Robust Face Recognition using Key-point Descriptors

Soeren Klemm, Yasmina Andreu, Pedro Henriquez and Bogdan J. Matuszewski
Robotics and Computer Vision Research Laboratory, School of Computing Engineering and Physical Sciences, University of Central Lancashire, Preston, U.K.

Keywords: Face Recognition, SIFT, SURF, ORB, Feature Matching, Face Occlusions.

Abstract: Key-point based techniques have demonstrated a good performance for recognition of various objects in numerous computer vision applications. This paper investigates the use of some of the most popular key-point descriptors for face recognition. The emphasis is put on the experimental performance evaluation of the key-point based face recognition methods against some of the most popular and best performing techniques, utilising both global (Eigenfaces) and local (LBP, Gabor filters) information extracted from the whole face image. Most of the results reported in literature so far, on the use of the key-points descriptors for the face recognition, concluded that the methods based on processing of the full face image have somewhat better performances than methods using exclusively key-points. The results reported in this paper suggest that the performance of the key-point based methods could be at least comparable to the leading "whole face" methods and are often better suited to handle face recognition in practical applications, as they do not require face image co-registration, and perform well even with significantly occluded faces.

1 INTRODUCTION

Face is one of the most frequently used biometric features. It is used for subject recognition and identity verification with the most commonly listed advantages, including ubiquitous, non-contact and non-invasive data acquisition. For many years, face recognition and related research have been of great interest to computer vision and image processing communities, with applications exploited in public security (Chellappa et al., 1995), fraud prevention (Jafri and Arabnia, 2009) and crime prevention and detection (Kong et al., 2005). Although a fair amount of research has been made on the development of face recognition systems based on range and 3D face scans (Zhang and Gao, 2009), the 2D image based systems are still being investigated due to prevalence of the relevant acquisition systems as well as simplicity and low cost of such systems. One of the most popular techniques applied to face recognition uses a specially designed image transformation model to represent faces in a more compact and/or discriminative manifold. The typical examples of this class of techniques include Principal Component Analysis (PCA) (Yang et al., 2004), Linear Discriminant Analysis (LDA) (Belhumeur et al., 1997) or Independent Component Analysis (ICA) (Bartlett et al., 2002). These techniques are often called global methods, as the transformation uses the whole face to find the corresponding representation in the target manifold and the representation in that manifold can strongly depend on the spatially distant points in the original face image. In recognition of the difficulties with global methods in dealing with images captured in real settings, for example with changing illumination condition, another class of methods have been proposed, focused on extraction of local discriminative information. These techniques usually do not require images to be transformed to a different domain, but rather calculate local descriptors for all the pixels representing the face. These descriptors are subsequently integrated in a form of localised histograms with the final feature vector created through concatenation of the local histograms. The most popular techniques in this class include Local Binary Patterns (LBP) (Ahonen et al., 2006) and Gabor filters (Xie et al., 2010). Yet another class of methods is based on calculation of the local descriptors but only in the image regions associated with anatomically significant and predefined facial landmarks, such as eye or mouth corners. In the realm of processing images in practical applications all these techniques exhibit some deficiencies. In case of global and local methods based on processing of the whole face images, the basic requirement is that the database and the query images are spatially co-registered and accu-

Klemm S., Andreu Y., Henriquez P. and Matuszewski B..
Robust Face Recognition using Key-point Descriptors.
DOI: 10.5220/0005314404470454
In *Proceedings of the 10th International Conference on Computer Vision Theory and Applications* (VISAPP-2015), pages 447-454
ISBN: 978-989-758-090-1

rately cropped. This process is prone to errors particularly when faced with constraints imposed by practical applications. Similarly the anatomical landmark based methods require that the landmarks are accurately identified. The methods investigated in this paper, based on key-point descriptors, do not require image co-registration or accurate image cropping or indeed accurate detection of anatomical landmarks as the selection of the relevant key points is inherently embedded in the methods's key point matching. The main premise of this paper is that the techniques based on the key points can achieve comparable results with other techniques, and are more robust when dealing with images acquired in practical applications. The paper proposes the testing methodology and presents the comparative test results between the most popular methods based on the whole image processing and popular key-point methods. The performance of these techniques with respect to facial occlusion is tested in some detail as an example of the robustness of these methods to the conditions often met in real applications.

