Face Recognition from Robust SIFT Matching

Massimiliano Di Mella and Francesco Isgrò(✉)

Dipartimento di Ingegneria Elettrica E Delle Tecnologie Dell'Informazione, Università Degli Studi di Napoli Federico II, Napoli, Italy
francesco.isgro@unina.it

Abstract. This paper presents a face recognition algorithm based on the matching of local features extracted from face images, namely SIFT. Some of the earlier approaches based on SIFT matching are sensitive to registration errors and usually rely on a very good initial alignment and illumination of the faces to be recognised. The method is based on a new image matching strategy between face images, that first establishes correspondences between feature points, and then uses the number of correct correspondences, together with the total number of matches and detected features, to determine the likelihood of the similarity between the face images.

The experimental results, performed on different datasets, demonstrate the effectiveness of the proposed algorithm for automatic face identification. More exhaustive experiments are planned in order to perform a fair comparison with other state of the art methods based on local features.

Keywords: Face recognition · Feature matching

1 Introduction

Automatic face recognition from digital still images has attracted much attention [12,31] in the research community because of its large number of applications. Formally the problem can be defined as follows: given an input face image and a database of face images of known individuals, determine the identity of the face in the input image. Despite of the vast literature on the topic to date face recognition it is still an unsolved problem, although some recent work has reported good results on difficult datasets [28].

Face recognition methods can be classified as either holistic or feature based. Holistic face recognition makes use of global information from the images of faces to perform face recognition. The global information is represented by a small number of features which are directly derived from the pixel information of face images. These features capture the variance among different faces. The Eigenfaces method [29] and the Fisher's Linear Discriminate (FLD) [30] belong to this class. Local feature based methods started being proposed more recently, as an alternative to the holistic method, and are currently an area of active research in the face recognition field. Among the various local feature used we

V. Murino and E. Puppo (Eds.): ICIAP 2015, Part II, LNCS 9280, pp. 299–308, 2015.
DOI: 10.1007/978-3-319-23234-8_28

remind Local Binary Patterns [2], Histogram of Oriented Gradients [27], and Gabor Wavelets [19].

From Lowe's work on object recognition using SIFT (Scale Invariant Feature Transform) descriptors [20], multiple authors have applied such descriptors in other fields, like robot navigation [26], scene classification [25], and also face recognition [5,8,13,14,22]. The general approach first extracts a number of key-points in the images, and then compute a local descriptor for each key-point. Recognition is performed matching each point descriptor in the test image against all descriptors extracted from all the images in the database. The input image is assigned to a class in the database depending on the output of the matching procedure.

One important thing that must be taken into account when using local features for face recognition are false matched key-points. Most of the local feature approaches to face recognition tackle this problem adopting a grid-based matching strategy [7] that works establishing a few sub-regions on the face images: only descriptors between corresponding sub-regions are compared for matching. This local matching help to reduce (without eliminating) the number or wrong matches, but requires that the images are somewhat preregistered, making it difficult the application on databases with arbitrary poses and image sizes as PubFig [15]. Moreover variable illumination still has significant influence on the detection of keypoints, since the keypoint detector intrinsic to the SIFT technique is not really invariant to illumination [14].

In this work we present an improvement on the standard method for face recognition using local features using matching. The proposed method instead of relying on the fact that the images are pre-aligned, exploits the fact that images can have different poses, and that the existence of a large number of wrong matches between two face images is likely to mean that the two persons are different. Observing that many parts of a face nearly lie on a plane, wrong matches are determined assuming a homographic transformation between faces, that is computed using a standard RANSAC [11] procedure.

The use of an outlier detection step assuming an homographic model has been already discussed in [7], where it was more used as a post-processing step for a grid matching, and for a RANSAC-based system combination for combining different descriptors. In [16] is proposed a system where a robust estimation of the fundamental matrix is used to refine the output of a battery of SVM classifiers.

The paper is structured as follows. The next Section briefly reminds how SIFT key-points and descriptors are computed. Section 3 presents the proposed face recognition strategy. The experimental results are shown and discussed in Section 4. Section 5 is left to the final remarks.

