Machine Vision and Applications (2009) 20:139–162
DOI 10.1007/s00138-007-0114-y

ORIGINAL PAPER

An efficient image-mosaicing method based on multifeature matching

Ezzeddine Zagrouba · Walid Barhoumi · Slim Amri

Received: 29 June 2006 / Revised: 23 June 2007 / Accepted: 4 July 2007 / Published online: 21 December 2007

Abstract Mosaicing is connecting two or more images and making a new wide area image with no visible seam-lines. Several algorithms have been proposed to construct mosaics from image sequence where the camera motion is more or less complex. Most of these methods are based either on the interest points matching or on theoretical corner models. This paper describes a fully automated image-mosaicing method based on the regions and the Harris points primitives. Indeed, in order to limit the search window of potential homologous points, for each point of interest, regions segmentation and matching steps are being performed. This enables us to improve the reliability and the robustness of the Harris points matching process by estimating the camera motion. The main originality of the proposed system resides in the preliminary manipulation of regions matching, thus making it possible to estimate the rotation, the translation and the scale factor between two successive images of the input sequence. This estimation allows an initial alignment of the images along with the framing of the interest points search window, and therefore reducing considerably the complexity of the interest points matching algorithm. Then, the resolution of a minimization problem, altogether considering the couples of matched-points, permits us to perform the homography. In order to improve the mosaic continuity around junctions, radiometric corrections are applied. The validity of the here-with described method is illustrated by being tested on several sequences of complex and challenging images captured from real-world indoor and outdoor scenes. These simulations proved the validity of the proposed method against camera motions, illumination variations, acquirement conditions, moving objects and image noise. To determine the importance of the regions matching stage in motion estimation, as well as for the framing of the search window associated to a point of interest, we compared the matching points results of this described method with those produced using the zero-mean normalized cross correlation score (without regions matching). We made this comparison in the case of a simple motion (without the presence of a rotation around optical axis and/or a scale factor), in the case of a rotation and in the general case of an homothety. For justifying the effectiveness of this method, we proposed an objective assessment by defining a reconstruction error.

Keywords Regions matching · Harris points · Correlation scores · Mosaic · Reconstruction error

E. Zagrouba · W. Barhoumi (✉) · S. Amri
Unité de Recherche Systèmes Intelligents en Imagerie et Vision Artificielle (URSIIVA),
Institut Supérieur d'Informatique d'El Manar,
2 Rue Abou Rayhane El Bayrouni, 2080 Ariana, Tunisia
e-mail: walid.barhoumi@laposte.net; walid_barhoumi@yahoo.fr

E. Zagrouba
e-mail: ezzeddine.zagrouba@fsm.rnu.tn

S. Amri
e-mail: amri.slim@gmail.com

1 Introduction

Humans are able to synthesize the structure of the surrounding world by making the necessary displacement succession. In reality, this structure is the simple result of image collections that complete each another according to the order defined by the displacement. This allows the production of one image with a large vision field. In order to construct complete panoramic views, the simplest method is based on the alignment of some regular photographs or video sequence that cover all the vision field. Currently, several methods allow the construction of a mosaic from a video sequence. Many special acquirement systems have been used for a

clean improvement of the constructed mosaic quality. A case in point is the manipulation of an embedded-filter camera allowing the acquirement of the scene with different optic regulating [56] and the integration of the global positioning system (GPS) to deduce a preliminary estimation of the motion [21]. These methods are very limited since they impose the manipulation of a sophisticated acquirement material (digital video camera, calibrated stereoscopic bench, etc.) Therefore, given a sequence of images acquired with hand-held camera, the research community is more interested in the development of validated mosaicing techniques which are independent of the acquirement modalities (none a priori knowledge on the camera intrinsic and extrinsic parameters).

A wide screen, usually the case in aerial photographs, can only be proportionally obtained through a set of images. The mosaic is hence commonly used to increase the view field with pasting successively overlapped images onto the same plan [44], onto a cylinder [36,41,51] or onto a sphere for more complicated motions [64]. The mosaicing process allows to synthesize the semantic content of an image sequence on a single image which is easier to understand. In practice, several mosaicing applications are introduced. For example:

- *video compression by alignment of different images of one sequence*: the two described mosaic types are the static mosaics and the dynamic ones. In both cases, the mosaic is built by using a global transformation. In the second case, the motions which were not taken into account for this transformation were analyzed and coded apart [26,28];
- *virtual-environment creation*: many beautiful landscapes are sometimes hidden by obstacles and cannot therefore be captured entirely from a particular view angle. A mosaic often solves the issue into a construction of completely navigatable "virtualized" [29] environments by creating arbitrary views from a certain number of nodes [16,67,68,78];
- assimilation of the eye retina mosaic by encoding it as a quadratic surface in order to simplify the eye surgery [8,9,72];
- *construction of a high resolution panoramic view* [80]. In fact, most mosaicing applications include the construction of high resolution images covering an endless field of view by using inexpensive equipments, and the creation of immersive environments for effective information exchange through the Internet [12,22,30,49].

Thus, image mosaicing seems to be a real challenging research field and there are yet various open problems to be solved. To be qualified as efficient, a mosaicing method should be robust against light change, moving objects and image noise [24]. Moreover, invariance to image rotation and camera zoom are very appreciated properties. The method suggested in this paper tries to address all the issues mentioned above. In the setting of this work we studied the case where the motion model between each two successive images of the studied sequence is a random transformation with unknown parameters. We noted that no a priori knowledge is necessary regarding manipulated images characteristics such as camera calibration parameters, approximation of the acquirement motion, etc.

The next part of this paper falls into four sections. Section 2 is devoted to a synthesis of some relevant previous works on image mosaicing. In Sect. 3, we describe the proposed method based on the regions and Harris points primitives. We present this method in terms of three levels: primitives extraction, primitives matching and mosaic construction. Region matching stage precedes to the Harris points matching in order to estimate the angle of rotation, the subsequent scale factor and thereafter the search window of the potential homologous of each point of interest. This improves the interest points matching reliability as well as proposing a first alignment of the studied images. Section 4 provides an assessment of the proposed method. In fact, the definition of a reconstruction error proves the validity of the suggested method. A summary of the results of this research is presented at the end with the shortcomings of it and some of the perspectives.

2 Related works

The literature of the field describes a growing number of introduced mosaicing processing approaches [77]. Each approach is characterized by the used image set (video sequence or succession of images) and by the information associated with the data (presence or absence of intrinsic and extrinsic parameters of the acquirement system). Mainly, we distinguish two major categories of approaches. The first one does not use a process of 2D primitives extraction (also known as direct methods). It generally treats the case of video sequences [26,27,51,53,54,60]. The second category is based on the extraction and the matching of 2D primitives (also known as feature based methods). It is mainly applied in the general case of overlapped images sets [3,5,15,16,20,37,40,78].

2.1 Mosaicing methods without primitives extraction

For this category of approaches, we can subdivide the used techniques into two classes. For the first class, the mosaic is constructed by minimization of a quadratic error, particularly the sum of squared differences (ssd), applied on the intensity values. This consists of estimating the motion parameters by quadratic error of intensity difference with the overlapping area. For the ssd minimization, the most used optimization technique is the Levenberg-Marquardt technique [27]. The

major drawback of this technique is its sensitivity for local minima [28]. In fact, all the mosaic reconstruction steps are strongly dependent on the results of the motion estimation which is itself based on the ssd minimization [60]. The main advantage of these methods is their accurate registration since the overall data can be fully exploited [24]. Taking into consideration the nature of the error function to be minimized, which takes into account all of pixels, the methods belonging to this first class are not very stable against light change and against the presence of moving objects.

For the second class, the mosaic is constructed by mapping the considered input images onto a cylinder or a sphere to obtain a 360° panorama. This consists in approaching the cylinder, or the sphere, and projecting one-dimensional strips [10,41,51,54,59,64]. In order to project images sequences onto a curved manifold surface, in accordance with the motion of the camera, general techniques were developed [50,79]. In [55], an automatic method of loop closure detection is proposed, thus making it possible to warp the conic mosaic into a cylindrical one. Indeed, the loop closure constraint is generally used to connect the first and the last matched frames. The advantage of these approaches lies in the good quality of the resulting mosaics and the redundancy of information on the studied images. However, the principal disadvantage is depending of the results on the studied sequence and the sophistication level of the acquirement material [51]. We have concluded that for mosaicing methods where no primitive extractions are required, the motion between two successive images has to be small and an initial estimation of the motion needs to be provided. Direct methods proved their efficiency for mosaicing large overlapping regions with small translations and rotations. Because of this limitation, this category is generally restricted to the case of video sequences (seen the small motion between two successive images) and the case of calibrated acquisition systems [17].

2.2 Mosaicing methods with primitives extraction

The second category of mosaicing approaches proposes that both processes of primitives extraction and their matching are possible in the overlapping area. Indeed, instead of using all the available data, these methods try to match primitives only in the overlapping area of the two images to be registered. Most mosaicing approaches based on primitives have usually the unified architecture depicted in Fig. 1. The presence of the calibration stage depends on the manipulated data (sequence of images either calibrated or not). The interaction between the three stages of primitive extraction, primitive matching and homography calculation is a main characteristic of these methods. In the literature, a lot of different features have been used [69], including region [46,47], line [38,65] and point features [57,75]. Most of the existing methods use point

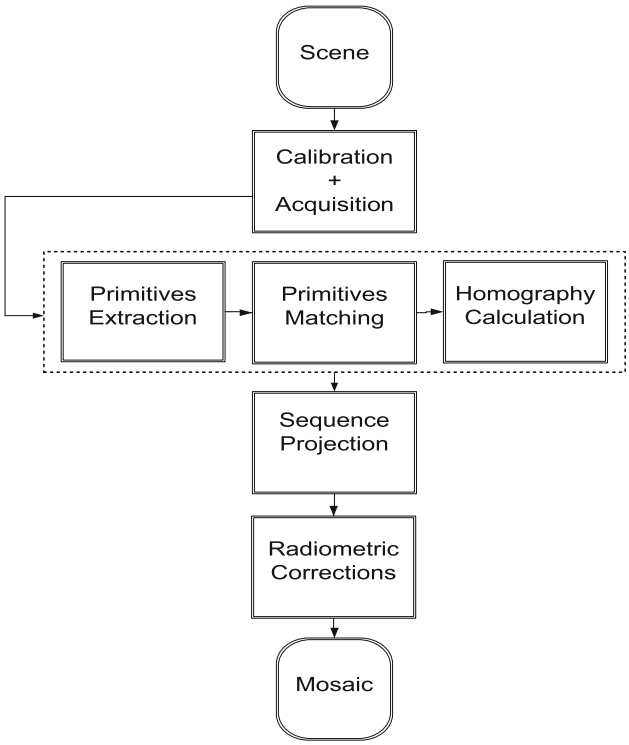

Fig. 1 Architecture of standard mosaicing system based on primitives extraction

features, such as corners, which will be then matched by using a correlation measure in the local area [52]. The mosaicing methods based on primitives extraction have many advantages over the ones without primitives extraction. In fact, they do not require initialization steps, they can handle small overlapping regions and they are very efficient against light change. Moreover, feature based methods are more flexible regarding to image rotation, zoom and moving objects, when features are appropriately used. However, they are computationally intensive [35].

According to the studied motion nature, this category of approaches can be subdivided into three classes:

– a first class corresponds to cases where the camera motion is limited to a translation or a rotation. Most of the existing approaches use only the points of interest [77], particulary the points of Harris [23] (also referred to as the Plessey detector), to construct the mosaic [16,40,44]. Given the difficulty of the interest points matching, this class often requires a priori knowledge (intrinsic and/or extrinsic parameters) or manual matching of the primitives. Each image of the studied sequence presents an overlap space with several other images and thereby a motion can be estimated given the presence of some correlated primitives between each pair of these images. To choose the best homography, the widespread idea is to use the least-squares optimization method in order to compensate

the error with all the images presenting an overlap space with the current image. Once the mosaic is built, another problem can appear if a digital camera, which regulates automatically its diaphragms (presence of automatic gain control system), is used. This problem is illustrated by the presence of large differences of intensities in two adjacent images. The most used solution consists in applying a transformation on the points belonging to the zone of overlap. This transformation changes the intensity value of the considered pixel by the average value weighted by the distances between the studied pixel and the borders of the images containing this point. This radiometric correction permits us to hide the discontinuities of the intensities around the junctions while replacing them by a non-real degradation between the fragments that make the mosaic. This category of approaches is very limited since it requires that the motion between two successive images is too small or that an evaluation of the motion is present as an input parameter for the mosaicing system.

