

# Automatic Remote-Sensing Image Registration Using SURF

Rochdi Bouchiha and Kamel Besbes

**Abstract**—Image registration is a key, essential element in analysis of Remote sensing images. Registration is critical both for initial processing and for end-user processing of those image products for data fusion, and change detection. This paper focused on the feature-based category of image registration algorithms. Many techniques for the detection and description of images' local characteristics have been proposed to register a set of images without user intervention. However, it is unclear which descriptors are more appropriate. The descriptors should be distinctive and at the same time robust to changes in viewing conditions as well as to errors of the sensor. In our evaluation, we have separated the detector from the descriptor as their performance depends on the interest point detector used. The descriptors are compared according to their recall and runtime efficiency and this deals with several geometric and photometric changes. We also propose an extension to the SURF descriptor and the results show the effectiveness of proposed improvements compared to base SURF method. Furthermore, we observe that the SURF descriptor outperforms the others' descriptors. Finally, based on the test results, we propose an approach to register automatically remotely sensed images.

**Index Terms**—Automatic registration, feature-based registration, remote sensing.

## I. INTRODUCTION

Image registration is a problem encountered in many applications of image processing where it's necessary to apply an analysis of two or more images of a scene obtained from various sensors, or the same sensor but at different times.

In remote sensing, there are two ways, which are used to register images, the first one is the parametric way using all required flight parameters of the remote sensing platform, and the other is the non-parametric way using only a set of GCPs (ground control points) for the image registration. These points are used to determine the parameters of a transformation function, which is then used to apply a deformation with an interpolation function on the new image without affecting the reference image.

In the second way, there are two methods: the first one is the manual registration [1], which is not feasible in the case where a large number of images must be registered, and with times cited from a few hours to 4-5 days for registering a single image [2]. And the second one is the automatic registration techniques that require little or no operator

supervision. These techniques have the advantage of multi-temporal and / or multi-sensor and / or multi-resolution information.

Recently, a number of works have been carried out, which propose techniques for the automatic registration of remotely sensed images [1]–[6], however, no comprehensive study has been carried out to analyze the performance of these techniques when dealing with different image transformation.

The registration process is usually done in four stages:

- 1) The first step is the feature detection. These features are: Points [1], [3], [7], Lines, edges and contours, Real objects in the scene such as buildings, Templates: Windows of predefined size.
- 2) The second step is the matching stage; it depends on the method used: area-based method or feature-based method or the combination of the two methods. The area-based methods [8] are not the object of the current work, called the correlation-like methods or template matching deal with the images without detecting salient points. The feature-based methods deal with two sets of features in the reference and in the sensed images. The aim is to find the pair-wise correspondence between them using their spatial relation or invariant descriptors. Each element in the reference image is compared with all elements in the sensed image. Pairs of points are accepted as relevant and are called control points when their characteristics are very similar. Others combine the two techniques as demonstrated in [4].
- 3) In the third step, the mapping function design, the parameters of the best transformation that models the deformation between the images can then be estimated using these tie points obtained in the previous step and the robust methods such as LMedS (Least Median of Square) and RANSAC (Random Sampling Consensus).
- 4) Finally, the image transformation and resampling step, the sensed image is transformed and thus to register the images using the estimated mapping function and a resampling technique (The nearest neighbor function, the bilinear and bicubic functions and others).

The use of feature-based methods is recommended if the images contain enough distinctive and easily detectable objects. This is usually the case of applications in remote sensing. The typical images contain a lot of details (towns, rivers, roads, forests, lakes, buildings, etc.).

In Section II, we present a state of the art on local descriptors used in our comparison as well as our contribution. Section III describes our evaluation criteria and the data sets. Experimental results and discussion are presented in Section IV; this is followed by the proposed

Manuscript received June 18, 2012; revised August 2, 2012

The authors are with the microelectronic and instrumentation laboratory, Faculty of Sciences, University of Monastir, Tunisia (e-mail: rochdi.bouchiha@gmail.com, kamel.besbes@fsm.mu.tn).

approach for registering images in Section V. Finally, section VI concludes the paper.

