ROBUST EXTRACTION OF CONTROL-POINT PAIRS FOR REGISTERING HIGH-RESOLUTION SATELLITE IMAGES

Vicente Arévalo and Javier González-Jiménez

MAchine Perception and Intelligent Robotics (MAPIR) research group
System Engineering and Automation Department, University of Málaga

Accurate registration of high-resolution images requires the extraction of an important number of reliable control-point pairs distributed over the entire images and localized with a high precision. The possibility of automatically extracting them from images is of notorious significance, since commercial packages only provide the user with some tools that assist him/her when carrying out their manual identification. In this paper we proposed a method to robustly and accurately extracting such matches in a full automatic way. This method consists of two steps: first, an initial set of control point pairs is identified in both images following a coarse-to-fine approach, and second, these pairs are processed in order to discard outliers that do not fit to the affine epipolar geometry inherent to the two views. To gain also in precision, the positions of remaining pairs (inliers) are refined on the basis of a Maximum Likelihood error minimization. As demonstrated in the paper through experimental results using panchromatic QuickBird images (0.6 meter/pixel), our approach is able to improve the registration accuracy up to 45% percent.

Keywords: Control point extraction; Image registration; High-resolution satellite image; Affine epipolar geometry; Ikonos; QuickBird.

1. Introduction

Image registration is the process of spatially aligning two images of the same scene acquired on different dates (multitemporal analysis), from different viewpoints (multiview analysis) and/or using different sensors (multimodal analysis). In this process, one image remains without modification (the *fixed* image) whereas the other one (the *moving* image) is spatially transformed until fitting with the fixed one. Image registration is a crucial step in those image analysis applications where the final result comes from the association of several data sources, for example, image fusion, change detection, 3D scene reconstruction, etc.

Most registration techniques applied within the remote sensing field deal with the registration process by means of a mapping function estimated from a set of representative pairs of control points (CP) identified in both images (Gonzalez *et al.*, 2001; Zitová & Flusser, 2003). Defining precise corresponding points in both images may require the user to spend a considerable amount of time in those situations where hundred of control points are necessary for capturing the geometric distortions, especially in high-resolution satellite images, as revealed in (Arévalo & González, 2006). To facilitate this process, current commercial packages include specific tools to assist their identification, but fail when the images present important radiometrical and geometrical differences. It is clear, then, the convenience of having *automatic* software tools to robustly and precisely extract pairs of CP despite the possible differences between the images, which is the main contribution of this paper.

In satellite remote sensing, an important simplification about the scene geometry can be made: the perspective projection, which is a non-linear mapping, can be approximated by a paraperspective one (so-called affine projection), which is a linear transformation. This simplification leads to a great reduction of computation in many vision problems and provides certain *a priori* knowledge about the scene geometry (Xu & Zhang, 1996). This work takes advantage of this fact in order to make more efficient and robust the automatic detection of corresponding control points, given a pair of high resolution images. More precisely, our method applies a corner detector to identify key points in the fixed image and a feature tracker that searches for their corresponding points in the moving one. The matcher that we propose consists of an enhanced extension of the well-known Lukas-Kanade feature tracker (Lucas & Kanade, 1981). To uniformly distribute the control point pairs throughout the images, the identification of keypoints in the fixed image is accomplished by selecting a point from every cell of a rectangular grid arranged on it. Robustness in the process is achieved by recovering the affine epipolar geometry of the two images to register. In particular, we exploit such information to discard the matches not consistent with the estimated 3D viewing geometry (labelled as outliers) and to accurately re-estimate the localization of the rest (inliers).

The proposed method has been successfully tested with panchromatic QuickBird images (0.6 meter/pixel) acquired on different seasons, which leads to significant radiometric profiles, and from different acquisition poses, which in combination with a high-relief terrain introduces important geometric differences between the images.

The remainder of this paper is organized as follows. In section 2, we state and justify the assumption of paraperspective camera projection for satellite images. In section 3, we describe in detail the proposed method and present some experimental results in section 4. Finally, some conclusions and future work are outlined.

2. The assumption of paraperspective camera projection for satellite image acquisition

When a camera field-of-view is narrow and the size of the sensed objects is small with respect to their distance to the camera, as it is the case of the satellite images, the perspective projection, which is a non-linear mapping, can be approximated by a paraperspective transformation, which is a linear one. This simplification leads to great reduction of complexity in many vision problems including the epipolar geometry estimation of different views (Xu & Zhang, 1996).

