A Combined SIFT/SURF Descriptor for Automatic Face Recognition

Ladislav Lenc, Pavel Král

Dept. of Computer Science & Engineering
Faculty of Applied Sciences
University of West Bohemia
Plzeň, Czech Republic

NTIS - New Technologies for the Information Society
Faculty of Applied Sciences
University of West Bohemia
Plzeň, Czech Republic
{llenc,pkral}@kiv.zcu.cz

ABSTRACT

This paper deals with Automatic Face Recognition (AFR). A novel approach which combines the SIFT and SURF features for the face representation is proposed. The obtained combined SIFT/SURF descriptor is then used for face comparison by the adapted Kepenekci matching method. The proposed method is evaluated on the FERET and CTK corpora. The obtained recognition rates are 98.4% and 64.6% respectively. These recognition scores show that our approach outperforms significantly all other methods on these corpora. The differences between recognition error rates of the proposed approach and the second best one are 41% and 7% in relative value respectively.

Keywords: Czech News Agency Corpus, Face Recognition, Scale Invariant Feature Transform, SIFT, Speeded-Up Robust Features, SURF

1 INTRODUCTION

Recognizing people from images or video sequences is an intensively studied domain. Unfortunately, there are still several reasons why the usage of face recognition system is not as broad as it could be. Especially the quality of facial images to be recognized is determining for the accuracy of the face recognition system. The early holistic methods such as Eigenfaces [1] or Fisherfaces [2] have reached a good performance just while using well aligned and equally illuminated images. However, in the case of poorly aligned or differently illuminated pictures the recognition rate decreases rapidly. These weaknesses of the holistic methods are partially solved using feature based methods. These approaches don't take an image as a whole but represent it as a set of features. The features representing one image are then compared with features representing the other images. Using features instead of holistic description brings lower sensitivity to variances in the images. There are a lot of approaches how to create the features. Particularly the features based on the Gabor wavelets [3, 4, 5, 6] achieved noteworthy results. In the last couple of years some new algorithms for feature detection and extraction have been proposed.

Especially the Scale Invariant Feature Transform (SIFT) [7] attracted a lot of attention. It has been widely used for object detection and for panorama stitching. Later, it also proved to perform well in the face recognition field. This method is scale, rotation and lighting invariant which is very important in the case of images acquired in the uncontrolled environment. Another efficient descriptor similar to the SIFT is the Speeded-Up Robust Features (SURF) [8]. It comes with some important advantages comparing to the SIFT. The creation of the SURF descriptors is computationally more efficient and the descriptor size is reduced from 128 to 64. It brings faster face registration and also faster recognition. Compared to the SIFT, it also includes larger surrounding of the key-point for the descriptor creation. The SURF features are thus less local than the SIFT ones.

This work presents a combination of these two descriptors introducing a new combined SIFT/SURF descriptor. This descriptor will be used for Automatic Face Recognition (AFR) with an adapted Kepenekci matching scheme.

The rest of this paper is organized as follows. Section 2 summarizes the SIFT and SURF algorithms and their usage for the face recognition. The following section describes our proposed approach, a combined SIFT/SURF descriptor with adapted Kepenekci matching. Section 4 describes face datasets used for evaluation and presents the results obtained on these datasets. Section 5 concludes this paper and gives some directions for future work.

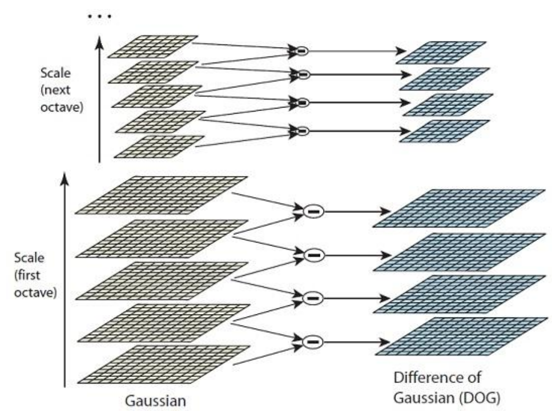

Figure 1 Difference of Gaussian filters at the different scales [9]

2 RELATED WORK

2.1 Scale Invariant Feature Transform (SIFT)

The scale invariant feature transform was first introduced by David Lowe [7]. It detects points of interest in images and creates local descriptors which characterize the neighbourhood of these points.