## II. AN OVERVIEW OF THE DESCRIPTORS

### A. State of the Art on Local Descriptors

Here, the main tasks are to bridge the gap between image semantics and pixels and to compute a descriptor vector for each interest point such that the descriptor is highly distinctive and partially invariant to the remaining variations such as illumination, rotation, etc. Different approaches have been proposed for describing such points:

**SIFT:** Given a feature point, the SIFT descriptor (Scale Invariant Feature Transform) [9] computes the gradient vector for each pixel in the feature point's neighborhood and builds a histogram of gradient directions. The SIFT descriptor creates a  $16 \times 16$  neighborhood that is partitioned into 16 subregions of  $4 \times 4$  pixels each. For each pixel within a subregion, SIFT adds the pixel's gradient vector to a histogram of gradient directions by quantizing each orientation to one of 8 directions and weighting the contribution of each vector by its magnitude. Each gradient direction is further weighted by a Gaussian function with  $\sigma = n/2$  where  $n$  is the width of the descriptor window. This leads to a SIFT feature vector with 128 elements ( $4 \times 4 \times 8 = 128$ ). This vector is then normalized to enhance invariance to changes in illumination.

**PCA-SIFT:** The PCA-SIFT descriptor (Principal Components Analysis SIFT) [10] is a variety of SIFT with two main differences: (1) the descriptor is calculated for a region of size  $39 \times 39$  sub-regions instead of  $4 \times 4$  used in SIFT and (2) instead of 8 bins for the orientation, PCA-SIFT calculates the orientation in the  $x$  and  $y$  directions. The result is a vector of dimension 3042 ( $39 \times 39 \times 2$ ), which is then reduced to 36 with the principal component analysis.

**GLOH:** the GLOH descriptor (Gradient Location and Orientation Histogram) [11] is an extension of the SIFT descriptor designed to increase the robustness and distinctiveness of the descriptors. Instead of using the Cartesian coordinate, GLOH makes use of the log-polar coordinate system and calculates descriptors in 17 subregions: 8 in angular direction and 3 in radial direction at 3 different radii (6, 11 and 15). For each subregion, 16 gradients are computed, giving a vector of dimension 272 ( $16 \times (8 + 3 \times 3)$ ), this is then reduced to 64 through principal components analysis.

**SURF:** the SURF descriptor (speeded Up Robust Features) [12] is partly inspired by the SIFT descriptor but integral images used in conjunction with filters known as Haar wavelets in order to increase robustness and decrease computation time. Haar wavelets are simple filters which can be used to find gradients in the  $x$  and  $y$  directions. Extraction of the descriptor can be divided into two distinct tasks. The first step is to construct a square window around the interest point. This window contains the pixels which will form entries in the descriptor vector and is of size  $20\sigma$ , where  $\sigma$  refers the scale at which the point was detected. Furthermore the window is oriented along a computed direction such that all subsequent calculations are relative to this direction it is

important that this direction is found to be repeatable under varying conditions. To determine the orientation, Haar wavelet responses of size  $4\sigma$  are calculated for a set of pixels within a radius of  $6\sigma$  of the detected point. The specific set of pixels is determined by sampling those from within the circle using a step size of  $\sigma$ . The responses are weighted with a Gaussian centered at the interest point. In keeping with the rest the Gaussian is dependent on the scale of the point and chosen to have standard deviation equal to  $2.5\sigma$ . Once weighted the responses are represented as points in vector space, with the  $x$ -responses along the abscissa and the  $y$ -responses along the ordinate. The dominant orientation is selected by rotating a circle segment covering an angle of  $\pi/3$  around the origin. At each position, the  $x$  and  $y$ -responses within the segment are summed and used to form a new vector. The longest vector lends its orientation the interest point. The descriptor window is divided into  $4 \times 4$  regular subregions. Within each of these subregions Haar wavelets of size  $2\sigma$  are calculated for 25 regularly distributed sample points. If we refer to the  $x$  and  $y$  wavelet responses by  $dx$  and  $dy$  respectively then for these 25 sample points (i.e. each subregion) we collect the vector  $[\sum dx, \sum dy, \sum |dx|, \sum |dy|]$ . Therefore each subregion contributes four values to the descriptor vector leading to an overall vector of length 64 ( $4 \times 4 \times 4 = 64$ ). The resulting SURF descriptor is invariant to rotation, scale, brightness and, after normalization, contrast.

