New approach to the identification of the easy expression recognition system by robust techniques (SIFT, PCA-SIFT, ASIFT and SURF)

Ahmed Chater, Abdelali Lasfar
Laboratory of System Analysis, Information Processing and industry Management,
High School of Technology SALE Mohammadia Engineering School in Rabat, Morocco

Article Info

Article history:

Received Jul 25, 2019
Revised Jan 16, 2020
Accepted Feb 5, 2020

Keywords:

Face recognition
Feature matching
Four database
Normalization of feature
RANSAC with SURF
Recognition rate

ABSTRACT

In recent years, facial recognition has been a major problem in the field of computer vision, which has attracted lots of interest in previous years because of its use in different applications by different domains and image analysis. Which is based on the extraction of facial descriptors, it is a very important step in facial recognition. In this article, we compared robust methods (SIFT, PCA-SIFT, ASIFT and SURF) to extract relevant facial information with different facial posture variations (open and unopened mouth, glasses and no glasses, open and closed eyes). The simulation results show that the detector (SURF) is better than others at finding the similarity descriptor and calculation time. Our method is based on the normalization of vector descriptors and combined with the RANSAC algorithm to cancel outliers in order to calculate the Hessian matrix with the objective of reducing the calculation time. To validate our experience, we tested four facial images databases containing several modifications. The results of the simulation show that our method is more efficient than other detectors in terms of speed of recognition and determination of similar points between two images of the same face, one belonging to the base of the text and the other one to the base driven by different modifications. This method, which can be applied on a mobile platform to analyze the content of simple images, for example, to detect driver fatigue, human-machine interaction, human-robot. Using descriptors with properties important for good accuracy and real-time response.

Corresponding Author:

Ahmed Chater,
Laboratory of System Analysis,
Information Processing and industry Management,
High School of Technology SALE,
Mohammed V University, Rabat, Morocco.
Email: ahmedchater11@gmail.com, ali.lasfar@gmail.com

1. INTRODUCTION

Easy recognition has been a field of research interest in previous years, which is used by different applications in computer vision. Currently addressing the problems that arise from the acquisition of images with different variations in pose and variation of easy expression. There are some traditional algorithms for face recognition such: as EigenFace [1], FisherFace [2], 2D-PCA [3], and Elastic Graph Matching [4]. The Scale Invariant Feature Transform (SIFT) proposed by David G. Lowe [5] [6], has been widely used in object detection and recognition. There are also some works on the use of SIFT features in face recognition, such as SIFT-GRID proposed by M. Bicego [7] and SIFT CLUSTER proposed by Jun Luo [8]. Many traditional methods can extract the remarkable points based on local methods, and research has shown that (SIFT) works well in facial recognition, but the disadvantages of the technique take a long time to extract the descriptors and change of illumination. The authors propose other methods which is already used to reduce computation time for example: Kd-tree is used in the search step for the nearest connection, and PCA is proposed to reduce the dimensions of SIFT characteristics is called PCA-SIFT.

However, the technique SIFT still do not allow to respond the requirements of online services. Then, the authors propose another detector of the same performance as SIFT, called (SURF) is a detector robust in term of different transformation of face like we mentioned in the fourth title. Firstly (SURF) is used to extract the remarkable points using matrix approximation of Hessian applicate on the integral images in order to locate descriptor that allow to decrease the analysis time of image, and then we use the wavelets in the x and y directions to describe the distribution of the intensity in the vicinity of the remarkable points, in addition the detector (SURF) used only (64) dimension to decrease the time of calculation.

In this article, we have presented the comparison between the different robust detectors tested in our previous work by different variations in viewpoints [9, 10]. Our method based on the detector (SURF) with the RANSAC [11] algorithm which estimates the data between three steps. The first one consists to extract descriptors. The second one chooses random entry points and then estimates these parameters by the adjustment module. The third step compares these parameters by the compatibility of the adjustment module based on a certain maximum error threshold for objective to cancel out outliers and noise, look at the algorithm below. The result of the comparison of our method gives a good result in terms of speed of correspondence calculation and recognition rate with different variations of face change of the same person.

