J Math Imaging Vis (2010) 37: 166–182
DOI 10.1007/s10851-010-0199-1

Mesh Topological Optimization for Improving Piecewise-Linear Image Registration

Javier González · Vicente Arévalo

Published online: 31 March 2010

Abstract This paper presents a mutual-information based optimization algorithm for improving piecewise-linear (PWL) image registration. PWL-registration techniques, which are well-suited for registering images of the same scene with relative local distortions, divide the images in conjugate triangular patches that are individually mapped through affine transformations. For this process to be accurate, each pair of corresponding image triangles must be the projections of a planar surface in space; otherwise, the registration incurs in errors that appear in the resultant registered image as local distortions (distorted shapes, broken lines, etc.). Given an initial triangular mesh onto the images, we propose an optimization algorithm that, by swapping edges, modifies the mesh topology looking for an improvement in the registration. For detecting the edges to be swapped we employ a cost function based on the mutual information (MI), a metric for registration consistency more robust to image radiometric differences than other well-known metrics such as normalized cross correlation (NCC). The proposed method has been successfully tested with different sets of test images, both synthetic and real, acquired from different angles and lighting conditions.

Keywords Image registration · Piecewise-linear functions · Mesh optimization · Greedy search · Mutual information

This work is supported by the Spanish Government under the research contracts CICYT DPI-2008-03527 and 2008-TEP-4016.

J. González · V. Arévalo (✉)
Dept. of System Engineering and Automation,
University of Málaga, Campus Teatinos–Complejo Tecnológico,
29071 Málaga, Spain
e-mail: varevalo@ctima.uma.es

J. González
e-mail: jgonzalez@ctima.uma.es

1 Introduction

Image registration is a crucial stage in many computer vision applications like image fusion, change detection, mosaicking, 3D scene reconstruction, etc. In this process one image remains fixed (the *fixed* or *reference* image) whereas the other (the *moving* or *input* image), acquired on a different time, from a different viewpoint and/or using a different sensor, is spatially transformed until fitting with the first one. A broad variety of mapping functions has been reported in the literature including polynomial [1], radial basis functions [2], piecewise-linear [3] or -cubic [4] functions, multiquadric functions [5], B-splines functions [6], etc. (see [7] for a survey).

Of particular significance is the case of piecewise-linear (PWL) functions, which are especially suitable for registering images of polyhedral scenes (typical in indoor and urban environments), since they divide the images into triangles which are individually registered through affine transformations (see Fig. 1(a)). For PWL registration to perform accurately every pair of corresponding triangles must lie on projections of a 3D planar surface, otherwise, the registration may generate undesirable artifacts, such as broken lines, which diminish the registration quality (see Fig. 1(a)).

Current implementations of PWL image registration included in scientific image processing software packages such as Matlab [8], Image Registration Software [9] (from Image Fusion Systems Research) or in remote sensing ones like ENVI/IDL [10], ERDAS [11], etc., generate the conjugate triangular meshes from a set of correspondence pairs by means of some triangulation technique, typically the Delaunay's method [12], which produces triangles of balanced size and shape, but which are not optimal for covering as many planar patches as possible.

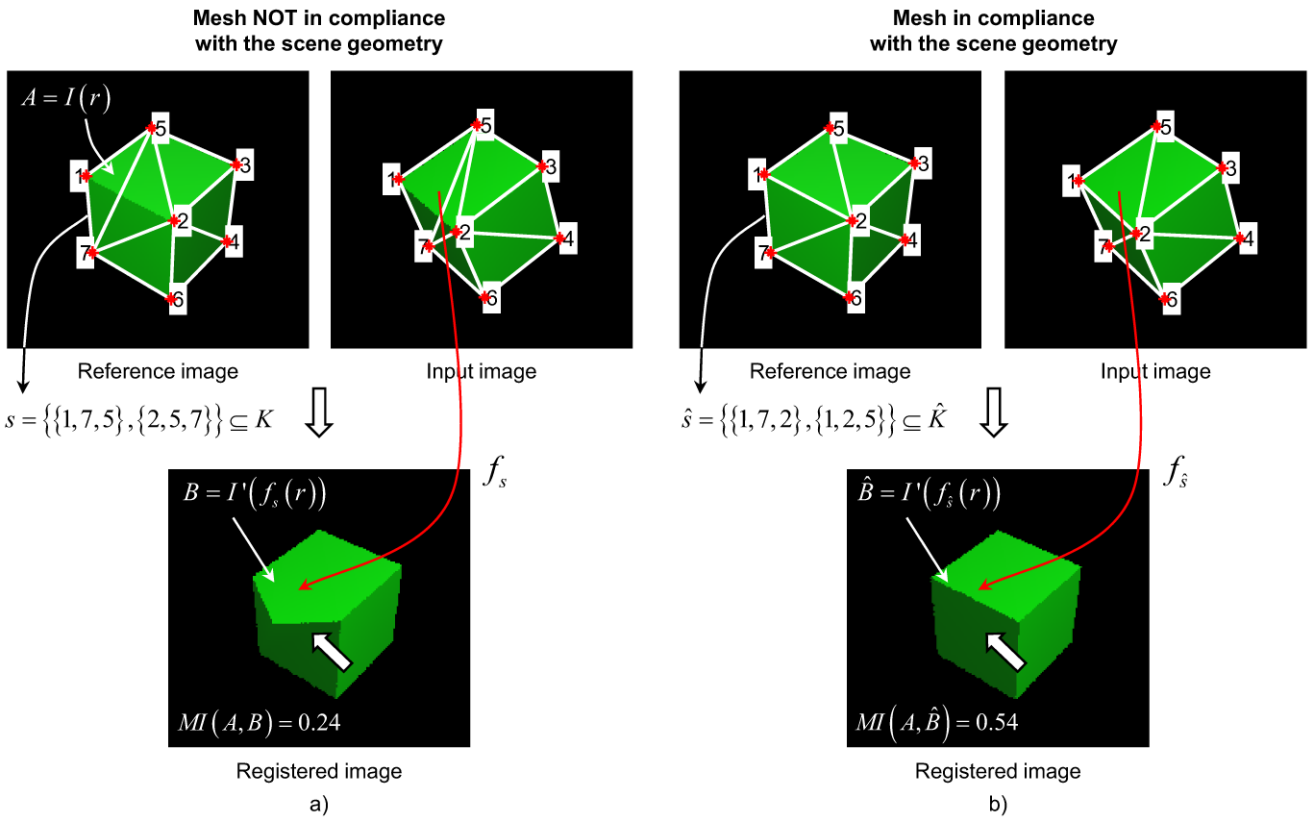

Fig. 1 For a piecewise-linear registration process to be successful, the triangles must be projections of one single plane surface of the scene, as the triangle {1, 2, 5} in (**b**); otherwise broken lines are produced and the registration consistency decreases, as in (**a**)

In this paper, we propose a method to modify the topology of a given initial triangular mesh by iteratively swapping its edges in order to improve the registration of a pair of images. We state this process as a *greedy* search [13] that, at each step, focuses on a particular quadrilateral and swap its central edge (shared by the two adjacent triangles) if the resultant triangles yield a better registration consistency. Though a global minimum is not guaranteed, the algorithm converges to an optimized mesh which produces a highly accurate PWL-registration between the images. A by-product of such improvement is the possibility of back-projecting the triangular mesh to space and to reconstruct an unscaled 3D model of the scene. Notice that, although the proposed method is especially suited for polyhedral scenes, it can be also applicable to curved surfaces. In that case, the method tries to find the triangular mesh that better approximate the surface, which is obviously limited by the geometrical realization (i.e. selected vertices) of the mesh.

A key aspect of our proposal is that of measuring how good the registration of a pair of conjugate quadrilateral image patches is. In this work we propose the use of the mutual information (MI) associated to the intensity values of the patches as a measure of their registration consistency [14, 15]. Unlike other well-known metrics such as normalized cross-correlation (NCC) [16] or sum of square differences (SSD) [17], MI can cope with non-linear differences in the image radiometry and, consequently results in clear advantage in many applications of registration.

Next, a review of the most representative techniques of mesh optimization is given, placing special emphasis in those proposed within the image registration field. In Sect. 3, several assumptions and definitions, as well as the formulation used in subsequent sections, are presented. Section 4 describes the proposed method, the consistency estimation function and the optimization process. In Sect. 5, we present and discuss some experimental results. Finally, some conclusions are outlined.

2 Related Works

The generation of optimal triangular meshes is a problem of significance in a variety of fields such as object modelling, surface approximation, image compression, image reconstruction, etc. Though its ultimate goal is to approximate, as well as possible, some data by a piecewise planar function, the concrete goal of a mesh optimization technique varies with the type of problem. Thus, in object modelling the optimization is aimed at generating 3D triangular meshes that

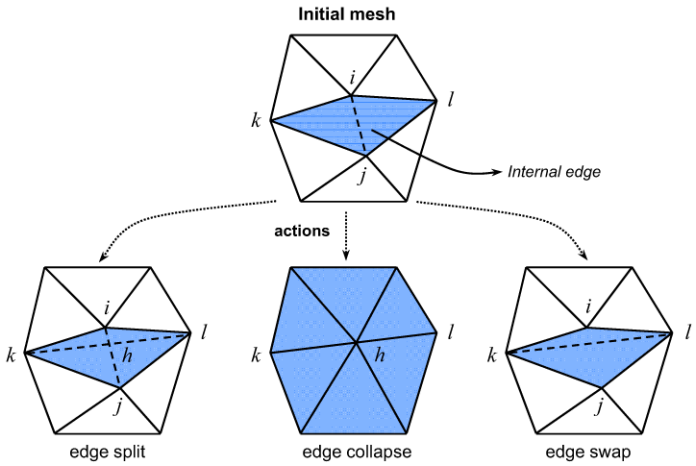

Fig. 2 Edge-based actions employed to modify the topological/geometrical realization of a given mesh

properly represent the 3D shape of a scene or object using the minimum number of triangles. In image processing we find interesting applications where the so-called Data-Dependent-Triangulation (DDT) approach is used to approximate the intensity function, either to reduce the amount of data of an image (as in image compression [18]), or to fit a continuous piecewise-planar surface upon the discrete image samples (image interpolation [19]).