2 BASELINE METHODS

For years, face recognition methods based on Eigenfaces and Local Binary Patterns have been widely employed reporting successful recognition rates under different scenarios. Over time, this has led to consider these two techniques as standard face recognition methods. Nevertheless, new face recognition approaches have continued to be published, such is the case for Local Gabor XOR Patterns. According to the latest survey on face recognition (Bereta et al., 2013), Local Gabor XOR Patterns is a state-of-the-art method due to the very good recognition rates reported. In this section these well-established face recognition techniques are briefly described.

Eigenfaces method is an early technique of describing images of faces with a relatively low dimensional vector. The method was Proposed by Sirovich and Kirby (Sirovich and Kirby, 1987). The feature vector of an image is calculated by first subtracting a mean face image and then performing a Karhunen-Loeve transform/Principal Component Analysis (PCA). Eigenfaces relies on well adjusted input images that have been normalized with respect to illumination and contrast and aligned with major landmarks that, if not perfectly aligned, introduce detection errors, as they will have an influence on scale, position and rotation of the image.

Local Binary Patterns (LBP) were originally defined to describe image textures (Ojala et al., 2002). However, they have been applied to describe face images achieving high face recognition rates (Ahonen et al., 2006). LBP based descriptors have become a state-of-the-art approach in face analysis not only for their successful results but for the fact that they are not affected by changes in mean luminance since they are invariant against greyscale shifts (Ojala et al., 2002). An LBP number provides information about the distribution of grey level values of a circularly symmetric set of neighbouring pixels. Given a face image and its division into equally sized regions, a histogram is computed for each region. The resulting descriptor comprises the concatenation of all histograms. LBP descriptors can be affected by face localisation errors because they are based on descriptions of local regions of the images.

Local Gabor XOR Patterns (LGXP) (Xie et al., 2010) are a descriptor based on Gabor filters and Local XOR Patterns. This method is one of the best methods for face recognition according to the most recent face recognition survey (Bereta et al., 2013). This descriptor utilises Gabor filters which have proved to be a powerful tool for addressing different computer vision tasks including face recognition (Xie et al., 2010). The idea behind LGXP is to use the Gabor phase information but alleviating its sensitivity to the varying positions. With that purpose, the Gabor phase is quantised to a set of ranges, so if various phases belong to the same interval they are treated as a similar descriptor. Then the Local XOR Pattern operator is applied to quantised phase maps. As for the previous descriptor, this method also requires the face images to be aligned.

3 FACE RECOGNITION BASED ON KEY POINTS

In the previous section, the most widely employed face recognition approaches were introduced. However, other methods can be applied to address this problem. A well-known set of techniques for general object detection and recognition is investigated in the following sections. These techniques search for characteristic points (usually known as key points) in the image that usually correspond to corners or blobs, which are regions of the image where some properties remain constant. Next, information about the surrounding of these key points is extracted and the image is characterised by the set of all extracted descriptors.

3.1 Key-point Detectors and Descriptors

One of the most popular techniques of this kind is Scale Invariant Feature Transform (SIFT) (Lowe, 2004), which has been used for solving face recognition problems achieving competitive results (Luo et al., 2007; Geng and Jiang, 2009; Dreuw et al., 2009; Kisku et al., 2010). These works propose modifications or extensions to SIFT in order to improve the recognition rates. In this paper a comparison of the robustness of the different baseline face recognition methods and key point based techniques is presented. Hence, any improvement that can be applied to key-point detectors and descriptors, could be applied to the described approach as well.

The SURF feature detector and descriptor was proposed as a faster but at least as powerful alternative to the SIFT detector/descriptor (Bay et al., 2008). Both descriptors have been designed with a view to image matching but are used in many different areas of computer vision like object detection, video tracking and face recognition (Geng and Jiang, 2009; Du et al., 2009). Both descriptors are invariant to scale changes and rotation and their performance is comparable in many cases. SURF appears to be a more preferable method as it is faster to compute and produces smaller feature vectors compared to SIFT.