2. DESCRIPTOR EXTRACTION BY ROBUST METHODS

Key point extraction techniques are based on invariant to affine transformations among these techniques (SIFT, ASIFT, PCA-SIFT and SURF). The quality of operation of a facial recognition system is linked to the choice of detector for feature extraction because each technique is adapted to a given context. We chose the Speeded-Up Robust Features (SURF) detector over other methods because of its robustness and the use of second-order Gaussian partial derivatives, which improve the time of real-time image analysis**.** The different steps of the algorithm (SURF) for the extraction of the key points follow the following steps (Hessian matrix-based interest points, Interest point description and descriptor components).

2.1. Theory of surf (speeded-up robust features)

In 2006, Bay *et al*. [12] propose a new method of local description of points of interest. Named SURF (Speed-Up Robust Features). Strongly influenced by the SIFT approach, it couples a step of registration of the analysis area with the construction of a histogram of oriented gradients. The calculation process consists in determining the rotation (or recording) angle to be applied to the local description window. To this, the authors apply Haar wavelets to the integral image, thus significantly reducing computation time. These wavelets make it possible to calculate the first derivatives of the image on a square neighborhood and thus to study the distribution of the horizontal and vertical gradients. The responses of the wavelets then make it possible to draw the graph of distribution of the gradients and to deduce there from the angle of registration. On the initial image the circle represents the region of interest whose radius is equal to 6s where s corresponds to the characteristic scale extracted from the fast-Hessian detector

2.2. Hessian matrix-based interest points

The SURF detector is based on the determinant of the Hessian matrix [12]. In order to motivate the use of the Hessian, we consider a continuous function of two variables such that the value of the function at (x, y) is given by $f(x, y)$. The Hessian matrix (H) is the matrix of partial derivate of the function $f(x, y)$. Where

$$H(f(x, y)) = \begin{bmatrix} \dfrac{\partial^2 f}{\partial x^2} & \dfrac{\partial^2 f}{\partial x \partial y} \\ \dfrac{\partial^2 f}{\partial x \partial y} & \dfrac{\partial^2 f}{\partial y^2} \end{bmatrix} \quad (1)$$

The determinant of this matrix, Called discrimination, is calculated as follows:

$$\det(H) = \frac{\partial^2 f \partial^2 f}{\partial x^2 \partial y^2} - \left(\frac{\partial^2 f}{\partial x \partial y} \right)^2 \quad (2)$$

The discriminating value is used to determine the maximum and minimum of (2). If the result of the product of the negative eigenvalues the points is not a local extrema, then if the products of the positive eigenvalues value the points classified as extrema. In [6] and [8] it describes that a Hessian matrix can be done as a great detector for its high production in computational time and precision. Scale range can be obtained through the determinant of the Hessian or Hessian–Laplace detector. Given a point $P\ (x, y)$ in the image I, the Hessian matrix H (p, σ) in p at scale σ is defined as follows (3):

$$H(X,\sigma) = \begin{bmatrix} L_{xx}(P,\sigma) & L_{xy}(P,\sigma) \\ L_{yx}(P,\sigma) & L_{yy}(P,\sigma) \end{bmatrix} \quad (3)$$

Where $L_{xx}(P,\sigma) = I(x) * \frac{\partial^2 g(\sigma)}{\partial x^2}$. The convolution of the second order Gaussian derivative $\frac{\partial^2 g(\sigma)}{\partial x^2}$ with the image at point $I(P)$ and similarly by L_{yy}, and $L_{xy}(P,\sigma) = I(P) * \frac{\partial^2 g(\sigma)}{\partial xy}$.

These derivatives are known as Laplacian of Gaussians. Based on the turn indicator, we can compute the determinant of the Hessian for each pixel in the image and use the power to and the remarkable points. Then the hessian determinant calculates to extract the remarkable. Lowe [4] found a performance increase in approximating the Laplacian of Gaussians by a difference of Gaussians. In a similar manner, Bay [13] proposed an approximation to the Laplacian of Gaussians by using box-lter representations of the respective kernels. The (SURF) approach exceeds (SIFT) in terms of speed to calculate points of interest and their accuracy. SURF uses the built-in image box filter against the (SIFT) approach to apply the filter to each image size in the image pyramid. In the SURF strategy, In the SURF approach, the box filter in (Figure 1) starts with a 9×9 size filter as the initial scale layer where it is referred to as scale s=1.2 (the approximated Gaussian derivative with the value σ=1.2) and instead of having image pyramids, the original image will be filtered by bigger masks, denoted them by D_{xx}, D_{xy}, and D_{yy}. Hessian determinant using the approximated Gaussians and it is expressed as follows (4):

$$\det(H_{approx}) = D_{xx} D_{yy} - (w D_{xy})^2 \quad (4)$$

The relative weight equals $w = 0.9$, to balance the expression of (4). Which allows the conservation of energy between Gaussian nuclei approximated and Gaussian nuclei.