Mesh optimization techniques can be classified according to different perspectives: the mechanism used for modifying the mesh (i.e. type and scope of the actions), the metric for evaluating the goodness of a given mesh modification (energy or cost functions), and the procedure for accomplishing the mesh refinement.

According to the type and scope of the actions applied to modify the mesh, we encounter techniques where (see Fig. 2):

(1) only the topological realization is modified, by swapping edges [16, 17, 19, 20],
(2) only the geometrical realization is modified, by refining the vertex coordinates (approach mostly employed in image registration) [21–24], and
(3) both the topological and geometrical realization are simultaneously refined, by splitting, collapsing, and/or swapping edges and refining the vertex coordinates [25–28].

Many of these methods were developed in the field of geometric modelling to simplify and refine an initial very-detailed 3D mesh obtained upon a dense set of vertices provided by a 3D sensor, for example, a laser range finder [25, 26]. Similar methods have been also used for 3D scene reconstruction [17, 20, 28], generation of compatible meshes (i.e. isomorphic meshes of the interiors of two polygons with correspondence between their vertices) for constructing swept volumes [29–31] and in different applications of the DDT concept to image processing [18, 19, 32].

Unlike these methods, which apply on 3D meshes, in piecewise-linear image registration we are provided with two conjugate 2D triangular meshes which must be modified in an attempt to maximize their image registration consistency. An example of this is the work in [21], which relocates the mesh vertices (the mesh topology remains fixed) in order to compensate for the affine motion in video streaming. Vertex coordinate refinement, though being suitable for smooth image distortions, does not provide enough correction power to accommodate the possibly important geometric differences between the images when they are acquired from very different angles, as happens in the example of Fig. 1.

Typically, topological mesh optimization techniques are formulated as minimization (or maximization) processes that range from random searches [20] to more complex procedures based on simulated annealing [33], Bayesian stochastic models [28], variational approaches[1] [34], etc.

Whatever the applied optimization technique, one of the key points is that of defining a convenient cost or energy function to evaluate the enhancement in the refined mesh if certain action is applied. Whereas in other fields (object modelling, surface fitting, image interpolation, etc.) measuring the quality of a mesh can be accomplished upon the available 3D points, in image registration we must rely only on the radiometric similarity between the reference and input image. So far, several metrics have been used for this purpose: sum of square differences (SSD) [17], normalized cross-correlation (NCC) [16], and some templates based on image differences [20].

None of these measurements are invariant to non-linear radiometric differences between images, as it may be the case of having images captured by different cameras, or

[1] In www.itk.org, we can find a broad variety of code which implements numerous of these techniques such as potential yields, elastic bodies, etc.

the same camera but with very different lighting conditions, which provokes shadows, intensity saturations, reflections, etc. In this paper, we propose an optimization process driven by a cost function based on the mutual information (MI) of the images being registered. The implemented algorithm relies on a greedy search that modifies the mesh topology by applying edge swap operations that entail an increase in the MI. Though MI has been used as an image registration consistency metric in some works [35–39], this is the first time it has been integrated into a mesh optimization framework for piecewise-linear registration.

3 Assumptions and Definitions

In this work we suppose that the scene projection onto the sensor can be approximated by a paraperspective transformation, also called affine or parallel camera [40]. This simplification is assumable in those computer vision setups where the perspective effects are negligible, that is, parallel lines in space almost keep their parallelism in image. Affine projection leads to a great reduction in complexity in many vision problems. In particular, for image registration, it implies that 3 points in correspondence (instead of the 4 ones required for its general form) suffice to estimate the homography (also called affinity under this assumption) which maps points from one image patch to another [41]. Thus, when performing an affine mapping between two conjugate image triangles, they must perfectly match; otherwise, the triangles are projections of a non-planar surface. Next, we introduce the notation employed in this work as well as some useful definitions.

A *triangular mesh* is a piecewise-linear structure consisting of triangular faces put together along their edges and vertices. Formally, a mesh is a pair $M = (K, V)$, where $V = \{v_i, i = 1, \ldots, m | v_i \in \mathbb{R}^2\}$ is a set of vertex positions which defines the shape of the mesh in $\mathbb{R}^2$ and K is a topological space, called simplicial complex, which determines the connectivity of the vertices, edges and faces. A *simplicial complex* K consists of a set of vertices $\{1, \ldots, m\}$ together with a set of non-empty subsets of the vertices, called the simplices of K: the 0-simplices $\{i\} \in K$ are *vertices*, the 1-simplices $\{i, j\} \in K$ are *edges*, and the 2-simplices $\{i, j, k\} \in K$ are *triangles* or *facets* [26, 42].

For a given simplicial complex K (see Fig. 3), the *topological realization*, denoted by $|K|$, results of identifying the vertices $\{1, \ldots, m\}$ with the standard basis vectors $\{e_1, \ldots, e_m\}$ of $\mathbb{R}^m$. Let $\phi : \mathbb{R}^m \mapsto \mathbb{R}^2$ be the linear mapping that sends the i-th standard basis vector $e_i \in \mathbb{R}^m$

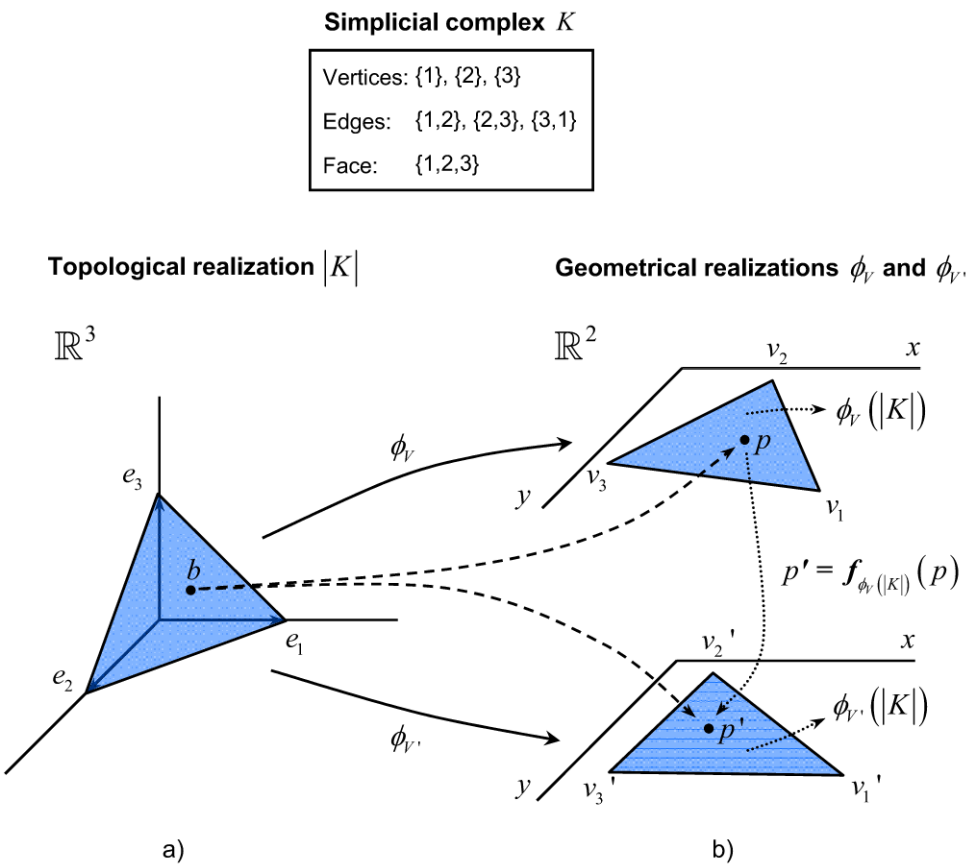

Fig. 3 Example of mesh representation: a mesh consisting of one face

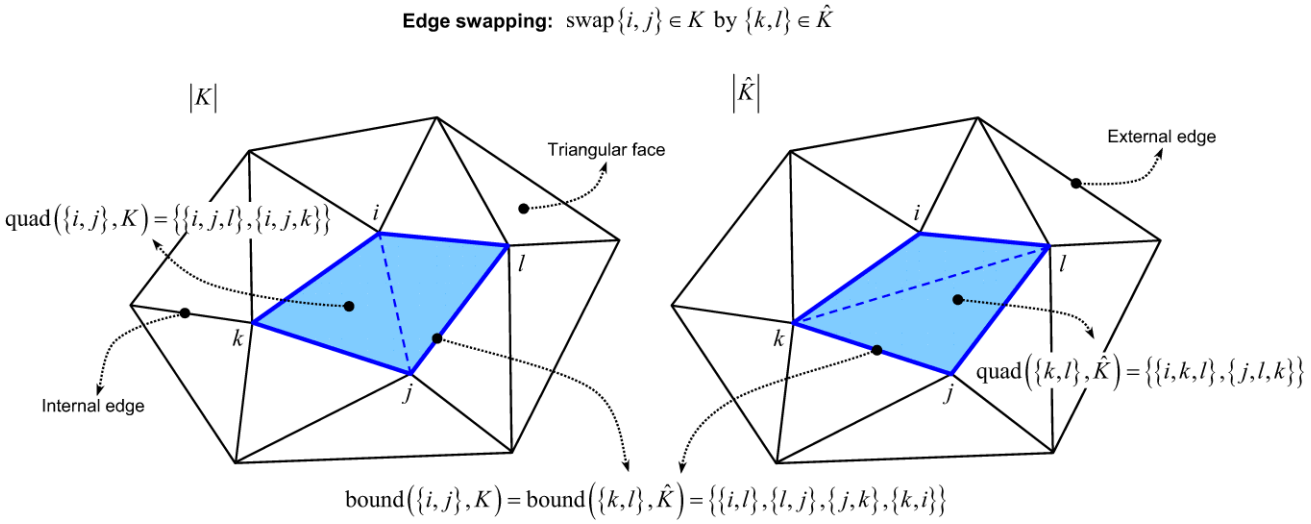

Fig. 4 The topological action of swapping an edge when all preconditions, as explained in Sect. 4.2, are verified. Figure also illustrates the topology elements that take part in a piecewise-linear image registration process. Observe that after applying the swapping action the boundary of the quadrilateral does not change

to $v_i \in \mathbb{R}^2$. The *geometrical realization* of M is given by $\phi_V(|K|)$, where we write the subscript V in ϕ_V to make explicit that it is specified by that particular vertex set. The map ϕ_V is called an *embedding* if it is 1-1, that is, if $\phi_V(|K|)$ is not self-intersecting.