– for the second class, the motion of the camera is a rigid transformation. In other words, the camera can realize more complicated motions such as translation + rotation (without doing a zoom) [2,3,5,15,34]. Concerning the existing methods for this class, the manipulation of the interest points is often combined with the manipulation of 2D structural primitives (structure in n-upletses and regions) in order to simplify the matching stage. These approaches suffer from the 2D primitives' characteristics instability in the case of an important motion, which does not guarantee a big area of overlap.
– the motion of the camera is random (presence of scale factor) for the third class which is the most generalized class and is correspondent to the case of any transformation [78]. This last class is often based on the combination of the interest points, especially the Harris points, and the invariant corners [57], mainly the Deriche–Blaszka (DB) corners [13]. To avoid the problem of interest points matching, all possible homographies are always combined. Unfortunately, it is always possible to fall into local optima because of the corners inaccuracy and the instability of their characteristics relative to the viewpoint modification [62]. In recent years, there has been a growing interest in using invariant features for images matching [2,4,45]. These features are designed to be invariant to translation, rotation and scaling of the images. In particular, given their geometrically invariance under similarity transforms and their partial invariance under affine changes in intensity, the scale invariant feature transform (SIFT) features [33] are considered as a powerful approach for matching procedure. In fact, in their comparative survey, Mikolajczyk and Schmid [45] evaluated a variety of these invariant descriptors and found that SIFT generally outperforms most of the previous approaches specially that it gives improved matching for images at different scales. SIFT features are located at scale-space maxima/minima of a difference of Gaussian function. At each feature location, a characteristic scale and orientation are established. This provides a similarity-invariant frame in which to make measurements. Illumination invariance is achieved by using gradients and normalized descriptor vectors [5]. However, each image can generate about one thousand SIFT keys which increases the processing time of the mosaicing system. In our case, the improved version of the Harris detector [58] is the most suitable since we dispose a preliminary estimation of the scale factor produced by the regions matching step. Moreover, in order to highlight the invariance to different geometric and photometric transformations, many different techniques for describing regions, which are covariant with a class of transformations, have been recently developed [19]. In [47], authors presented a snapshot of the state of the art in affine covariant region detectors used essentially for image registration. They compared the performance of six types of detectors on a set of real test images under varying imaging conditions. The realized experiments allowed to deduce, in terms of repeatability and accuracy, the slight superiority of the "maximally stable extremal regions" detector (MSER) [39] with regards to the other studied detectors. This detector is based on the watershed algorithm to find intensity regions, called extremal regions, which have either higher or lower intensity than the pixels on its boundary. The set of these extremal regions is stable under monotonic change of image intensities [47], which ensures that common photometric changes leave this set unaffected. The extremal regions are also unaffected under affine geometric transformations since extremal regions are composed of connected topology-pixels which are preserved against geometric transformations [39]. To the best knowledge of the authors, covariant regions detectors have not been widely utilized in image mosaicing yet [61]. This is maybe because these methods are relatively new [25,43,70,71].

2.3 Discussion

As indicated by the recent history of newly developed applications, image mosaicing has become a major field of research. However, the existing mosaicing approaches differ according to the type of the studied sequences (video sequences or series of overlapping images) and the type of the acquisition device motion (simple or complex). We concluded that for the conception of a mosaicing approach treating the case of a sequence of overlapping images, it is necessary to specify the type of the images to treat (acquisition information). Then, the adopted strategy should be specified, as well as the

used primitives, while specifying the use-goal of each one. As a conclusion, we can mention that merging automatically collections of images having a partially overlapping content where the motion effected by the acquisition equipment is unspecified and important, is still a problem which was not entirely solved. In fact, to build mosaic, it is necessary to find the model of the motion to be applied on each image of the input sequence. However, the motion model estimate requires the matching of particular points in the space of overlap. The majority of the existing approaches use the points of interest due to the precision of their detection. Nevertheless, their matching is the most critical task because it needs huge computational cost. Indeed, the only radiometric attribute associated to a point is its intensity which is not discriminatory and can vary a lot according to the luminance variation. The only geometric attribute associated to a point is its position which moves randomly when the motion is undefined and when a priori knowledge is not available. Some matching techniques consider the neighborhood of an interest point in order to find its potential homologous. Unfortunately, the computed correlation scores are not steady for light change and for any motion, particularly for the rotation around the optical axis or the scale factor change. On the other hand, regions are the 2D structures with the richest semantic attributes and the most stable under any transformation. However, the methods of regions extraction are not sufficiently precise on the borders. Thus, our idea is to exploit the duality between the regions and the points of interest in order to benefit from the complementarity of these two primitives advantages. In fact, we used a rotation and scale partially invariant matching procedure based on regions in order to provide a first estimation of the motion and therefore to define a reduced search window of homologous coordinates, for each point of interest. This has avoided blind search of correspondence between the points of interest in two images. Simulations show that the proposed method generates mosaics with accuracy without being computationally intensive.

3 Overview of the proposed method

The construction of a mosaic from an ordered image sequence $(I_1,\ldots,I_i,\ldots,I_n)$ can be reduced to the problem of two successive images alignment. Therefore, as we suppose that the sequence is sorted out, the problem consists in the estimation of the homography describing the motion between two successive images I_i and I_{i+1}. The proposed method of automatic mosaicing is based on three main levels: the primitives extraction, the primitives matching and the mosaic construction (Fig. 2).

The primitives extraction level involves two stages. Initially, input images are segmented into regions using a region growing process followed by a no-meaningful regions elimination step. Then, the second stage is dedicated to the extraction of Harris points which coincide with abrupt intensities [57]. The second level is made up of three sub-stages. First, a prediction/validation matching process is applied to the obtained two regions maps, relative to the couple of images (I_i, I_{i+1}). The prediction is made by measuring correlation scores, and the validation is completed while verifying the relative position constraint. Then, from the analysis of the matched-regions relative positions, we can partially estimate the angle of rotation, the scale factor and the search window (position and dimension) of the potential homologous relatively to each interest point. Given these estimations, the matching of the interest points is based on computing zero-mean normalized cross correlation scores (zncc) [23] between the potential homologous points followed by the verification of the uniqueness constraint. Besides, the RANSAC consensus [18] was applied in order to reject outliers matches. In fact, given a point of interest, the framing of the search window of the potential homologous points increases the matching reliability while reducing the combinatory complexity. The last level of the proposed method looks for the optimal homography H minimizing the Euclidean error. This minimization problem has been solved by the QR-factorization technique followed by a relaxation algorithm. The obtained homography is applied to all of the image I_{i+1} points in order to express them in reference to the image I_i. To improve the visual quality of the mosaic, two processes of intensities interpolation and radiometric correction are successively applied. However, there is still some visible lines along the boundary of the overlapping area despite the care brought to the previous stages. For this, we proceed with a blending strategy by taking into account the history of pixels values. Blending consists in assigning specified colors to the overlapping regions. Blending is also important when there are moving objects in the sequence in focus [11]. Ultimately, the mosaic is built while projecting all the images of the studied sequence on the plan of one image chosen as a reference.

4 Primitives extraction

This level is composed of two stages which can be performed simultaneously. The first stage consists in the segmentation into regions of each couple (I_i, I_{i+1}) of successive images belonging to the studied sequence. We chose to use the regions primitives given the wealth of information offered by them which will consequently permit to estimate correctly the motions (rotation and scale factor) and to limit, thereafter, the search space of homologous coordinates of each interest point. To achieve the segmentation into regions, we used the regions growing technique [42]. Indeed, as the segmentation goal in our case is only a first estimation of the motion, we opted for the region growing technique due to its rapidity and

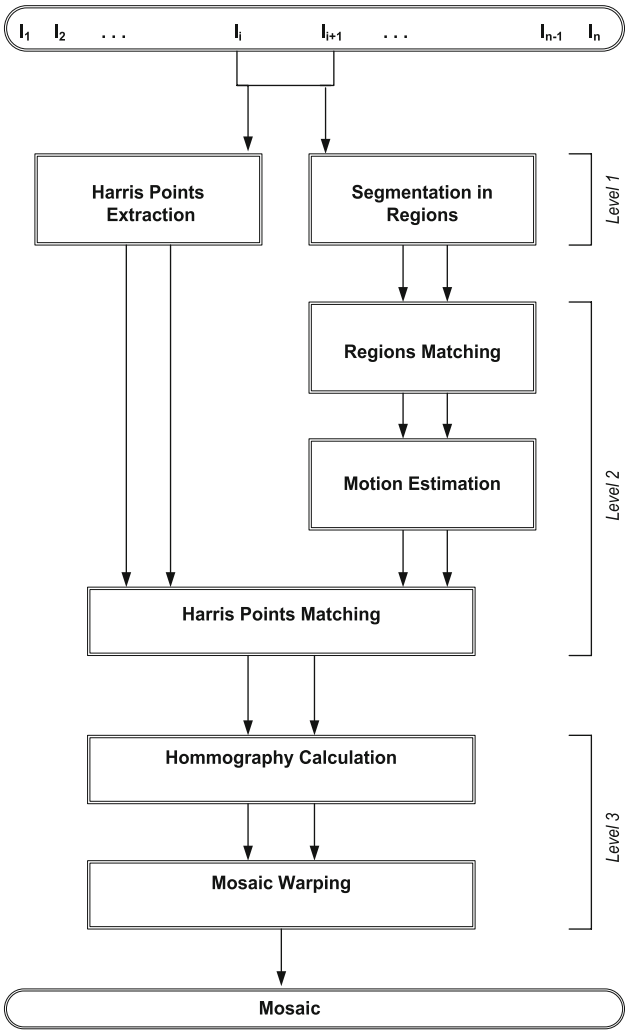

Fig. 2 The proposed image mosaicing method

its efficiency for various images types. This technique is one of the conceptually simplest approaches to image segmentation. Neighboring pixels of similar amplitude are grouped together to form a segmented region. However, in practice, constraints, some of which are reasonably complex, must be placed on the growth pattern to achieve acceptable results. In our case, we used an unseeded technique where the homogeneity criterion measures the similarity between the pixel intensity and the intensity mean of the current region. This unsupervised segmentation process is preceded by a pre-treatment step permitting to improve the images quality while preserving regions' contours. In fact, we apply the filter of Nagao [48] which is a well adapted pre-treatment technique for regions segmentation by pixels aggregation [73]. It permits the minimization of the noise effects on the final built mosaic. The inherent over-segmentation effects have been treated by a post-treatment. It consists in the merging of the small no meaningful regions to the including regions or to the regions maximizing the length of the common border with these small regions (Fig. 3).

The second stage of this level is reserved for the interest points extraction. Following a bibliographic survey on the interest points detectors [57,77], we chose to apply the improved version of the Harris detector (ImpHarris) [23,58]. This detector is more steady and more robust than many other detectors [76] (Fig. 4). In particular, in [57] the authors proved that the improved Harris detector outperforms considerably other detectors in terms of two novel relevant criteria for the evaluation of points interest detectors: repeatability and information criteria. Repeatability compares interest points detected on images taken under varying conditions (camera parameters, its position relative to the scene, illumination conditions, etc.) However, information content is a measure of the distinctiveness of an interest point, which is based on the likelihood of a local gray value descriptor at the point within the population of all the observed interest point descriptors. The Harris detector [23] is based on a matrix M related to the auto-correlation function. This matrix averages the derivatives of the signal in a window W around a point (i, j) while using a Gaussian in order to weight the derivatives inside this window (1). Then, interest points, particulary the corners, are detected if the matrix M has two significant eigenvalues. In the improved version [58], derivatives are computed more precisely by replacing the $[-2 \ -1 \ 0 \ 1 \ 2]$ mask with derivatives of a Gaussian ($\sigma_w = 1$). In order to guarantee a fast detection, Deriche's recursive implementation of the Gaussian filter was used [14]. This allows us to extract these points faster than many other detectors.

$$M_{ij} = \mathrm{e}^{\frac{-x^2-y^2}{2\sigma_w^2}} \otimes \begin{pmatrix} I_x(i,j)^2 & I_x(i,j)I_y(i,j) \\ I_x(i,j)I_y(i,j) & I_y(i,j)^2 \end{pmatrix}. \quad (1)$$

5 Primitives matching

This level starts with regions matching in order to produce a first estimation of the rotation around optical axis and of the scale factor. This estimation makes a preliminary alignment of successive images (I_i and I_{i+1}) possible. To match each point of interest, the estimated homography relating the two images permits the framing of the search area, which reduces considerably the combinatory complexity of the interest points matching.