The first step of this algorithm is the determination of extrema in the image filtered by the Difference of Gaussian (DoG) filter. The input image is gradually down-sampled and the filtering is performed at several scales, which ensures the scale invariance. Figure 1 demonstrates the process of creation of the DoG filters at different scales [9]. Each pixel is then compared with its neighbours. Neighbours in its level as well as in the two neighbouring (lower and higher) levels are examined. If the pixel is maximum or minimum of all neighbouring pixels, it is considered to be a potential key-point.

In the next step, the detected key-points are further examined to choose the "best" candidates. Stability is determined for each key-point. Locations with low contrast and unstable locations along edges are discarded. Then, orientation is assigned to each of the remaining key-points. The computation is based upon gradient orientations in the neighbourhood of the pixel. The values are weighted by the magnitudes of the gradient.

The resulting set of points is then used to create the feature vectors (descriptors). The computation involves the 16×16 neighbourhood of the pixel. Gradient magnitudes and orientations are computed in each point of the neighbourhood. Their values are weighted by a Gaussian. For each sub-region of size 4×4 (16 regions), the orientation histograms are created. Finally, a vector containing 128 (16×8) values is created. The SIFT algorithm is described in detail in [7, 9, 10]. An implementation example can be found in [11].

2.2 Speeded-Up Robust Features (SURF)

Speeded-Up Robust Features [8] is another efficient method for key-point detection and descriptor creation. The method is also invariant to image rotation. The integral image [12] is utilized in this approach to speed-up the key-point detection process. The detector called by the authors a "Fast-Hessian" detector is based on the Hessian matrix[1]. Box filters are used as an approximation of second order Gaussian derivatives. Together with the integral image, it allows very fast computation. Contrary to the SIFT, no image pyramid by the sub-sampling is created. Instead, the box filters are up-scaled and applied to the original image.

To ensure the rotation invariance, the orientation is assigned to each key-point. The computation is based on the circular neighbourhood of the key-point. The resulting SURF descriptor is a vector of the length of 64. Also an upright

[1] a square matrix of the second order partial derivatives of a given function

version of SURF (U-SURF) was proposed. It doesn't compute the orientation (is not orientation invariant) and simplifies and accelerates the computation process.

Recently, the SURF features were also utilized for the face recognition. One of the first applications of SURF in face recognition domain was proposed in [13]. It applies a geometrically constrained point matching scheme. It matches only key-points inside corresponding regions in the compared images which also reduces the computational coasts. A point pair with minimal distance within given rectangular surrounding of key-point is found. The difference between this distance and the distance of the second nearest key-point is calculated. If the difference is higher than a predefined threshold, the point pair is considered to be matched. As a similarity measure, a number of the matched key-points is used. If the number of matched points is lower than a predefined threshold, the matching is considered as not reliable. In this paper, several variants of SURF features are evaluated. The reported rates for the FERET database range from 95.2 to 96.5%.

In [14] another method using SURF features is described. In this paper, a grid-based feature extraction is proposed. It means that the descriptors are extracted in points on regular grid instead of using the key-point detector. Many combinations of feature extraction methods and matching schemes are evaluated. It also compares the use of SIFT and SURF features. Interesting results on AR-Face and CMU-PIE datasets are reported. Surprisingly, the best results are reached using the Upright version of SURF (U-SURF).

In [15] a SURF based face recognition approach combined with cell similarity [16] is proposed. Different cell division strategies are evaluated. Very good results on the ORL and on author's own database are reported.