Thus, to refer to any point within a part s of the mesh ($s \subseteq K$), we employ the notation $p \in \phi_V(|s|) \in \mathbb{R}^2$. For example, $p \in \phi_V(|t|)$ refers to one point within the triangle $t = \{i, j, k\} \in K$; $p \in \phi_V(|q|)$ refers to one point within a quadrilateral of M consisting of two adjacent triangles $q = \{\{i, j, k\}, \{i, j, l\}\} \in K$, and so on.

In addition to the above general definitions, we introduce the following particular ones, of interest for stating our method in the next section:

- An edge $\{i, j\} \in K$ is *external* or *boundary* if it is a subset of only one face in K, and *internal* or *shared* otherwise.
- Given an internal edge $e = \{i, j\} \in K$, we define the following functions (see Fig. 4):
  - quad$(\{i, j\}, K) = \{\{i, j, l\}, \{i, j, k\}\}$ which delivers the two triangles that share the edge e. We will also refer to these triangles as the "triangles associated to the edge e".
  - bound$(\{i, j\}, K) = \{\{i, l\}, \{l, j\}, \{j, k\}, \{k, i\}\}$ which gives the four edges of the quadrilateral.
- Let $M = (K, V)$ and $M' = (K, V')$ be the triangular meshes defined onto two images to register. M and M' have the same topological realization $s \subseteq K$ (they are isomorphic) and present different geometric realizations, given by the set of vertices' pairs $\{(v_i, v_i'), i = 1, \ldots, n | v_i \in V, v_i' \in V\}$. We define the piecewise-linear transformation as the embedding $\boldsymbol{f} : \phi_V(|s|) \mapsto \phi_{V'}(|s|)$ which geometrically maps a point $p = (x, y)^\top \in \phi_V(|s|)$ to another point $p' = (x', y')^\top \in \phi_{V'}(|s|)$ as follows:

$$p' = \boldsymbol{f}_{\phi_V(|s|)}(p) = \begin{cases} f_{\phi_V(|t_1|)}(p) & \text{if } p \in \phi_V(|t_1|) \\ f_{\phi_V(|t_2|)}(p) & \text{if } p \in \phi_V(|t_2|) \\ \vdots & \\ f_{\phi_V(|t_m|)}(p) & \text{if } p \in \phi_V(|t_m|) \end{cases} \quad (1)$$

where m is the number of triangles, and $f_i : \phi_V(|t_i|) \mapsto \phi_{V'}(|t_i|)$ is an affine mapping estimated from the geometrical realization of the three vertices of $t_i \in s$ in both meshes, which can be expressed by the transformation:

$$p' = f_{\phi_V(|t_i|)}(p) \equiv \begin{cases} x' = a_{i,1} + a_{i,2}x + a_{i,3}y \\ y' = b_{i,1} + b_{i,2}x + b_{i,3}y \end{cases} \quad (2)$$

being $a_{i,j}$ and $b_{i,j}$, with $j = 1, 2, 3$, the affine transformation coefficients. Notice that once $\boldsymbol{f}_{\phi_V(|s|)}$ (for clarity, $\boldsymbol{f}_s$ from now on) has been applied $\phi_V(|s|) = \phi_{V'}(|s|)$, that is, the corresponding faces of both meshes must overlap perfectly (remember that $\phi_V(|s|)$ represents all the points—pixels—within the mesh given by the simplicial s).
- An edge $\{i, j\} \in K$ is said to be *3D-compatible* if it lies on a projection of a 3D plane surface, and *3D-incompatible* otherwise. Since 3D scene information is not available, we assume that edge *3D-incompatibility* manifests as an error in the piecewise-linear registration of its "associated faces": the larger the error, the higher the *3D-incompatibility* of the edge.

4 Description of the Proposed Method

The method presented in this paper modifies the connectivity of the conjugate meshes by iteratively swapping the mesh

edge that leads to the greatest improvement in the image registration consistency. This procedure only affects the mesh connectivity since the number of vertices and their coordinates remain without modification.

4.1 Mutual Information as a Metric for Registration Consistency

The metric we employ for measuring the registration consistency is the mutual information (MI) [43]. MI measures the statistical dependency or information redundancy of two random variables. Unlike other similarity measures such as the sum of square differences (SSD) or the normalized cross-correlation (NCC) which allow for a functional relationship between the gray-levels of the image patches to register, the MI responses to their statistical relationship, which can be estimated from the joint entropy. The advantage of this metric is that it is more robust to possible image radiometric differences that are difficult (or impossible) to model by a function, which happens to be the case of noise, shadows and specular reflections or those stemmed from images acquired from different angles, with different sensors or at different moments in time [36].

Mathematically, the MI of two equal-sized[2] image patches A and B can be written as:

$$MI(A,B) = H(A) + H(B) - H(A,B) \tag{3}$$

where $H(A)$ and $H(B)$ are the entropies of A and B, and $H(A,B)$ their joint entropy:

$$\begin{aligned} H(A,B) &= -\sum_{a,b} P_{A,B}(a,b)\log_2 P_{A,B}(a,b) \\ H(A) &= -\sum_{a} P_A(a)\log_2 P_A(a) \\ H(B) &= -\sum_{b} P_B(b)\log_2 P_B(b) \end{aligned} \tag{4}$$

being $P_A(a)$, $P_B(b)$ and $P_{A,B}(a,b)$ the probability distribution functions estimated from the intensity joint histogram h of A and B

$$\begin{aligned} P_{A,B}(a,b) &= \frac{1}{N} h_{A,B}(a,b) \\ P_A(a) &= \sum_{b} P_{A,B}(a,b) \\ P_B(b) &= \sum_{a} P_{A,B}(a,b) \end{aligned} \tag{5}$$

where

$$h_{A,B} = \begin{pmatrix} h(0,0) & h(0,1) & \dots & h(0,M-1) \\ h(1,0) & h(1,1) & \dots & h(1,M-1) \\ \dots & \dots & \dots & \dots \\ h(M-1,0) & h(M-1,1) & \dots & h(M-1,M-1) \end{pmatrix}, \tag{6}$$

N is the number of pixels in the image patch, and M is the number of histogram bins, respectively.

The value $h_{A,B}(a,b)$, $a \in [0, M-1]$, $b \in [0, M-1]$, is the number of corresponding pairs having the intensity value a in the first image and the intensity value b in the second one. In the case of 8-bit gray-scale images, the original value of M is 256; however, in practice, it is convenient to use a lower value (e.g. 128, 64, 32, ...), for three reasons:

(1) to make more reliable the estimation of the joint probability from the joint histogram when N is not very large (i.e. the joint histogram needs to be representative enough);
(2) to make the process less time consuming, that is, less terms in the sum of (3); and
(3) to provide the method robustness against intensity noise.

In our implementation, given that the size of the image patches is usually in the range between some hundreds and a few thousands (triangles too small are previously discarded from the initial mesh provided by the Delaunay triangulation), we have used 16 gray-level bins. An alternative way of overcoming the problem of small triangles is to use non-parametric estimation methods as the Parzen window, though it entails a higher computational cost (reader may refer to [15] for more detail about this approach).

Observe from (4) that, when all the intensity values of the two images are independent one from another (that is, without correlation) the argument of the logarithm becomes one, and the MI achieves its minimum at zero (MI is always non-negative). To illustrate how the joint histogram captures the idea of statistical dependency, Fig. 5 shows the joint histograms of two pairs of equal-sized, synthetic images of just four gray-levels each. When one of them is rotated, the joint histogram exhibits more dispersion than when they overlap perfectly.

Finally, with the purpose of illustrating the performance of the MI in comparison to the NCC we have conducted the two experiments shown in Fig. 6. In these experiments we evaluate the similarity of two image quadrilaterals of a synthetic cube sensed from different points of view using both MI and NCC. The image triangles are given by the two possible topological configurations of the vertices $\{1,2,3,4\}$. In the first experiment (Fig. 6(a)), the cube sides have been randomly coloured, whereas in the second one (Fig. 6(b)), the position of the cube respect to the illumination source

[2]Notice that the requirement that the two image patches have the same size always holds since the moving image triangle is mapped to the reference one, having then the same number of pixels.