5.1 Regions matching

After the regions segmentation of the images I_i and I_{i+1}, we proceed to the regions matching in order to produce a first approximation of the transformation between these two images while analyzing the relative positions of the potential-matched regions centers of gravity. The camera motion is in

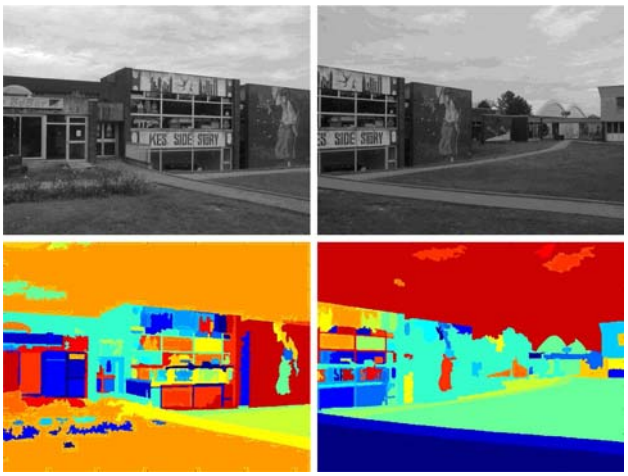

Fig. 3 Segmentation in regions. At the *top*: two successive images of the input sequence (the same input images are used for Figs. 4, 5 and 7). At the *bottom*: correspondent region maps produced by region growing

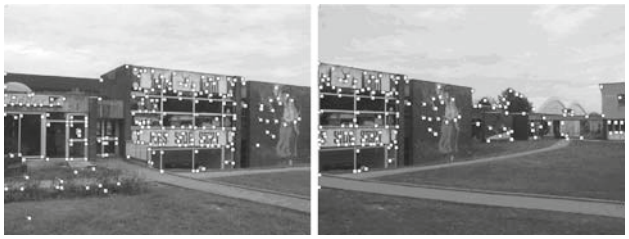

Fig. 4 Interest points extraction with ImpHarris detector

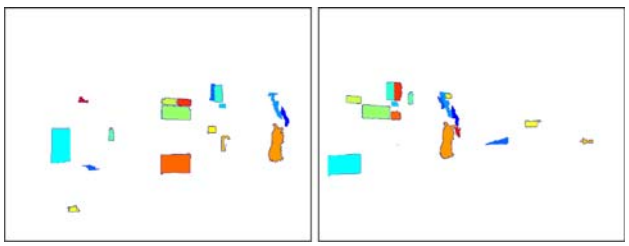

Fig. 5 Regions matching (the set Λ of couples performing high correlation scores)

general important. Therefore, the regions having a part of their edges on one border of the image do not have, in general, correspondents in the next acquired image. Indeed, these regions perform often false matching which must be eliminated. Hence, we consider for matching only the regions which do not "touch" the image borders (Fig. 8). Then, a process of matching by prediction/validation puts in correspondence the two gotten region maps produced by the segmentation step. The prediction is made while measuring, for every pair of regions (R, R') $(\in I_i \times I_{i+1})$, a correlation score Cor calculated according to the Eq. (2). Thus, only the set Λ $(\subset I_i \times I_{i+1})$ of couples performing high correlation scores (Cor $\simeq 1$) is kept (Fig. 5).

$$\mathrm{Cor}(R, R') = \frac{1}{|\Omega|} \cdot \sum_{A \in \Omega} \frac{\mathrm{Min}(A(R), A(R'))}{\mathrm{Max}(A(R), A(R'))}, \tag{2}$$

where, $|\,|$ is the set cardinality operator and Ω is the set of the following normalized geometric and radiometric attributes used for region characterization [74]: compactness, occupation rate, second invariant moments, intensity mean. As in our case we suppose that the scene is sufficiently far from the used acquirement device (hypothesis of scene planarity), the matching procedure is partially invariant with respect to translation (Fig. 7), rotation (Fig. 8) and scale factor (Fig. 9).

In order to be validated, these couples must thereafter verify the constraint of the relative position [1]. We start by focusing on four couple of regions $(\subset \Lambda)$ verifying a spatial likeness between quadrilaterals formed by their centers of gravity (Fig. 6). This returns to the checking of the equations system (3) and of the orientation similarity between the two angles $\alpha_i = (\overrightarrow{G1_iG2_i}, \overrightarrow{G1_iG4_i})$ and $\alpha_{i+1} = (\overrightarrow{G1_{i+1}G2_{i+1}}, \overrightarrow{G1_{i+1}G4_{i+1}})$, where Gn_k is the gravity center of the region Rn belonging to the image I_k.

$$\forall\, (j,k) \in \{1,\ldots,6\}^2, \quad \frac{d_j^i}{d_j^{i+1}} \approx \frac{d_k^i}{d_k^{i+1}}. \tag{3}$$

Then, we browse the remaining couples belonging to Λ in order to define the elements verifying the constraint of the relative position according to the already validated regions (Figs. 7, 8, 9). The set Λ_f of these couples will form the basis of the initial estimation of the motion between the two images I_i and I_{i+1}. Note that we do not target to match all the regions. But, we look only at match some regions correctly. The proposed mosaicing method fails only if it could not match at least four couples of regions correctly. While testing our method on various simulations, we had not seen this case even for small overlapping areas. Theoretically, if it was the case we will focus our interest on dealing only with the points of interest matching. Indeed, the matching is immediate for the residuals points in the small overlapping area since the number of these points is extremely reduced and there is almost no ambiguity about the matching of each interest point.

5.2 Transformation estimation

In order to carry out the preliminary homography estimation, the set Λ_f of matched couples is now used to estimate the motion between I_i and I_{i+1}, (I_{i+1} being the current image). To align the image I_{i+1} with the image I_i, we start by a first estimation of the rotation angle and of the scale factor between the two considered images (I_i and I_{i+1}). We consider, among the set Λ_f, the two couples of regions (R_i, R'_i) and (R_{i+1}, R'_{i+1}) presenting the best correlation scores Cor.

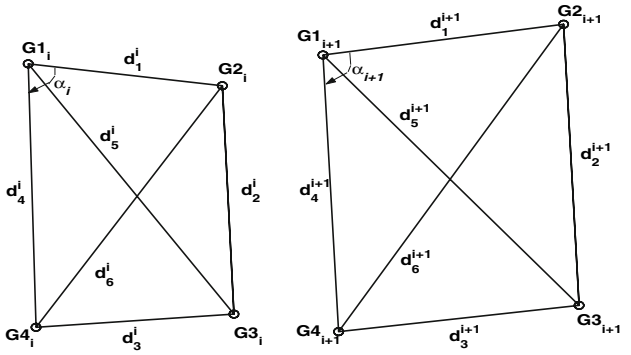

Fig. 6 Constraint of the relative position

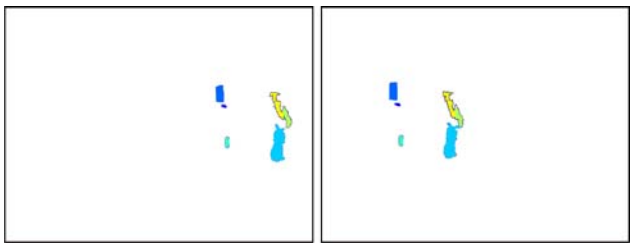

Fig. 7 The set Λ_f of matched regions verifying the relative position constraint

In fact, the angle α between the two vectors $\overrightarrow{G_i G'_i}$ and $\overrightarrow{G_{i+1} G'_{i+1}}$ deals with a first estimation of the rotation angle (Fig. 10). Besides, the ratio f of the distances between their centers of gravity (G_i, G'_i) and (G_{i+1}, G'_{i+1}) can be considered as a first estimation of the scale factor between the two images. We applied the combination $\Im$ of the rotation by the angle α with the scale factor f on each pixel (x, y) of the image I_{i+1} according to the Eq. (4). However, taking into consideration the discreet nature of digital images, the application of the combination $\Im$ can generate some points without antecedents. This causes the apparition of some points without corespondent intensities. To resolve this problem, we applied, to each of these points, a bi-cubic intensity interpolation while considering among its neighbors those already having antecedents. This new image, noted RI_{i+1}, will be used instead of the image I_{i+1} for the interest points matching. The composition $\Im$ is also applied successively on the gravity centers coordinates of the matched regions and on the interest points detected in the first level, in order to cancel the rotation and the zoom effects and then to optimize the reduction of the search window of the homologous interest points and thereafter the interest points matching results.

$$(x\ y\ 1).\begin{pmatrix} f.\cos(\alpha) & -f.\sin(\alpha) & 0 \\ f.\sin(\alpha) & f.\cos(\alpha) & 0 \\ 0 & 0 & 1 \end{pmatrix}. \tag{4}$$

We seek a framing of the search space of the possible homologous according to each interest point that we detected

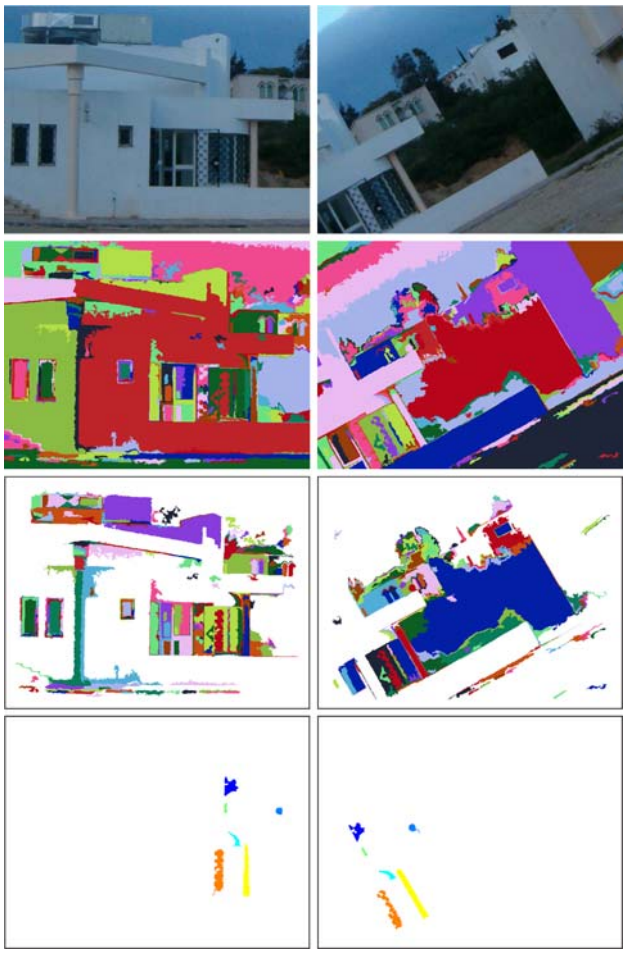

Fig. 8 Application of a rotation. At the *top*: two successive images with a rotation of 45°. Second row: correspondent region maps. Third row: correspondent region maps after remove of regions touching the image borders. At the *bottom*: the set Λ_f of matched couples

in the first level. When the motion around the optical axis of the camera is significant, the interest points matching by using one of the traditional scores of correlation is impossible. Even for a reduced rotation angle, the interest points matching remains a very difficult task because of the blind search of the interest points homologous in all the image. However, given the analysis of the relative positions of the gravity centers $G_j(X_j, Y_j)$ and $G'_k(X'_k, Y'_k)$, corresponding to each couple of homologous regions (R_j, R'_k), we can estimate the dimensions L_x and L_y of the search window of the homologous interest points. We compute for each couple of homologous regions $(\in \Lambda_f)$ a disparity vector $D_{jk} = (d_x^{jk}, d_y^{jk})$ according to the following Eq. (5):

$$d_x^{jk} = X_j - X'_k \quad \text{and} \quad d_y^{jk} = Y_j - Y'_k. \tag{5}$$

The value of L_x (resp. L_y), horizontal (resp. vertical) dimension of the search window of the homologous interest points belonging to the image I_i, corresponds to the length of

Fig. 9 Application of a scale factor to a sequence containing a raster of identical windows. At the *top*: two successive images where the value of the scale factor approximates 2. In the *middle*: correspondent region maps after remove of regions touching the image borders. At the *bottom*: the set Λ_f of matched couples

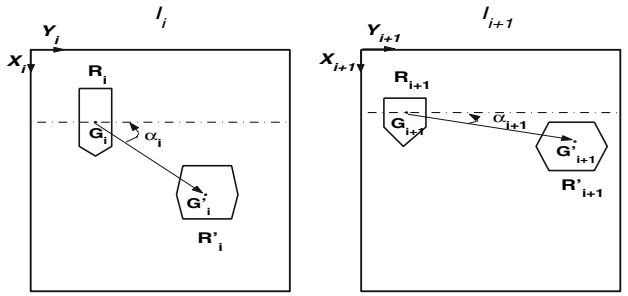

Fig. 10 Estimation of the rotation

the X-disparity interval (resp. Y-disparity) (Fig. 11). These values are estimated according to the following Eq. (6):

$$\begin{cases} L_x = |L_x^1 - L_x^2|, \text{ with,} \\ L_x^1 = \max_{(R_j,R_k)\in\Lambda_f} d_x^{jk}, \\ L_x^2 = \min_{(R_j,R_k)\in\Lambda_f} d_x^{jk}. \\ \\ L_y = |L_y^1 - L_y^2|, \text{ with,} \\ L_y^1 = \max_{(R_j,R_k)\in\Lambda_f} d_y^{jk}, \\ L_y^2 = \min_{(R_j,R_k)\in\Lambda_f} d_y^{jk}. \end{cases} \tag{6}$$

For a large set of test sequences composed of 640 × 480 images, a reduction of the search window to the mean size of 5 × 5 was achieved.

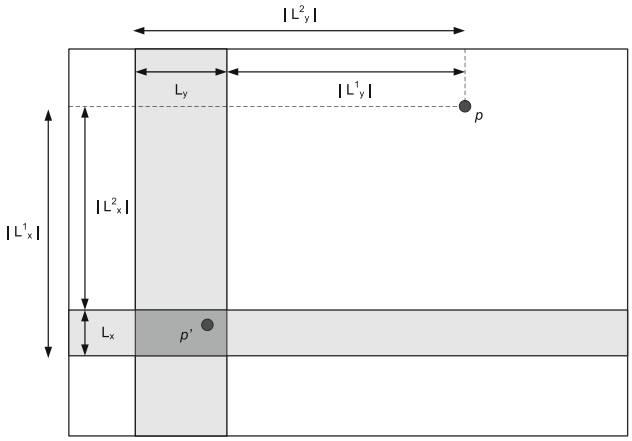

Fig. 11 Homologous search window

We evaluated the performance of the proposed preliminary estimation of the motion under the following four changes in imaging conditions: illumination, image blur, scale (+ in-planar rotation) changes and viewpoint changes. This was applied to the same reference set of test images used in [47]. Each of the test sequences contains six images with a defined gradual geometric or photometric transformation. These images are either of planar scenes or the camera position is fixed during acquisition, so that in all cases the images are related by homographies. This means that the mapping relating images can be correctly defined which permits the determination of known ground truth matches. The ground truth in all cases is provided by mapping the regions detected on the images to a reference image using the computed homographies. More details about the used test sets can be found in [47] (Note: all the images as well as the computed homographies are available on http://www.robots.ox.ac.uk/~km). Two examples from each sequence of the used test set are shown in Fig. 12.