We will not present the descriptor "Differential Invariants for Color Images" because it is very expensive in terms of computation time than the other descriptors.

### B. The Proposed Enhancement

To increase the performance of the SURF descriptor, we suggest taking advantage of the richness of color in images. The standard implementation of the SURF computes the descriptor using only an eight-bit grayscale image. Our improvements to the SURF descriptor use other color spaces as follows:

**SURF-HSV3:** the computation of the descriptor is made for points in the third plan of the HSV (Hue, Saturation, Value) color space.

**SURF-NTSC2:** is the same as the first one but with a different space: NTSC (National Television System Committee) and exactly with the second plan of this space.

**SURF-HISTEQ:** In this case, there is no transition to another space but a histogram equalization of color is performed before transforming the color image to greyscale.

## III. TEST DATA AND EVALUATION CRITERIA

### A. Test Data

We used three sets of images. Each test set consists of a reference image and more than four sensed images.

Fig. 4. presents in the first line the reference images for the three sets and in the second line some sensed images.

The image database used is a combination of satellite scenes captured over the city of Monastir and cropped to  $512 \times 512$  pixels. These images show several geometric transformations such as changes of scale (scale factor between 1 and 5), the rotation changes where the angles vary

between 25 degrees and 175 degrees with a step of 25 degrees and the photometric transformations such as changing illumination where the equations of the affine transformation are:

$$\begin{cases} R_i = R_{ref} \times a + b \\ G_i = G_{ref} \times a + b \\ B_i = B_{ref} \times a + b \end{cases} \quad (1)$$

where:

$R_{ref}$ ,  $G_{ref}$  and  $B_{ref}$  are the gray level of the Red, Green and Blue of the reference image.

$R_i$ ,  $G_i$  and  $B_i$  are the gray level of the Red, Green and Blue of the sensed image  $i$ .

$a$ : from 0.1 to 1.9. Where, the value 0.1 generates the more contrasted image. And the value 1.9 generates the image brighter. The 1.0 generates the image closest to the reference image, a reason to have a summit at this level in the curves of the next sections.

$b$ : a constant set to 2 in our evaluations.

### B. Evaluation Criteria

The performance of the local descriptor is evaluated using the number of correct matches obtained for each pair of images. The Recall demonstrated in [10] is based on the number of points of interest properly matched and the number of actually existing matches:

$$recall = \frac{\text{number of correct matches}}{\text{number of correspondences}} \quad (2)$$

A pair of points is called correctly matched, only if it respects the homography matrix with a maximum error of  $\epsilon$ .

For the purpose of the tests carried out in this study the  $\epsilon$ -error is set at 3 pixels in the case of scale change, rotation angle change and is set at 0.5 pixels in the case of photometric transformation.

### C. Experimental Results and Discussion

To evaluate the descriptors, we used the points of interest detected by the same detector, and we have disabled the detection process implemented in those descriptors.

In the case of scale changes, Fig. 1 shows the recall plot versus the scale changes for the five descriptors. These curves are decreasing, indicating that the descriptors change with the change in scale. These curves show that the enhanced descriptor SURF-HSV3 was ranked first. They also show a great similarity in the performance of descriptors GLOH, SURF and SURF-HSV3. However, the SIFT and the PCA-SIFT could not resist the large scale changes in the satellite scenes. The results degrade rapidly for scale factor above 2.

In the second case of geometric transformation, Fig. 2. shows the same ranking of the descriptors, but it confirms the clear superiority of the SURF-NTSC2 descriptor.