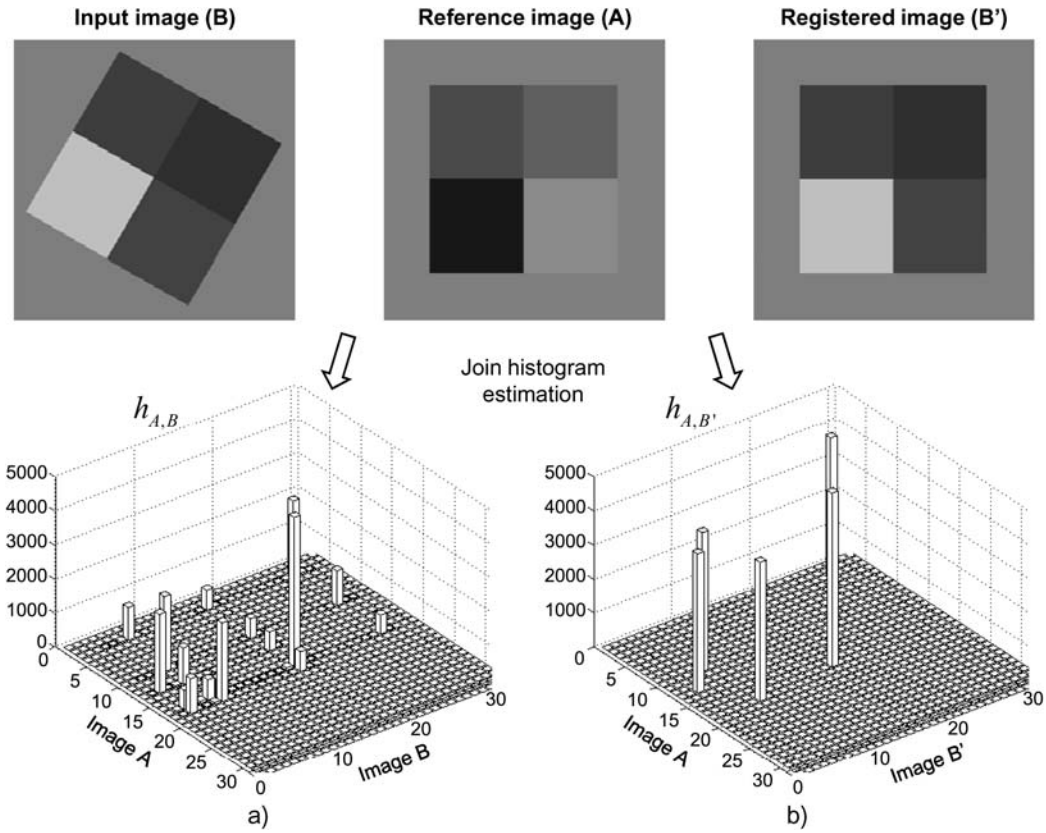

Fig. 5 Joint histograms of two pairs of synthetic images: (**a**) a misaligned pair, and (**b**) a perfectly overlapped pair. Observe that the joint histogram (of 32 gray-levels) presents less dispersion when the images are aligned. Notice that this fact is independent of the intensity values of the two images being identical or not

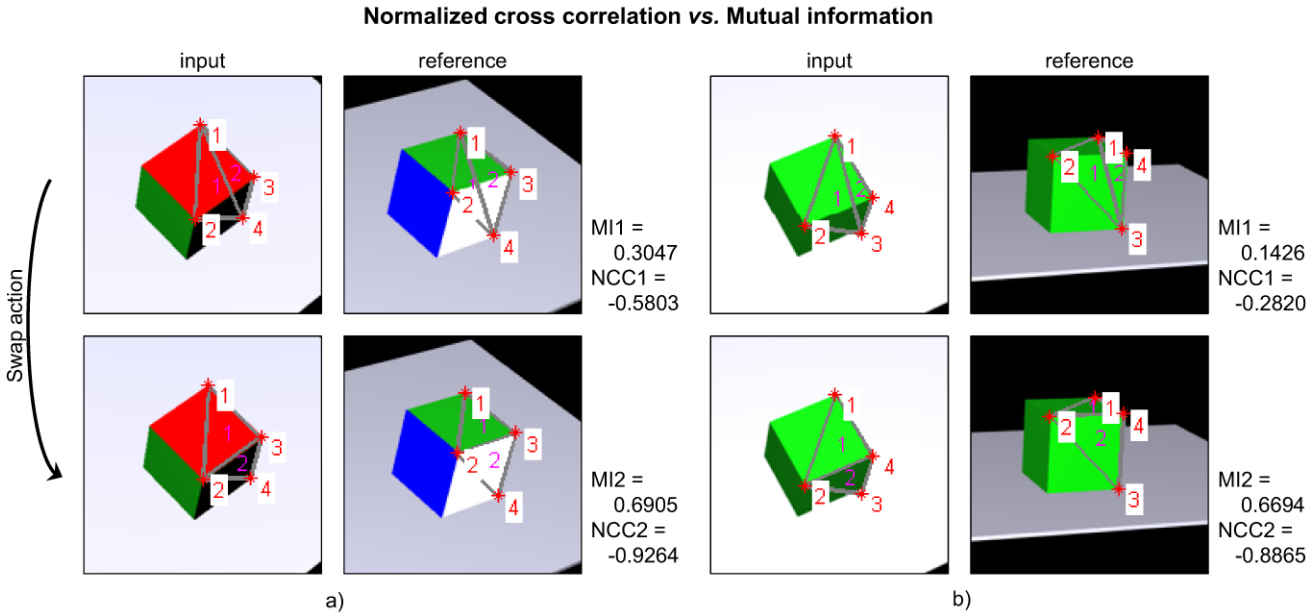

Fig. 6 Two experiments illustrating the suitability of the Mutual Information (MI) as a consistency measure for image registration. The figures show the MI in comparison to the Normalized Cross-Correlation (NCC) when (**a**) the object is coloured in a different manner and (**b**) the illumination of the scene is changed. The measures are computed from the image triangles given by the two possible topological realizations of the vertices $\{1, 2, 3, 4\}$. Observe how the NCC fails in both experiments, giving lower values in topological configurations compatible with the scene. On the contrary, the MI delivers a clear improvement in both cases

has been changed. We see how the MI outputs a larger value when the topological configuration is 3D-compatible, while the NCC fails. These simple examples show the effectiveness of the MI when applied to image pairs with non-functional radiometric changes, in contrast to NCC, which can manage image intensity differences but only if they follow a linear function (that is, intensity shift and/or contrast scaling).

4.2 Checking Edges to Swap

We take advantage of the robustness of the MI for checking the 3D-compatibility of the mesh edges. Thus, given the images I and I' to register and their corresponding meshes defined by $M = (K, V)$ and $M' = (K, V')$, we determine the 3D-compatibility of an edge $\{i, j\} \in K$ by computing the registration consistency of the quadrilaterals $s = \text{quad}(\{i, j\}, K)$ and $\hat{s} = \text{quad}(\{k, l\}, \hat{K})$, that is, before ($\omega_s$) and after ($\omega_{\hat{s}}$) the edge swap, respectively (see Fig. 4). Formally, that improvement is measured by:

$$\Delta\omega(\{i,j\}) = \underbrace{MI(I(r), I'(\boldsymbol{f}_{\hat{s}}(r)))}_{\omega_{\hat{s}}} - \underbrace{MI(I(r), I'(\boldsymbol{f}_{s}(r)))}_{\omega_s} \tag{7}$$

where r refers to the pixels contained in $\phi_V(|s|)$ (or $\phi_V(|\hat{s}|)$, since both geometrical realizations are identical). Thus, $I(r)$ represents the quadrilateral region of the reference-image defined by s, and $I'(\boldsymbol{f}_s(r))$ and $I'(\boldsymbol{f}_{\hat{s}}(r))$ the transformations of its input-image counterparts according to the two possible topological configurations s and $\hat{s}$, before and after the edge swap, respectively.

Before evaluating the 3D-compatibility of any edge, say $\{i, j\} \in K$, it must be checked if it verifies the following preconditions:

(1) $\{i, j\}$ must be an internal edge,
(2) the resultant swapped edge is a new one ($\{k, l\} \notin K$), and
(3) the action does not produce a patch reversal in $\hat{K}$. A patch reversal is a mesh inconsistency produced when the shared edge of two adjacent faces, which make up a concave quadrilateral, is swapped (see Fig. 7).

Once these preconditions are met, the swap of the edge is accepted if $\Delta\omega(\{i, j\}) > 0$, that is, when it leads to some increase in the MI.

In practice, we only accept an action to be applied if the increase is above a given threshold δ. The aim of this threshold is to prevent the application of actions on quads that lie on projections of planar surfaces and, because of the image resampling and the computational errors, may yield small consistency differences that should not be considered as true improvements. As consequence of this threshold, actions that entail real registration improvements less than δ will not be detected.

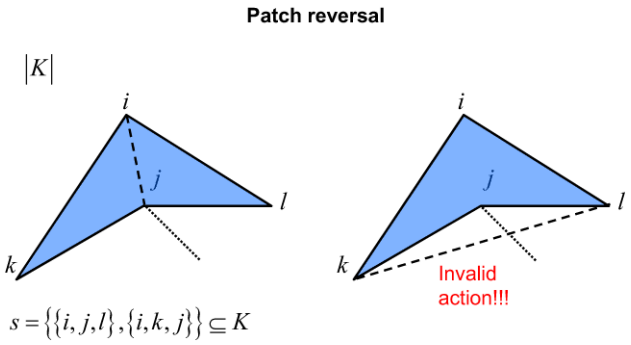

Fig. 7 A patch reversal produced by a edge swap action in a concave quadrilateral

The estimation of the MI (and so, the estimation of $\Delta\omega$) is sensitive to the number samples (i.e. the number of pixels in the quadrilateral), to the number of output bins, and, in the case of the registration process, to the resampling function used (i.e. nearest neighbour, bilinear, or bicubic); since it does not exist an analytical expression that relates $\Delta\omega$ with these parameters, we have experimentally analysed the behaviour of $|\Delta\omega|$ when applied to quadrilaterals that lie on projections of planar surfaces in order to properly choose the value of δ (the reader can find a brief description of this experiment in Fig. 8). From this study, we have verified that $|\Delta\omega|$ keeps below 0.008. According to this result, we have set the value of δ to 0.01.

4.3 Mesh Optimization

The overall optimization process can be formulated as a greedy search [44], which starts with two corresponding triangular meshes M and M' resulting, for example, from the application of a Delaunay's triangulation method over a set of the conjugate points identified in both images (either manually or by automatic methods [45]). Though greedy search may fall in local maximums, its computational cost is significatively lower than other optimization alternatives based on genetic algorithms or simulated annealing.

Formally, the optimization process can be expressed as finding the simplicial complex $\hat{K}$ that maximizes the registration consistency of the whole images, that is

$$\hat{K} = \arg\max_{K} MI(I(m), I'(\boldsymbol{f}_K(m))) \tag{8}$$

where $m = \phi_V(|K|)$ represents all image pixels within the mesh given by K.