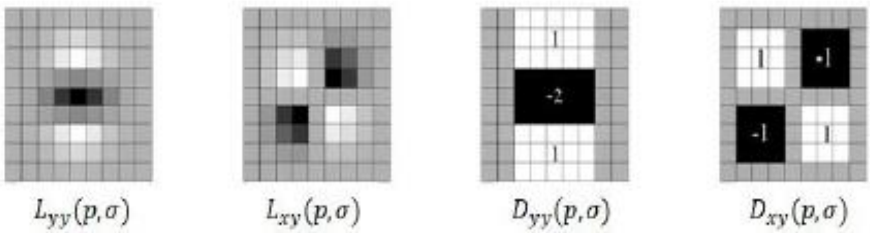


Figure 1. Laplacian of Gaussian Approximation. Top Row: The discretized and cropped second order Gaussian derivatives in the x, y and xy -directions. We refer to these as L_{xx}, L_{xx}, and L_{xy}.

2.3. Interest point description and descriptor components.

Haar wavelet that allows extracting the properties of the points of interest, then determine the orientation in both the (x and y) direction as follows. First, create a square region localized on the points of interest, and then determine the different direction that introduced in [14]. Second, we divide the main region into an equal sub-region (4x4) as shown in Figure 2, which preserves the data of interest for each sub-region. Then we determine the wavelet responses of Haar at 5x5 by equidistance points.

With dx: wavelet response in the horizontal direction and dy: wavelet response in the vertical direction (filter size 2s). To increase the robustness towards geometric deformations and localization errors, the responses dx and dy are first weighted with a Gaussian $\sigma = 3.3s$ centered at the interesting point. Third, before providing data on the polarity of the intensity variations, we also determined the sum of the absolute values of the responses, $|dx|$ and $|dy|$. So, each sub-region has a four-dimensional descriptor vector (v) for its underlying intensity construction as follows:

$$v = \left(\sum d_x, \sum d_y, \sum |d_x|, \sum |d_y|\right) \tag{5}$$

By concatenating all sub-regions of dimension (4x4), to obtain a descriptor vector of length 64. The wavelet responses are invariant in illumination. Invariance to contrast (a scale factor) is obtained by converting the descriptor into a unit vector.

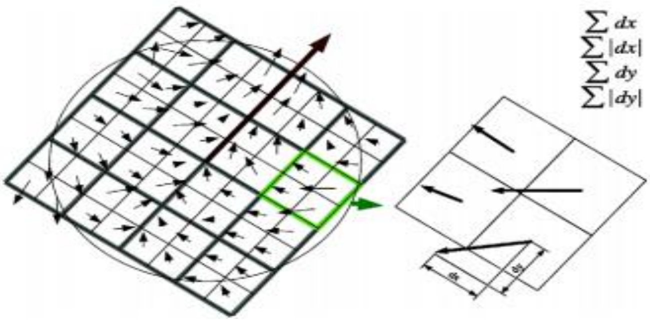


Figure 2. Construction the 4-dimensional descriptor (SURF)

3. DATABASE AND ALGORITHM PROPOSED

One of the most important aspects of the development of new recognition system or facial expression detection is the choice of the database that will be used to test this system. In addition, common databases are necessary to benchmark our approach. In this article, we will present four (ORL, Grimace, Faces95 and Faces96). Popular easy expression databases that are publicly and freely available to be used to evaluate our algorithm.

3.1. Database ORL

The algorithms were evaluated to validate our experience, we used database a reference ORL in [15] which contains 40 subjects and each subject contains 10 different faces of the same person. The resolution of all 8-bit images is 112×92. Pixel, with 256 grayscales per pixel, the different variations that are: change of lighting and variation of easy expression (open mouth and not open, glasses and not glasses, eyes open and closed). As shown in Figure 3.