The aim of this evaluation is to measure the repeatability of the proposed motion estimator, and implicitly regions segmentation and matching, under different geometric and photometric transformations. The basic measure of repeatability $\Re$ is based on the relative amount of overlap between the homologous regions used for the motion estimation (Fig. 13). This gives a good indication of the chance that the regions can be matched correctly. For each estimation, we considered the set of matched-couples Λ_f used for the definition of the search window (and implicitly the set Λ_{me} used for the motion estimation). We applied the defined homography H_{ref} on the set Λ_f (and implicitly Λ_{me}) of the regions belonging to I_{ref}. Two matchable regions, R_r ($\in I_{\text{ref}}$) and R_t ($\in I_{\text{test}}$), are deemed to correspond if their similarity degree S (7) is sufficiently elevated.

$$S(R_r, R_t) = \frac{1}{2}\cdot\left(\frac{|R_t \bigcap \Psi(R_r)|}{|R_r|} + \frac{|R_t \bigcap \Psi(R_r)|}{|R_t|}\right), \tag{7}$$

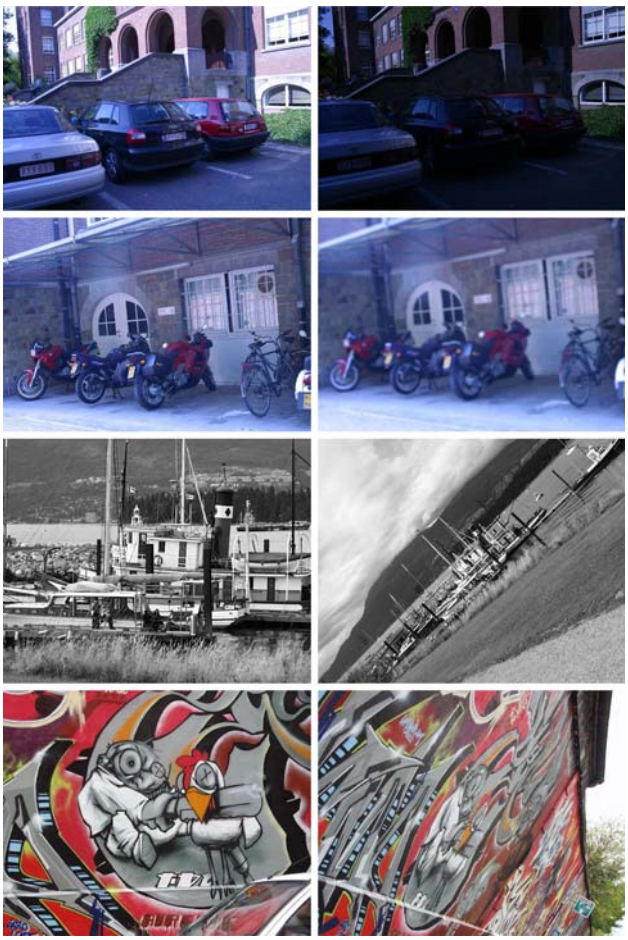

Fig. 12 Outline of the used data set. At the *top*: light change. Second row: image blur. Third row: zoom (+ in-planar rotation). At the *bottom*: viewpoint change. (Note: all images are of medium resolution, approximately 800 × 640 pixels)

where, H_{ref} is the given homography relating the two images I_{ref} and I_{test} and $\Psi(R_{\text{ref}}) = H_{\text{ref}}^{t}.R_{\text{Ref}}.H_{\text{ref}}$ is the projection operator.

Given the set Λ_f (resp. Λ_{me}) of the used matched-couples for the search window (resp. for the preliminary motion estimation), the repeatability $\Re$ is defined according to the Eq. (8). It reflects, in absolute terms, the average of similarity degrees of matchable regions detected in images under gradual increase of different geometric and photometric transformations (Fig. 14). In general we would an estimator to have a high repeatability score (ideal plot for repeatability would be a horizontal line at 100%).

$$\Re = \frac{1}{|\Lambda_{f/me}|} \sum_{(R_r, R_t) \in \Lambda_{f/me}} S(R_r, R_t), \tag{8}$$

where, $\Lambda_{f/me}$ is either the set Λ_f or the set Λ_{me}.

The results of the realized tests are illustrated in Fig. 14. We note that the best performance achieved for the motion estimation based on Λ_{me} (resp. search window framing based on Λ_f) is 96.9% (resp. 92.7%) for light change. For small transformations, the highest obtained repeatability scores indicate how well the used estimator performs for this case. The repeatability scores recorded with Λ_{me} outperforms the ones recorded with Λ_f (except one case that we explain

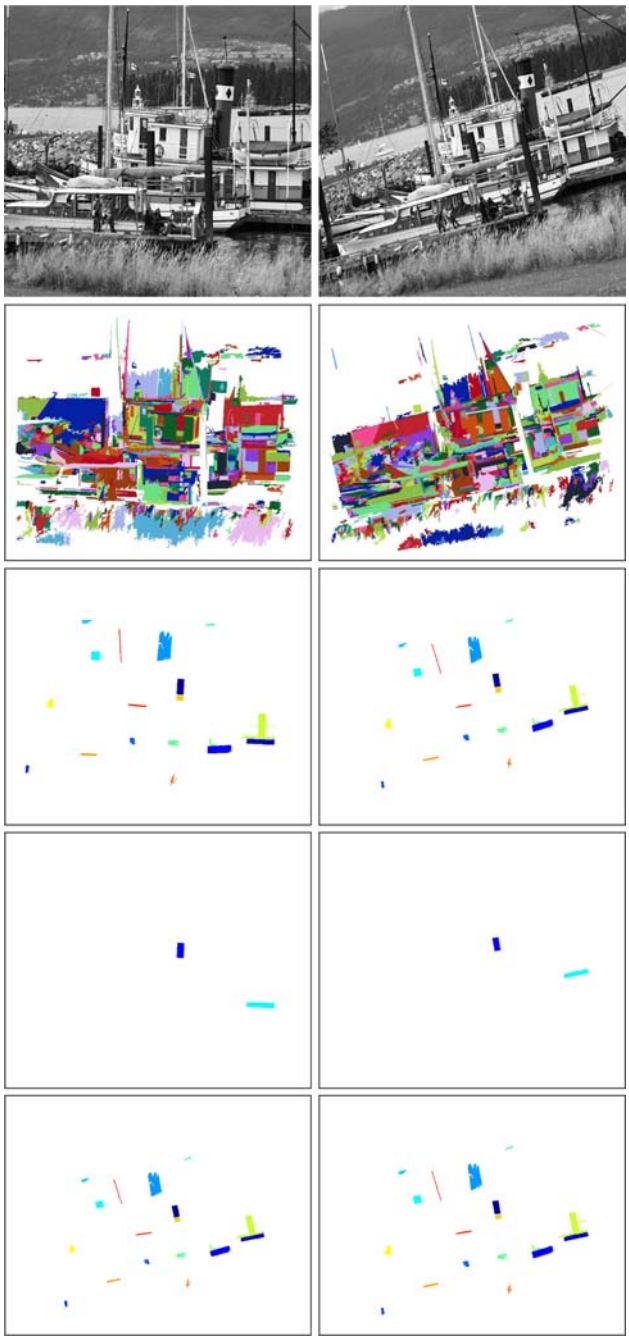

Fig. 13 Evaluation of the proposed motion estimator. At the *top*: original images (*left*: I_{ref}, *right*: I_{test}). Second row: correspondent region maps after remove of regions touching the image borders. Third row: the set Λ_f of matched couples. Fourth row: the set Λ_{me} of the used matched couples for preliminary motion estimation. At the *bottom*: the set Λ_f after applying the given homography

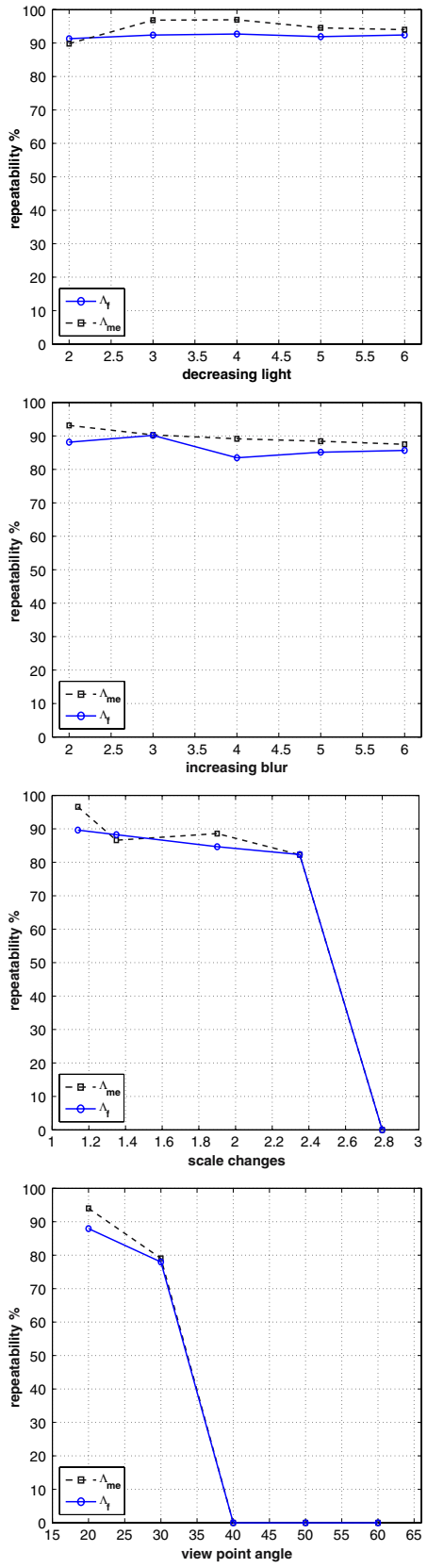

Fig. 14 Evaluation of the proposed motion estimator. At the *top*: viewpoint change. Second row: scale change (+ in-planar rotation). Third row: image blur. At the *bottom*: light change

below). This can be accounted for by the fact that Λ_{me} is composed of the two better matched couples among those belonging to Λ_f. The results for image blur and light change are better than for viewpoint and scale changes, showing a high level of invariance to image blur and light change. In what follows we will discuss only the recorded results with Λ_f ($\Lambda_{me} \subset \Lambda_f$). The repeatability score for a light change (resp. image blur) varies between 91.3 and 92.7% (resp. 83.4 and 90.1%). Indeed, for image blur and light change all curves are nearly horizontal, showing good robustness to these changes. However, repeatability varies between 82.3 and 89.6% below scale change of 2.4° and decreases for large viewpoint changes. For a scale change of 1.35, the repeatability score recorded with Λ_f is slightly higher than the one recorded with Λ_{me}. This can be explained by the lack of stable edges on which the region extraction is based. In other words, the presence of tiny defects of the segmentation around the regions borders is reflected by the fact that the two best correlated regions (Λ_{me}) can not record the best similarity degree (S) among the strongly correlated regions (Λ_f). In fact, a small irregularity on the borders modify appreciably the shape descriptors (from where the correlation measurement) without too much deterioration of the overlapping rate. For viewpoint changes under 40°, the repeatability score is going from 77.9 down to 87.9% and decreases considerably for larger viewpoint changes. As a conclusion, we can mention that viewpoint changes are the most difficult type of transformation to cope with, followed by scale (+ in-planar rotation) changes. Furthermore, the absolute score shows that beyond a certain value, a small variation falls directly the repeatability of the proposed estimator. This is because the region boundaries become smooth, and the segmentation process is less accurate. Concerning the null repeatability scores recorded with the high viewpoint and scale changes, this is due to the nonexistence of matchable regions for these cases. Indeed, the used descriptors are not projective invariant measures and therefore a significant change of the viewpoint and/or scale factors (outside the range for which the estimator is designed) limits the proposed approach. This can be understood by the fact that in most cases larger transformations result in lower quality images and/or smaller commonly visible parts between I_{ref} and I_{test} [47]. In this case, regions are considered as inefficient candidates for the preliminary homography estimation and the proposed system uses only the interest points to estimate the final homography. Thus, interest points matching can be achieved without significant ambiguity in the small overlapping area since the number of these points is extremely reduced in this area.

5.3 Interest points matching

In order to match efficiently the interest points, we start with the generation of the matching hypotheses while calculating

the zero-mean normalized cross correlation score zncc (9) [23]. In practice, this measure provides, in a reduced calculation time, one of the best matching performance and it is invariant according to the local linear changes of intensity and therefore light changes do not affect the track [31,32].

$$\begin{cases} \text{zncc}(x, y, x', y') = \dfrac{\sum_{l,k} Q^1_{l,k}.Q^2_{l,k}}{\sum_{l,k} Q^3_{l,k}.\sum_{l,k} Q^4_{l,k}}, \\ \text{with,} \\ Q^1_{l,k} = (I(x+l, y+k) - \overline{I(x, y)}), \\ Q^2_{l,k} = (I'(x'+l, y'+k) - \overline{I'(x', y')}), \\ Q^3_{l,k} = \sqrt{(I(x+l, y+k) - \overline{I(x, y)}^2)}, \\ Q^4_{l,k} = \sqrt{(I'(x+l, y+k) - \overline{I'(x, y)}^2)}, \end{cases} \tag{9}$$

where, $\overline{I(x, y)}$ and $\overline{I'(x', y')}$ are the intensities means of the two windows centered respectively on (x, y) $(\in I_i)$ and (x', y') $(\in RI_{i+1})$.