3 COMBINED SIFT/SURF DESCRIPTOR WITH ADAPTED KEPENEKCI MATCHING

3.1 Combined SIFT/SURF Descriptor

The proposed descriptor is based on the combination of the SIFT and SURF features. The SIFT descriptor has 128 dimensions whereas the SURF descriptor has 64 dimensions. We normalize and concatenate these two descriptors in order to create a vector of the length 192.

The representation of a face is created as follows:

1. SIFT key-points detection
2. SURF key-points detection
3. Calculation of the SIFT and SURF descriptors in all detected key-points
4. Normalization of the SIFT and SURF descriptors
5. Concatenation of these two vectors in order to create a combined descriptor

We assume that combining the key-point locations found by both algorithms, better robustness is achieved.

3.2 Adapted Kepenekci Matching Scheme

The adapted Kepenekci matching scheme proposed in [17] is used. It combines two matching methods and uses a weighted sum of the two values as a result.

Let T be a test image and G a gallery image. For each feature vector t of the face T we determine a set of relevant vectors g of the face G. A vector g is relevant iff:

$$\sqrt{(x_t - x_g)^2 + (y_t - y_g)^2} < distanceThreshold \tag{1}$$

where x and y are coordinates of the feature vector points.

If no relevant vector to vector t is identified, the vector t is excluded from the comparison procedure. The overall similarity of two faces OS is computed as an average of similarities between each pair of the corresponding vectors as:

$$OS_{T,G} = mean\{S(t,g), t \in T, g \in G\} \tag{2}$$

Then, the face with the most similar vector to each of the test face vectors is determined. The C_i value informs how many times the gallery face G_i was the closest one to some of the vectors of test face T. The similarity is computed as C_i / N_i where N_i is the total number of feature vectors in G_i. Weighted sum of these two similarities is used for similarity measure:

$$FS_{T,G} = \alpha OS_{T,G} + \beta \frac{C_G}{N_G} \tag{3}$$

The face is recognized as follows:

$$\hat{FS}_{T,G} = \arg\max_{G}(FS_{T,G}) \tag{4}$$

The cosine similarity is used for vector comparison.

4 EXPERIMENTAL SETUP

4.1 Corpora

The proposed method will be evaluated on two corpora. The first one is the well-known FERET dataset [18]. It is used mainly for comparison of the accuracy of the proposed method with other most efficient face recognition methods. The second dataset is the Czech News Agency[2] (ČTK) corpus. This dataset is unique because it is created semi-automatically from the ČTK Photobank[3] and is composed of real photos. This corpus is thus suitable for evaluation of automatic face recognition approaches in real conditions.

4.1.1 FERET Dataset

The FERET dataset [18] contains 14051 images of 1199 individuals. The images were collected between December 1993 and August 1996. The resolution of the images is by 256 × 384 pixels. The images are divided into the following categories according to the face pose: frontal, quarter-left, quarter right, half-left, half-right, full-left and full-right, and are stored in the .tiff format. The images are also grouped into several probe sets. The main probe sets of the frontal images are summarized in the Table 1.

There is usually only a few seconds between the capture of the gallery-probe pairs in the f* sets. The individuals in the fb set differ in facial expressions, while the images in the fc set differ in illumination conditions. The images in the dup1 probe set were obtained in the three years period and the dup2 set is a sub-set of the dup1. Note that only one image per person is available.

Type	Images no.
fa	1196
fb	1195
fc	194
dup1	722
dup2	234

Table 1 Images numbers in the the main frontal probe sets of the FERET dataset

4.1.2 Czech News Agency (ČTK) Database

This corpus is composed of the images of individuals in uncontrolled environment that were randomly selected from the large ČTK Photobank. All images were taken during a long time period (20 years or more). The detection and extraction of faces was realized automatically using our algorithm proposed in [19]. They were automatically resized to the size of 128 × 128 pixels and transformed to gray-scale. The resulting corpus contains images of 638 individuals. 10 images for each person are available. Note that orientation, lighting conditions and background of images differ significantly. A correct face recognition on this dataset is thus very difficult.