Figure 3. Example of images of faces of two individuals from the database (ORL)

3.2. Database Grimace

The algorithm is evaluated to validate our experience, we used database a reference Grimace in [16] which contains 18 subjects and each subject contains 20 different faces of the same person. The resolution of all 8-bit images is 180x200 pixel, with 256 grayscales per pixel, the different variations that are: (Backgrounds, Head Scale, Head turn, tilt and slant, Position of the face in the image, Image lighting variation, Expression Variation, Additional comment). As shown in Figure 4.

Figure 4. Example of face images of two individuals from the Grimace dataset

3.3. Database Faces95

We used the Faces95 reference database in [17] which contains 72 subjects and each subject contains 20 different faces of the same person. The resolution of all 8-bit images is 180x200, the different variations being: (Backgrounds: Head Scale, Head turn, tilt and slant, Position of face in image, Image lighting variation, Expression Variation). As shown in Figure 5.

Figure 5. Examples of face images of two individuals from the Faces96 dataset

3.4. Database description Faces96

We used database a reference Faces96 in [18] which contains 152 subjects and each subject contains 20 different faces of the same person. The resolution of all 8-bit images is 196x196 (Variation of individual's images Backgrounds). Head Scale, Head turn, tilt and slant, Position of face in image, Image lighting variation, Expression Variation: some expression variation, Additional comment, images were taken in a single session. As shown in Figure 6.

Figure 6. Examples of face images of two individuals from the Faces96 dataset

4. ALGORITHM PROPOSED

This algorithm is based on our work in [19, 20]. We decompose the database into two databases which are the test database and the training database by different percentages. For example, 50% test base and 50% training base, 40% test base and 60% training base and 30% test base and 70% training base. Then between the two images, one belongs to the test base and the other to the training base, then extract the remarkable points for each one by robust methods (SIFT, PCA-SIFT, ASIFT, and SURF) after normalizing data then selects the best points by the RANSAC algorithm. Then determines the similar points by the metric of the Euclidean distance between two vectors. The steps of the proposed method summarize in the algorithm below as shown in Figure 7. The proposed technique (the point normalization detected by SURF and then the association with the RANSAC algorithm) is illustrated in the Figure 8 below. The validation of our experience allows us to have text by four databases (ORL, Faces95, Face96, and Grimace) according to different variations of faces.

Start.

Step 1 : Input two images one belonging to the test database and the other to the training database

Step 2 : Extract the remarkable points of the face by robust methods (SIFT, PCA-SIFT, ASIFT, and SURF).

Step 3 : Normalize data detects

Step 4 : Identification of outliers by the RANSAC algorithm, then cancellation of these points to establish the correct matching

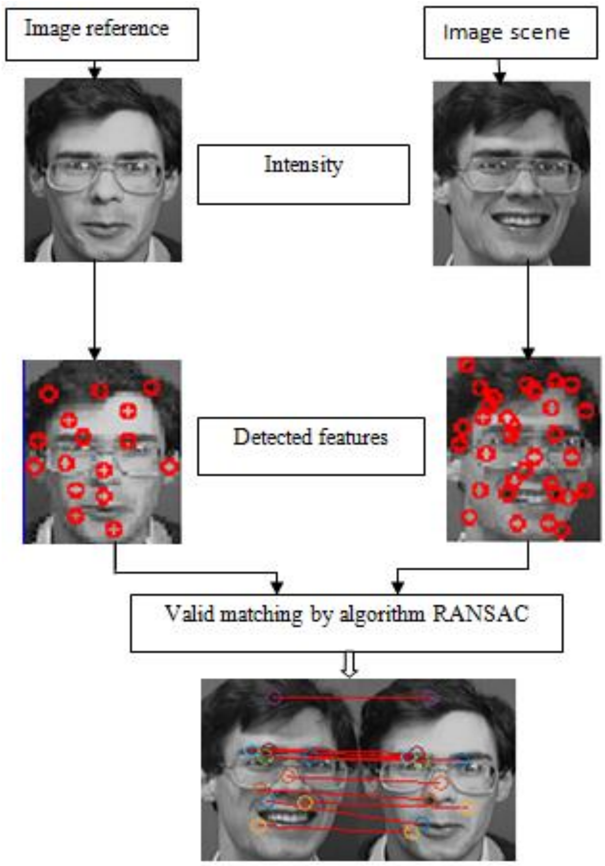
Step 5 : Calculate the recognition rate by the formula often

$$Recognition\ rate = \frac{correct.\ matches}{total.\ matches} x100 \tag{6}$$

Step 6 : Repeat the steps for each image in the query set and gallery apply the different detectors to assign a simulated image

Stop.

