

Method of optimization of the fundamental matrix by technique speeded up robust features application of different stress images

Ahmed Chater, Hicham Benradi, Abdelali Lasfar

Laboratory of System Analysis, Information Processing Management and Industry, University Mohammed V, Rabat, Morocco

Article Info

Article history:

Received Feb 27, 2021

Revised Jul 19, 2021

Accepted Aug 14, 2021

Keywords:

Epipolar geometry

Feature matching

Fundamental matrix

Optimization function

Standard eight point algorithm

ABSTRACT

The purpose of determining the fundamental matrix (F) is to define the epipolar geometry and to relate two 2D images of the same scene or video series to find the 3D scenes. The problem we address in this work is the estimation of the localization error and the processing time. We start by comparing the following feature extraction techniques: Harris, features from accelerated segment test (FAST), scale invariant feature transform (SIFT) and speed-up robust features (SURF) with respect to the number of detected points and correct matches by different changes in images. Then, we merged the best chosen by the objective function, which groups the descriptors by different regions in order to calculate F. Then, we applied the standardized eight-point algorithm which also automatically eliminates the outliers to find the optimal solution F. The test of our optimization approach is applied on the real images with different scene variations. Our simulation results provided good results in terms of accuracy and the computation time of F does not exceed 900 ms, as well as the projection error of maximum 1 pixel, regardless of the modification.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Ahmed Chater, Hicham Benradi, Abdelali Lasfar

Laboratory of System Analysis, Information Processing Management and Industry, University

Mohammed V

Rabat, Morocco

Email: ahmedchater11@gmail.com, benradi.hicham@gmail.com, ali.lasfar@gmail.com

1. INTRODUCTION

The relationship that links two or more images with different changes is called epipolar geometry [1]. The relation that links the scenes with different modifications named fundamental matrix. The determination of F is based on the calibrations of the intrinsic and extrinsic parameter of the camera, among these application of the latter and the processing in real time [2]–[5]. The determination of F is based on the remarkable points detected in the images. The detection of these points is done by two techniques which are manual and automatic. The manual technique of selection of the points gives many errors when we have several images or videos. On the other hand, the automatic technique performs better than the manual technique.

There are several automatic image feature extraction methods, for example, Harris [6], features from accelerated segment test (FAST) [7], scale invariant feature transform (SIFT) [8], and speed-up robust features (SURF) [9]. In this work, we use automatic techniques to extract key points from images with different variations. In the literature, we find that both techniques (FAST and Harris) give good results in terms of computation time but are sensitive to orientation and lighting variations. On the other hand, both algorithms (SIFT and SURF) are acceptable regardless of the scene modifications [10]. Then the step of determining the correspondence of key points by extraction techniques, the calculation of F by standard robust methods with eight points in [11]. The points received a response for the calculation of F [12]. This technique is tricky for

the determination of the F matrix [13]. Among the techniques of the evaluation of the extracted points we find in the literature (LMedS) [14], random sample consensus (RANSAC) [15] and m-estimator [16]. In this paper we present a technique for determining the relationship between two images with different modifications in order to optimize the computation time and the projection error. The results of the simulations obtained is better can be applied to real time processing.

2. RESEARCH METHOD

2.1. Feature extraction techniques

The totality of the F estimation techniques require a number of input match points [17]. In the literature, there are two categories for extracting the descriptors, one based on the gray level intensity function and the other based on the binary descriptors (LBP, TP-LBP and FP-LBP) [18]–[21]. In this article, we have extracted the characteristics by the techniques previously cited in the introduction.

Then we compare them in terms of the number of points extracted and the similarity and processing time by different variations of the scene. Mikolajczyk and Schmid in [22], he compared Harris [6], FAST [7], and SIFT [10] and other detection techniques and noticed that SIFT has provided very good returns in terms of rotation and change of views as well as scale. Then [11], the SIFT descriptor which allows to detect a larger number of descriptors than the SURF method [21], but the latter is more efficient, hat SIFT has provided very good returns in terms of processing.

