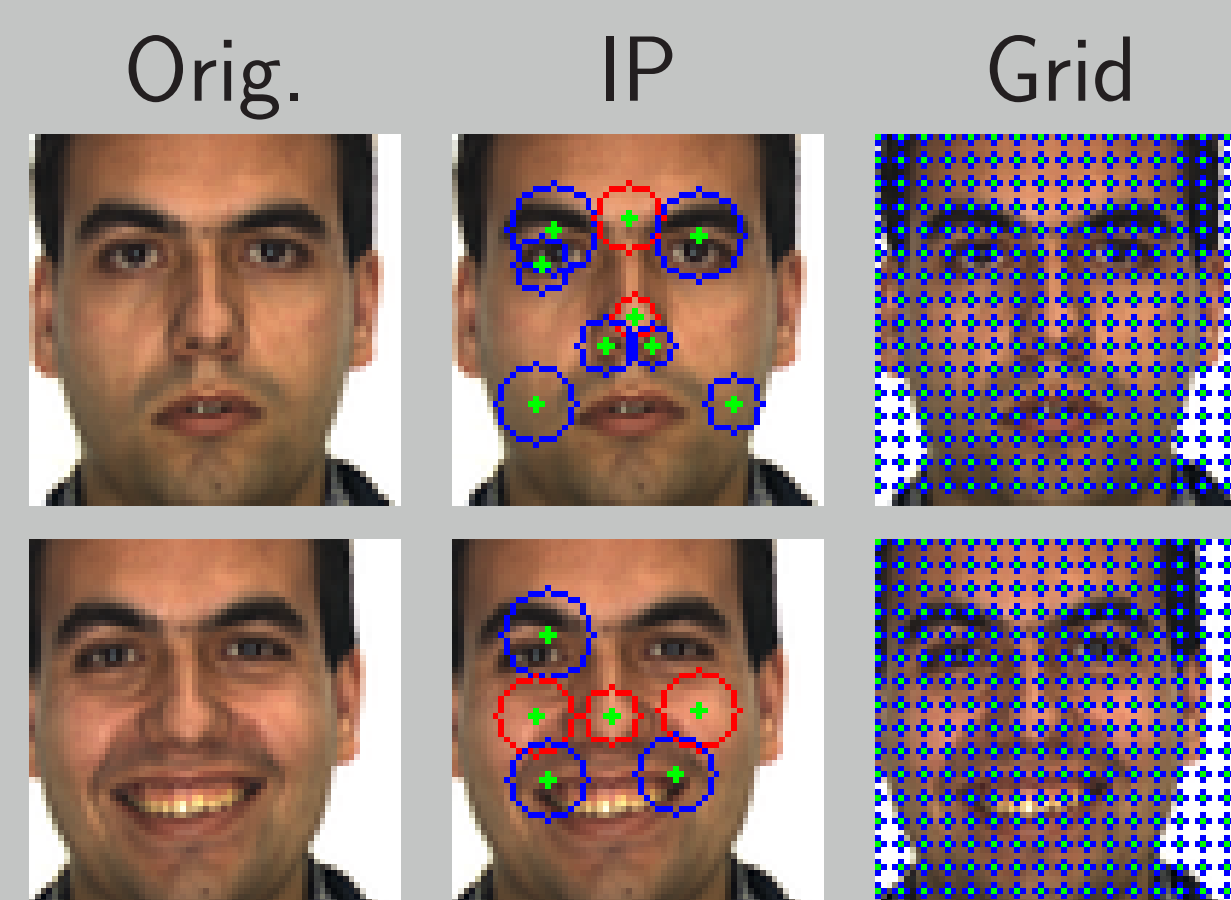
**RWTH**AACHEN  
UNIVERSITY

## Introduction

- ▶ Most face recognition approaches are sensitive to registration errors
  - ▷ rely on a very good initial alignment and illumination
- ▶ We propose/analyze:
  - ▷ grid-based and dense extraction of local features
  - ▷ block-based matching accounting for different viewpoints and registration errors

## Feature Extraction

- ▶ Interest point based feature extraction
  - ▷ SIFT or SURF interest point detector
  - ▷ leads to a **very sparse** description
- ▶ Grid-based feature extraction
  - ▷ overlaid regular grid
  - ▷ leads to a **dense** description



## Feature Description

- ▶ Scale Invariant Feature Transform (SIFT)
  - ▷ 128-dimensional descriptor, histogram of gradients, scale invariant
- ▶ Speeded Up Robust Features (SURF)
  - ▷ 64-dimensional descriptor, histogram of gradients, scale invariant
- ▶ face recognition: invariance w.r.t. rotation is often not necessary
  - ▷ rotation dependent upright-versions U-SIFT, U-SURF-64, U-SURF-128