For the photometric transformation, all algorithms can cope with changing illumination conditions up to certain extend where simply not enough interest points are generated. This breakdown is clearly visible in Fig. 3. Illumination level 0.1 corresponds to the darkest image. Level 1.9 is the clearest image, where hardly any feature was described, except for

SURF-HISTEQ. In this case, the SIFT is ranked second in the Recall.

Note that the recall value when the illumination level is equal to one is computed from two very similar images.

In the speed side, the SURF and the improved version of the SURF have the shorter time of computation, which can be 1/3 and 1/4 of the others as in Table I:

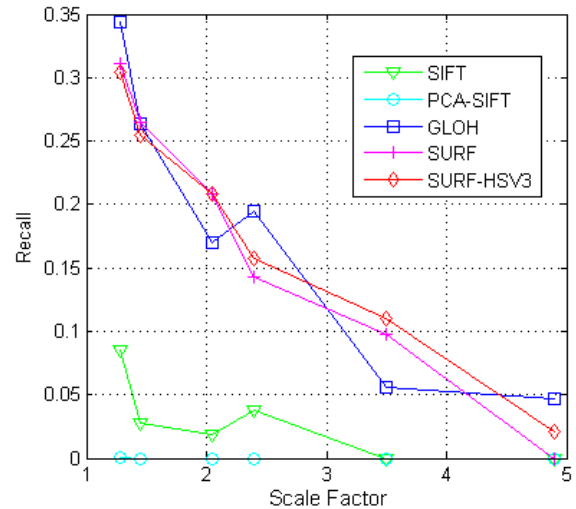


Fig. 1. Recall versus to the scale factor.

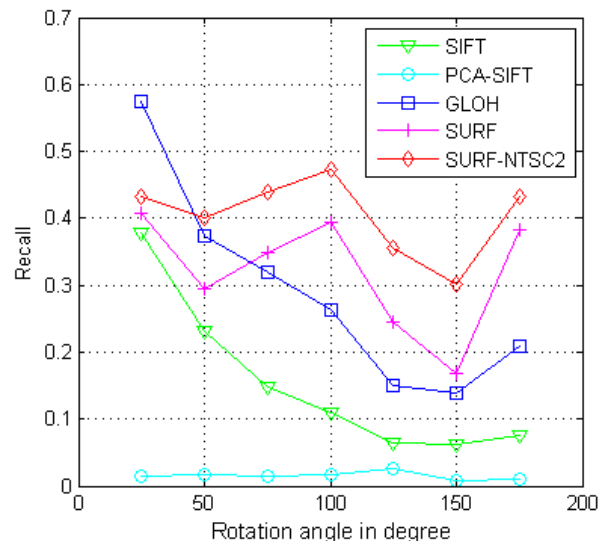


Fig. 2. Recall versus to the angle of the rotation.

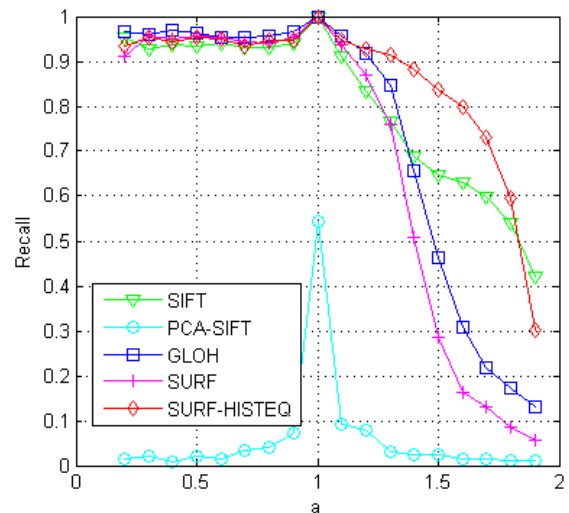


Fig. 3. Recall versus to the coefficient of illumination.

TABLE I: THE MEAN TIME (IN SECONDS) TO COMPUTE THE DESCRIPTORS OF ALL POINTS IN ONE IMAGE.