2.2. Detector SURF (speed-up robust features)

Detector SURF [21] proposes a new method for local description of points of interest, called SURF and it is strongly influenced by the SIFT approach where it couples a recording step of the analysis area with the construction of a histogram of the oriented gradients. The calculation process consists of two steps which are the extraction of the points of interest and the description of the points of interest then each step contains three sub-steps. i) The extraction of points of interest: create an approximation of the Hessian matrix: is done by the product of the convolution of the integral image by the Gaussian filter with the scale factor increasing. Calculate the responses of the kernels used: is done by subtracting between two neighboring images belonging to the same octave in order to determine the stable points. Finding maxima at scale and in space: this step is based on the Taylor development of the Gaussian difference function (DoG) in order to determine the position and scale of the detected points. ii) Description of the points of interest: determine the size of the study descriptor window in circle form which is characterized by the circle radius and scale. Get the dominant orientation: The calculation process consists of determining the rotation (or registration) angle to be applied to the local description window. To do this, the authors apply Haar wavelets to the integral image, which considerably reduces the computation time. These wavelets make it possible to calculate the first derivative of the image on a square neighborhood and thus to study the distribution of horizontal and vertical gradients. Then, the responses of the wavelets are used to plot the distribution of the gradients and to deduce the angle of alignment on the initial image Extract the SURF descriptor: at each position, the x and y responses within the segment are summed and used to form a new vector with a size of 64. The different steps of the algorithm are summarized as shown in the Figure 1.

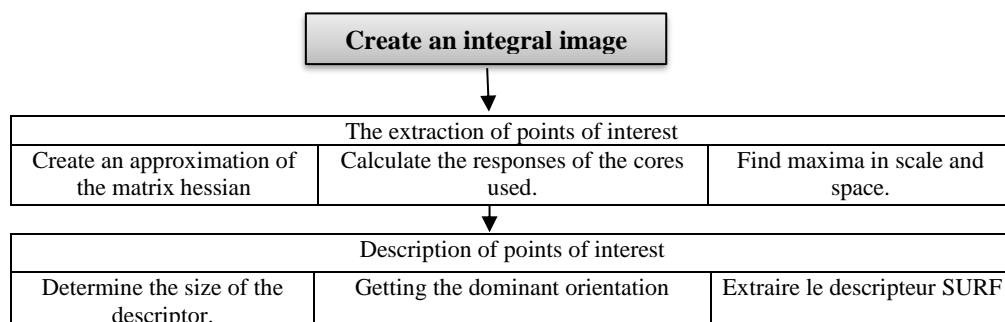


Figure 1. The stages of the creation of descriptor vectors by the SURF technique

2.3. Calculus fundamental matrix F

The matrix F is calculated from points extracted between two or more images with different variations of the scene. the determination of the selected points between the images is based on the Euclidean distance

metric. The different points chosen represent straight lines which are called epipolar line, equation (1) represents the points selected by different variation of the scene.

$$m_i^T F m_i = 0 \quad (1)$$

This relation connects by the extracted points between two images of the same scene $m_i = (x_i, y_i, w_i)$ and $m'_i = (x'_i, y'_i, w'_i)$ of the corrdonnes in 3D and F the fundamental matrix which connects the two images with different changes.

2.4. Statistical technique of calibration and evaluation

There are several techniques for determining the F, which is based on the matches in the matrix. Ces dernières se divisent en deux catégories: les classiques lineaire et non linéaires [14]. The work revealed that the linear technical methods according to [22], are at the limit of noise tolerance variation. Then the nonlinear methods (M-estimators, RANSAC, and LMeds) which are robust to noise according to [23]. The M-estimators technique is able to provide an excellent classification for noisy images, can also take into account the following aberrations and techniques (RANSAC and LMeds) unlike the M-estimators which are sensitive to noisy images and do not retain the aberrations when computing F. The (2) represents the link between the points detected by different variations of the scene.