After the search window estimation on the image RI_{i+1} for each interest point belonging to the image I_i, we can conclude that all the points of the image I_i which generate windows of search placed outside of the image RI_{i+1} are points of interest without homologous. For all the remaining interest points (which can have homologous), we calculate the correlation score zncc between each of these points and those which are inside the corresponding search area in the image RI_{i+1}. Then, only the hypotheses of matching having a correlation score zncc greater than a threshold γ (≈ 1) and which do not rape the uniqueness matching constraint, are kept.

A point which is matched with two other points localized in a small zone can construct an error source. This case is not so frequent since the search window dimensions are in general sufficiently small. In order to increase the interest points detection accuracy, we keep the site of those detected on the image I_i and we execute an alignment process for the corresponding points of interest in the image I_{i+1}. An iterative process considers each couple of matched points $c_k = (P_k, P'_k)$ $(\in I_i \times I_{i+1})$ and if one of the 3×3 neighbors of P'_k achieves with P_k a better correlation score, then P'_k is replaced by this neighbor (Fig. 15).

Given that there is a non-zero probability that some matched points are outliers, we refine the kept matches using an outlier rejection procedure based on RANSAC (RANdom SAmpling Consensus) [18], which has been by far the most widely adopted method for the treatment of external sources of error. RANSAC is a convenient tool to aim refinement since it considers outliers detection prior to the homography estimation. The application of RANSAC ameliorates the homography estimate robustness, by keeping only the set of feature matches which are loosely consistent (Fig. 16). Since

Fig. 15 Extraction of the homologous points of interest

RANSAC failed when the fraction of outliers is too great [4], we opted for its use as the last step of the interest points matching procedure [4].

6 Mosaic construction

Given the couples of the matched points produced by the second level, the first stage of the mosaic construction level consists on the homography computation. Then, two processes of intensities interpolation and radiometric correction are applied in order to improve the achieved mosaic quality. At the end, given the produced results for the set of successive image couples forming the studied sequence, the general mosaic is warped.

6.1 Homography computation

In the cases of a planar scene images acquired from different viewpoints (general camera motion) or an image collection of 3D scene acquired from the same viewpoint (motion of rotation), the transformation between two images I_i and I_{i+1} is a linear transformation in homogenous coordinates according to the following Eq. (10):

$$\begin{cases} [x^{i+1}_k,\ y^{i+1}_k,\ z^{i+1}_k]^t = T.[x^i_k,\ y^i_k,\ z^i_k]^t, \\ \text{with,} \quad T = \begin{bmatrix} H_{1,1} & H_{1,2} & H_{1,3} \\ H_{2,1} & H_{2,2} & H_{2,3} \\ H_{3,1} & H_{3,2} & H_{3,3} \end{bmatrix}. \end{cases} \tag{10}$$

(x^i_k, y^i_k, z^i_k) and $(x^{i+1}_k, y^{i+1}_k, z^{i+1}_k)$ are the homogenous coordinates of the points p^i_k $(\in I_i)$ and p^{i+1}_k $(\in I_{i+1}$ which is the current image).

Hence, we deduce the following system (11) that expresses the cartesian coordinates (x^{i+1}_k, y^{i+1}_k) $(\in I_i)$ according to (x^i_k, y^i_k) $(\in I_{i+1})$:

$$\begin{cases} x^{i+1}_k = \dfrac{H_{1,1}.x^i_k + H_{1,2}.y^i_k + H_{1,3}}{H_{3,1}.x^i_k + H_{3,2}.y^i_k + H_{3,3}} \\ y^{i+1}_k = \dfrac{H_{2,1}.x^i_k + H_{2,2}.y^i_k + H_{2,3}}{H_{3,1}.x^i_k + H_{3,2}.y^i_k + H_{3,3}}. \end{cases} \tag{11}$$

Fig. 16 Application of a rotation. At the *top*: interest points extraction with ImpHarris detector from the two successive images used in Fig. 8. In the *middle*: homologous interest points. At the *bottom*: kept homologous interest points after applying RANSAC

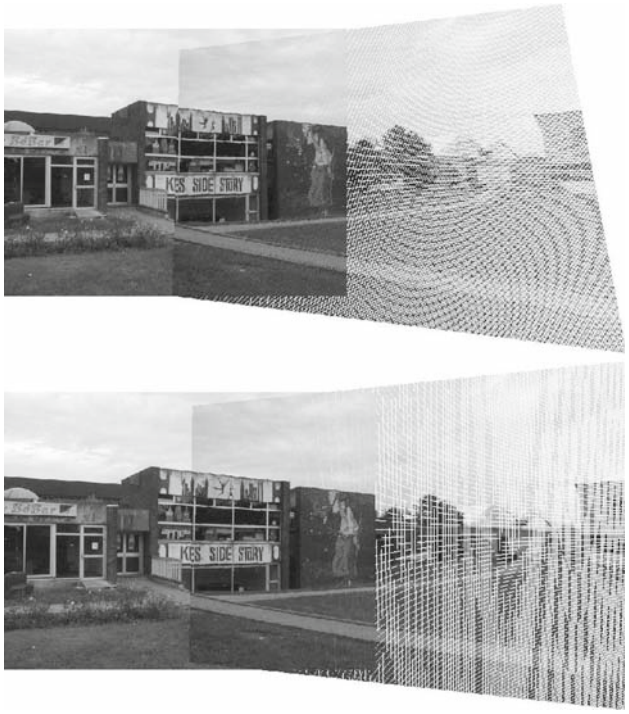

Fig. 17 Application of the homography onto two successive images (without interpolation). At the *top*: mosaic before error minimization. At the *bottom*: mosaic after error minimization

As the matrix T is defined up to scale factor, we obtain, for $H_{3,3}$ equals to 1 and for a given couple $c_k = (p_k^i, p_k^{i+1})$, the following system of equations (12):

$$\begin{cases} A_k.H = B_k, \text{ with,} \\ A_k = \begin{bmatrix} x_k^i & y_k^i & 1 & 0 & 0 & 0 & -x_k^i x_k^{i+1} & -y_k^i x_k^{i+1} \\ 0 & 0 & 0 & x_k^i & y_k^i & 1 & -x_k^i y_k^{i+1} & -y_k^i y_k^{i+1} \end{bmatrix}, \\ B_k = \begin{bmatrix} x_k^{i+1} \\ y_k^{i+1} \end{bmatrix}, \text{ and} \\ H = [H_{1,1}\ H_{1,2}\ H_{1,3}\ H_{2,1}\ H_{2,2}\ H_{2,3}\ H_{3,1}\ H_{3,2}]^t. \end{cases} \tag{12}$$

While writing the Eq. (12) for four couples of no-collinear points (c_1, c_2, c_3, c_4), we obtain the following Eq. (13):

$$[A_1\ A_2\ A_3\ A_4]^t.H = [B_1\ B_2\ B_3\ B_4]^t. \tag{13}$$

The matrix H can be obtained by solving the given 8×8 Cramer system (13). However, our experimentations showed that the resulting mosaic quality is very sensible to the false matching effects as it is illustrated by Fig. 17. The final number of the matched couples, verifying the all considered constraints, is largely superior to four. Consequently, the available equations' number is greater than the unknown variables number of the studied system. Thus, the homography H can be seen as a solution of the following optimization problem (14):

$$\min_{H \in \Re^8} \|AH - B\|_2. \tag{14}$$

where, $A = [A_1 \cdots A_m]^t$, $B = [B_1 \cdots B_m]^t$ and m is the final number of the matched interest points couples.

In practice, the system (14) has been solved by using a valid method based on the QR-factorization technique [63] followed by the relaxation-algorithm. This algorithm refines the obtained homography H by iterative elimination of the matched points couples belonging to the set ϑ (15).

$$\vartheta = \{(p_k^i, p_k^{i+1}) \,/\, d(H(p_k^i) - p_k^{i+1}) \geq \kappa\}, \tag{15}$$

where, κ is a dynamically descendant threshold and d is the Euclidian distance.

Then, the homography is computed again given the kept couples. This process stops when a sub-pixellic level of precision (mean distance lower than one pixel) is reached. Figure 17 shows the contribution of the error-minimization step for the improvement of the reconstruction process.

6.2 Intensities interpolation and radiometric correction

Once the image I_{i+1} is mapped $(H^t.I_{i+1}.H)$ in the reference of the image I_i, an interpolation process completes

Fig. 18 The mosaic illustrated in Fig. 17 after interpolation and radiometric correction

the built mosaic by interpolating the 3×3 neighbors of each pixel without antecedent. Despite this interpolation, the mosaic remains discontinuous around junctions. Indeed, in the best conditions, each pixel along a ray would have the same intensity in every image that it intersects [6]. However, this is not always true. This is mainly due to the automatic gain control system often integrated in the manipulated camera. In addition, a number of unmodelled effects cause the presence of some image edges in the resulting mosaic, for instance vignetting, parallax effects, mis-registration errors, radial distortion and so on [5]. For all that, a good blending strategy is important in order to improve the mosaic continuity around junctions. We opted for the multi-bend blending scheme used by [5] and developed by Burt and Adelson [7]. This scheme ensures smooth transitions between images despite illumination differences, whilst preserving high frequency details. An assumption of this blending strategy is described as: the image whose centre is closest to a given pixel in the rendering has the best information about that pixel, and therefore the highest blending weight [6]. We thus assign a weight function to each image pixel that varies linearly from 1 at the centre of the image to 0 at the edge. This process permits the information merge from multiple images. Besides, to avoid the blurring of high frequency details (if there are small registration errors) the low frequencies details are blended over a large spatial range, and the high ones are blended over a short range. Figures 18 and 35 show the improvement of the produced results after the application of the interpolation and the radiometric correction processes.

6.3 Mosaic warping

In the previous sections, we presented the different steps to build mosaic from two successive images of the studied sequence. In general, the number of images in the sequence is widely higher than two. Therefore, we must generalize the described mosaicing process for any number of images. We suppose that every image I_i ($i \in \{2, \ldots, n-1\}$) of the sequence overlaps only with the image I_{i-1} (for $i \in \{2, \ldots, n\}$) and the image I_{i+1} (for $i \in \{1, \ldots, n-1\}$). Each image I_i is bound to the following image (resp. previous) by an homography H_i (resp. H_{i-1}). We suppose that the final mosaic is a projection of all the sequence on the plan of only one image chosen as reference. To minimize distortions on the mosaic borders, we chose the image $I_{\lfloor n/2 \rfloor}$ as reference. Hence, the final mosaic can be computed by computing, for each image I_i, the composition of the necessary homographies in order to align the considered image on the reference image plan (16). This minimizes the distortions and the calculation time. The interpolation and the radiometric correction processes are necessary only after the projection of all the images of the sequence. Indeed, our experiments allowed us to deduce that the produced mosaic quality is much better compared to the one recorded, if these two processes are applied separately after each image projection (Fig. 24).

$$\begin{aligned} &\text{if } (i > \lfloor n/2 \rfloor) \text{ then } H = (H_{\lfloor n/2 \rfloor} \circ \cdots \circ H_{i-1}). \\ &\text{if } (i < \lfloor n/2 \rfloor) \text{ then } H = (H_i \circ \cdots \circ H_{\lfloor n/2 \rfloor}). \end{aligned} \tag{16}$$

7 Experiments and results

The new method presented in this paper allows us to stitch efficiently multiple panoramas fully automatically, without an intervention of the user. This method is valid for large camera translation and zoom, orientation of the input images and partially to moving objects (Fig. 19). Moreover, it is tolerant against light change (Fig. 20), change in acquirement conditions (Fig. 21) and for image contrast change (Fig. 22) (Note: all figures are scaled for a better fit to the pages).

In order to prove the importance of the regions matching stage for the motion estimation (large translation, rotation and zoom) as well as for the framing of the search window associated with each point of interest, we compared the points matching results produced by our method (with regions matching) versus those produced while using only the *zncc* score (without regions matching). We perform this comparison in the case of simple motion (translation without presence of rotation and/or a scale factor), in the case of rotation around the optical axis and in the case of a scale factor. For a subjective assessment of the proposed method, we defined a reconstruction error which allows us to prove the effectiveness of the proposed method.