To generate the two isomorphic meshes $M = (K, V)$ and $M' = (K, V')$:

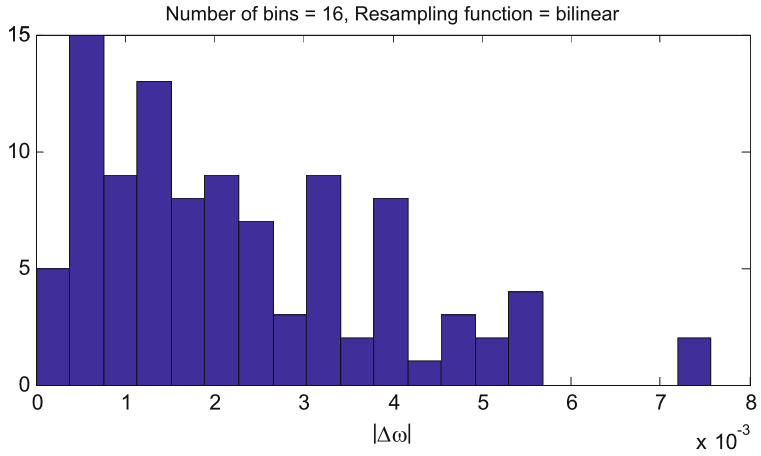

Fig. 8 Histogram of the $|\Delta\omega|$ values obtained from swap actions on planar surfaces, which ideally should output a zero value. To accomplish the experiment, we have selected 100 pairs of image quads of planar surfaces (20 of them manually selected and the other 80 ones obtained by modifying the original size of the selected ones); next, the value of $|\Delta\omega|$ is computed for each pair, leaving the number of bins and the resampling function fixed. Notice that these parameters do not change along the optimization process, so we only analyse the influence of the number of samples on the estimation of the MI

(1) the reference-image point set V is triangulated by means of some triangulation technique, then, the topological $|K|$ and geometrical realization $\phi_V(|K|)$ are generated;
(2) the conjugate mesh $\phi_{V'}(|K|)$ is generated by mapping $|K|$ to the corresponding input-image point set V'. Though $\phi_V(|K|)$ is an embedding, $\phi_{V'}(|K|)$ may not be, since self-intersections in $\phi_{V'}(|K|)$ (so called patch reversals) may appear because of occlusions or large camera displacements (see Fig. 9(a)). Therefore, the assumption that the expression (1) is a one-to-one mapping is not true. To overcome such inconsistencies, we analyze the initial topology applying the following modifications:
   (a) if one of the edges of the patch reversal is external, the triangle that contains it is removed, as shown in Fig. 9(b);
   (b) if, on the contrary, none of the edges is external, we swap the shared edge of the two affected triangles, as shown in Fig. 9(c).

Once the mesh has been checked for patch reversals, each edge of the mesh is analyzed following the greedy search depicted in Algorithm 1. It starts by creating a sorted list (in descending order) of the improvement in registration consistency $\Delta\omega$ for all the edges of the mesh. Notice that for properly sorting the list, the values of MI in (7) must be normalized, so, we employ the normalized mutual information (NMI) (reader may refer to [14] for other possible MI normalized variants), mathematically:

$$NMI(A,B) = \frac{MI(A,B)}{H(A,B)} \tag{9}$$

This list is computationally expensive to generate, but this is done just once, at the beginning. At each iteration of the optimization process, the first edge of the current list is swapped, and the list is updated by checking only those edges affected by the swapping action (its *boundary*). It is clear that such improvement in the local consistency leads to an improvement in the global one. The algorithm stops when the list is empty, that is, when all the mesh edges have been explored and no further improvement can be achieved by swapping actions.

Unlike other optimization techniques employed in 3D scene reconstruction, as the random search formulated in [20], this procedure guarantees the iterative improvement of the image registration consistency up to the degree that the geometrical realization of the mesh tolerates. Notice that, without relocating the vertices and, possibly, introducing additional ones, the mesh may be not good enough to completely avoid 3D-incompatible edges. Figure 10 illustrates the proposed optimization method when applied to an initial topological configuration containing several 3D-incompatible edges.[3]

5 Experimental Results

This section shows some experimental results that illustrate the performance of our approach. We compare it to other two methods [17, 20], which also employ both swap actions for exploring the search space and image similarity functions

[3] Illustrative videos can also be downloaded from http://babel.isa.uma.es/mapir/index.php/theoretical/46-mesh-optimization-paper.

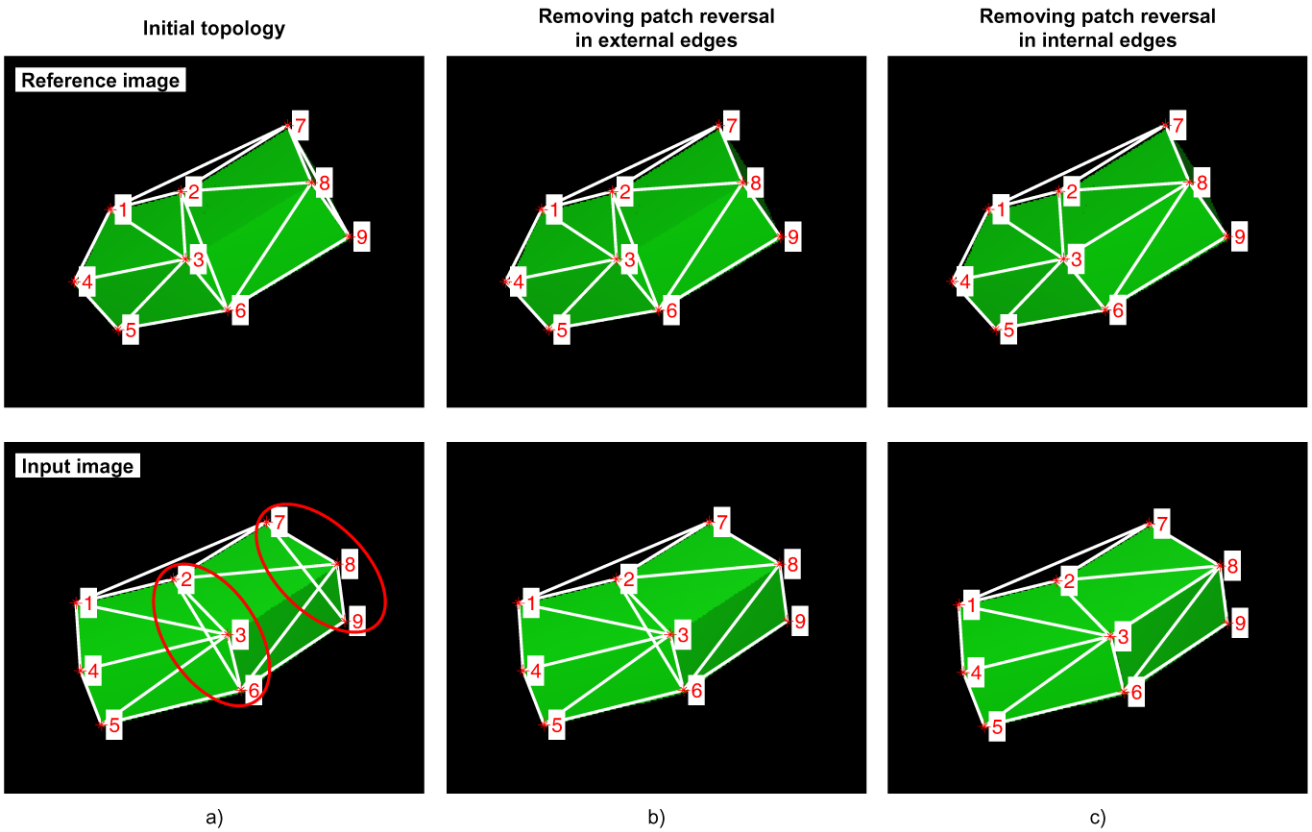

Fig. 9 Topological changes proposed for solving the patch reversals in the initial meshes. (**a**) The initial meshes and the two situations where patch reversal occurs. (**b**) Solution for a patch reversal of an external edge: the triangle that contains the external edge $\{7, 9\}$ is removed. (**c**) Patch reversal of an internal edge: the internal edge $\{2, 6\}$ shared by the two affected triangles is swapped

for driving the search process. One is the work by Morris and Kanade [17] which aims at detecting inconsistencies in a 3D mesh by measuring the similarity of the image patches that results from projecting it onto the images. It applies a greedy search driven by the sum of square differences of the whole images (a global approach). The other one is the work by Nakatuji et al. [20] which proposes a random search that pursues the refinement of the topological realization of two conjugate 2D meshes for an optimal 3D image reconstruction. For detecting 3D-incompatible edges the authors employ a square template of fixed size that is correlated with the image patches (a local approach). Since the implementation of these methods are not available online we have implemented them in Matlab following the indications of both papers as faithfully as possible. We have also included a comparison of the proposed method using both MI and NCC-based cost functions.

5.1 Datasets and Methodology

We have employed images belonging to the ALOI library [46], which includes real images of a broad variety of objects, as well as synthetic images generated from VRML models and real images of urban scenes (e.g. building facades). The purpose of selecting this diversity of images is to test the method under different types of illumination, image contents, and observation poses.

The conjugate points (CP) that define the geometrical realization of the mesh have been obtained in two ways: manually and automatically. For the latter the following procedure has been implemented: the Harris corner detector [47] identifies distinctive feature points in the reference image, and then the Lucas-Kanade feature tracker [48] detects their corresponding points in the input image. For both, the manual and automatic case, given the set of original conjugate points, the affine epipolar geometry of the two images was robustly estimated applying the MAPSAC algorithm [49], which allowed us to discard spurious pairs.

Although consistently matched, some of the identified CPs give rise to 3D-incompatible edges once the Delaunay's triangulation algorithm [12] is applied on them (as in Fig. 1). The objective of our optimization method is to correct, as much as possible, those situations. Whether this improvement is achieved or not is assessed in two different ways:

Algorithm 1 Given two images I and I' to register, and two initial triangular meshes on them defined by $M = (K, V)$ and $M' = (K, V')$, determine a new topological realization by iteratively swapping edges which improve the consistency of PWL image registration.

```
//Build a sorted list (indexed by edge) with the Δω of each edge
Δω_list ⇐ ∅
for all {i, j} ∈ K do
    if {i, j} verifies the preconditions then
        Δω_list[{i, j}] ⇐ Δω({i, j}) //from expression (7)
    end if
end for
sort Δω_list in descending order

//Iterate while there exist an edge swap that improves the consistency
while the first element of Δω_list > δ do
    swap its corresponding edge, say {i, j} ∈ K by {k, l} ∈ K̂
    Δω_list[{i, j}] ⇐ ∅ //remove {i, j} from the list
    //Update Δω_list with the Δω of the boundary edges of {i, j}
    for all {m, n} ∈ bound({i, j}, K) do
        if {m, n} verifies the preconditions then
            Δω_list[{m, n}] ⇐ ω({m, n}) //from expression (7)
        end if
    end for
    sort Δω_list in descending order

    //Update de topological realization
    K ⇐ K̂
end while
```

(a) by evaluating the goodness of the image registration, that is, by measuring the MI of the complete reference and registered images (expression (8)), and
(b) by checking if the unscaled 3D scene reconstructed from the two resulting meshes is more accurate than that obtained from the initial ones. In the event that 3D scene structure was known, it could be used for that evaluation. Otherwise, as it is the case here, we make the evaluation by visual inspection.