$$A f = 0 \quad (2)$$

With f: the elements of the F in the form of the chosen descriptors:

$$f = [f_{11}, f_{12}, f_{13}, f_{21}, f_{22}, f_{23}, f_{31}, f_{32}, f_{33}]^T A = \begin{pmatrix} x'_1 x_1 & x'_2 x_2 & \cdot & x'_n x_n \\ x'_1 y_1 & x'_2 y_2 & \cdot & x'_n y_n \\ x'_1 w_1 & x'_2 w_2 & \cdot & x'_n w_n \\ y'_1 x_1 & y'_2 x_2 & \cdot & y'_n x_n \\ y'_1 y_1 & y'_2 y_2 & \cdot & y'_n y_n \\ y'_1 w_1 & y'_2 w_2 & \cdot & y'_n w_n \\ w'_1 x_1 & w'_2 x_2 & \cdot & w'_n x_n \\ w'_1 y_1 & w'_2 y_2 & \cdot & w'_n y_n \\ w'_1 w_1 & w'_2 w_2 & \cdot & w'_n w_n \end{pmatrix}^T$$

When we develop (2) we find nine equations and each equation presents the relationship between two images of different variations. The calculation of F is based on the best choices of the detected points.' but the disadvantage of this technique is when the detected points are incorrectly located [24], [25]. If the rank of F is different from two, the solution of the optimization (3) is performed by the least squares technique.

$$\min \sum_i (m_i^T F m_i)^2 \quad (3)$$

The solution of (3) is done by the singular value decomposition (SVD) technique. This technique was developed by Hartley according to [26]–[28] to make the method even more efficient. This way of proceeding allowed a notable improvement of the performances of the eight points method.

2.5. Robust methods

The iterative statistical techniques give good results when the data is noisy among these techniques are [15], [29], [30], LMeds [14], and M-estimators [27]. The ransac technique is based on the setting of thresholds to take the right descriptors for the estimation of F [25], [31]. The choice of the points by the (LMeds) technique is based on the distance between the point and the epipolar lines in order to determine the best F, the last technique of the M-estimator is integrated with the two previous techniques, it allows to divide the studied points into two groups (inliers and quasi-inliers). Equation (4) presents the problem to be addressed.

$$\min_F \sum_i w_i r_i^2 \quad (4)$$

$$r_i = (x_i, y_i, w) \begin{bmatrix} f_{12} & f_{12} & f_{13} \\ f_{21} & f_{22} & f_{23} \\ f_{31} & f_{32} & f_{33} \end{bmatrix} (x'_i, y'_i, w')^T$$

w_i : Is the optimization function. $r_i = f_{12}x_i x'_i + f_{21}y_i x'_i + f_{31}w x'_i + f_{21}x_i y'_i + f_{22}y_i y'_i + f_{23}w y'_i + f_{13}x_i w' + f_{23}y_i w' + f_{33}w w'$ (5) is referred to by [13], [25]. w_i :

$$w_i = w_i(p'_i, p''_i) = \begin{cases} 1 & r \leq \phi_i \sigma \\ \theta & \phi_i \sigma < r < \sigma \\ \frac{\theta \sigma}{|r_i|} & \sigma \leq r_i < \varphi \sigma \\ 0 & \varphi \sigma < r \end{cases} \quad (5)$$

ϕ : separator between two outlier and quasi-outlier regions and σ can be expressed as $\sigma = \frac{\text{median}(r_i)}{\lambda}$, the scale of the error and θ , take the random values between (0, 1);

$$E_{\text{sampson}} = \sum_i^n w_i \frac{(m_i^T F m_i')^2}{(F m_i)_1^2 + (F m_i)_2^2 + (F^T m_i')_1^2 + (F^T m_i')_2^2} \quad (6)$$

where $(F m_i)_j^2, j=1, 2 (F m_i)$. In the reference article [12], we find that the statistical method (LMeds) outperforms the RANSAC method in terms of the estimation of F. The M-estimator method is robust when the images are noisy.

2.6. The proposed technique

First, we load two images of the same scene by different variation, then we apply the following algorithms (Harris, FAST, SIFT, and SURF) and after the comparison between that, we find SURF the most robust by different variation. Secondly, we take the descriptor of the latter and normalize by all and then choose the eight random points to find the dimension matrix (8x9) and after decomposed by the SVD method to find the matrix of 3x3 of property following of equal rank 2, determines and zero. Finally, we apply the function of optimization to find the optimal solution (F) under the iterative algorithm. The different steps are represented in the Figure 2.