7.1 Simple motion

Figure 23 illustrates relative statistics to the interest points matching process, according to the applied threshold of correlation (C_s). We recorded, for various values of C_s, the

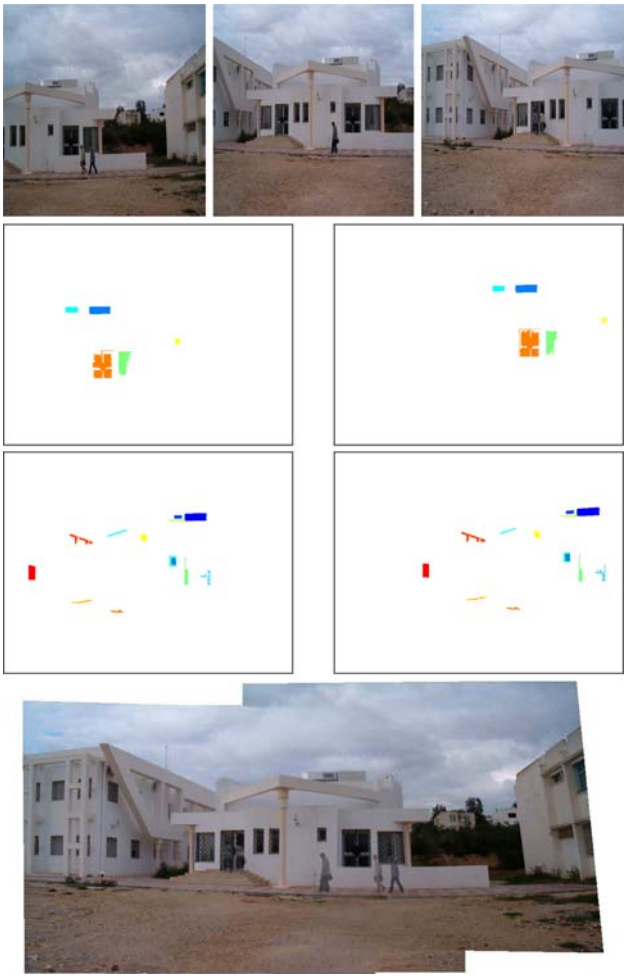

Fig. 19 Robustness of the proposed method against moving objects. At the *top*: three successive images (the moving objects are the two persons). In the *middle*: the set Λ_f of matched couples relatively to, up to down, first and second images, second and third images. At the *bottom*: the built mosaic (the reference image is the second one)

number N of the well-matched points of interest while using respectively the zncc score followed by the elimination of the ambiguous couples and the proposed method. We noticed that below a threshold value of 0.85, the number of matched points produced by the proposed method is extensively superior to the same number while using only the zncc score. This can be explained by the fact that the framing of the search space reduces the number of ambiguous points whose apparition is bound to the existence of repetitive structures on two successive images. The application of low correlation scores (<0.65) does not permit to match any point of interest when we use only the zncc score. However, the proposed method matches 68 points, for the same studied pair of images (Figs. 24, 25).

We calculated for the same data (input sequence and correlation thresholds) the mean distance between the real homologous and the correspondant points of interest determined

Fig. 20 Robustness of the proposed method against light change. At the *top*: two successive images. In the *middle*: the built mosaic before blending. At the *bottom*: the built mosaic after blending

Fig. 21 Robustness of the proposed method against change under acquisition conditions. At the *top*: two images acquired in different conditions (sunny and raining days). At the *bottom*: the built mosaic

by application of the estimated homography, and this while using respectively the zncc score (d_1) and the proposed method (d_2) (Table 1). A sub-pixellic precision is recorded

Fig. 22 Robustness of the proposed method against image contrast change. At the *top*: two different images acquired with different devices (mobile phone and hand-held camera). At the *bottom*: the built mosaic

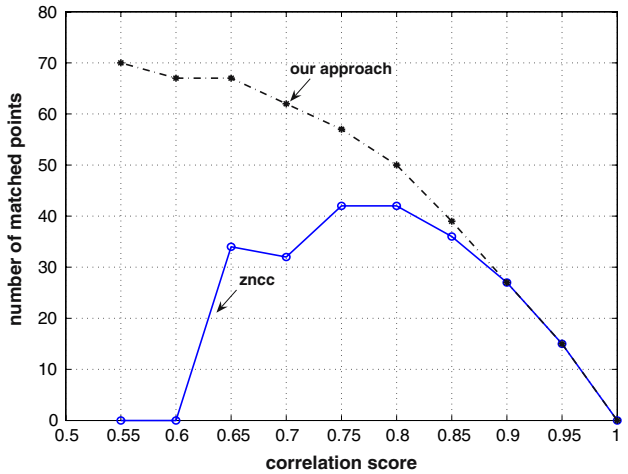

Fig. 23 Number of matched points relatively to the correlation score

while using the introduced method (d_2), even for low correlation scores, for which we could not match any point while using only the zncc score (d_1). Let's note that a high number of matched couples guarantees a good quality of reconstruction. In fact, the presence of no precise matching does not have perceptible influence when the number of couples is in the order of tens. However, this can generate a defect of reconstruction when this number is low.

7.2 Rotation around the optical axis

In Fig. 26, we recorded the same number N in the case of a rotation around the optical axis. This was done by applying a gradually increasing rotation angle to an image of the input pair. We deduced that below a rotation angle value equals to

Fig. 24 Application of a simple motion (translation). At the *top*: four original images of a sequence (outdoor scene) where the overlapping area is about 80%. At the *bottom*: the built mosaic (the reference image is the first one). This figure illustrates also the distortion when the interpolation is performed after each image projection

Fig. 25 Application of a simple motion (translation). At the *top*: three successive images where the overlapping area is about 50%. At the *bottom*: the built mosaic (the reference image is the second one)

$\pi/8$, the use of the zncc score (without regions matching) does not permit to match any point of interest. This can be explained by the no-invariance of the *zncc* score against rotation. The preliminary evaluation of the rotation angle is very profitable, hence.

Table 1 Mean distance between real homologous and correspondant points of interest determined by application of the estimated homography

C_s	0.9	0.85	0.8	0.75	0.7	0.65	0.6	0.55
d_1	0.76	0.89	0.79	0.7	0.72	0.6	–	–
d_2	0.76	0.8	0.79	0.8	0.8	0.77	0.77	0.79

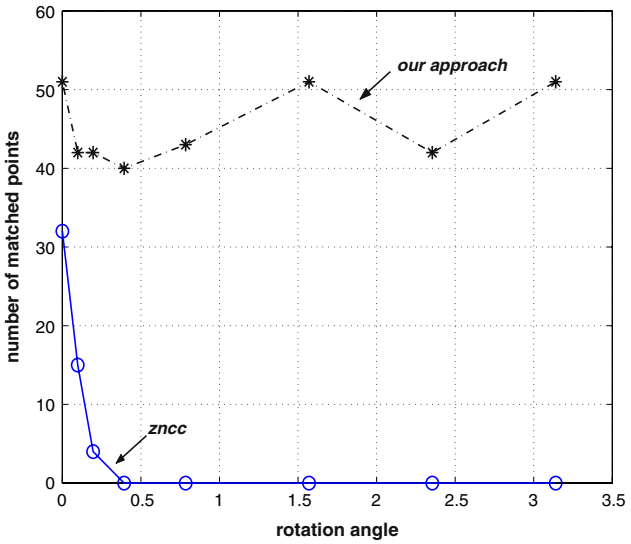

Fig. 26 Number of matched points in the presence of rotation

Fig. 27 Application of planar rotation. At the *top*: two successive images where the value of the rotation angle is 180°. At the *bottom*: the built mosaic

Fig. 28 Application of non-planar rotation around the optical axis. At the *top*: two successive images where the value of the rotation is approximately 60°. At the *bottom*: the built mosaic

In Table 2 we calculated for an increasing rotation angle (e_1) and with a fixed correlation threshold (=0.8), the estimated rotation angle (e_2) as well as the error (a_{ee}) of this estimation. This table shows that a_{ee} is always low ($\leq$0.023). Moreover, the estimation of e_2 and its application allow the interest points matching independently of the applied rotation angle. This permits to deduce that the proposed method minimizes clearly the dependance between the number of matched points and the rotation angle e_1 even if this rotation is planar (Fig. 27) or it is non-planar (Fig. 28).

7.3 Scale factor

In Fig. 29, we evaluate the interest points matching against scale change. We present the statistical results of the interest points matching relatively to an increasing scale factor (with a correlation threshold equals to 0.8) applied to one image of the considered pair. We deduce that the number of matched points is lowly bound to the scale factor when we use our method, contrarily to the only use of the zncc score. Besides, we can conclude that below a scale factor of 1.2, the determination of the zncc score does not permit to match a sufficient number of interest points. This is explained by the no-invariance of the zncc score by homothety. The scale factor must be thus estimated before performing the interest points matching.

Table 2 Estimation of the rotation angle

e_1	0	$\frac{\Pi}{32}$	$\frac{\Pi}{16}$	$\frac{\Pi}{8}$	$\frac{\Pi}{2}$	$\frac{3\Pi}{4}$	Π
e_2	0.022	0.107	0.179	0.385	1.569	2.334	3.118
a_{ee}	0.022	0.009	0.016	0.007	0.001	0.022	0.023

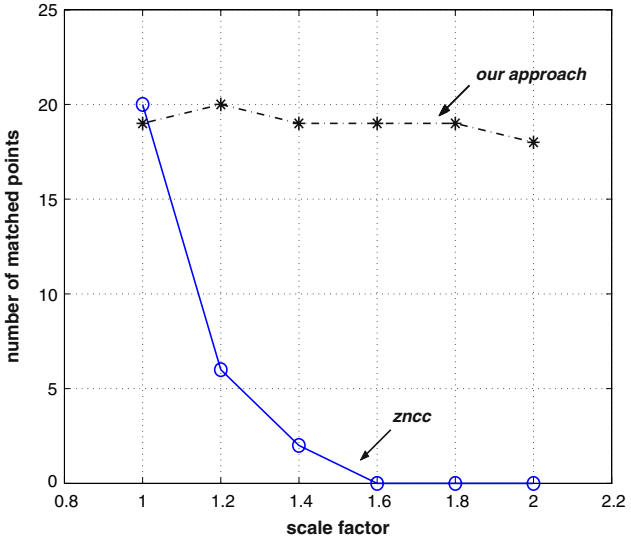

Fig. 29 Number of matched points in the presence of a scale factor

Fig. 30 Application of a scale factor. The built mosaic of the two images used in Fig. 9

In Table 3 we measure for an increasing scale factor (z_1) ($\in [1, 2]$) and with a fixed correlation threshold ($= 0.8$), the estimated scale factor (z_2) as well as the error (z_{ee}) of this estimation. This table shows that z_{ee} is relatively low (≤ 0.06). Thus, we deduce that the proposed method minimizes strongly the dependance between the number of matched points and the scale factor z_1, contrarily to the only use of the *zncc* score (Fig. 30).

7.4 Objective assessment of the built mosaic quality

For an objective assessment concerning the interest points matching and the radiometric correction processes, we defined a reconstruction error E (17) measuring the mean of the intensity difference between two successive images I_i and I_{i+1}, on the overlapping area O.

$$\mathsf{E} = \frac{1}{|O|} \cdot \sum_{(x,y)\in O} |I_{i+1}(x, y) - I_i(x, y)|. \tag{17}$$

The error E was calculated for 11 pairs of indoor and outdoor images in order to plot four curves (Fig. 31) defining:

1. the reconstruction error (error $\mathsf{E1}$) when only the best four couples, of the matched points of interest, are used;
2. the reconstruction error (error $\mathsf{E2}$) when n couples ($n \in]4, 12]$) are used;
3. the reconstruction error (error $\mathsf{E3}$) when more than 12 couples are used;
4. the reconstruction error (error $\mathsf{E4}$) after the application of the radiometric corrections ($n > 12$).

The relative curve of the error $\mathsf{E1}$ shows the insufficiency of the best four couples for the homography estimation. It is important to introduce an optimization technique to calculate therefore the desired homography. However, the manipulation of more than four points (error $\mathsf{E2}$) decreases considerably the reconstruction error. The use of a greater number of interest points couples (error $\mathsf{E3}$) improves further the results. This confirms our anterior interpretation concerning the necessity of increasing the matched interest points number to the detriment of a slight decrease of the correlation score. The smallest reconstruction error has been recorded after application of the radiometric correction (error $\mathsf{E4}$). For the images pairs 4, 6, 8 and 10, the application of radiometric correction did not improve the reconstruction quality. This can be explained by the fact that these pairs do not require a such stage, since the two correspondant images have been practically acquired in similar conditions. For all other pairs, the reconstruction has been clearly improved.