The epipolar geometry is the intrinsic projective geometry between two images (views) of the same scene. It is independent of scene structure, and only depends on the cameras' internal parameters and their relative poses. The fundamental matrix $\mathbf{F}$ encapsulates this intrinsic geometry, that is, $\mathbf{F}$ is the algebraic representation of epipolar geometry. If a point in 3-space $\mathbf{X}$ is imaged at $\mathbf{x} = (x, y, 1)^{\mathsf{T}}$ in a certain image, and at $\mathbf{x}' = (x', y', 1)^{\mathsf{T}}$ in another one (in homogeneous coordinates), then both points satisfy the relation $\mathbf{x}'^{\mathsf{T}}\,\mathbf{l}' = 0$, where $\mathbf{l}' = \mathbf{F}\,\mathbf{x}$ is named the *epipolar line* of $\mathbf{x}$ (see figure 1). The fundamental matrix $\mathbf{F}$ is the unique 3×3 rank 2 homogeneous matrix which maps a point of one image to a line of the other $(\mathbf{x} \mapsto \mathbf{l}')$. This relation is called *epipolar constraint* and can be expressed as follows:

$$\mathbf{x}'^{\mathsf{T}}\,\mathbf{F}\,\mathbf{x} = 0 \tag{1}$$

which holds for any pair of corresponding image points (pixels) $\langle \mathbf{x}, \mathbf{x}' \rangle$.

The fundamental matrix is independent of the scene structure, and has 8 independent ratios (degrees of freedom or *dof*) that can be estimated from correspondences identified in both images (7 pairs for its general form), without requiring knowledge of the cameras' internal parameters or relative pose.

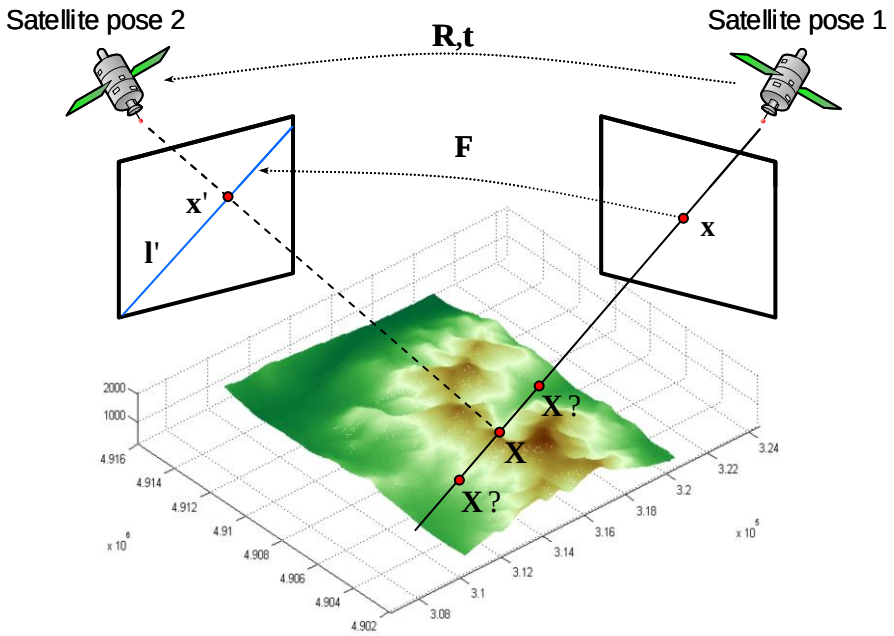

Figure 1: Intrinsic projective geometry between two images (epipolar constraint).

On the other hand, if a set of 3D points $\{\mathbf{X}_i, i=1,\cdots,n\}$ are coplanar (see figure 2), their projections $\{\langle \mathbf{x}_i, \mathbf{x}_i' \rangle, i=1,\cdots,n\}$ must verify the expression:

$$\mathbf{x}_i' = \mathbf{H}\,\mathbf{x}_i \tag{2}$$

where $\mathbf{H}$ is a projective transformation (also called projectivity or homography) induced by the planar surface which maps any point from one image plane to another $(\mathbf{x} \mapsto \mathbf{x}')$. $\mathbf{H}$ is a 3×3 non-singular homogeneous matrix of the form:

$$\mathbf{H} = \begin{pmatrix} \mathbf{A} & \mathbf{t} \\ \mathbf{v}^{\mathsf{T}} & v \end{pmatrix} \tag{3}$$

being $\mathbf{A}$ a 2×2 non-singular matrix, $\mathbf{t} = (t_x, t_y)^{\mathsf{T}}$ a translation 2-vector; and the vector $\mathbf{v}^{\mathsf{T}} = (v_1, v_2)$.