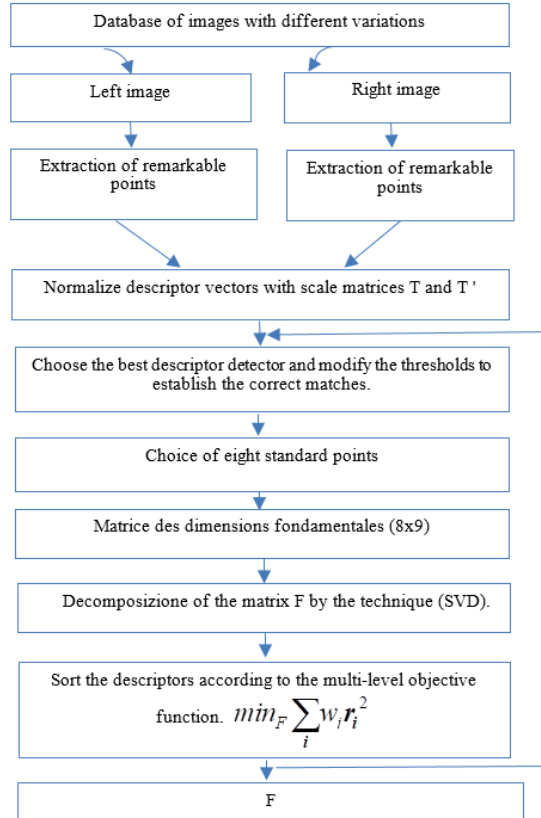


Figure 2. Proposed algorithm

3. RESULTS AND DISCUSSION

In this part, we study the four key point detectors (Harris, FAST, SIFT and SURF) and then we associate the RANSAC technique with seven normalized points by varying the brightness, rotation and the moving object. Then F is determined by two tests without limit the points extracted and with limit. The results of our simulations are shown in the Figure 3, Figure 4, and Figure 5. We illustrate the first test without limitation of the points extracted by different detector with different modifications. Then the second test with limitation of the points in order to optimize the calculation time and the projection error the Figure 6 represents that.

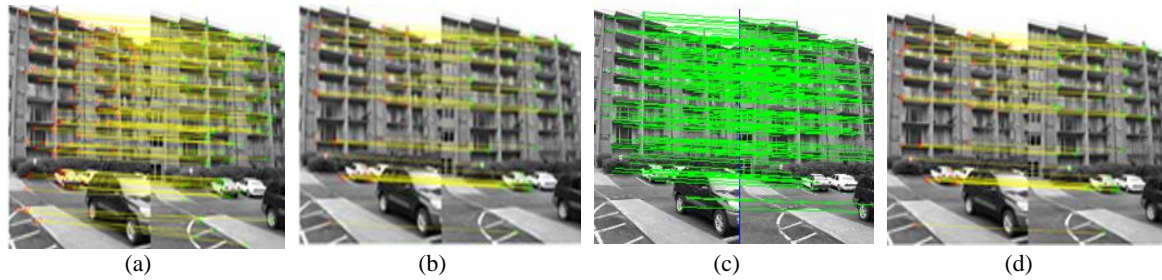


Figure 3. The primary test with motion variation: (a) SURF, (b) Harris, (d) SIFT, and (c) FAST

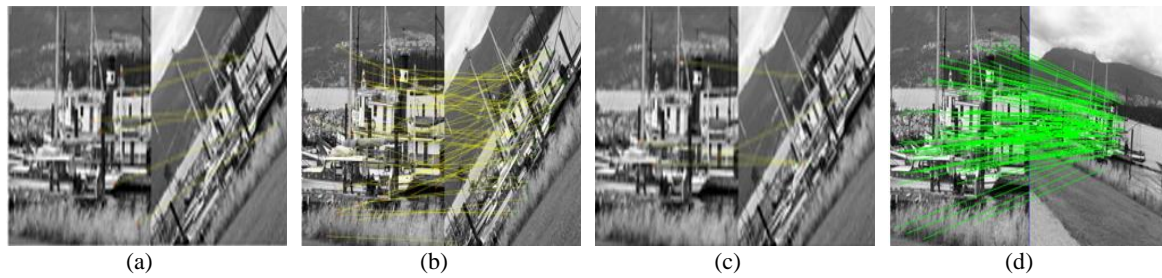


Figure 4. The primary test with rotation variation: (a) Harris, (b) SURF, (c) FAST(c), and (d) SIFT