[2] http://www.ctk.eu
[3] http://multimedia.ctk.cz/en/foto/

Figure 2 shows the examples of one face from this corpus. This corpus is available for free for the research purposes at http://home.zcu.cz/~pkral/sw/ or upon request to the authors.

4.2 Recognition Results on the FERET Dataset

Table 2 compares and evaluates our approaches with the other very efficient methods on a large sub-set of the FERET corpus. The *fa* set is used for training, while the *fb* set for testing.

The first reported method is an implementation of the Elastic Bunch GraphMatching (EBGM) algorithm by Bolme [3]. The second algorithmis proposed by Ahonen in [20]. This approach is based on Local binary patterns. Wagner et al. propose in [21] another efficient approach based on the Sparse Representation and Classification (SRC) algorithm whose score is reported next. The fourth method is introduced by Kepenekci in [6] and is based on the Gabor wavelets. The next approach is our implementation of the SIFT based Kepenekci method [19] with an adapted Kepenekci matching proposed in [17]. The last method is the combined SIFT/SURF descriptor with adapted Kepenekci matching proposed in this paper.

Figure 2 Examples of one face from the ČTK face corpus

The table 2 shows that the recognition rates of all approaches are very good and close one to another. This experiment also shows that the proposed combined SIFT/SURF descriptor with adapted Kepenekci matching method outperforms the other approaches.

No.	Method	Recognition Rate [%]
1.	Elastic bunch grpah matching (Bolme)	96.4
2.	Local binary patterns (Ahonen)	96.6
3.	SRC algorithm (Wagner)	95.2
4.	Kepenekci method (Kepenekci)	96.3
5.	SIFT based Kepenekci method (Lenc)	97.3
6.	**Combined SIFT/SURF (proposed)**	**98.4**

Table 2 Recognition results of several AFR methods on the FERET dataset

4.3 Recognition Results on the ČTK Corpus

The table 3 compares the previously proposed SIFT based Kepenekcimethod, the SURF based Kepenekcimethod, U-SURF based Kepenekci method and the proposed combined SIFT/SURF descriptor. Note, that SURF and U-SURF based Kepenekci methods are the approaches that used SURF or U-SURF features with an adapted Kepenekci matching.

This experiment shows that the proposed combined SIFT/SURF descriptor with adapted Kepenekci matching method significantly outperforms the other approaches on this real corpus. The difference between recognition error rate of the "best" proposed approach and the second best one is 7% in relative value.

Method	Recognition Rate [%]
SIFT based Kepenekci	61.9
SURF based Kepenekci	57.1
U-SURF based Kepenekci	61.9
Combined SIFT/SURF	**64.6**

Table 3 Recognition results of some most efficient AFR methods on the ČTK dataset

5 CONCLUSIONS AND PERSPECTIVES

A novel approach which combines the SIFT and SURF features for the face representation was proposed. The proposed combined SIFT/SURF descriptor was then used for AFR together with the adapted Kepenekci matching method. The proposed method was evaluated on the FERET and the CTK corpora. The obtained recognition rates were 98.4% and 64.6% respectively. These recognition scores showed that our approach outperforms significantly all other methods on these corpora. The differences between recognition error rates of the proposed approach and the second best method were 41% and 7% in relative value respectively.

In our future work, we would like to propose more sophisticated combinations of the SIFT and SURF descriptors. We will evaluate several combination methods as already shown in [22]. The key issue of the AFR is to create a more compact representation of a face. We would like to reduce the total amount of features by selecting only the most discriminating ones. Several feature selection techniques will be compared and evaluated for this task.