2 Features Detection in Scale-Space

SIFT key-points were first proposed in [20] and attracted the attention of the computer vision community for their tolerance to scale changes, illumination variations, and image rotations. These features are also claimed robust to affine distortion, change of viewpoints and additive noise.

The process of building SIFTs [21] is heavily inspired by the scale-space framework, but it keeps all the information related to the different levels of resolution. The process can be sketched in two phases: the first is key-points detection in scale-space pyramid and the second is key-points description using the image gradient at the right level of resolution.

SIFT features have always been known for being computationally intensive, so that less expensive features (e.g., SURF [4] and BRIEF [6]) have been proposed. In fact, when considering one of the many C/C++ SIFT implementations available on the Internet that run on standard CPU, it is true that it is not possible to extract features at a high frame rate. On the other hand, for applications where computational performance is an issue, it is possible to consider the GPU-based implementation provided in the library OpenVIDIA [1]. It exploits the processing power of the graphics card to achieve a significant speedup (10x) over traditional software versions. In [17] it is reported that speeds around 60 frames per second (640×480 pixels in size) have been reached with an off-the-shelf nVIDIA graphics board that carried out both feature extraction and matching while, at the same time, relieving the main CPU.

3 Proposed Method

The strategy proposed in the paper is depicted in the pseudo-code given in Algorithm 1. The procedure takes as an input a face image P and a gallery $\mathcal{G}$ of face images the identity of which is known. The system assigns an identity to the unknown person.

In a nutshell the method first extract the SIFT features from all the images, and then find robust feature correspondences between P and the images in the gallery assuming a homographic model. The quantities of matches returned is used into a scoring function that gives a measure of similarity between two faces.

Algorithm 1. Pseudo-code of the face recognition strategy proposed in this paper.

identity Face Recognition(SingleImage P, Gallery $\mathcal{G}$)

$descSiftProbe \leftarrow extractSift(P)$
for all image $\in \mathcal{G}$ **do**
  $descSiftImage \leftarrow extractSift(image)$
  $matchPoints \leftarrow match(descSiftProbe, descSiftImage)$
  $inlier \leftarrow RansacHomography(matchPoint)$
  $score \leftarrow getScore(matchPoint, inlier, descSift)$
  $evalVector(image) \leftarrow score$
end for
$image \leftarrow searchMaxScoreInVector(evalVector)$
$identity \leftarrow getIdentity(image)$

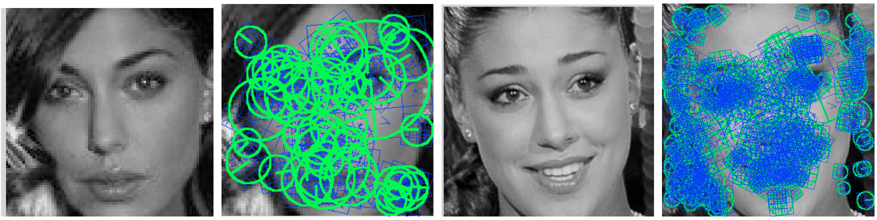

Fig. 1. Two examples of face images with detected key-points.

The identity of the image in the gallery most similar to the unknown image P is returned.

The first step of the procedure is extracting the SIFT key-points and descriptors (see Figure 1) from the face image P. Then each image in the gallery is compared to the input image P determining the correspondences between the features extracted from P and the ones extracted from the image in the gallery, that were precomputed. The similarity measure used for the matching between SIFT descriptor is the one proposed by Lowe in [21]: a feature vector f_i in the image P is matched to a feature vector g_j in the image from the gallery if the Euclidean distance d_{ij} between the two vectors is such that

$$d_{ij} = \min(D_i) < 0.6 \min(D_i - \{d_{ij}\})$$

where $D_i = \{d_{ih} = d(f_i, g_h), \forall g_h\}$.

As it can be seen from the image in Figure 2(left) even when matching images of the same person wrong matches can occur. The presence of wrong matches between the two face images can lead to errors when computing the similarity if they are not taken into account properly. This can be particularly true when comparing face images of different individuals. This is evident when looking at Figure 3, where three face images are matched against the same one: the last two matches, that are between faces of different persons, return worst matches than the first one, where different face images of the same person are compared.