## Feature Matching

- ▶ Recognition by Matching
  - ▷ nearest neighbor matching strategy
  - ▷ descriptor vectors extracted at keypoints in a test image  $\mathbf{X}$  are compared to all descriptor vectors extracted at keypoints from the reference images  $\mathbf{Y}_n, n = 1, \dots, N$  by the Euclidean distance
  - ▷ decision rule:

$$\mathbf{X} \rightarrow \mathbf{r}(\mathbf{X}) = \arg \max_{\mathbf{c}} \left\{ \max_{\mathbf{n}} \left\{ \sum_{\mathbf{x}_i \in \mathbf{X}} \delta(\mathbf{x}_i, \mathbf{Y}_{n,c}) \right\} \right\}$$

- ▷ additionally, a ratio constraint is applied in  $\delta(\mathbf{x}_i, \mathbf{Y}_{n,c})$
- ▶ Viewpoint Matching Constraints
  - ▷ maximum matching: unconstrained
  - ▷ grid-based matching: absolute box constraints
  - ▷ grid-based best matching: absolute box constraints, overlapping
- ▶ Postprocessing
  - ▷ RANSAC-based outlier removal
  - ▷ RANSAC-based system combination

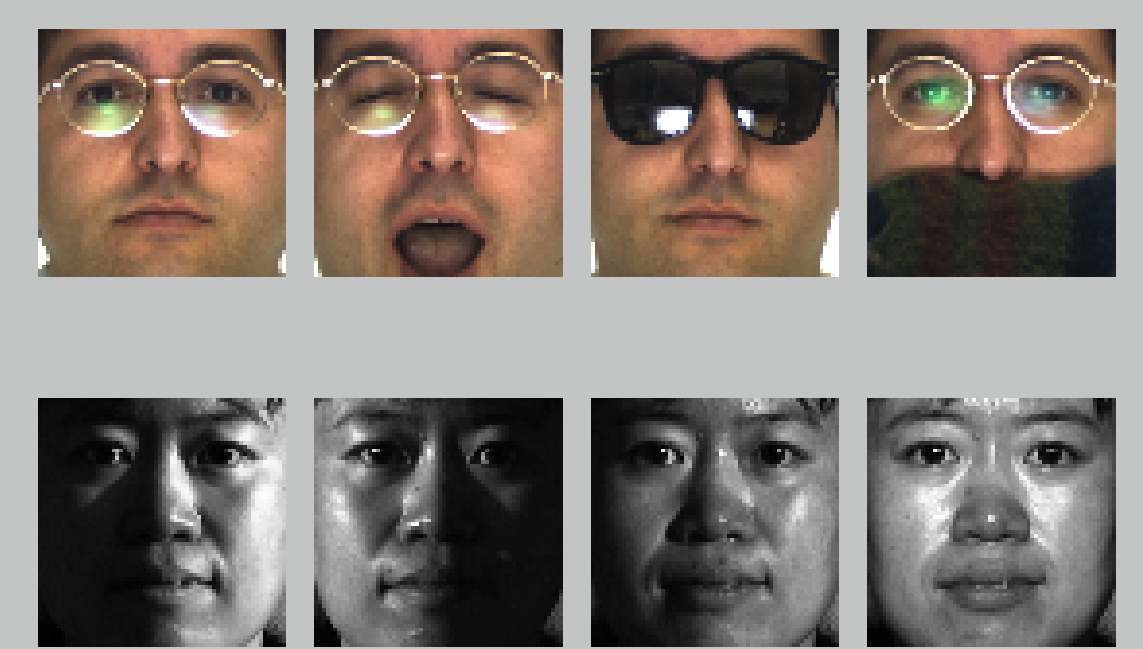
## Matching Examples for the AR-Face and CMU-PIE Database

| Feature | Maximum | Grid | Grid-Best | Feature | Maximum | Grid | Grid-Best | Feature |
|---------|---------|------|-----------|---------|---------|------|-----------|---------|
| SIFT    |         |      |           | SURF    |         |      |           | SURF    |
|         |         |      |           |         |         |      |           |         |
|         |         |      |           |         |         |      |           |         |
| U-SIFT  |         |      |           | U-SURF  |         |      |           | U-SURF  |
|         |         |      |           |         |         |      |           |         |
|         |         |      |           |         |         |      |           |         |

- ▶ Matching results for the AR-Face (left) and the CMU-PIE database (right)
  - ▷ maximum matching show false classification examples
  - ▷ grid matchings show correct classification examples
  - ▷ upright descriptor versions reduce the number of false matches

## Databases

- ▶ AR-Face
  - ▷ variations in illumination
  - ▷ many different facial expressions
- ▶ CMU-PIE
  - ▷ variations in illumination (frontal images from the illumination subset)