While SIFT utilises local minima and maxima of a Difference of Gaussians operator to detect key points, SURF uses easier to compute Box Filters and integral images to further speed up the process. It should be noted that there is also a U-SURF implementation that interleaves the orientation assignment and leads to more distinct features and faster computation times though is no longer providing rotation invariance.

The SURF feature descriptor is created by rotating the neighbourhood according to the calculated orientation of the feature and then dividing this area into 4×4 square sub-regions. Then box filter responses in horizontal and vertical direction at 5×5 equally spaced sampling points within each of the sub-regions are calculated. These responses and their absolute values are summed up separately for each sub-region. The 4-dimensional descriptor vectors of all 16 sub-regions are concatenated and form a 64-dimensional feature vector. The SIFT descriptor comprises of approximately the same number of features, however it contains 8-bin histograms of gradient orientation per sub-region resulting in a descriptor of 128 elements.

Another key point based approach that will be evaluated is Oriented Fast and Rotated BRIEF (ORB) (Rublee et al., 2011). This method is composed of the FAST key point detector and BRIEF feature descriptor. FAST key-point detector counts the number n of 16 consecutive pixels in a circular pattern around the candidate that are brighter or darker than the candidate by a certain threshold t. If $n \geq 12$ the candidate is treated as a corner. The BRIEF descriptor is a concatenation of the results of 256 binary tests. Each test is a comparison of the intensity values of two pixels within a patch surrounding the key point. To make the descriptor more robust against noise and small rotations, the compared values are the average of a 5×5 window around the test pixels.

3.2 Feature Matching

All key point based methods require a measure to quantify the likeliness of two faces based on the results of the comparison of their key points. In contrast to the methods described in Section 2 for key point based methods there is no pair of equally sized descriptors but two sets, generally containing a different number of descriptors, to compare.

For the experiments performed the similarity of the query face to each of the N faces in the database is calculated as follows.

For every query face a set of I key points

$$K_Q = \{K_{Q1}, K_{Q2}, \ldots, K_{QI}\} \tag{1}$$

is detected and a feature descriptor D_{Qi} is calculated for every K_{Qi} such that the descriptor set

$$D_Q = \{D_{Q1}, D_{Q2}, \ldots, D_{QI}\} \tag{2}$$

can be assumed to describe the face in all necessary detail.

For every face k in the database a set of key points K_k and descriptors D_k where $k = 1 \ldots N$ are detected and calculated using exactly the same methods. Let $J_k = |K_k| = |D_k|$ be the number of key points in the k-th image of the database.

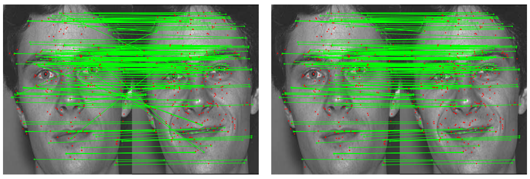
The indices of the key point $K_{k_i j_i}$ with the closest match to the query key point Q_i are given as:

$$(k_i, j_i) = \arg\min_{k,j} \|D_{Qi} - D_{kj}\| \tag{3}$$

These key points were determined using the FLANN (Muja and Lowe, 2009) algorithm for SIFT and SURF features and a Brute Force nearest neighbour search for ORB. FLANN cannot be used for ORB, as the Hamming distance has to be applied for matching BRIEF descriptors.

To exclude randomly matched feature descriptors, that do not belong to the same part of the face, the image of the face is divided using a 3×3 grid pattern. The resulting 9 subsets of pixels are defined as:

$$A_{r,c} = \{(x,y) : c-1 \leq x_a\, 3/w < c, \text{ and } r-1 \leq y_a\, 3/h < r\} \tag{4}$$



(a) (b)

Figure 1: Matched (green) and un-matched (red) key points in corresponding faces using the SURF descriptor. Figure (a) shows false matched key points (diagonals), that are eliminated by matching only key points that belong to the same sub-area of the images as can be seen in Figure (b).