Figure 7. Basic steps involved in the application of the proposed method



Identical face

Figure 8. Illustration of the proposed method by databases ORL

4.1. Some simulation results and discussion

In this section, we have studied the different robust detectors using the (SIFT, PCA-SIFT, ASIFT) [21, 22, 23, 24] and SURF methods by varying the expression of the faces of the same person. Then, we have measured the number of descriptors, the number of matches and the processing time by different detectors according to the change in facial expression. Using the database images of real faces, the Figures 9-14 below show some examples of the extraction of descriptors and similar points by different robust detectors. After having several simulations texts, we find that the detectors (ASIFT) exceed the detector (SIFT) in terms of descriptor. As illustrated in the Figure 11.

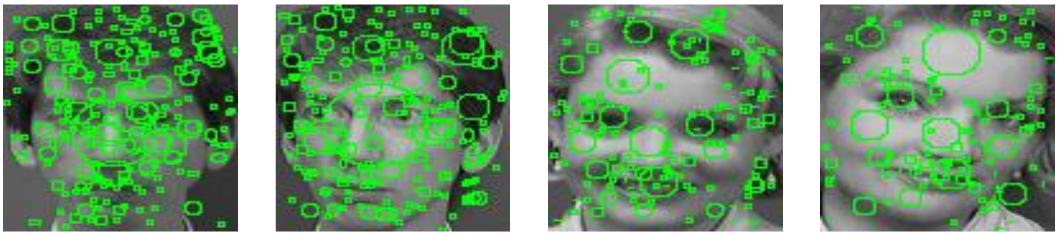


Figure 9. The two left faces the extraction by feature ASIFT and the two right faces by feature SIFT