ACKNOWLEDGEMENTS

This work has been partly supported by the UWB grant SGS-2013-029 Advanced Computer and Information Systems and by the European Regional Development Fund (ERDF), project "NTIS - New Technologies for Information Society", European Centre of Excellence, CZ.1.05/1.1.00/02.0090. We also would like to thank Czech News Agency for support and for providing the photo data.

REFERENCES

[1] Matthew A. Turk and Alex P. Pentland. Face recognition using eigenfaces. In IEEE Computer Society Conference on In Computer Vision and Pattern Recognition. Computer Vision and Pattern Recognition, 1991.

[2] 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, 1997.

[3] David S. Bolme. Elastic Bunch Graph Matching. PhD thesis, Colorado State University, 2003.

[4] L. Shen. Recognizing Faces - An Approach Based on Gabor Wavelets. PhD thesis, University of Nottingham, 2005.

[5] L. Shen and L. Bai. A review on gabor wavelets for face recognition. Pattern Analysis & Applications, 2006.

[6] B. Kepenekci. Face Recognition Using Gabor Wavelet Transform. PhD thesis, The Middle East Technical University, 2001.

[7] David G. Lowe. Object recognition from local scale-invariant features. In International Conference on Computer Vision, 1999.

[8] Herbert Bay, Andreas Ess, Tinne Tuytelaars, and Luc Van Gool. Speeded-up robust features (surf). Comput. Vis. Image Underst., 110(3):346–359, June 2008.

[9] David G. Lowe. Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 2, 2004.

[10] J. Krizaj, V. Struc, and N. Pavesic. Adaptation of sift features for robust face recognition. 2010.

[11] D.G. Lowe. Software for sift, 2004.

[12] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In Conference on Computer Vision and Pattern Recognition, 2001.

[13] Geng Du, Fei Su, and Anni Cai. Face recognition using surf features. pages 749628–749628–7, 2009.

[14] Philippe Dreuw, Pascal Steingrube, Harald Hanselmann, and Hermann Ney. Surf-face: Face recognition under viewpoint consistency constraints. In Proceedings of the British Machine Vision Conference, pages 7.1–7.11. BMVA Press, 2009. doi:10.5244/C.23.7.

[15] Song Cao. A fast surf way for human face recognition with cell similarity. In Industrial Electronics and Applications (ICIEA), 2011 6th IEEE Conference on, pages 166–169, 2011.

[16] Manuele Bicego, Andrea Lagorio, Enrico Grosso, and Massimo Tistarelli. On the use of sift features for face authentication. In In: Conf. on Computer Vision and Pattern Recognition Workshop (CVPRW). (2006, page 35. IEEE Computer Society, 2006.

[17] L. Lenc and P. Král. Novel matching methods for automatic face recognition using sift. In 8th AIAI (Artificial Intelligence Applications and Innovations) Confence, Halkidiki, Greece, September 27-30 2012.

[18] 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.

[19] L. Lenc and P. Král. Face recognition under real-world conditions. In International Conference on Agents and Artifitial Intelligence, Barcelona, Spain, February 14-18 2013.

[20] Timo Ahonen, Abdenour Hadid, and Matti Pietikinen. Face recognition with local binary patterns. In Tomáš Pajdla and Jiří Matas, editors, Computer Vision - ECCV 2004, volume 3021 of Lecture Notes in Computer Science, pages 469–481. Springer Berlin Heidelberg, 2004.

[21] A. Wagner, J. Wright, A. Ganesh, Zihan Zhou, H. Mobahi, and Yi Ma. Toward a practical face recognition system: Robust alignment and illumination by sparse representation. Pattern Analysis and Machine Intelligence, IEEE Transactions on, 34(2):372–386, 2012.

[22] P. Král, C. Cerisara, and J. Klečková . Combination of Classifiers for Automatic Recognition of Dialog Acts. In Interspeech'2005, pages 825–828, Lisboa, Portugal, September 2005. ISCA.