Where w and h denote width and height of the image in pixels, $x = [0, 1, \ldots, w-1]$ and $y = [0, 1, \ldots, h-1]$ represent the positions of one pixel in the image and $1 \leq r, c \leq 3$ are the row and column within the grid. The similarity for the k-th face in the database is now defined as:

$$S_k = \sum_{i=1}^{I} \left[\delta_{k,k_i} \sum_r \sum_c \left(\delta_{K_{k_i j_i}}(A_{r,c}) \delta_{K_{Qi}}(A_{r,c}) \right) \right] \quad (5)$$

Where $\delta_{i,j}$ denotes the Kronecker delta function and $\delta_{x_0}(A)$ denotes the Dirac measure. The face with the highest similarity is chosen as the correct match.

Figure 1 shows the matches before (a) and after (b) the pruning based on Eq. 4.

4 EXPERIMENTAL STUDY

The aim of this experimental study is to test the described face recognition method based on key points (Section 3) together with the presented baseline face recognition techniques (Section 2).

First, a set of experiments using non occluded images is performed in order to have baseline recognition rates. Following, experiments with occluded faces are carried out considering simulated and real occlusions.

4.1 Experimental Setup

The experiments can be grouped as follows i) experiments with non-occluded faces, ii) experiments with simulated occlusions and iii) experiments with real occlusions. These three groups of experiments are carried out using each of the previously introduced face recognition methods: key points, LGXP, LBP and Eigenfaces. Before carrying out all these experiments, an experiment considering only non-occluded faces is used to decide what type of key-point detector and descriptor (SURF, SIFT or ORB) is more suitable for the task.

For all the face recognition methods, except the ones based on key points, the faces need to be aligned so the facial features (eyes, nose and mouth) are approximately in the same location in all images. In order to perform such alignment, the manually annotated coordinates of the eyes, nose tip and mouth of all images are requiered. Given the desired position of the eyes, nose tip and mouth, a rigid transformation is applied to the coordinates of those facial features in each image to obtain the aligned face.

4.2 Face Image Data Sets

Two face image databases are used in the experiments, those are FERET (Phillips et al., 1998; Phillips et al., 2000) database and the AR (Martinez and Benavente, 1998) database.

The FERET database is one of the most widely used databases for evaluating face recognition performance. Different partitions of the database are provided with it, resulting in different data sets. For our experiments the *fa* data set is used as the gallery set, and the testing data sets are *fb*, *dup1* and *dup2*.

The AR database consists of images taken from 126 individuals on 2 sessions with a separation of 14 days. As there are not available the same number of images per subject, a subset of 50 female and 50 male individuals is used. Out of the 13 images that were captured per subject and session, only three images of each session are used in this work: frontal face, face with sunglasses and face with a scarf. For our experiments, the non-occluded images of the first session are used as the gallery set, and the testing datasets are i) the non-occluded images of the second session, ii) the images of the subjects wearing sunglasses from both sessions, and iii) the images of the subjects wearing a scarf from both sessions.

The performance of the different methods is tested using the originally occluded data sets from the AR database as well as simulated occlusions applied to the FERET data sets (see Section 4.4). By performing the tests on both data sets, the performance with respect to a controllable amount of simulated occlusion can be tested and the findings can be supported with more realistic conditions of the AR images.

4.3 Design Parameters

In this section the parameter settings for each of the methods used in the experiments are indicated.

Key point based methods use the SURF, SIFT and ORB implementations provided with the Open Source Computer Vision library[1] (OpenCV). It is

[1]OpenCV library is available online at http://opencv.org/

worth noting that these methods use no aligned face images, however the location of the face in the image needs to be detected. The well-known Viola & Jones face detector (Viola and Jones, 2001) is used, with the implementation provided with the OpenCV library.

LGXP method is based on the work by (Xie et al., 2010). Hence, the values of the design parameters are those given in that work. Although, the results reported in that publication have not been replicated, the recognition rates achieved with our implementation are useful for comparing the performance of this method under different conditions tested in the experiments.

LBP method is based on regions of 9×9 pixels from which LBPs are calculated taking 8 sample points with a radius of 2 pixels. Each of the histograms extracted from each region has 59 bins, since uniform LBPs are used (Ahonen et al., 2006).

Eigenfaces method is implemented using the default parameters of the FaceRecognizer provided with the OpenCV library.