	SIFT	PCA-SIFT	GLOH	SURF	SURF-X
S	3.84	3.80	5.70	0.85	1.25 (HSV3)
R	2.38	2.33	3.72	1.44	0.79 (NTSC2)
I	3.59	3.17	5.47	0.83	0.78 (HISTEQ)

S: Scale changes, R: Rotation angle changes, I: Illumination

#### IV. PROPOSED APPROACH

We conclude from these results that current state of the art techniques can successfully be used to register remotely sensed images obtained under different conditions. The proposed method is suitable for on-the-fly processing of multi-source images, while addressing the issues of varying illuminations and resolutions of images, local deformations and different perspectives.

The process operates in five stages (see Fig. 5.) Given the sensed and the reference image, the registration is achieved through the two parallel processes, converting the image to the others' color spaces and the detection of the interest points in each image by the MSER (Maximally Stable Extremal Regions) detector proposed in [13], followed by the calculation of the SURF descriptors in each detected point.

In the second step, the three descriptors are applied to all obtained points to compute 3 vectors for each point.

The third step in this approach is the feature matching (see Fig. 6.) For matching of two extracted point data sets no further assumptions about the orientation are made. The matching relies on a descriptor for each point and is performed by calculating the Euclidian distance between the vectors. The shorter the distance is, the more likely a homologous point pair is present. The use of that kind of feature description instead of image patches and image correlation allows a very computational efficient and fast matching.

In this approach, we used the Best-bin-first search method [14] that can identify the nearest neighbors with high probability using only a limited amount of computation. The BBF algorithm uses a modified search ordering for the k-d tree algorithm so that bins in feature space are searched in the order of their closest distance from the query location. This search order requires the use of a heap (data structures) based priority queue for efficient determination of the search order. The best candidate match for each descriptor vector is found by identifying its nearest neighbor in the set of points from the sensed image. The nearest neighbors are defined as the points with minimum Euclidean distance from the given descriptor vector. The probability that a match is correct can be determined by taking the ratio of distance from the closest neighbor to the distance of the second closest.

We rejected all matches in which the distance ratio is greater than 0.71, which eliminates 90% of the false matches while discarding less than 5% of the correct matches.

Then the algorithm calculates the homography matrix based in the correctly matched points. The final process then rectifies the sensed image by superposing it in the reference image.

An improvement can be done in the future by adding a subtask in the feature matching step, consisting in the

calculation of the RMSE and then remove of point pairs causing the increase of the RMSE value. If the total RMSE is still too large, another round of consistency checks is carried out. The iteration continues until a RMSE is less than the threshold (such as 0.5 pixels).

Based in the calculation of the RMSE, the algorithm can decide in the third step which is the best descriptor for the treated case, the SURF-NTSC2 or the SURF-HSV3 or the SURF-HISTEQ descriptor.

It should be noted that the proposed approach is not evaluated in this work; however, we have visibly observed quite good results from this technique. Few examples of the registration obtained by applying our algorithm on images from different sensors are shown in Fig. 7. The evaluation of this technique, for the interested, can be found in [3].



Fig. 4. Scenes used for the evaluation of the performance.

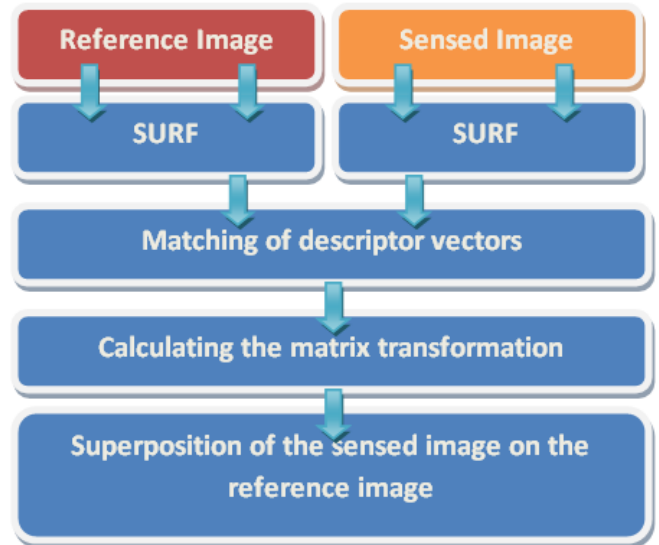


Fig. 5. Proposed workflow for the automatic image registration.