Figure 11(a–f) shows some of the test images employed in this work, as well as the initial meshes generated from the identified CP sets (automatically in the cases (c, f), and manually for the rest).

5.2 Results

Figures 11(g–l) show the final meshes obtained when running our method for the image pairs shown in Fig. 11(a–f). The effectiveness of the method in these experiments is demonstrated in Fig. 12, and Table 1. In the first, the improvement in the global (whole image) registration consistency is displayed along the different swapping actions. Table 1 shows the percentage of 3D-compatible edges (determined by visual inspection) which are not boundary edges, that is, the mesh *correctness*. The algorithm stops, in less than 25 iterations for all these cases, when all the mesh edges have been explored and no further improvement can be done by applying swap actions.

Notice that, during the optimization process, there are actions that apparently do not improve the image registration consistency (revealed as small flat stretches in the curves of Fig. 12(b, c, f)). We say "apparently", since in fact such an improvement exists (otherwise the action is not applied) but its contribution to the global registration consistency is small. According to the Algorithm 1, despite producing a small improvement, these actions are the best candidates at each iteration, leading to topological configurations that in subsequent iterations substantially improve the consistency. This situation is typical in those configurations where several adjacent edges are 3D-incompatible, as for example, the image pair shown in Fig. 11(b).

Table 1 summarizes the correctness, and the computational time for each of the compared methods. We can see that the performance of the three methods decrease for the same image pair, particularly, the method of Morris and Kanade, which employs the SSD of the whole images as cost function (less robust to illumination changes than MI). Another drawback of their approach is its high computational

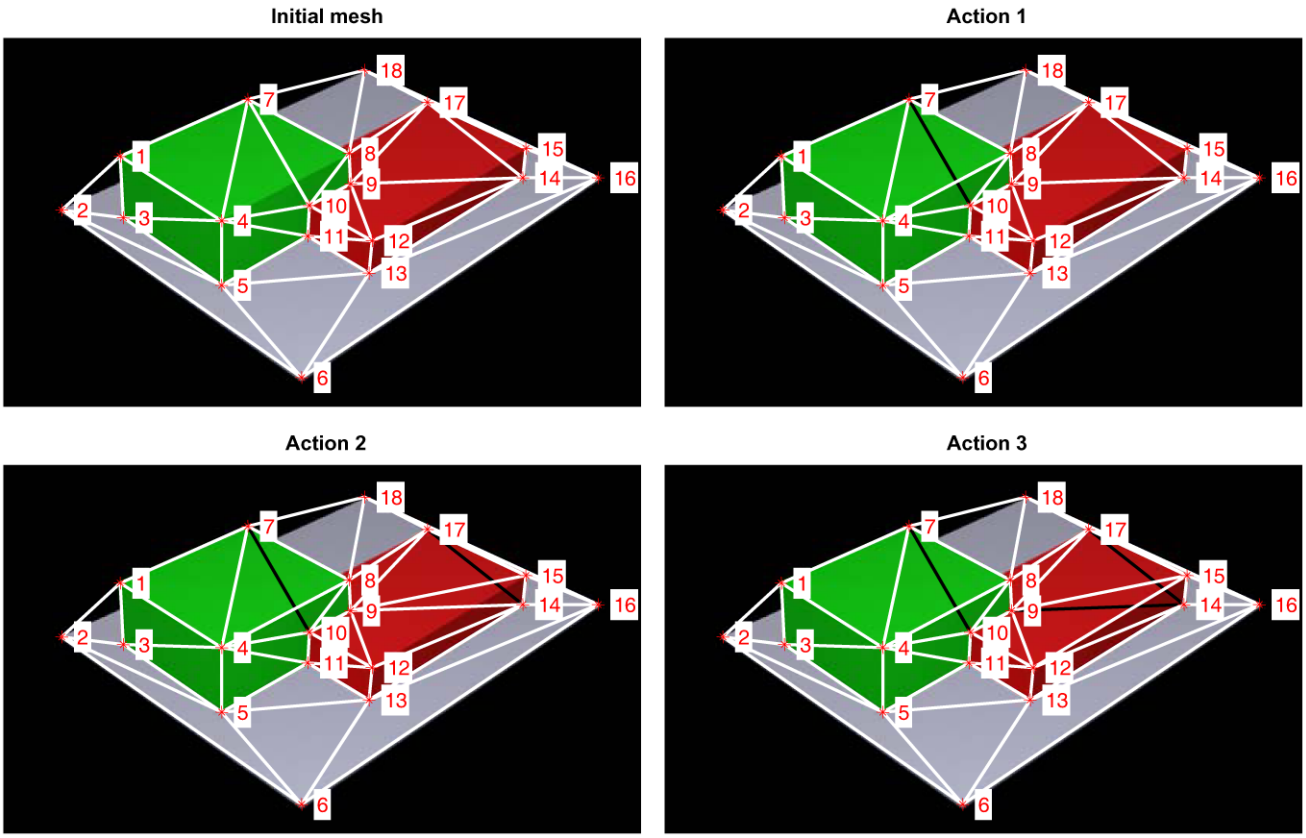

Fig. 10 Illustration of the optimization process. The initial mesh shows a topological configuration containing several 3D-incompatible edges (edges {9, 14}, {14, 17} and {7, 10}). As depicted in the Algorithm 1, when the edge {14, 17} is swapped (second action), its adjacent edges are considered again for swapping, which leads to the edge {9, 14} to be swapped (third action)

cost, with prohibitive times for large number of edges (as for example, for the image pair of Fig. 11(c)). A similar behaviour is observed in the method of Nakatuji et al., where the cost of computing the affinities for mapping the corresponding image patches to the template also slows down the optimization process. When the conjugate points do not lie on the vertices or edges of the polyhedral scene (which happens in practice), the methods of Morris and Kanade and Nakatuji et al. present an important decrease of their effectiveness: the number of actions significatively grows and the correctness decreases. This fact can be clearly observed in the results for the image pair shown in Fig. 11(c).

It should be remarked the robustness of our method to changes in illumination, as can be appreciated in the results for the image pair of the Fig. 11(b), a pair of urban scenes acquired under very different lighting conditions. These changes go also unnoticed for the NCC variant of the proposed method (see the column 2 of the Table 1). The computational cost (per action) of both variants are quite similar.

Finally, with the aim of illustrating the possibility of generating consistent 3D reconstructions from a pair of registered conjugate images, we have reconstructed an unscaled 3D surface of the sensed scene by projecting back the two meshes[4] (see Fig. 13). Apart from its interest in 3D scene reconstruction, it allows us to check the effectiveness of the proposed procedure by contrasting the 3D models associated to the initial and refined meshes. Our analysis has limited to visually contrasting the initial and optimal 3D reconstructions.

6 Conclusions

Image registration is an essential step in a broad variety of image processing applications where the final result comes from the combination of several sources, as for example change detection, image fusion, 3D scene reconstruction, etc.

In this paper we have proposed a technique for automatically optimizing the conjugate triangular meshes employed by a piecewise-linear registration process: having more suitable meshes means that the registration is more accurate. To

[4]For example, by applying the factorization algorithm for affine reconstruction proposed in [41] (p. 437).