Regarding the size of the images, the key point based method uses the images as they are provided in the face databases. For the other face recognition methods the images from the FERET database are resized to 80×88 pixels following the recommendation given in (Xie et al., 2010). However, an experiment using larger image sizes is also perfomed in order to check if the reduced size affects the recognition results. For the AR data sets the ready available set of warped images (Martinez and Kak, 2001) is used, each image has a size of 120×165 pixels.

4.4 Simulated Occlusions

In order to have a controlled amount of occlusions in the images, simulated occlusions were added to the FERET data sets used for testing. These occlusions were introduced randomly by replacing a square shaped area of the detected face. To provide a more realistic scenario the occlusions were not only black or white masks, but taken from the Colored Brodatz Texture database. The occluded images were created by first detecting the face in the original FERET image. A square shaped mask of 10% and 25% of the size of the facial area was created by randomly selecting one of the 112 textures in the Brodatz database and an equally sized area within each of the templates. This area was then copied to a random position inside the facial area of the FERET image. The process was repeated 5 times and the reported results are averaged.

Table 1: Face recognition rates (%) over FERET data sets achieved by different key-point methods.

	SIFT	SURF	ORB
fb	93.75	97.88	81.15
dup1	52.31	57.61	29.48
dup2	44.30	52.63	21.93

Table 2: Face recognition rates (%) over FERET data sets without occlusions. Images have been aligned and cropped to 80×88 pixels for LGXP, LBP and Eigenfaces methods.

	SURF	LGXP	LBP	Eigenfaces
fb	97.88	80.24	92.74	69.56
dup1	57.61	47.01	47.96	23.51
dup2	52.63	39.04	28.51	6.14

5 RESULTS

The evaluation of the tested face recognition methods is first performed using non-occluded data. The results provide a certain level of confidence in the implemented algorithms as they can be compared to formerly reported results. Those results are then compared with the scaled simulated occlusion results and verified using the results from real occlusions.

5.1 No Occlusions

Table 1 shows that on non-occluded faces of the FERET database, the SURF descriptor performs as expected based on the results formerly reported (Geng and Jiang, 2009) the same applies to SIFT (Liu et al., 2011). SURF constantly provides better results than the other key point based descriptors, especially ORB does not seem to be suited for face recognition, at least for the given test environment. For this reason, the results of SIFT and ORB will no longer be reported for other tests.

Table 2 shows the results for all the described methods on the same database. In this Table as well as in Table 3, the SURF results are duplicated (from Table 1) for reference only. The results for the LBP algorithm are seconded by previously reported results on the FERET images (Bereta et al., 2013; Yang and Chen, 2013). On the used implementation of LGXP, the performance achieved is worse than previously reported (Xie et al., 2010).

It is worth highlighting at this point that images of 80×88 pixels were used in order to follow the same experimental setup as in (Xie et al., 2010). To assess the influence of the comparatively small size of the aligned images, a second experiment was performed using larger images of 300×330 pixels (see

Table 3: Face recognition rates (%) over FERET data sets without occlusions, but with images aligned and cropped to 300×330 pixels for LGXP, LBP and Eigenfaces.

	SURF	LGXP	LBP	Eigenfaces
fb	97.88	93.65	82.16	69.46
dup1	57.61	52.17	51.22	23.37
dup2	52.63	41.23	48.25	5.70

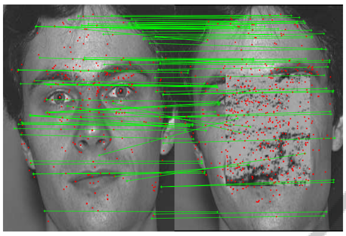


Figure 2: Example of matched key points in a face with 25% of simulated occlusion using the SURF descriptor. Despite a few wrong matches, there is still enough information left to recognise the subject correctly.

Table 3). While the change in size has no influence on the results of the Eigenface descriptor, there is even a decrease in the performance of the LBP method. It should be noted, that SURF detection has not been performed on resized images, therefore the figures given in Table 3 are for reference only.

The additional image details due to the increased resolution do not provide any significant changes to the position of the image vectors in the PCA feature space. Therefore, the separability of the classes does not improve when using the Eigenfaces method.