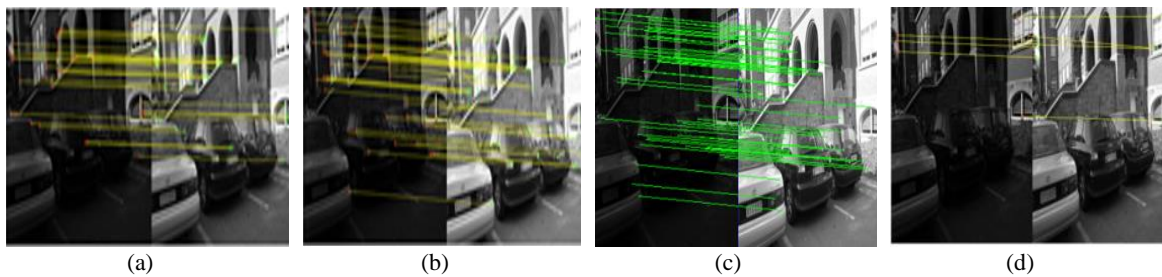


Figure 5. The primary test with lighting variation: (a) SURF, (b) Harris, (c) SIFT, and (d) FAST

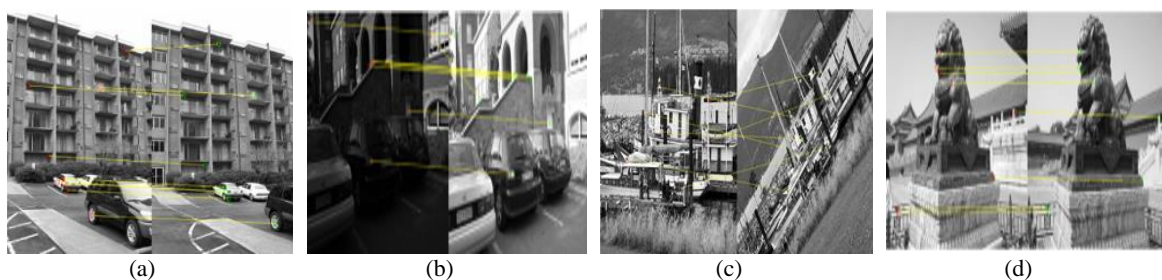


Figure 6. The secondary test with by changing the point of view: (a) moving object, (b) lighting, (c), rotation, and (d) uniqueness threshold

After limiting the number of points detected by (SURF) to have the best eight points to calculate (F), the Figures 3-6 show that. The different results of the simulation are grouped in the Table 1. It represents four algorithms of the extraction of the characteristics of the images and our technique. The results of our simulations show in the table above that the SIFT detector is better than the others. Then a monk projection error that we show in the result of the simulation in the following as shown in Figure 7. Figure 8 represents the comparison of our proposal and other techniques based on scene variation and feature extraction technique to optimize processing time. The results of our simulation show that our technique is better in terms of processing time with different modification of the scene does not exceed 900 ms.