The strategy we propose is actually very simple. We assume that the most of the face is nearly planar, therefore there is a homography [11] between two face images, that can be estimated from at least four point correspondences, as an homography depends on 8 free parameters. In this way we assume that there is a parametric model θ linked to the point correspondences, therefore we can use robust estimation methods [24] to detect all the wrong correspondences, that are the outliers for the estimated model.

The *Random Sample Consensus* (RANSAC) selects random subsets of the data set. It proceeds as follows:

1. randomly selects a subset L of 4 correspondences and estimate the model θ;
2. computes the subset L' of correspondences that are within some error tolerance from the estimated model θ; L' is called the *consensus* set of L;

3. if the number of correspondences in L' is larger than a given threshold then L' is used to compute the model using Least Squares and exit;
4. if the number of observations in L' is smaller than a given threshold goto 1

Having a reliable estimation of the model, wrong matches can be determined using an outlier rejection rule [24].

The similarity score between the unknown face image P and each image in the gallery takes into account the results of the wrong matches feature matching step described above. Exploiting the fact that face images of different person should return a larger number of wrong matches than images of the same person, we investigated the use of different score function between two face images P, the unknown probe image, and G, one of the images in the gallery. For this study we compared the performance of the following scores:

$$\begin{aligned} M_2(P,G) &= \frac{I_{PG}}{M_{PG}} \\ M_1(P,G) &= \frac{I_{PG}}{M_{PG}} + \frac{M_{PG}}{N_P} \\ M_3(P,G) &= \frac{M_{PG} * I_{PG}}{N_P^2} \end{aligned} \tag{1}$$

where M_{PG} is the number of matches between the two images, I_{PG} is the number of correct matches determined by the robust estimation of the homography between P and G, and N_P is the number of feature point detected in the image P.

The scores are maximised when $M_{PG} = I_{PG} = N_P$, that can happen, for instance, when the two images are identical. M_2 is the simplest one, and consider as score only the ratio of inliers. The other two also consider the portion of points of interest detected in the image P for which a match in the image G is found. In practice the second one considers the sum of the two components, while the last proposed score function weighs the ratio $\frac{M_{PG}}{N_P}$, that gives already an idea of similarity between faces [3], with the ratio $\frac{I_{PG}}{N_P}$, that is the fraction of feature points in the unknown image that can be considered good matches.

After iterating the process for each image in the gallery, the recognition is then performed determining the image $\hat{G}$ such that

$$\hat{G} = \underset{G \in \mathcal{G}}{\operatorname{argmax}} \, M_h(P,G) \tag{2}$$

where M_h is any of score functions defined in Equation (1), and then assigning to P the identity of $\hat{G}$.

4 Experimental Assessment

In this Section we report a preliminary assessment of the method proposed in this paper. As a measure of the goodness of the method we adopted the widely used recognition rate. For the experiments we considered two different data-sets: ORL, and Extended Yale B [10].

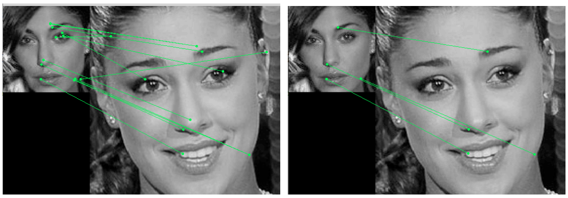

Fig. 2. Feature matching for face images of the same person. Left image shows correspondences obtained directly, there is a wrong match. Right image shows only correct matches after the outlier rejection from RANSAC.

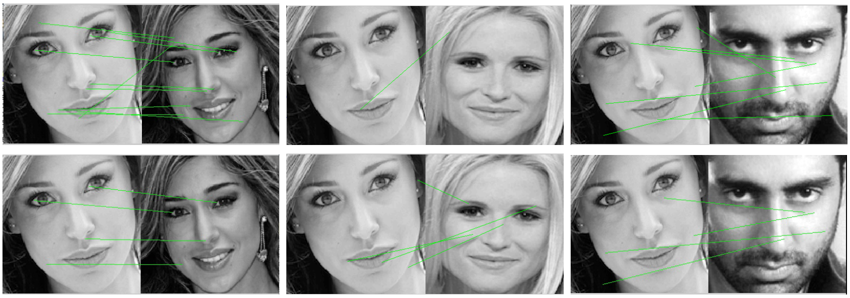

Fig. 3. Top row point matches before the RANSAC procedure. Bottom row matches determined as correct after the RANSAC procedure. Notice how the inliers determined in the case of different persons may still contain wrong matches, due to the number of wrong matches larger than 50% of the all matches.