7.5 Discussion

As mentioned above, to be qualified as efficient, a mosaicing approach should be effective against camera motion,

Table 3 Estimation of the scale factor

z_1	1	1.2	1.4	1.6	1.8	2
z_2	1.01	1.18	1.38	1.59	1.75	2.06
z_{ee}	0.01	0.02	0.02	0.01	0.05	0.06

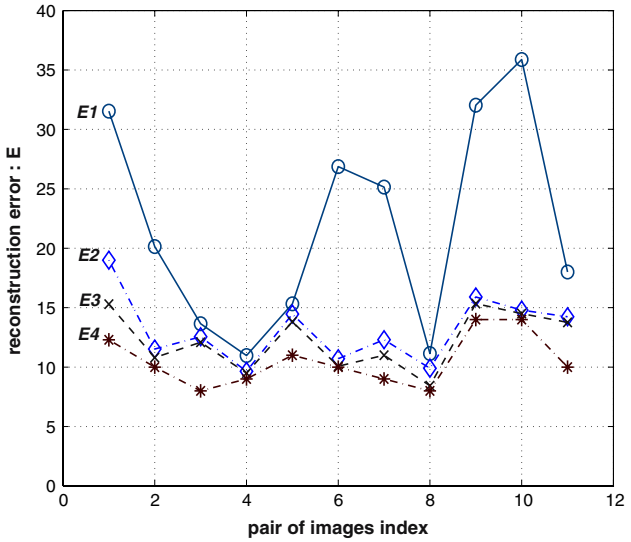

Fig. 31 Evaluation of the reconstruction errors

Fig. 32 Application of a rotation and a scale factor. At the *top*: two successive images (outdoor scene) where the scale factor is approximately 2 and the rotation angle around optical axis is approximately 50°. At the *bottom*: the built mosaic

light change, moving objects and image noise. The proposed method tried to address all of these issues. In fact, in the previous subsections, we showed the tolerance of the introduced mosaicing method against image motion (large translation, rotation and scale factor) and radiometric changes. The testing of the proposed method on the various complex and challenging images sequences, of outdoor (Figs. 32, 33) and indoor scenes (Fig. 34), confirms the high quality of the visual evaluation of the obtained results. The mosaicing is performed without camera calibration or prior knowledge about the studied scene. The regions matching step minimizes the effects of noise presence, illumination variation and changes in the scene. Indeed, most of the used sequences are held by a simple device (hand-held camera without tripod) and they do not present necessarily a wide overlap spaces. Scenes with moving objects can be partially handled by our method since there are at least four matchable regions in the overlapping space. Our method is fully automated and is tolerant against the segmentation errors since we consider only the best matchable regions.

Thus, experimental assessment stage allows us to summarize the contribution of our method in four points:

1. fully automated method requiring no a priori knowledge;
2. stability relative to the camera motion and partially to the presence of moving objects (Fig. 35);
3. the high quality of the rebuilt generalized mosaic, which is tolerant against changes in the acquirement conditions;
4. the reduced complexity of our method thanks to the framing of the potential homologous search window relative to each point of interest.

The study of some mosaicing approaches based on Harris points matching permitted us to justify the reason of the proposed method. Some approaches [16,44,66] impose that the acquirement instrument motion is a rotation or a translation. Some more general approaches [3,15] treat the case of rigid motion, but they require a slight motion guaranteeing a very big space of overlap. However, our method does not require the verification of any constraint on the camera motion which can be complex. Compared to others approaches treating the case of complex motion, our method produces more relevant results. The manipulation of the region primitive, which is steady against viewpoint change, permitted the estimation of the desired homography from the correctly matched points

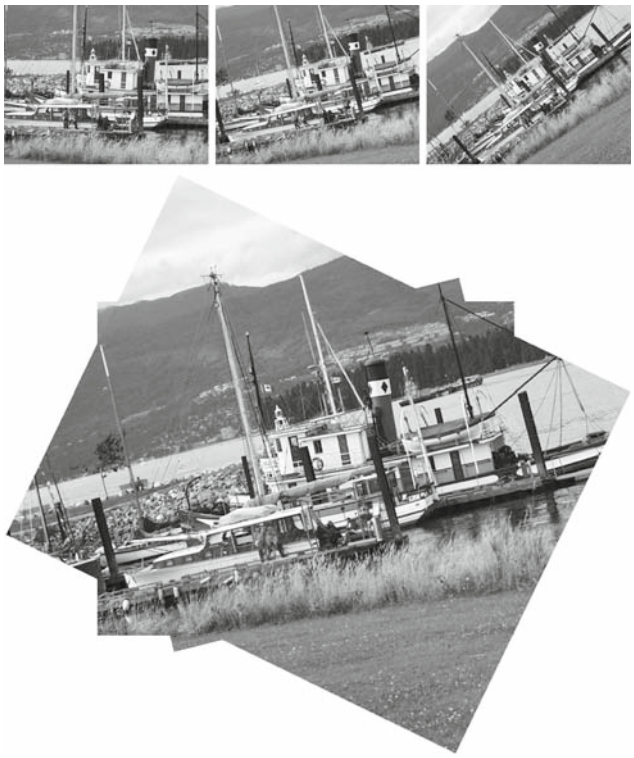

Fig. 33 Application of a scale factor and in-planar rotation. At the *top*: three successive images (outdoor scene). At the *bottom*: the built mosaic (the reference image is the second one)

Fig. 34 Mosaic obtained from two images (indoor scene) where the overlapping area is about 60%

of interest with no need to explore all the possible matching combinations. This reduces considerably the combinatory complexity.

8 Conclusions and perspectives

This paper has presented a new method for fully automatic image mosaicing permitting the integration of information from an ordered image sequence in order to create a comprehensive view of the scene. The suggested method is made of three levels. First, regions and invariant Harris points are extracted form each image. Then, the extracted primitives are matched by measuring correlation scores. The estimation of the camera motion and the framing of the search window of the potential homologous interest points are carried out by analyzing the matched-regions correspondant positions. The reduction of the search window of the potential homologous points decreases considerably the complexity of the interest points matching. Two processes of intensities interpolation and radiometric correction are then applied in order to improve the visual quality of the resulting mosaic. Finally, images are projected onto a common image grid to form a clear view. The proposed mosaicing method has been well tested on real-world indoor and outdoor scenes. It is shown to be accurate and adequate and runs in acceptable time within a timeout. Indeed, experimental simulations and objective assessment show that the proposed method is stable against camera motion, light change, moving objects and image noise.

The essential limitation of the suggested method is its restriction in the case of a relatively far scene, compared to the used acquirement devices, in order to guarantee the planarity hypothesis of the scene. Otherwise, with significant camera translation, it will be impossible to align images without using dense 3D reconstruction (3D mosaicing). One of the most crucial problems in image mosaicing is the local matching error accumulation in the case of pairwise stitching. In our case, as the images are captured and ordered in advance, the global matching stage reduces the accumulated errors by minimizing simultaneously the mismatching between all successive pairs of images. Moreover, the blending does not only remove the visual discontinuities, but also make the method more robust against small matching errors accumulation. Finally, we noticed that even under the emitted planarity hypothesis, the suggested method is partially invariant against view point change (Fig. 36) and scale change. Our method fails only in the case of images in natural environment. This is mainly due to the limited efficiency of the used segmentation with textured images.

We are now working on the extension of the proposed method to the case of an unordered image dataset in which an image can overlap with a part of the mosaic which is not seen in its predecessor [5]. We are also working on the application of the proposed method to combine a serie of video frames in the framework of video indexation and wide-area video surveillance. In addition, we are also trying to improve the treatment of moving objects. The idea is to detect

Fig. 35 Typical output of the proposed panoramic mosaicing method. At the *top*: five successive images characterized by many camera motions and the presence of moving objects. In the *middle*: the built mosaic before blending. At the *bottom*: the built mosaic after blending (the reference image is the third one)

these objects during the segmentation and matching steps in order to remove them from the final mosaic. Possible future directions would be made to attempt the use of the proposed method for 3D image registration applications such as robot navigation in 3D environment. Furthermore, a multi-agent implementation of the herewith described method can clearly reduce its calculation time. This helps propose a global consistency estimation of alignment specially in the case of full view panoramic mosaic generation. To summarize, the proposed method allows us to construct, automatically and in a reliable way, high-quality seamless full view panoramic mosaics from arbitrary collections of images acquired using

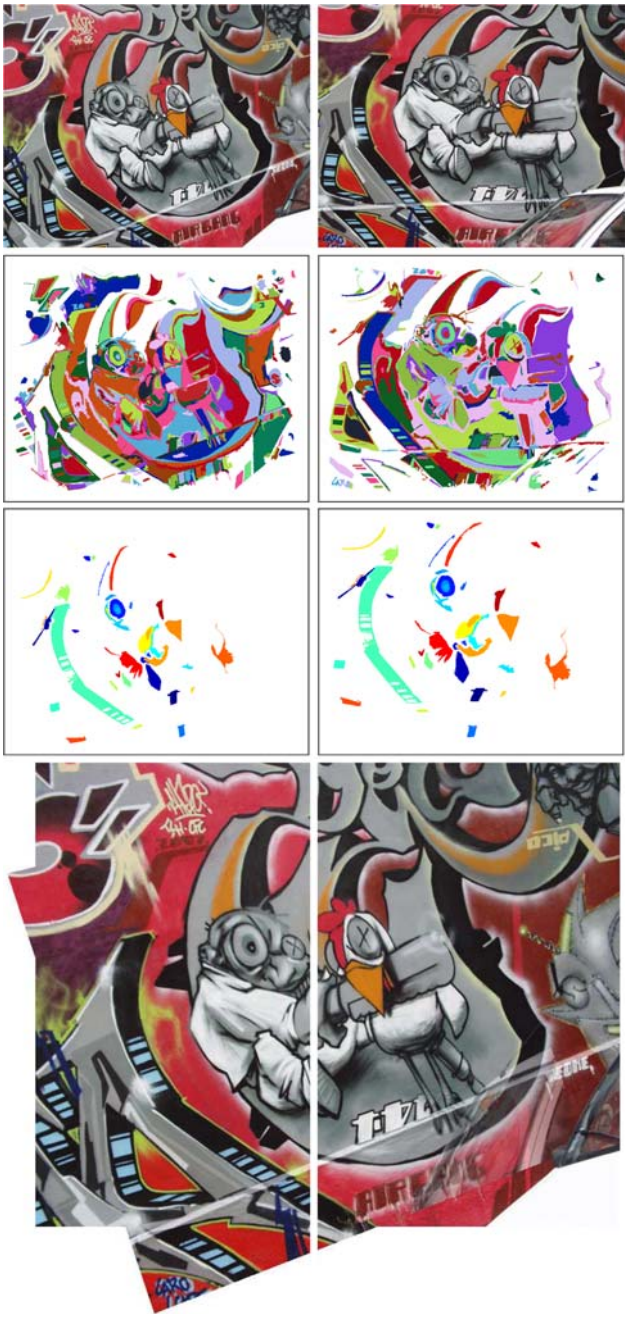

Fig. 36 Application of viewpoint changes. At the *top*: two successive images affected by viewpoint change. Second row: correspondent region maps after remove of regions touching the image borders. Third row: the set Λ_f of matched couples. At the *bottom*: the built mosaic

inexpensive simple photographic equipments and relatively free hand motions. The final mosaic is not a full 360° view, nor is 3D geometrical correctness guaranteed.

Acknowledgments We would sincerely like to thank and acknowledge Dr. Krystian Mikolajczyk, Robotics Research Group in the Department of Engineering Science, University of Oxford; who generously allowed us to use their reference set of test images for the evaluation of the preliminary motion estimation (Fig. 12).

References

1. Barhoumi, W., Zagrouba, E.: Segmentation en régions guidée par l'appariement pour un couple stéréoscopique non calibré. In: Proceedings of the 3eme Workshop Traitements et Analyse d'Information: Méthodes et Applications, Hammamet, Tunisia, pp. 287–292 (2003)
2. Baumberg, A.: Reliable feature matching across widely separated views. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, South Carolina, pp. 774–781 (2000)
3. Bhosle, U., Chaudhuri, S., DuttaRoy, S.: Fast method for image mosaicing using geometric hashing. IETE J. Res. Spec. Issue Visual Media Process. **48**(4), 317–324 (2002)
4. Brown, M., Lowe, D.: Invariant features from interest point groups. In: Proceedings of the 13th British Machine Vision Conference, Cardiff, pp. 253–262 (2002)
5. Brown, M., Lowe, D.G.: Recognising panoramas. In: Proceedings of the IEEE International Conference on Computer Vision, Nice, pp. 1218–1225 (2003)
6. Brown, M.: Multi-image matching using invariant features. PhD thesis, The University of British Columbia (2005)
7. Burt, P.J., Adelson, E.H.: A multiresolution spline with application to image mosaics. ACM Trans. Graph. **2**(4), 217–236 (1983)
8. Can, A., Stewart, C.V., Roysam, B.: Robust hierarchical algorithm for constructing mosaic from images of the curved human retina. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Colorado, pp. 286–292 (1999)
9. Choe, T.E., Cohen, I., Lee, M., Medioni, G.: Optimal global mosaic generation from retinal images. In: Proceedings of 18th International Conference on Pattern Recognition, Hong Kong, pp. 681–684 (2006)
10. Colombo, C., DelBimbo, A., Pernici, F.: Image mosaicing from uncalibrated views of a surface of revolution. In: Proceedings of British Machine Vision Conference, London (2004)
11. Davis, J.: Mosaics of scenes with moving objects. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Santa Barbara (1998)
12. Dellaert, F., Thrun, S., Thorpe, C.: Mosaicing a large number of widely dispersed, noisy, and distorted images: a bayesian approach. Technical Report, CMU-RI-TR-9934, Robotics Institute, Carnegie Mellon University, USA (1999)
13. Deriche, R., Blaszka, T.: Recovering and characterizing image features using an efficient model based approach. In: Proceedings of the Conference on Computer Vision and Pattern Recognition, New York, pp. 530–535 (1993)
14. Deriche, R.: Recursively implementing the Gaussian and its derivatives. Technical Report, RR 2422 INRIA, France (1993)
15. Douze, M., Charvillat, V., Thiesse, B.: Mosaïques d'images par approximations successives. In: Proceedings of Journées Francophones des Jeunes Chercheurs en Analyse d'Images et Perception Visuelle (ORASIS'01), Cahors, pp 97–102 (2001)
16. Duplaquet, M.L.: Building large image mosaics with invisible seam-lines. In: Proceedings of the SPIE Congres AeroSence, Orlando, pp. 369–377 (1998)
17. Feldman, D., Zomet, A.: Generating mosaics with minimum distortions. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshop, Washington DC, pp. 163–170 (2004)
18. Fischler, M.A., Bolles, R.C.: Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Comm. ACM **24**, 381–395 (1981)
19. Fraundorfer, F., Bischof, H.: Evaluation of local detectors on non-planar scenes. In: Proceedings of the 28th Workshop of the Austrian Association for Pattern Recognition, Hagenberg, pp. 125–132 (2004)