Table 1. Comparison of the performance of the four algorithms of feature extraction and similar points, applied on the basis of images with different variation of the scene

The detectors	Without modification				Limitation of points extracted by surf
	Harris	FAST	SIFT	SURF	In motion
	In motion	In motion	In motion	In motion	Lighting
	Lighting	Lighting	Lighting	Lighting	Rotation
	Rotation	Rotation	Rotation	Rotation	
Kypnt1	302	449	54	423	82
	273	112	24	152	18
	264	76	16	111	43
Kypnt2	328	512	480	439	63
	271	110	142	126	21
	325	91	142	88	35
Inliers	40	54	153	150	8
	36	24	105	73	8
	32	16	86	46	8

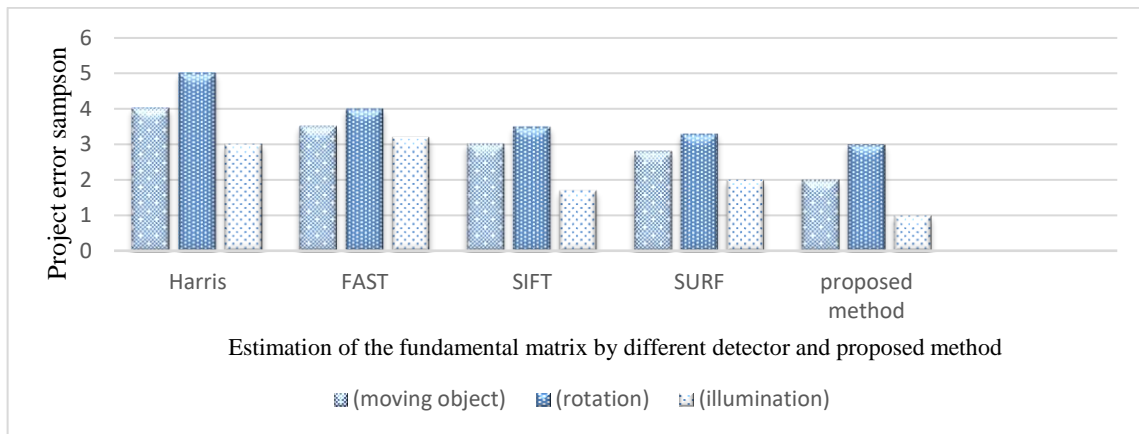


Figure 7. Error project Sampson

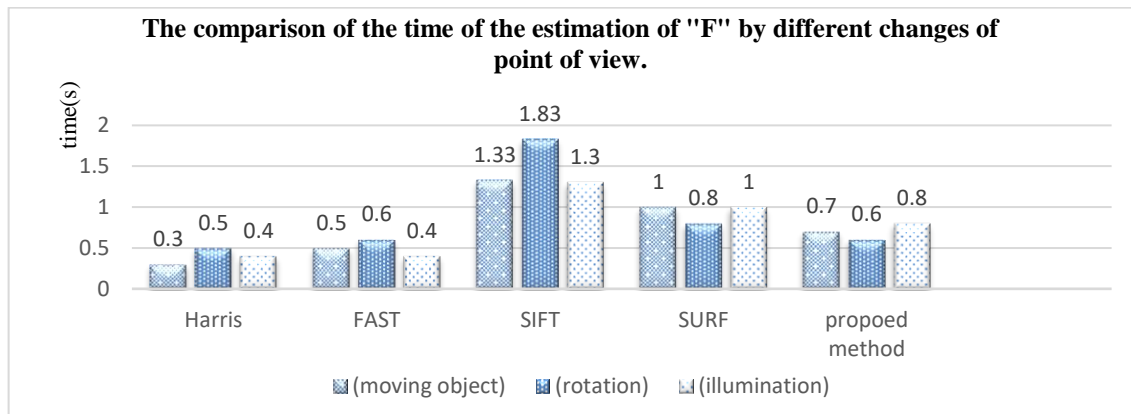


Figure 8. Estimation of the processing time according to the different changes of the scene and the detector

4. CONCLUSION

In this paper, we deal with the comparison between our approach and that of the literature according to the following parameters: the detected points (descriptor), the correct match, the computation time of the F estimate and the Sampson projection errors. This work is divided into two techniques. The first technique consists in merging, between the different techniques of feature extraction which are (SIFT, FAST and Harris), with the standardized eight-point algorithm of RANSAC. It obtained the number of student descriptors and the number of correspondents low compared to the number of extracted points. The major disadvantage of this technique, which eliminates outliers and student computing time, is that it does not apply to real-time applications. The second technique (our technique) based on the fusion between the detector (SURF) with the modification of the uniqueness threshold and the standardized eight-point M-estimation algorithm to obtain the fundamental matrix by different levels of the objective function.