.

ORL contains 400 images of 40 subjects taken at different times, varying the lighting, facial expressions (e.g., open or closed eyes, smiling or not smiling) and facial details (e.g., glasses or no glasses). The images of two subjects from the ORL database are shown in Figure 4.

The Extended Yale B contains 38 subjects and each subject has approximately 64 frontal view images taken under different illuminations conditions. Because of time reason we did not manage to run experiments on the whole data-set. The experiments shown in this paper were run on a randomly selected subset of 15 subjects for 10 illumination conditions, randomly chosen among the set of relatively good lighting conditions. The images used in the experiment for two subjects are shown in Figure 5.

The experiments were run as follows. For each subject j we randomly divided the data-set into a gallery $\mathcal{G}_j$ and a test set $\mathcal{P}_j$, with the size of the gallery $N_{\mathcal{G}_j}$ increasing from 2 to 9, and computed the recognition rate for each $N_{\mathcal{G}_j}$. We run this procedure 2500 times for each gallery size, and returned the mean recognition rate and standard deviation.

In Figure 7 the results returned on the ORL data-set are shown. The graph shows that the score M_1 returns very bad results, while with the other two scores a good recognition rate is achieved, with better results obtained with $M-3$, for which we have a recognition rate above 90% with a gallery of at least 4 images for each subject. Of course the performance improve with the gallery size.

Our results are compared with recent algorithms based on SIFT matching reporting results on the ORL data-set. In [23] it is reported a recognition rate of 93.5 on a single run with a gallery size of 5 for each subject. Under similar experimental conditions [8] report a recognition rate of 95.5. The recognition rate returned by our algorithm with a gallery of the size of 5 images for each subject is 95.06 with a standard deviation of 1.5. A gallery consisting of a single sample for subject is used in [9] reporting a recognition rate of 77.72 averaged over 10 runs, with a standard deviation of 1.64. The closest condition we can report is with a gallery size of 2, for which we obtain an average recognition rate of 77.85 with a standard deviation of 2.67. The comparative results are summarized in Table 1. In [18] are reported experiments on the ORL data-set using a simple matching strategy where performance close to the one returned by our experiments are obtained using less standard metrics for the matching.

The results on the Yale data-set are reported in Figure 6. For the comparison among the three score we have results similar to the ORL data-set, with M_1 performing poorly, and MM_3 returning the best results. For M_3 we have a recognition rate above the 98% with a gallery size as small as 3 samples per subject. Similar results are reported in [14], where a similar experiment run using all the subjects, and dividing the best lighting conditions into a gallery of around 7 images per subject (exact figures are not reported), and around 4 test images per subject. The recognition rate reported is 100%, as in our case.

The much better performance on the Yale data-set are probably due to the fact that the images, in this case, are all well cropped frontal views.

Among the three scores considered in this work, M_3 is the one returning the best results, as it is the one which mixes in a better way both the ratios involved in M_2 and M_3. What it is worth commenting at this point is the poor results returned by M_1. This is mainly due to the fact that, as it is possible to see from the examples in Figure 3, two images of the same face can return a very small number of matches, so that even a small set of inliers (say four) is returned by the RANSAC procedure, this can produce a large value for M_1, causing an erroneous recognition.

Table 1. Comparison with recent literature using SIFT matching on the ORL data-set. See text for more details.

	$N_{\mathcal{G}}$	Reported	Our result
[9]	1	77.22	77.85
[8]	5	95.5	95.06
[23]	5	93.5	95.06

Fig. 4. The images of two subjects included in the ORL data-set.

Fig. 5. The images of two subjects included in the Yale data-set.