LBP seems to benefit from the smaller image sizes due to the noise reduction resulting from the subsampling as it requires the application of a low-pass filter. In contrast to LBP, the Gabor filter allows one feature descriptor to cover multiple scales, therefore the results benefit from the greater detail but still cover an area large enough to generate distinct features.

Nevertheless, it should be noted, that the SURF descriptor performs best in all experiments without occlusions. This does also include the robustness against appearance changes, that can be observed by the drop in recognition performance in between the *fb* and *dup2* data sets.

5.2 Simulated Occlusions

Table 4(a) and Figure 3(a) show the results obtained on the *fb* FERET data set with 10% of simulated occlusions. As those results confirm there is only a small drop in recognition performance for all methods when the facial area is occluded. This is due to the high similarity of the faces in both sets, that allows for a constant quality of the single features extracted from the whole facial area. The experiments using the *dup1* and *dup2* (see Figure 3(b)) sets show a much larger drop for all descriptors.

Table 4: Face recognition rates (%) over FERET data sets with artificial occlusions. Images have been aligned and cropped to 80×88 pixels for LGXP, LBP and Eigenfaces.

(a) 10% of occlusion

	SURF	LGXP	LBP	Eigenfaces
fb	96.31	70.16	86.05	53.31
dup1	51.11	36.11	45.92	16.68
dup2	40.18	29.91	27.98	4.56

(b) 25% of occlusion

	SURF	LGXP	LBP	Eigenfaces
fb	92.07	44.98	61.63	27.30
dup1	39.76	17.74	24.62	8.75
dup2	24.65	13.51	15.00	2.19

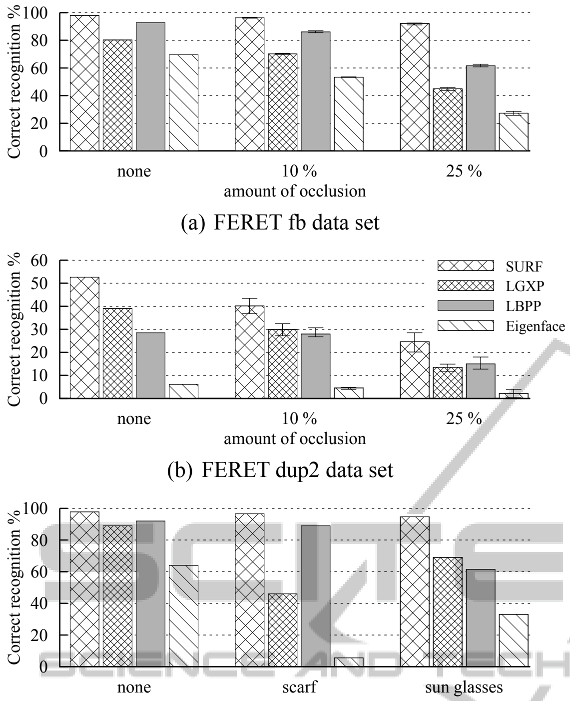
Table 5: Face recognition rates (%) over AR images without and with real occlusions.

	SURF	LGXP	LBP	Eigenfaces
no occlusion	97.75	89.00	92.00	64.00
sunglasses	94.71	69.00	61.50	33.00
scarf	96.63	46.00	89.00	5.50

Table 4(b) shows the recognition rates obtained for each of the methods when the occlusion is 25% of the facial area. This is a more challenging task and so the performances drop more significantly if compared to the results using non-occluded images. However, the SURF method seems to be less affected than the rest which could be due to the quality of the keypoint descriptors. As can be seen in Figure 2, which depicts the matches of a non-occluded gallery image against an image with 25% of occlusion, the number of wrong matches to the occluded area is negligible. Therefore, it can be assumed that there are enough remaining matches to allow good results even when 25% of the facial area is occluded.

The performance of LBP in the *fb* tests shows the lack of flexibility of the methods based on local descriptors to deal with the loss of information. As the occlusions are located randomly and the squares may appear rotated in the aligned images the occlusions degrade the information of a larger part of the histograms than 25%. The same reasoning applies to the LGXP descriptor.