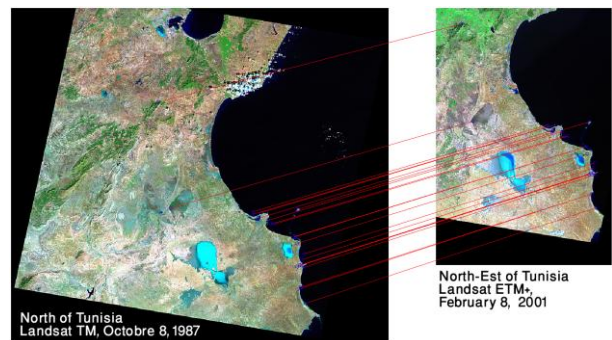


Fig. 6. Automatic matching result of images from two different sensors.

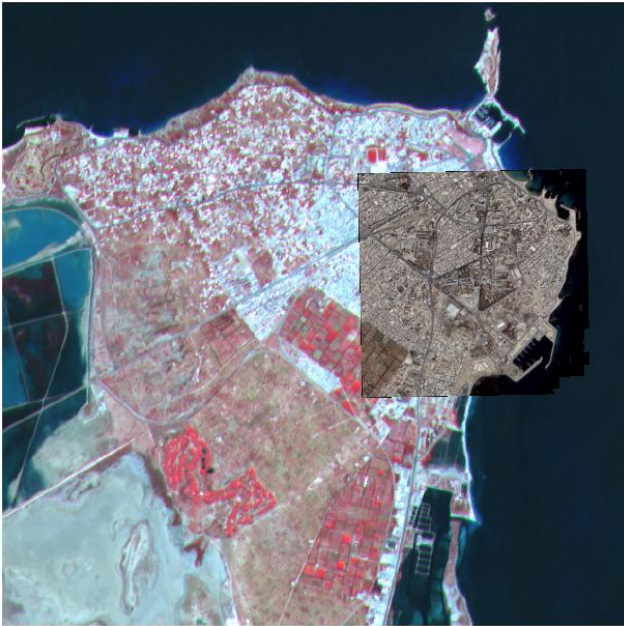


Fig. 7. Registration of images from two different sensors: result of our approach.

## V. CONCLUSION

In this study, we presented the competition that lies between five descriptors, and we used 3 data sets of images to assess the performance of the investigated algorithms. The main focus is on the performance of the techniques when applied to register remote-sensing images. From this experimental comparison, it is concluded that the descriptor SURF-X (Where X: HSV3, NTSC2 or HISTEQ depending on the nature of the scene) outperforms the others' techniques and is robust to rotational changes, scale changes and illumination changes.

Relying on these findings, we proposed the registration algorithm characterized by:

The accuracy: sub-pixel registration,

The robustness: invariance versus different changes (change of scale, rotation, brightness change, resolution change, sensor change: aerial or satellite image),

The speed: shown by the low CPU time calculations and

The automation: no need for the intervention of the supervisor.

The performance of the descriptor degrades rapidly as the scale increases, but with a scale factor lower than 5 the algorithm can register images that have a significant difference in spatial resolution, for example, scenes from Spot and Landsat, from Ikonos and Quickbird, From Eros and GeoEye.

## REFERENCES

- [1] A. Wong and D. A. Clausi, "Arrsi: Automatic registration of remote-sensing images," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 45, no. 5, pp. 1483-1493, 2007.
- [2] R. D. Eastman, J. L. Moigne, and N.S. Netanyahu, "Research issues in image registration for remote sensing," in *Proc. Computer Vision and Pattern Recognition*, vol., no. 17-22 June 2007, pp.1-8.
- [3] Y. Bentoutou, N. Taleb, K. Kpalma, and J. Ronsin, "An automatic image registration for applications in remote sensing," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 43, no. 9, pp. 2127-2137, 2005.