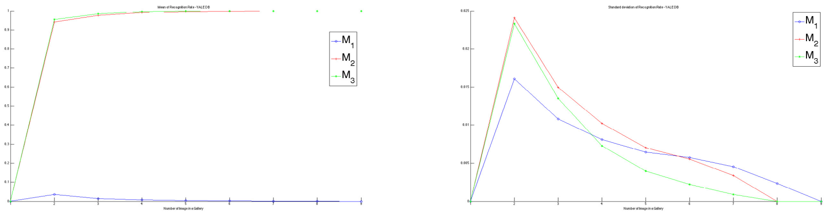

Fig. 6. Results on the Yale data-set. On the left the average recognition rate plotted against the size of the gallery $N_{\mathcal{G}}$. The recognition rate has been averaged over 2500 run for Each gallery size. On the left the graph of the standard deviation.

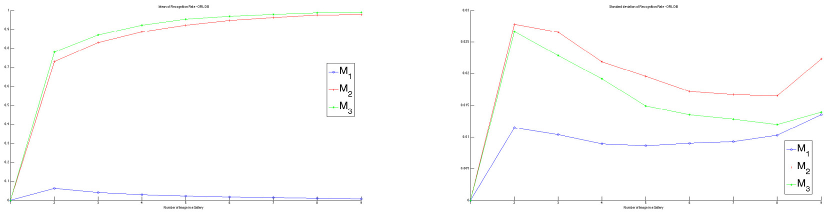

Fig. 7. Results on the ORL data-set. On the left the average recognition rate plotted against the size of the gallery $N_{\mathcal{G}}$. The recognition rate has been averaged over 2500 run for each gallery size. On the left the graph of the standard deviation.

5 Conclusions

In this paper we presented a face recognition system based on SIFT features. The method first establishes correspondences between feature points extracted from the face images, and then uses the number of correct correspondences, obtained assuming an homographic transform between the images, to build measurement to determine the likelihood of the similarity between the face images. The recognition is then performed using a rank-1 approach.

The result reported are comparable with recent result with other algorithms based on SIFT matching, showing that the method is promising. However more experiment are still needed for a definitive analysis of the performance of the method. The algorithm must be validated on the whole Extended Yale data-set. Further exhaustive experiments on the Feret data-set, and on the more difficult LFW data-set are planned.

Further work needs also to be done exploring how the performance change using local features different from the SIFT.

Acknowledgments. This work has been partially supported by the Smart Health 2.0 Project - PON04a2_C, funded by the Italian government (MIUR) in the program PON R&C 2007-2014.

References

1. OpenVIDIA: Parallel GPU computer vision. http://openvidia.sourceforge.net
2. Ahonen, T., Hadid, A., Pietikäinen, M.: Face description with local binary patterns: Application to face recognition. IEEE Transactions on Pattern Analysis Machine Intelligence **28**(12), 2037–2041 (2006)
3. Aly, M.: Face recognition using sift features. CNS/Bi/EE report 186 (2006)
4. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Speeded-up robust features (surf). Computer Vision and Image Understanding **110**(3), 346–359 (2008)
5. Bicego, M., Lagorio, A., Grosso, E., Tistarelli, M.: On the use of sift features for face authentication. In: Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop, pp. 35–41 (2006)
6. Calonder, M., Lepetit, V., Strecha, C., Fua, P.: BRIEF: binary robust independent elementary features. In: Daniilidis, K., Maragos, P., Paragios, N. (eds.) ECCV 2010, Part IV. LNCS, vol. 6314, pp. 778–792. Springer, Heidelberg (2010)
7. Dreuw, P., Steingrube, P., Hanselmann, H., Ney, H.: SURF-face: face recognition under viewpoint consistency constraints. In: Proceedings of the British Machine Vision Conference, pp. 1–11 (2009)
8. Geng, C., Jiang, X.: Face recognition using SIFT features. In: Proceedings of the International Conference on Image Processing, pp. 3313–3316 (2009)
9. Geng, C., Jiang, X.: Face recognition based on the multi-scale local image structures. Pattern Recognition **44**(10), 2565–2575 (2011)
10. Georghiades, A., Belhumeur, P., Kriegman, D.: From few to many: Illumination cone models for face recognition under variable lighting and pose. IEEE Trans. Pattern Anal. Mach. Intelligence **23**(6), 643–660 (2001)
11. Hartley, R.I., Zisserman, A.: Multiple view geometry. Cambrige University Press (2000)
12. Jafri, R., Arabnia, H.R.: A Survey of Face Recognition Techniques. Journal of Information Processing Systems **5**(2), 41–68 (2009)