As the occluded images are no longer in their expected subspace in the Eigenfaces feature space, the projection vectors calculated by both methods do no longer provide relevant features.



(c) AR database

Figure 3: Average recognition performance of the tested methods under different occlusions. The error bars indicate the minimal and maximal performance achieved for the different versions of the occluded data sets.

5.3 Real Occlusions

The results shown in row 1 of Table 5 define a baseline for the experiments with real occlusions. As can be seen, they are comparable to the results of the *fb* test set without any occlusions (Table 2). This is due to the only 14 days difference in between the first and the second session of the AR data gathering process.

The two types of occlusions in the AR sets appear to be less challenging than the simulated occlusions. The overall performance of all the descriptors is confirmed (see Table 5 and Figure 3(c)). However, when looking at the recognition rates achieved by all the methods, it seems that the key-point descriptor was more robust than the rest, as there is no much drop from the baseline results and those obtained with real occlusions. In addition, the key-point method had also an overall better performance. Regarding the LGXP method, its performance over scarf images is not as good as the performance over sunglasses occlusions which seem counterintuitive as eyes tend to provide more discriminant information than other areas of the face.

6 CONCLUSIONS

The experiments show that, although the recognition rates of key point based face recognition methods are slightly lower than reported results for Local Descriptors (Bereta et al., 2013; Xie et al., 2010), they outperform the tested implementations when the faces are partially occluded.

Furthermore, the use of key-point detectors makes exact alignment redundant, which does not only eliminate the need for annotated data sets but also saves time during the pre-processing stage of the recognition process.

Out of the tested range of key-point detectors and descriptors, SURF appears to be the most successful one. Not only does it show higher matching speeds (Du et al., 2009) it does also outperform the well known SIFT on the occluded faces of the AR database. ORB does, in the described setup, not achieve results that are comparable to the other methods. The same applies to the holistic method Eigenfaces, that was not capable of delivering good performance with partially occluded faces.

There are many possible ways to further increase the performance of key point based face recognition methods. The described matching method (Section 3.2) is very generic. It seems promising to use more advanced methods (Geng and Jiang, 2009) or include a bayesian approach for evaluating the strength of a matched key point pair.

As face images can, in many cases, be assumed to be roughly aligned along the vertical axis, the U-SURF descriptor might deliver even better performance than rotation invariant SURF.

In the area of binary descriptors, there are other methods that show performance comparable to SIFT when it comes to feature matching. A combination of FAST detector and BRISK descriptor was reported to perform very well (Bekele et al., 2013).

Finally, Gabor filters were able to increase the performance in combination with LBPs (Bereta et al., 2013; Xie et al., 2010) and have spatial locality, in contrast to DFT or DCT, it seams reasonable to combine them with key-point descriptor methods as well.

ACKNOWLEDGEMENTS

This work has been supported by the EU FP7 Project SEMEOTICONS (Grant agreement no: 611516). Portions of the research in this paper use the FERET database of facial images collected under the FERET program, sponsored by the DOD Counterdrug Technology Development Program Office.

REFERENCES

Ahonen, T., Hadid, A., and Pietikinen, M. (2006). Face description with local binary patterns: Application to face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 28:2037–2041.

Bartlett, M. S., Movellan, J. R., and Sejnowski, T. J. (2002). Face recognition by independent component analysis. *IEEE Trans. Neural Networks*, pages 1450–1464.

Bay, H., Ess, A., Tuytelaars, T., and van Gool, L. (2008). Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding*, 110(3):346–359.

Bekele, D., Teutsch, M., and Schuchert, T. (2013). Evaluation of binary keypoint descriptors. In *IEEE Int. Conf. Image Processing (ICIP)*, pages 3652–3656.

Belhumeur, P. N., Hespanha, J. P., and Kriegman, D. J. (1997). Eigenfaces vs. Fisherfaces: recognition using class specific linear projection. *IEEE Trans. Pattern Anal. Mach. Intell.*, 19(7):711–720.

Bereta, M., Pedrycz, W., and Reformat, M. (2013). Local descriptors and similarity measures for frontal face recognition: A comparative analysis. *Journal of Visual Communication and Image Representation*, 24(8):1213–1231.