20. Fusiello, A., Aprile, M., Marzotto, R., Murino, V.: Mosaic of a video shot with multiple moving objects. In: Proceedings of the International Conference on Image Processing, Barcelona, pp. 307–310 (2003)
21. Gong, Y., Proietti, G., LaRose., D.: A robust image mosaicing technique capable of creating integrated panoramas. In: Proceedings of the IEEE International Conference on Information Visualization, London, pp. 24–29 (1999)
22. Gracias, X., Negahdaripour, S.: Underwater mosaic creation using video sequences from different altitudes. In: Proceedings of OCEANS'05 MTS/IEEE Conference, Boston, pp. 1295–1300 (2005)
23. Harris, C., Stephens, A.: Combined corner and edge detector. In: Proceedings of the 4th Alvey Vision Conference, Manchester, pp. 147–151 (1988)
24. Heikkila, M., Pietikainen, M.: An image mosaicing module for wide-area surveillance. In: Proceedings of the Third ACM International Workshop on Video Surveillance & Sensor Networks, Singapore, pp. 11–18 (2005)
25. Huang, X., Sun, Y., Metaxas, D., Sauer, F., Xu, C.: Hybrid image registration based on configural matching of scale-invariant salient region features. In: Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshop, Washington DC, pp. 167–177 (2004)
26. Irani, M., Hsu, S., Anandan, P.: Video compression using mosaic representations. Elsevier-Sig. Process. Image Commun. **7**, 529–552 (1995)
27. Jethwa, M., Zisserman, A., Fitzgibbon, A.W.: Real-Time panoramic mosaics and augmented reality. In: Proceedings of the Britsh Machine Vision Conference, London, pp. 852–862 (1998)
28. Jianchao, Y., Chern, C.T.: The practice of automatic satellite image registration. Asian J. Geoinf. **3**(4), 11–18 (2003)
29. Kanade, T., Rander, P.W., Narayaman, P.J.: Virtualized reality: constructing virtual worlds from real scenes. IEEE Trans. Multimedia **4**(1), 34–47 (1997)
30. Richmond, K., Rock, S.M.: An operational real-time large-scale visual mosaicking and navigation system. In: Proceedings of OCEANS'06 MTS/IEEE Conference, Boston, pp. 1–6 (2006)
31. Krotkov, E., Hebert, M., Simmons, R.: Stereo perception and dead reckoning for a prototype lunar rover. Auton. Robots **2**(4), 313–331 (1995)
32. Lhuillier, M., Quan, L.: Edge-constrained joint view triangulation for image interpolation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, South Carolina, pp. 218–224 (2000)
33. Lowe, D.: Object recognition from local scale- invariant features. In: Proceedings of the International Conference on Computer Vision, Corfu, pp. 1150–1157 (1999)
34. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004)
35. Mallick, S.P.: Feature based image mosaicing. Technical Report, CSE 252C University of California, San Diego (2002)
36. Mann, S., Picard, R.: Virtual bellows: Constructing high quality stills from video. In: First IEEE International Conference on Image Processing, Austin, pp. 363–367 (1994)
37. Marzotto, R., Fusiello, A., Murino, V.: High resolution video mosaicing with global alignment. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Washington, pp. 692–698 (2004)
38. Matas, J., Burianek, J., Kittler, J.: Object recognition using the invariant pixel-set signature. In: Proceedings of the British Machine Vision Conference, London, pp. 606–615 (2000)
39. Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust wide baseline stereo from maximally stable extremal regions. In: Proceedings of the British Machine Vision Conference, Cardiff, pp. 384–393 (2002)
40. McLauchlan, P.F., Jaenicke, A.: Image mosaicing using sequential bundle adjustment. In: Proceedings of the British Machine Vision Conference, Bristol, pp. 616–625 (2000)
41. McMillan, L., Bishop, G.: Plenoptic modeling: an image-based rendering system. In: Proceedings of the 22nd Annual Conference on Computer Graphics (SIGGRAPH'95), Los Angeles, pp. 39–46 (1995)
42. Mehnert, A., Jackway, P.: An improved seeded region growing algorithm. Pattern Recogn. Lett. **18**, 1065–1071 (1997)
43. Merino, L., Wiklund, J., Caballero, F., Moe, A., De Dios, J.R.M., Forssen, P.E., Nordberg, K., Ollero, A.: Vision-based multi-UAV position estimation: Localization based on blob features for exploration missions. IEEE Robot. Autom. Magaz. **13**(3), 53–62 (2006)
44. Meunier, L., Borgmann, M.: High-resolution panoramas using image mosaicing. Final Project, EE368 (Digital Image Processing), Stanford University, USA (2000)
45. Mikolajczyk, K., Schmid, C.: A performance evaluation of local descriptors. In: Proceedings of the Interational Conference on Computer Vision and Pattern Recognition (2003)
46. Mikolajczyk, K., Schmid, C.: Comparaison of affine-invariant local detectors and descriptors. In: Proceedings of the 12th European Signal Processing Conference, Vienna (2004)
47. Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., Van Gool, L.: A comparaison of affine region detectors. Int. J. Comput. Vis. **65**(1–2), 43–72 (2006)
48. NagaoM. Matsuyama, T.: Edge preserving smoothing. Comput. Graph. Image Process. **9**, 394–407 (1979)
49. Nicolas, H.: New methods for dynamic mosaicing. IEEE Trans. Image Process. **10**(8), 1239–1251 (2001)
50. Peleg, S., Rousso, B., Rav-Acha, A., Zomet, A.: Mosaicing on adaptive manifolds. IEEE Trans. Pattern Anal. Mach. Intell. **22**(10), 1144–1154 (2000)
51. Peleg, S., Herman, J.: Panoramic mosaics by manifold projection. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Juan, pp. 338–343 (1997)
52. Remondino, F.: Detectors and descriptors for photogrammetric applications. In: Proceedings of the Symposium on Photogrammetric Computer Vision, Bonn, pp. 49–54 (2006)
53. Robinson, J.A.: A simplex-based projective transform estimator. In: Proceedings of the International Conference on Visual Information Engineering, Guildford, pp. 290–293 (2003)
54. Rousso, B., Peleg, S., Finci, I., Rav-Acha, A.: Universal mosaicing using pipe projection. In: Proceedings of the Sixth IEEE International Conference on Computer Vision, Bombay, pp. 945–952 (1998)
55. Sawhney, H.S., Kumar, R., Gendel, G., Bergen, J., Dixon, D., Paragano, V.: VideoBrushTM: experiences with consumer video mosaicing. In: Proceedings of the Workshop on Applications of Computer Vision, Los Alamitos, pp. 56–62 (1998)
56. Schechner, Y.Y., Nayar, S.K.: Generalized mosaicing. In: Proceedings of the IEEE International Conference on Computer Vision, vol. 1, pp. 17–24, Vancouver (2001)
57. Schmid, C., Mohr, R., Bauckhage, C.: Evaluation of interest point detectors. Int. J. Comput. Vis. **37**(2), 151–172 (2000)
58. Schmid, C., Mohr, R.: Local grayvalue invariants for image retrieval. IEEE Trans. Pattern Anal. Mach. Intell. **19**(5), 530–534 (1997)
59. Shum, H.Y., Szeliski, R.: Systems and experiment paper: construction of panoramic mosaics with global and local alignment. Int. J. Comput. Vis. **36**(2), 101–130 (2000)
60. Shum, H., Szeliski, R.: Panoramic image mosaics. Technical Report MSR-TR-97-23, Microsoft Research, USA (1997)
61. Singh, R., Vatsa, M., Ross, A., Noore, A.: Performance enhancement of 2D face recognition via mosaicing. In: Proceedings of the Fourth IEEE Workshop on Automatic Identification Advanced Technologies, MorganTown, pp. 63–68 (2005)

62. Smith, P., Sinclair, D., Cipolla, R., Wood, K.: Effective corner matching. In: Proceedings of the 9th British Machine Vision Conference, London, pp. 545–556 (1998)
63. Sun, C.: Dealing with dense rows in the solution of sparse linear least squares problems. Technical Report CTC95TR227, Advanced Computing Research Institute, Cornell University, Ithaca (1995)
64. Szeliski, R., Kang, S.B.: Direct methods for visual scene reconstruction. In: IEEE Workshop on Representation of Visual Scenes, Cambridge, pp. 26–33 (1995)
65. Tell, D., Carlsson, S.: Combining appearance and topology for wide baseline matching. In: Proceedings of the 7th European Conference on Computer Vision, Copenhagen, pp. 68–81 (2002)
66. Teyssier, P.: Reconstruction panoramique. Technical Report, Ecole Polytechnique, France (1999)
67. Trucco, E., Doull, A., Odone, F., Fusiello, A., Lane, D.: Dynamic video mosaicing and augmented reality for subsea inspection and monitoring. In: Oceonology International 2000, Bringhton (2000)
68. Unnikrishnan, R., Kelly, A.: Mosaicing large cyclic environments for visual navigation in autonomous vehicles. In: Proceedings of the IEEE International Conference on Robotics and Automation, Washington DC, pp. 4299–4306 (2002)
69. Van Gool, L., Tuytelaars, T., Turina, A.: Local features for image retrieval. In: State-of-the-Art in Content-Based Image and Video Retrieval, pp. 21–41. Kluwer, Dordercht (2001)
70. Wang, J., Quan, L., Sun, J., Tang, X., Shum, H.Y.: Picture Collage. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, New York, pp. 347–354 (2006)
71. Winter, M., Bischof, H., Fraundorfer F.: Maximally stable corner clusters: a novel distinguished region detector and descriptor. In: Proceedings of the Austrian Cognitive Vision Workshop, pp. 59–66 (2005)
72. You, X., Fang, B., Tang, Y.Y.: Mosaicing the retinal fundus images: a robust registration technique based approach. Lect. Notes Comput. Sci. Adv. Natural Comput. **3612**, 663–667 (2005)
73. Zagrouba, E., Belhassen, S.: Construction Robuste de Mosaiques Généralisées. In: Proceedings of the International Conference on Image and Signal Processing (ICISP'2003), Agadir, pp. 519–526 (2003)
74. Zagrouba, E.: 3D facets construction for stereovision. In: International Conference on Artificial Intelligence pp. 116–126. Springer, Rome (1997)
75. Zuliani, M., Kenney, C., Manjunath, B.S.: A mathematical comparison of point detectors. In: Proceedings of the IEEE Image and Video Registration Workshop (IVR), Washington DC (2004)
76. Zitova, B., Flusser, J., Kautsky, J., Peters, G.: Feature point detection in multiframe images. In: Proceedings of the Czech Pattern Recognition Workshop, Czech Pattern Recognition Society, Perslak, pp. 117–122 (2000)
77. Zitova, B., Flusser, J.: Image registration methods: a survey. Image Vis. Comput. **21**, 977–1000 (2003)
78. Zoghlami, I., Faugeras, O., Deriche, R.: Using geometric corners to build a 2D mosaic from a set of image. In: Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Puerto Rico, pp. 420–425 (1997)
79. Zomet, A., Peleg, S., Arora, C.: Rectified mosaicing: mosaics without the curl. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, South Carolina, pp. 2459–2465 (2000)
80. Zomet, A., Peleg, S.: Applying super-resolution to panoramic mosaics. In: Proceedings of the IEEE Workshop on Applications of Computer Vision, Princeton, pp. 286–287 (1998)

Author biographies

E. Zagrouba is a Computer Science full Professor in the Higher Institute of Computer Sciences (ISI) - University of El Manar, Tunisia. He received a Ph.D. degree in Computer Sciences from the Polytechnic National Institute of Toulouse (INPT-ENSEEIHT, France) in 1994. He is the head of the Research Unit of Imaging Intelligent Systems and Computer Vision (URSIIVA, Tunisia) and he is also the general secretary of the Tunisian Research Association on Sciences for Images (ARTS-PI, Tunisia). Dr. Zagrouba's research interests, within the laboratory URSIIVA-ISI, are in Computer Vision and Image processing.

W. Barhoumi received a Master degree from the Polytechnic National Institute of Toulouse (INPT-ENSEEIHT, France) in July 1998 and a Ph.D. degree from the National School of Computer Science (ENSI), University of Mannouba (Tunisia) in January 2006. He is currently an Assistant Professor at the Superior School of Technology and Computer Science (ESTI), University of 7 November, Carthage, Tunisia. Dr. Barhoumi's research interests, within the laboratory URSIIVA-ISI, focuses on the themes of Medical Image Analysis and Computer Vision.

S. Amri received his Master degree from the Faculty of Sciences of Tunis (FST), University of El Manar (Tunisia) in 2005. He is currently a Ph.D. candidate in the FST. His research interests, within the laboratory URSIIVA-ISI, are Video Surveillance and Spatio-temporal image segmentation.