13. Kisku, D., Rattani, A., Grosso, E., Tistarelli, M.: Face identification by SIFT-based complete graph topology. CoRR abs/1002.0411 (2010)
14. Križaj, J., Štruc, V., Pavešić, N.: Adaptation of SIFT features for robust face recognition. In: Campilho, A., Kamel, M. (eds.) ICIAR 2010. LNCS, vol. 6111, pp. 394–404. Springer, Heidelberg (2010)
15. Kumar, N., Berg, A., Belhumeur, P., Nayar, S.: Attribute and simile classifiers for face verification. In: Proceedings of the International Conference on Computer Vision, pp. 365–372 (2009)
16. Kuo, C.H., Lee, J.D.: A two-stage classifier using SVM and RANSAC for face recognition. In: Proceedings/TENCON IEEE Region 10 Annual International Conference, pp. 1–4 (2007)
17. Lalonde, M., Byrns, D., Gagnon, L., Teasdale, N., Laurendeau, D.: Real-time eye blink detection with GPU-based SIFT tracking. In: Proceedings of the Canadian Conference on Computer and Robot Vision, pp. 481–487 (2007)
18. Lenc, L., Král, P.: Novel matching methods for automatic face recognition using SIFT. In: Iliadis, L., Maglogiannis, I., Papadopoulos, H. (eds.) AIAI 2002. IFIP AICT, vol. 381, pp. 254–263. Springer, Heidelberg (2012)
19. Liu, C.: Capitalize on dimensionality increasing techniques for improving face recognition grand challenge performance. IEEE Transactions on Pattern Analysis and Machine Intelligence **28**(5), 725–737 (2006)
20. Lowe, D.G.: Object recognition from local scale-invariant features. In: Proceedings of the International Conference on Computer Vision, pp. 1150–1157 (1999)
21. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision **60**(2), 91–110 (2004)
22. Luo, J., Ma, Y., Takikawa, E., Lao, S., Kawade, M., Lu, B.L.: Person-specific SIFT features for face recognition. In: IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 2 (2007)
23. Majumdar, A., Ward, R.K.: Discriminative sift features for face recognition. In: Electrical and Computer Engineering, CCECE 2009. Canadian Conference on, pp. 27–30 (2009)
24. Meer, P., Mintz, D., Rosenfeld, A., Kim, D.Y.: Robust regression methods for computer vision: a review. International Journal of Compuer Vision **6**(1), 59–70 (1991)
25. Pham, T., Waillot, N., Lim, J., Chevallet, J.: Latent semantic fusion model for image retrieval and annotation. In: Proceedings of the ACM Conference on Information and Knowledge Management, pp. 439–444 (2007)
26. Se, S., Lowe, D., Little, J.: Vision-based mobile robot localization and mapping using scale-invariant features. In: Proceedings of the IEEE Conference on Robotics and Automation, pp. 2051–2058 (2001)
27. Shu, C., Ding, X., Fang, C.: Histogram of the oriented gradient for face recognition. Tsinghua Science and Technology **16**(2), 216–224 (2011)
28. Taigman, Y., Yang, M., Ranzato, M., Wolf, L.: Deepface: closing the gap to human-level performance in face verification. In: Conference on Computer Vision and Pattern Recognition (CVPR), pp. 1701–1708 (2014)
29. Turk, M., Pentland, A.: Eigenfaces for recognition. Cognitive Neuroscience **3**(1), 71–86 (1991)
30. Xiang, C., Fan, X., Lee, T.: Face recognition using recursive Fisher linear discriminant. IEEE Transactions on Image Processing **15**(8), 2097–2105 (2006)
31. Zhao, W., Chellappa, R., Philips, P., Rosenfeld, A.: Face recognition: A literature survey. ACM Computing Survey, 399–458 (2003)