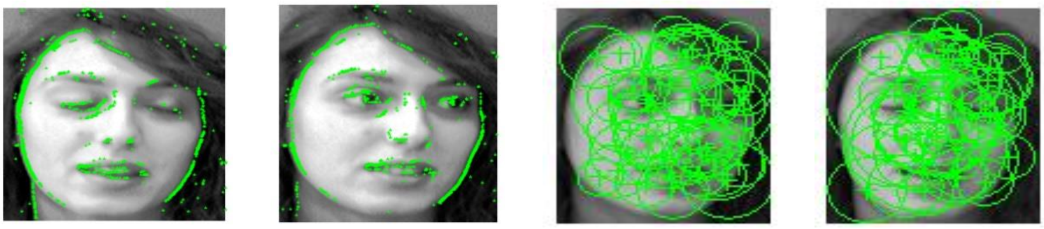


Figure 10. Feature extraction by PCA-SIFT and SURF

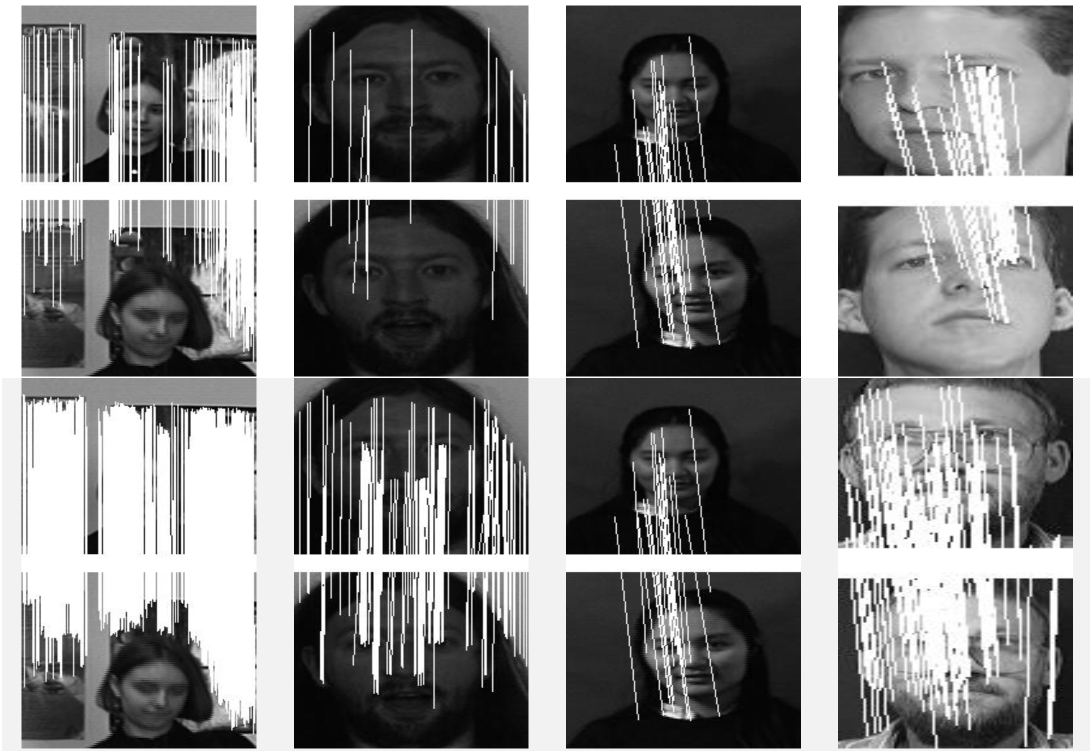


Figure 11. Correspondence between learning and evaluation faces, in the first line (SIFT) and the second line (ASIFT)

The simulation result shows that the ASIFT detector gives the number of descriptors higher than SIFT as illustrated in Figure 11. Hereafter, we use the PCA-SIFT detector, in Figure 12 which allows significant improvement in terms of corresponding power. The simulation results by the proposed method. We did some tests on the basis (ORL, Faces95, Faces96, and Grimace). First test: the two faces of different people, the Figure 13, below shows the result. When we put two different faces, we do not find any similar descriptor, as illustrated in the Figure 13. Second test: the two faces of the same person belong to four databases with different position, Figure 14 below shows the result.

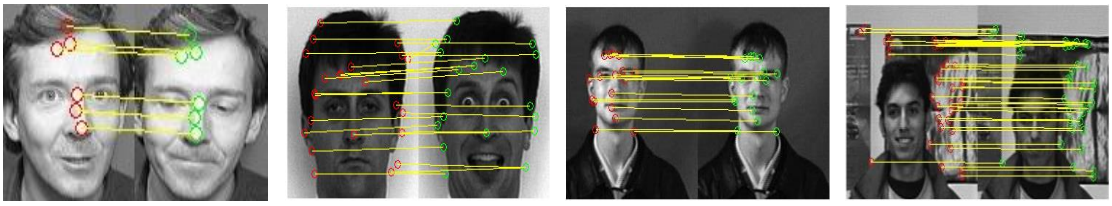


Figure 12. PCA-SIFT matches between learning and evaluation faces

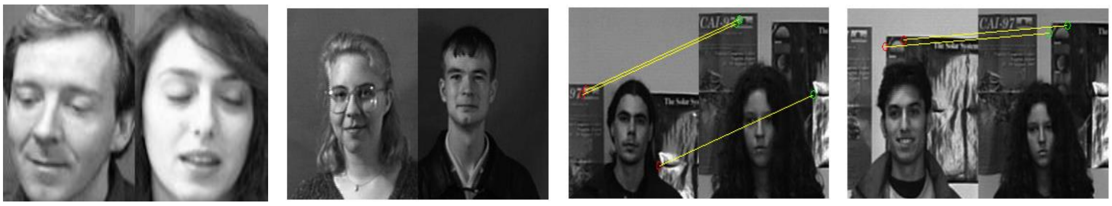


Figure 13. 0 key point matches normalization by SURF+RANSAC with different faces

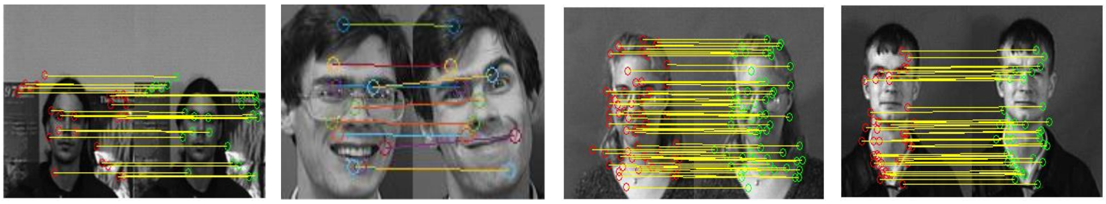


Figure 14. Matches of key points normalize by SURF+RANSAC

Validation of our approach based on a comparison between the following methods (SIFT, PCA-SIFT, ASIFT) and the proposed technique, we have clarified our method, increased the recognition rate compared to other existed techniques. The results of our simulation can be summarized in the Table 1 by four databases. The results of our database (Faces95) simulation can be summarized in the Table 2. The results of our database (Faces96) simulation can be summarized in the Table 3. The results of our database (Grimace) simulation can be summarized in the Table 4. The test on the four databases, as illustrated in the tables above, shows that our method offers satisfactory results in terms of recognition rates as shown in Table 5.