- [4] G. Hong and Y. Zhang, "The Image Registration Technique for High Resolution Remote Sensing Image in Hilly Area," in *Proc. 3rd Int Sym Remote Sensing and Data Fusion Over Urban Areas*, Tempe, AZ, USA, March 14-16, 2005.
- [5] G. J. Wen, J. J. Lv, and W. X. Yu, "A high-performance feature-matching method for image registration by combining spatial and similarity information," *IEEE Trans. on Geoscience and Remote Sensing*, vol. 46, no. 4, pp. 1266-1277, 2008.
- [6] L. Yu, D. Zhang, and E. J. Holden, "A fast and fully automatic registration approach based on point features for multi-source remote-sensing images," *Computers & Geosciences*, vol. 34, no. 7, pp. 838-848, July 2008.
- [7] H. Buiten, "Quality assessment of remote sensing image registration — analysis and testing of control point residuals," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 52, no. 2, pp. 57-73, April 1997.
- [8] P. Radhadevi, S. Solanki, M. Jyothi, V. Nagasubramanian, and G. Varadan, "Automated co-registration of images from multiple bands of liss-4 camera," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 64, no. 1, pp. 17-26, January 2009.
- [9] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proc. the International Conference on Computer Vision*, Corfu, 1999, pp. 1150-1157.
- [10] Y. Ke and R. Sukthankar, "Pca-sift: a more distinctive representation for local image descriptors," in *Proc Computer Vision and Pattern Recognition*, vol. 2, pp. 506-513, 2004.
- [11] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 27, no. 10, pp. 1615-1630, 2005.
- [12] H. Bay, A. Ess, T. Tuytelaars, and L. Vangool, "Speeded-up robust features (surf)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346-359, June 2008.
- [13] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," in *Proc. British Machine Vision Conference*, vol. 1, 2002, pp. 384-393.
- [14] J. S. Beis and D. G. Lowe, "Shape indexing using approximate nearest-neighbour search in high-dimensional spaces," in *Proc. Computer Vision and Pattern Recognition*, Washington, DC, USA, June 1997, pp. 1000-1006.



**Rochdi Bouchiha** was born in Sousse, Tunisia in 1980. He received the engineering and M.S. degrees from the National Engineering School of Sfax, Tunisia, in 2004 and 2007, respectively, both in computer engineering. Currently, he has finished his PhD thesis on "Image processing and data fusion for Geomatic systems" at uEi Laboratory, Faculty of Sciences of Monastir. His current research interests include image processing, data fusion and remote sensing.



**Kamel Besbes** was born in 1960 in Monastir, Tunisia. He received the B.S. degree from the Faculty of Sciences of Monastir (Tunisia) in 1985, the M.S. degree from the Ecole Centrale de Lyon (France) in 1986, the PhD degree from the Institut National des Sciences Appliquées de Lyon (INSA), France, in 1989 and the "State Doctorate Degree" from the Faculty of Sciences of Tunis (Tunisia) in 1995.

In 1989, he joined the Faculty of Sciences of Monastir as an Assistant Professor of Physics and Electronics. He has established teaching and research laboratories in microelectronics since 1990. Research efforts are focused on microelectronics from devices to embedded micro-systems and Instrumentation for detection and navigation for space application.

He has 120, published and presented numerous papers at workshops and conferences and in technical journals. He participated to the scientific and organization committee of several workshops and conferences as International Conference on Microelectronics since 1992 until now and Smart Systems & Devices since 2001. He was the vice-dean of the Faculty since six years (2000-2005). He is elected as the Dean of Sciences Faculty for three years 2008-2011. He is a member of the university council since 2005 until now. He is now a Professor and the head of the Microelectronics and Instrumentation Lab.