Chellappa, R., Wilson, C., and Sirohey, S. (1995). Human and machine recognition of faces: a survey. *Proc. IEEE*, 83(5):705–741.

Dreuw, P., Steingrube, P., Hanselmann, H., and Ney, H. (2009). SURF-face: Face recognition under viewpoint consistency constraints. In *Proc. British Machine Vision Conf.*, pages 1–1.

Du, G., Su, F., Cai, A., Ding, M., Bhanu, B., Wahl, F. M., and Roberts, J. (2009). Face recognition using SURF features. In *SPIE Proc. Int. Symp. Multispectral Image Proc. and Pattern Recog.*, pages 749628–1 – 7.

Geng, C. and Jiang, X. (2009). Face recognition using sift features. In *Image Processing (ICIP), 2009 16th IEEE International Conference on*, pages 3313–3316.

Jafri, R. and Arabnia, H. R. (2009). A Survey of Face Recognition Techniques. *Journal of Information Processing Systems*, 5(2):41–68.

Kisku, D. R., Gupta, P., and Sing, J. K. (2010). Face recognition using sift descriptor under multiple paradigms of graph similarity constraints. In *Int. J. Multimedia and Ubiquitous Engineering*, volume 5, pages 1–18.

Kong, S. G., Heo, J., Abidi, B. R., Paik, J., and Abidi, M. A. (2005). Recent advances in visual and infrared face recognitiona review. *Computer Vision and Image Understanding*, 97(1):103 – 135.

Liu, N., Lai, J., and Qiu, H. (2011). Robust Face Recognition by Sparse Local Features from a Single Image under Occlusion. In *Graphics (ICIG)*, pages 500–505.

Lowe, D. G. (2004). Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2):91–110.

Luo, J., Ma, Y., Takikawa, E., Lao, S., Kawade, M., and Lu, B.-L. (2007). Person-specific sift features for face recognition. In *Int. Conf. Acoustics, Speech and Signal Processing*, volume 2, pages II–593–II–596.

Martinez, A. and Benavente, R. (1998). The AR Face Database: CVC Technical Report #24.

Martinez, A. M. and Kak, A. C. (2001). PCA versus LDA. *IEEE Trans. Pattern Anal. Mach. Intell.*, 23(2):228–233.

Muja, M. and Lowe, D. G. (2009). Fast approximate nearest neighbors with automatic algorithm configuration. In *In VISAPP International Conference on Computer Vision Theory and Applications*, pages 331–340.

Ojala, T., Pietikainen, M., and Maenpaa, T. (2002). Multiresolution grayscale and rotation invariant texture classification with local binary patterns. *IEEE Trans. Pattern Anal. Mach. Intell.*, pages 971–987.

Phillips, P. J., Hyeonjoon Moon, Rizvi, S. A., and Rauss, P. J. (2000). The FERET evaluation methodology for face-recognition algorithms. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(10):1090–1104.

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

Rublee, E., Rabaud, V., Konolige, K., and Bradski, G. (2011). ORB: An efficient alternative to SIFT or SURF. In *Int. Conf. Computer Vision*, pages 2564–2571.

Sirovich, L. and Kirby, M. (1987). Low-dimensional procedure for the characterization of human faces. *Journal of the Optical Society of America A*, 4(3):519.

Viola, P. and Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. In *Computer Vision and Pattern Recognition.*, pages I–511–I–518.

Xie, S., Shan, S., Chen, X., and Chen, J. (2010). Fusing Local Patterns of Gabor Magnitude and Phase for Face Recognition. *IEEE Trans. Image Processing*, 19(5):1349–1361.

Yang, B. and Chen, S. (2013). A comparative study on local binary pattern (LBP) based face recognition: LBP histogram versus LBP image. *Neurocomputing*, 120:365–379.

Yang, J., Zhang, D., Frangi, A. F., and Yang, J. (2004). Two-dimensional PCA: A new approach to appearance-based face representation and recognition. *IEEE Trans. Pattern Anal. Mach. Intell.*, 26(1):131–137.

Zhang, X. and Gao, Y. (2009). Face recognition across pose: A review. *Pattern Recognition*, 42(11):2876 – 2896.