Table 1. The simulation result in terms of the average of the detected descriptors, the average of correct matches, recognition accuracy on database (ORL)

Method	The average of the detected descriptors	The average of correct matches	Recognition accuracy (%)
SIFT	40	37.6	94
ASIFT	60	57	95
PCA-SIFT	20	19.12	95.6
The method proposes	30	29	96.6

Table 2. The simulation result in terms of the average of the detected descriptors, the average of correct matches, recognition accuracy on database (Faces95)

Method	The average of the detected descriptors	The average of correct matches	Recognition accuracy (%)
SIFT	45	37.6	95.16
ASIFT	70	67	95.73
PCA-SIFT	25	24.13	95.52
The method proposes	36	34.77	96.6

Table 3. The simulation result in terms of the average of the detected descriptors, the average of correct matches, recognition accuracy on database (Faces96)

Method	The average of the detected descriptors	The average of correct matches	Recognition accuracy (%)
SIFT	48	45.95	95.74
ASIFT	80	76.72	95,9
PCA-SIFT	37	35.37	95.6
The method proposes	28	27.16	97

Table 4. The simulation result in terms of the average of the detected descriptors, the average of correct matches, recognition accuracy on database (Grimace)

Method	The average of the detected descriptors	The average of correct matches	Recognition accuracy (%)
SIFT	36	34.18	94.97
ASIFT	74	71.48	96,6
PCA-SIFT	18	17.22	95.67
The method proposes	25	24	97

Table 5. Average processing time between two images of the same face according to different variants, application on the following databases (ORL, Faces96, Faces96, and GRIMACE)

Method	SIFT	PCA-SIFT	ASIFT	The method proposes
Average time of comparison between two images (s) by (ORL)	0.78	0.69	1.956	0.363
Average time of comparison between two images (s) by Faces96	0.78	0.74	1.586	0.572
Average time of comparison between two images (s) by Grimace	1.2	0.9	1.8	0.675
Average time of comparison between two images (s) by Faces96	0.54	0.52	1.406	0.493

After comparing our method with other techniques, we find that our method outperforms the others in terms of time to identify the corresponding face, as shown in the Table 5. The results of our simulation show that the proposed method responds faster than other detectors, and give good results by different changes, as it is shown in the previous simulation. So that we can apply to track moving objects in real time. In the future works follow we text our method on the following database public. Ck [25], Oulu-CASIA [26].

4 CONCLUSION

In this article, we have compared the robust detectors with the proposed method to determine the different parameters. The test is made on the four databases (ORL, Faces95, Face96, and Grimace). The simulation results show that the proposed method (the point normalization detects by SURF and then the association with the RANSAC algorithm algorithm) gives good results in terms of determining similar descriptors by different variations in facial pose, recognition rate and average time of comparison between two, one belongs to the test database and the other one to the training database, compared to (SIFT, ASIFT, PAC-SIFT). This approach that we proposed gave good results in terms of recognition rate and image analysis time. So, this approach can analyze images in real time. In the future work, the evaluation of our technique is applicate dataset at public. Ck, Oulu-CASIA compared to the result found.

REFERENCES

[1] H-F. Huang and S-C. Tai, "Facial expression recognition using new feature extraction algorithm," *Electron. Lett. Comput. Vis. Image Anal.*, vol. 11, no. 1, pp. 41–54, 2012.

[2] Yang, Ming-Hsuan. "Kernel Eigenfaces vs. Kernel Fisherfaces: Face Recognition Using Kernel Methods." Fgr. Vol. 2. 2002.

[3] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," *In Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 586-591, 1991.

[4] J. Yang, D. Zhang, A. F. Frangi, and J. Yang, "Two-dimensional PCA: A new approach to appearance-based face representation and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 1, pp. 131-137, 2004.