REFERENCES

- [1] G. Xu and Z. Zhang, *Epipolar Geometry in Stereo, Motion and Object Recognition*. USA: Kluwer Academic Publishers, 1996.
- [2] Q.-T. Luong and O. D. Faugeras, "Self-calibration of a stereo rig from unknown camera motions and point correspondences," Springer Berlin Heidelberg, 2001, pp. 195–229.
- [3] G. F. Mani, P. M. Feniiosky, and S. Savarese, "D4AR-A 4-dimensional augmented reality model for automating construction progress monitoring data collection, processing and communication," *Electronic Journal of Information Technology in Construction*, vol. 14, pp. 129–153, Jun. 2009.
- [4] H. Fathi and I. Brilakis, "Automated sparse 3D point cloud generation of infrastructure using its distinctive visual features," *Advanced Engineering Informatics*, vol. 25, no. 4, pp. 760–770, Oct. 2011, doi: 10.1016/j.aei.2011.06.001.
- [5] J. Teizer and P. A. Vela, "Personnel tracking on construction sites using video cameras," *Advanced Engineering Informatics*, vol. 23, no. 4, pp. 452–462, Oct. 2009, doi: 10.1016/j.aei.2009.06.011.
- [6] H. Wang, M. M. Ullah, A. Kläser, I. Laptev, and C. Schmid, "Evaluation of local spatio-temporal features for action recognition," 2009, doi: 10.5244/C.23.124.
- [7] Q. Zhu, S. Avidan, M. C. Yeh, and K. T. Cheng, "Fast human detection using a cascade of histograms of oriented gradients," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2006, vol. 2, pp. 1491–1498, doi: 10.1109/CVPR.2006.119.
- [8] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the IEEE International Conference on Computer Vision*, 1999, vol. 2, pp. 1150–1157, doi: 10.1109/iccv.1999.790410.
- [9] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 3951 LNCS, Springer Berlin Heidelberg, 2006, pp. 404–417.
- [10] P. Kumar, S. Henikoff, and P. C. Ng, "Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm," *Nature Protocols*, vol. 4, no. 7, pp. 1073–1082, Jun. 2009, doi: 10.1038/nprot.2009.86.
- [11] N. Snavely, S. M. Seitz, and R. Szeliski, "Modeling the world from Internet photo collections," *International Journal of Computer Vision*, vol. 80, no. 2, pp. 189–210, Dec. 2008, doi: 10.1007/s11263-007-0107-3.
- [12] J. F. Huang, S. H. Lai, and C. M. Cheng, "Robust fundamental matrix estimation with accurate outlier detection," *Journal of Information Science and Engineering*, vol. 23, no. 4, pp. 1213–1225, 2007, doi: 10.6688/JISE.2007.23.4.16.
- [13] A. Chater and A. Lasfar, "New approach to calculating the fundamental matrix," *International Journal of Electrical and Computer Engineering*, vol. 10, no. 3, pp. 2357–2366, Jun. 2020, doi: 10.11591/ijece.v10i3.pp2357-2366.
- [14] Z. Zhang, "Determining the epipolar geometry and its uncertainty: a review," *International Journal of Computer Vision*, vol. 27, no. 2, pp. 161–195, 1998, doi: 10.1023/A:1007941100561.
- [15] R. Szeliski, *Computer Vision*. London: Springer London, 2011.
- [16] C. B. Xiao, D. Z. Feng, and M. D. Yuan, "An efficient fundamental matrix estimation method for wide baseline images," *Pattern Analysis and Applications*, vol. 21, no. 1, pp. 35–44, Jun. 2018, doi: 10.1007/s10044-016-0561-z.
- [17] K. Zhang, X. Z. Li, and J. X. Zhang, "A robust point-matching algorithm for remote sensing image registration," *IEEE Geoscience and Remote Sensing Letters*, vol. 11, no. 2, pp. 469–473, Feb. 2014, doi: 10.1109/LGRS.2013.2267771.
- [18] A. Chater and A. Lasfar, "Comparison of robust methods for extracting descriptors and facial matching," Apr. 2019, doi: 10.1109/WITS.2019.8723858.
- [19] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF," in *Proceedings of the IEEE International Conference on Computer Vision*, Nov. 2011, pp. 2564–2571, doi: 10.1109/ICCV.2011.6126544.
- [20] W. Wang, "Detection of panoramic vision pedestrian based on deep learning," *Image and Vision Computing*, vol. 103, p. 103986, Nov. 2020, doi: 10.1016/j.imavis.2020.103986.
- [21] Z. Yang and L. S. C. Pun-Cheng, "Vehicle detection in intelligent transportation systems and its applications under varying environments: A review," *Image and Vision Computing*, vol. 69, pp. 143–154, Jan. 2018, doi: 10.1016/j.imavis.2017.09.008.
- [22] G. Yao, J. Cui, K. Deng, and L. Zhang, "Robust harris corner matching based on the quasi-homography transform and self-adaptive window for wide-baseline stereo images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 1, pp. 559–574, Jan. 2018, doi: 10.1109/TGRS.2017.2751567.
- [23] M. Ballerini *et al.*, "Interaction ruling animal collective behavior depends on topological rather than metric distance: Evidence from a field study," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 105, no. 4, pp. 1232–1237, Jan. 2008, doi: 10.1073/pnas.0711437105.
- [24] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 11, pp. 1330–1334, 2000, doi: 10.1109/34.888718.
- [25] 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, doi: 10.3923/jeasci.2018.1835.1843.
- [26] Y. Li, S. Velipasalar, and M. C. Gursoy, "An improved evolutionary algorithm for fundamental matrix estimation," in *2013 10th IEEE International Conference on Advanced Video and Signal Based Surveillance, AVSS 2013*, Aug. 2013, pp. 226–231, doi: 10.1109/AVSS.2013.6636644.