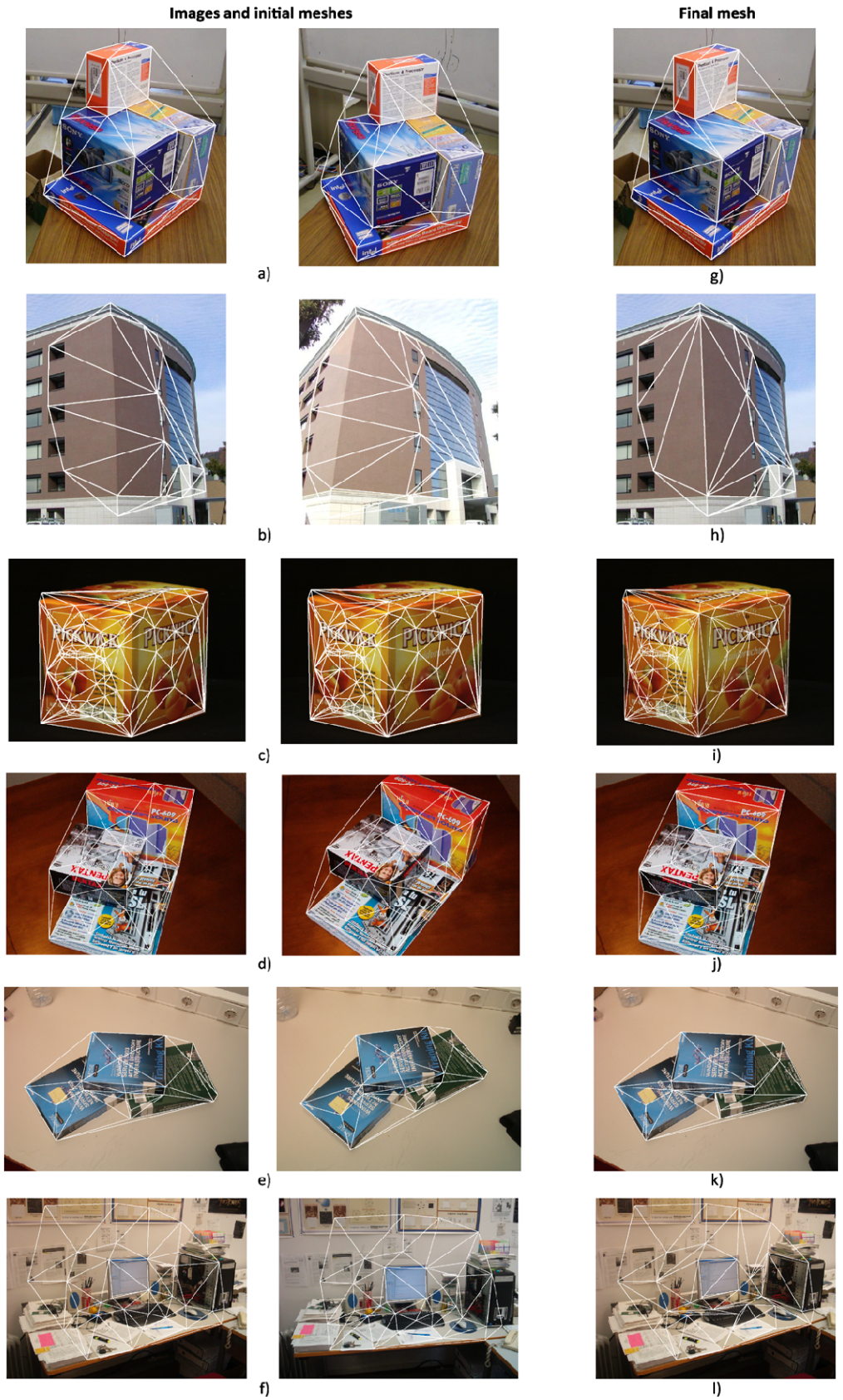

Fig. 11 (**a**–**f**) Pairs of real images of polyhedron scenes and their corresponding Delaunay triangular meshes. (**g**–**l**) Optimized triangular meshes provided by our method. Observe how the proposed process swaps those edges which go from one plane surface to another

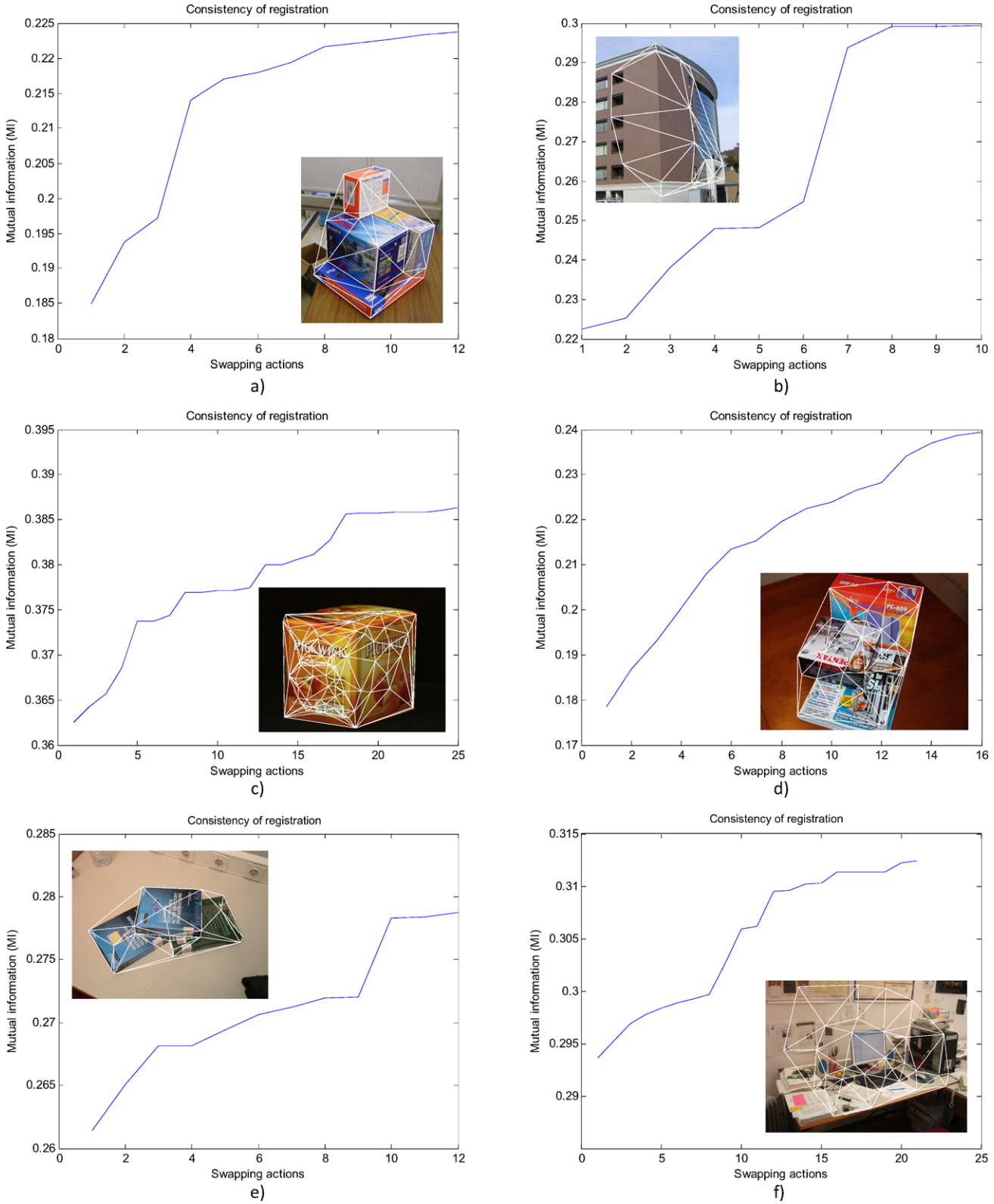

Fig. 12 Results of the experimental tests for each of the image pairs shown in Fig. 11(a–d)). The plots show a significant improvement in the registration consistency for all analyzed pairs. Notice that in plots (**b**), (**c**), and (**f**) there are some iterations where the registration consistency seems to remain steady after applying a swapping action. This is because the swapped edges lie on projections of almost-plane surfaces of the scene

Table 1 Mesh data, number of applied actions (*#A*), computational time (*T*), and correctness (*%C*) for the image pairs in Fig. 11

Test[a]		MI-based method			NCC-based method			Morris and Kanade			Nakatuji et al.		
Fig. 11	# edges	*#A*	*T* (sec)	*%C*	*#A*	*T* (sec)	*%C*	*#A*	*T* (sec)	*%C*	*#A*	*T* (sec)	*%C*
a	56	12	1.20	100	9	1.08	100	12	9.59	100	15	36.31	97.8
b	47	10	0.94	97.30	9	0.89	94.5	12	8.25	67.56	10	30.16	94.5
c	275	25	2.78	100	34	2.77	99.25	105	115.22	99.62	52	211.14	96.64
d	65	16	2.08	98.27	13	1.92	93.10	17	33.36	98.27	12	37.75	84.48
e	46	12	1.22	94.73	7	0.59	84.21	15	9.14	84.21	10	26.95	78.95
f	88	20	1.56	84.42	15	1.31	77.92	17	34.61	83.12	17	57.95	81.82

[a]We have employed Matlab R2008a on a Intel Core 2 Quad Q6600 for implementing the tests

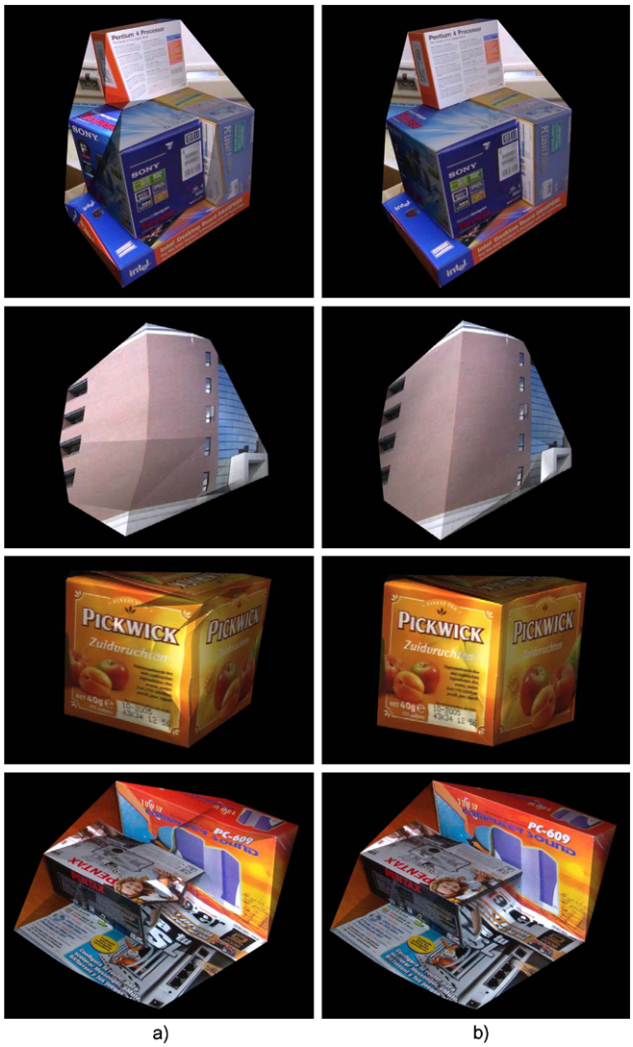

Fig. 13 3D scene reconstructions generated from two pairs of conjugate meshes: (**a**) the initial meshes and (**b**) the refined ones. In plots (**a**) we can observe some artifacts (e.g. broken lines), in places where 3D-incompatible edges exist. These artifacts disappear when these edges are conveniently swapped, as shown in plots (**b**)

achieve that, we iteratively modify the connectivity of both meshes through edge swapping actions. The function employed for evaluating the edge to be swapped is based on the mutual information, which is notoriously more robust than other well-known metrics such as NCC or SSD, since it is less sensitive to changes in lighting conditions or noise. The optimization procedure is formulated as a greedy search which finishes when the mesh topology can no longer be refined, that is, when all mesh edges have been successfully checked and no further improvement is possible through edge swap actions.