[5] D. G. Lowe, "Distince image features from scale-invariant keypoints," *Int. Journal of Computer Vision*, vol. 60, no. 2, pp. 91-110, 2004.

[6] J. Luo, Y. Ma, E. Takikawa, S. H. Lao, M. Kawade, and B. L. Lu, "Person-specific SIFT features for face recognition," *International Conference on Acoustic, Speech and Signal Processing (ICASSP2007)*, Hawaii, vol. 2, pp. 593-596, 2007.

[7] X. Qu *et al*., "Evaluation of SIFT and SURF for vision based localization," *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, vol. 41-B3, pp. 685-692, 2016.

[8] P. M. Kumar *et al.*, "Intelligent face recognition and navigation system using neural learning for smart security in Internet of Things," *Cluster Computing*, vol. 22, no. 4, pp. 7733-7744, 2017.

[9] *Wang, Hao, et al. "Cosface: Large margin cosine loss for deep face recognition." Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018.*

[10] A. Chater and A. Lasfar, "Detection of image descriptors and modification of the weighting function for the estimation of the fundamental matrix using robust methods," *Journal of Engineering and Applied Sciences*, vol. 13, no. 7, pp. 1835-1843, 2018.

[11] M. Ghergherehchi, S. Y. Kim, H. Afarideh, and Y. S. Kim, "RANdom sample consensus (RANSAC) algorithm for enhancing overlapped etched track counting," *IET Image Process.*, vol. 9, no. 2, pp. 97–106, 2015.

[12] H. Bay, A. Ess, T. Tuytelaars, L. Van Gool, "Speeded-up robust features (SURF)," Comput. Vis. Image Underst., 110(3), 346-359 (2008).

[13] L. Shao *et al.*, "Spatio-temporal Laplacian pyramid coding for action recognition," IEEE Transactions on Cybernetics, vol. 44, no. 6, pp. 817-827, 2014.

[14] A. Chater and A. Lasfar, "Robust Harris detector corresponding and calculates the projection error using the modification of the weighting function," *International Journal of Machine Learning and Computing (IJMLC)*, vol. 9, no. 1, pp. 62-66, 2019.

[15] AT&T Database of Faces 'ORL Face Database' AT&T Laboratories, Cambridge: http://cam-orl.co.uk/facedatabase.html

[16] Libor spacek's Facial image databases 'Grimace faceDatabase: http://cswww.essex.ac.uk/mv/allfaces/ Grimace.html

[17] Libor Libor spacek's Facial image databases 'face95 ImageDatabase: http://cswww.essex.ac.uk/mv/allfaces/ ace95.html

[18] Libor Libor spacek's Facial image databases 'face96 Image Database: http://cswww.essex.ac.uk/mv/allfaces/ ace96.html

[19] A. Chater and A. Lasfar, "Comparison of robust methods for extracting descriptors and facial matching," *International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS)*, Morocco, pp. 1-4, 2019.

[20] A. Chater and A. Lasfar, "New approach to calculating the fundamental matrix," *Int. J. Electr. Comput. Eng. IJECE*, vol. 10, no. 3, pp. 2357–2366, Jun. 2020.

[21] A. Vinay, V. S. Shekhar, A. Kumar C., S. Natarajan, and K. N. B. Murthy, "Affine-scale invariant feature transform and two-dimensional principal component analysis: A novel framework for affine and scale invariant face recognition," *IET Computer Vision*, vol. 10, no. 1, pp. 43-59, 2016.

[22] D. Mishkin, J. Matas, and M. Perdoch, "Mods: Fast and robust method for two-view matching," *Computer Vision and Image Understanding*, vol. 141, pp. 81-93, 2015.

[23] Jiang, DaYou, and Jongweon Kim. "Artwork painting identification method for panorama based on adaptive rectilinear projection and optimized ASIFT." Multimedia Tools and Applications 78.22, pp. 31893-31924, 2019.

[24] Sabharwal, Tanupreet, et al. "Recognition of surgically altered face images: an empirical analysis on recent advances." Artificial Intelligence Review 52.2, pp. 1009-1040, 2019.

[25] T. Kanade, J. F. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 46-53, 2000.

[26] G. Zhao, X. Huang, M. Taini, S. Z. Li, and M. Pietikainen, "Facial expression recognition from near-infrared videos," *Image and vision Computing*, vol. 29, no. 9, pp. 607-619, 2001.