- [27] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2004.
- [28] G. M. Jog, H. Fathi, and I. Brilakis, "Automated computation of the fundamental matrix for vision based construction site applications," *Advanced Engineering Informatics*, vol. 25, no. 4, pp. 725–735, Oct. 2011, doi: 10.1016/j.aei.2011.03.005.
- [29] A. Chater and A. Lasfar, "New approach to the identification of the easy expression recognition system by robust techniques (SIFT, PCA-SIFT, ASIFT and SURF)," *Telkommika (Telecommunication Computing Electronics and Control)*, vol. 18, no. 2, pp. 695–704, Apr. 2020, doi: 10.12928/TELKOMNIKA.V18I2.13726.
- [30] A. Chater and A. Lasfar, "Automated calculation of fundamental matrix from stereo images from a different point of view," in *Biometric Identification Technologies Based on Modern Data Mining Methods*, Springer International Publishing, 2021, pp. 105–118.
- [31] 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*, vol. 9, no. 1, pp. 62–66, 2019, doi: 10.18178/ijmlc.2019.9.1.766.

BIOGRAPHIES OF AUTHORS






Ahmed Chater    born on December 30, 1986 in Taounate. Degree in Engineering Sciences and Techniques, speciality: Image Processing, Laboratory of Systems Analysis, Information Processing and Industrial Management, Mohamed V University Rabat (Mohammadia School). My research focuses on: Segmentation and restoration of different types of color and gray scale images. Classification and recognition of facial expressions and machine learning. He can be contacted at email: ahmedchater11@gmail.com.



Hichman Benradi    born on October 07, 1985. He has a master degree in Data Engineering and Software Development from the Faculty of Science at Mohamed V University. Degree student at the Mohammadia School of Engineering in Rabat. His research focuses on facial recognition methods and image processing. He can be contacted at email: benradi.hicham@gmail.com.



Abdelali Lasfar    was born on January 10, 1971 in Salé. He is Professor of Higher Education at Mohammed V Agdal University, Salé Higher School of Technology, Morocco. His research focuses on compression methods, indexing by image content and image indexing and knowledge extraction from images. He can be contacted at email: ali.lasfar@gmail.com.