## Results: Manually Aligned Faces

- ▶ AR-Face: 110 classes, 770 train, 770 test

| Descriptor | Extraction   | # Features                 | Error Rates [%] |             |             |
|------------|--------------|----------------------------|-----------------|-------------|-------------|
|            |              |                            | Maximum         | Grid        | Grid-Best   |
| SURF-64    | IPs          | $64 \times 5.6$ (avg.)     | 80.64           | 84.15       | 84.15       |
| SIFT       | IPs          | $128 \times 633.78$ (avg.) | 1.03            | 95.84       | 95.84       |
| SURF-64    | 64x64-2 grid | $64 \times 1024$           | 0.90            | 0.51        | 0.90        |
| SURF-128   | 64x64-2 grid | $128 \times 1024$          | 0.90            | 0.51        | 0.38        |
| SIFT       | 64x64-2 grid | $128 \times 1024$          | 11.03           | 0.90        | 0.64        |
| U-SURF-64  | 64x64-2 grid | $64 \times 1024$           | 0.90            | 1.03        | 0.64        |
| U-SURF-128 | 64x64-2 grid | $128 \times 1024$          | 1.55            | 1.29        | 1.03        |
| U-SIFT     | 64x64-2 grid | $128 \times 1024$          | <b>0.25</b>     | <b>0.25</b> | <b>0.25</b> |

- ▶ CMU-PIE: 68 classes, 68 train (“one-shot” training), 1360 test

| Descriptor | Extraction   | # Features                 | Error Rates [%] |             |             |
|------------|--------------|----------------------------|-----------------|-------------|-------------|
|            |              |                            | Maximum         | Grid        | Grid-Best   |
| SURF-64    | IPs          | $64 \times 6.80$ (avg.)    | 93.95           | 95.21       | 95.21       |
| SIFT       | IPs          | $128 \times 723.17$ (avg.) | 43.47           | 99.33       | 99.33       |
| SURF-64    | 64x64-2 grid | $64 \times 1024$           | 13.41           | 4.12        | 7.82        |
| SURF-128   | 64x64-2 grid | $128 \times 1024$          | 12.45           | 3.68        | 3.24        |
| SIFT       | 64x64-2 grid | $128 \times 1024$          | 27.92           | 7.00        | 9.80        |
| U-SURF-64  | 64x64-2 grid | $64 \times 1024$           | <b>3.83</b>     | <b>0.51</b> | <b>0.66</b> |
| U-SURF-128 | 64x64-2 grid | $128 \times 1024$          | 5.67            | 0.95        | 0.88        |
| U-SIFT     | 64x64-2 grid | $128 \times 1024$          | 16.28           | 1.40        | 6.41        |

## Results: Unaligned Faces

- ▶ Automatically aligned by Viola & Jones

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

- ▶ Manually aligned faces



- ▶ Unaligned faces



## Results: Partially Occluded Faces

- ▶ AR-Face: 110 classes, 110 train (“one-shot” training), 550 test

| Descriptor          | Error Rates [%] |               |                  |                 |               |              |
|---------------------|-----------------|---------------|------------------|-----------------|---------------|--------------|
|                     | <i>AR1scarf</i> | <i>AR1sun</i> | <i>ARneutral</i> | <i>AR2scarf</i> | <i>AR2sun</i> | Avg.         |
| SURF-64             | 2.72            | 30.00         | 0.00             | 4.54            | 47.27         | 16.90        |
| SURF-128            | 1.81            | 23.63         | 0.00             | 3.63            | 40.90         | 13.99        |
| SIFT                | 1.81            | 24.54         | 0.00             | 2.72            | 44.54         | 14.72        |
| U-SURF-64           | 4.54            | 23.63         | 0.00             | 4.54            | 47.27         | 15.99        |
| U-SURF-128          | 1.81            | <b>20.00</b>  | 0.00             | 3.63            | 41.81         | 13.45        |
| U-SIFT              | <b>1.81</b>     | 20.90         | <b>0.00</b>      | <b>1.81</b>     | <b>38.18</b>  | <b>12.54</b> |
| U-SURF-128+R        | 1.81            | 19.09         | 0.00             | 3.63            | 43.63         | 13.63        |
| U-SIFT+R            | 2.72            | <b>14.54</b>  | 0.00             | <b>0.90</b>     | 35.45         | 10.72        |
| U-SURF-128+U-SIFT+R | <b>0.90</b>     | 16.36         | <b>0.00</b>      | 2.72            | <b>32.72</b>  | <b>10.54</b> |

## Conclusions

- ▶ Grid-based local feature extraction instead of interest points
- ▶ Local descriptors:
  - ▷ upright descriptor versions achieved better results
  - ▷ SURF-128 better than SURF-64
- ▶ System robustness: manually aligned/unaligned/partially occluded faces
  - ▷ SURF more robust to illumination
  - ▷ SIFT more robust to changes in viewing conditions
- ▶ RANSAC-based system combination and outlier removal