The proposed method has been successfully tested with different image pairs of urban scenes, polyhedral objects, both real and synthetic, which have been acquired from different angles and/or under different lighting conditions. The method outperforms two previously published approaches, which also employ swapping actions for exploring the search space and image similarity functions for driving the search process.

References

1. González, J., Ambrosio, G., Arévalo, V.: Automatic urban change detection from the IRS-1D PAN. In: IEEE-ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas, Rome, Italy, 2001, pp. 320–323
2. Bookstein, F.: Principal warps: Thin-plate-splines and the decomposition of deformations. IEEE Trans. Pattern Anal. Mach. Intell. **11**(6), 567–585 (1989)
3. Goshtasby, A.: Piecewise linear mapping functions for image registration. Pattern Recogn. **19**(6), 459–466 (1986)
4. Goshtasby, A.: Piecewise cubic mapping functions for image registration. Pattern Recogn. **20**(5), 525–533 (1987)
5. Ehlers, M., Fogel, D.: High-precision geometric correction of airborne remote sensing revisited: The multiquadric interpolation. In: Image and Signal Processing for Remote Sensing, vol. 2315, pp. 814–824. SPIE, Bellingham (1994)
6. Kybic, J., Unser, M.: Fast parametric elastic image registration. IEEE Trans. Image Process. **12**(11), 1427–1442 (2003)
7. Zitová, B., Flusser, J.: Image registration methods: A survey. Image Vis. Comput. **21**(11), 977–1000 (2003)
8. The Mathworks: http://www.mathworks.com
9. Image Fusion Systems Research: Transformation functions for image registration. IFSR Technical Report (2003). http://www.imgfsr.com/ifsr_tf.pdf. Accessed 27/04/2007

10. ITT Visual Information Solutions: http://www.ittvis.com/envi
11. Leica Geosystems: http://gi.leica-geosystems.com
12. Shewchuk, J.: Lecture notes on Delaunay mesh generation. Tech. Rep. 3, University of California at Berkeley (1999)
13. Cormen, T., Leiserson, C., Rivest, R., Stein, C.: Introduction to Algorithms, 2nd edn. MIT Press, Cambridge (2001)
14. Maes, F., Collignon, A., Vandermeulen, D., Marchal, G., Suetens, P.: Multimodality image registration by maximization of mutual information. IEEE Trans. Med. Imag. **16**(2), 187–198 (1997)
15. Viola, P., Wells, W.: Alignment by maximization of mutual information. Int. J. Comput. Vis. **24**(2), 137–154 (1997)
16. Perrier, J., Agin, G., Cohen, P.: Image-based view synthesis for enhanced perception in teleoperation. In: Enhanced and Synthetic Vision, vol. 4023, pp. 332–338. SPIE, Bellingham (2000)
17. Morris, D., Kanade, T.: Image-consistent surface triangulation. In: IEEE International Conference on Computer Vision and Pattern Recognition (CVPR'00), Hilton Head, SC, USA, 2000, pp. 332–338
18. Lehner, B., Umlauf, G., Hamann, B.: Image compression using data-dependent triangulations. In: 3rd International Symposium on Visual Computing (ISVC'07), Part I 4841, pp. 351–362 (2007)
19. Yu, X., Bryan, B., Sederberg, T.: Image reconstruction using data-dependent triangulation. Comput. Graph. Appl. **21**(3), 62–68 (2001)
20. Nakatuji, A., Sugaya, Y., Kanatani, K.: Mesh optimization using an inconsistency detection template. In: IEEE International Conference on Computer Vision (ICCV'05), vol. 2, Beijing, China. 2005, pp. 1148–1153
21. Servais, M., Vlachos, T., Davies, T.: Bi-directional affine motion compensation using a content-based, non-connected, triangular mesh. In: European Conference on Visual Media Production (CVMP'04), London, UK, 2004, pp. 49–58
22. Yip, B., Jin, J.: Image registration using triangular mesh. Adv. Multimed. Inf. Process. **3331**, 298–303 (2004)
23. Schlesinger, D., Flach, B., Shekhovtsov, A.: A higher order MRF-model for stereo-reconstruction. Pattern Recogn. **3175**, 440–446 (2004)
24. Matuszewski, B., Shen, J., Shark, L.: Elastic image matching with embedded rigid structures using spring-mass system. In: IEEE International Conference on Image Processing, vol. 2, pp. 937–940 (2003)
25. Hoppe, H.: Progressive meshes. Comput. Graph. **30**, 99–108 (1996) (Annual Conference Series)
26. Hoppe, H., DeRose, T., Duchamp, T., McDonald, J., Stuetzle, W.: Mesh optimization. Comput. Graph. **27**, 19–26 (1993) (Annual Conference Series)
27. de Bruin, P., van Meeteren, P., Vos, F., Vossepoel, A., Post, F.: Accurate and high quality triangle models from 3D grey scale images. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI'02), Part II 2489, pp. 348–355 (2002)
28. Vogiatzis, G., Torr, P., Cipolla, R.: Bayesian stochastic mesh optimisation for 3D reconstruction. In: British Conference on Machine Vision (BMVC'03), Norwich, UK, 2003, pp. 711–718
29. Surazhsky, V., Gotsman, C.: High quality compatible triangulations. In: Proceedings of 11th International Meshing Roundtable, pp. 183–192 (2002)
30. Surazhsky, V., Gotsman, C.: High quality compatible triangulations. Eng. Comput. **20**(2), 147–156 (2004)
31. Wang, Y., Wang, C.C.L., Yuen, M.M.F.: Duplicate-skins for compatible mesh modelling. In: SPM '06: Proceedings of the 2006 ACM Symposium on Solid and Physical Modeling, pp. 207–217. ACM, New York (2006)
32. Battiato, S., Gallo, G., Messina, G.: SVG rendering of real images using data dependent triangulation. In: Computer Graphics (SCCG'04), 20th Spring Conference on, pp. 185–192. ACM, New York (2004)
33. Schumaker, L.: Computing optimal triangulations using simulated annealing. Comput. Aided Geom. Des. **10**(3–4), 329–345 (1993)
34. National Library of Medicine Insight Segmentation and Registration Toolkit: http://www.itk.org
35. Cole-Rhodes, A., Johnson, K., LeMoigne, J., Zavorin, I.: Multiresolution registration of remote sensing imagery by optimization of mutual information using a stochastic gradient. IEEE Trans. Image Process. **12**(12), 1495–1511 (2003)
36. Chen, H., Varshney, P., Arora, M.: Performance of mutual information similarity measure for registration of multitemporal remote sensing images. IEEE Trans. Geosci. Remote Sens. **41**(11), 2445–2454 (2003)
37. Xie, H., Pierce, L., Ulaby, F.: Mutual information based registration of SAR images. In: IEEE International Conference on Geoscience and Remote Sensing Symposium (IGARSS'03), vol. 6, pp. 4028–4031 (2003)
38. Inglada, J., Giros, A.: On the possibility of automatic multisensor image registration. IEEE Trans. Geosci. Remote Sens. **42**(10), 2104–2120 (2004)
39. Kern, J., Pattichis, M.: Robust multispectral image registration using mutual-information models. IEEE Trans. Geosci. Remote Sens. **45**(5), 1494–1505 (2007)
40. Xu, G., Zhang, Z.: Epipolar geometry in stereo, motion, and object recognition: A unified approach, 1st edn. Kluwer Academic, Norwell (1996)
41. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision, 2nd edn. Cambridge University Press, Cambridge (2004)
42. Spanier, E.: Algebraic Topology, 1st edn. McGraw-Hill, New York (1966)
43. Cover, T., Thomas, J.: Elements of Information Theory, 1st edn. Wiley, New York (1991)
44. Strang, G.: Introduction to Applied Mathematics, 1st edn. Wellesley-Cambridge Press, Wellesley (1986)
45. Arévalo, V., Gonzalez, J.: An experimental evaluation of non-rigid registration techniques on QuickBird satellite imagery. Int. J. Remote Sens. **29**(2), 513–527 (2008)
46. Geusebroek, J., Burghouts, G., Smeulders, A.: The Amsterdam library of object images. Int. J. Comput. Vis. **61**(1), 103–112 (2005)
47. Harris, C., Stephens, M.: A combined corner and edge detector. In: Alvey Vision Conference, vol. 4, Manchester, UK, 1988, pp. 147–151
48. Bouguet, J.: Pyramidal implementation of the Lucas-Kanade feature tracker. Description of the algorithm. Tech. rep., Intel Corporation, Microprocessor Research Labs (1999). http://robots.stanford.edu/cs223b04/algo_tracking.pdf. Accessed 26/06/2006
49. Torr, P.: A structure and motion toolkit in Matlab: Interactive adventures in S and M. Tech. Rep. MSR-TR-2002-56, Microsoft Research (2002). ftp://ftp.research.microsoft.com/pub/TR/TR-2002-56.ps. Accessed 27/03/2007

Javier González received the B.S. degree in Electrical Engineering from the University of Seville in 1987. He joined the Department of "Ingeniería de Sistemas y Automática" at the University of Malaga in 1988 and received the Ph.D. from this University in 1993. In 1990–1991 he was at the Field Robotics Center, Robotics Institute, Carnegie Mellon University (USA) working on mobile robots. Since 1996 he has being heading Spanish and European projects on mobile robotics and perception. Currently, he is a full professor at the University of Málaga and leads the Machine Perception and Intelligent Robotics (MAPIR) group. His research interest includes mobile robot navigation, olfactory robotics and computer vision. In these fields he has published three books and more than 100 papers.

Vicente Arévalo received the M.S. and Ph.D. degrees in Computer Science from the University of Málaga, Málaga, Spain, in 2001 and 2008, respectively. Since 2003, he has been with the Department of "Ingeniería de Sistemas y Automática" at the University of Malaga. He was granted by the Spanish Government as Predoctoral Researcher in 2003, getting the position of Assistant Professor in 2009. His research mainly focuses on mobile robots, computer vision, and remote sensing, being the author of more than 20 journal and conference papers.