Just as the fundamental matrix, the homography has 8 *dof* and can be estimated from 4 corresponding points identified in both images, without requiring knowledge of the cameras' parameters or the relative pose of the plane.

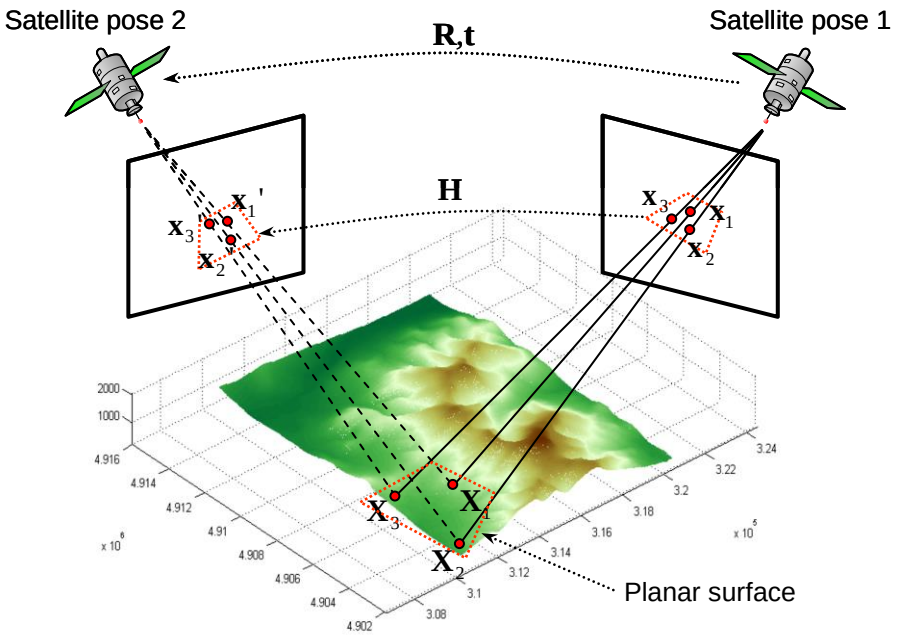

Figure 2: Projective transformation between two images induced by a planar surface (world plane).

In satellite remote sensing, because of the distance between the camera and the observed scene, we can assume a paraperspective (also called *affine*) projection. Under this assumption, the epipolar geometry can be simplified as follows:

- 4 pairs of points suffice to compute the epipolar geometry which relates both images, instead of the 7 pairs required for the general form. In this case, the matrix $\mathbf{F}$ (called now the affine fundamental matrix) has only 4 *dof* and is of the form:

$$\mathbf{F}_{\mathrm{A}} = \begin{pmatrix} 0 & 0 & a \\ 0 & 0 & b \\ c & d & e \end{pmatrix} \tag{4}$$

  This forces the epipolar lines to be parallel and, under the hypothesis that the errors in the image point localization follow a known model (typically Gaussian), their positions can be accurately re-estimated (Xu & Zhang, 1996).

- Regarding the estimation of the homographies, 3 pairs of control points are enough to compute the affine homography, instead of the 4 pairs required for its general form. The affine homography (also called affinity) has 6 *dof* and can be written as:

$$\mathbf{H}_{\mathrm{A}} = \begin{pmatrix} \mathbf{A} & \mathbf{t} \\ 0^{\mathsf{T}} & 1 \end{pmatrix} \tag{5}$$

Exploiting such simplifications, the identification of corresponding point pairs (in our case, for registering high resolution images) can be stated as follows: given the image point $\mathbf{x}_i = (x_i, y_i)^{\mathsf{T}}$ in the fixed image ($\mathbf{I}$), locate its corresponding points $\mathbf{x}'_i = (x'_i, y'_i)^{\mathsf{T}}$ in the moving one ($\mathbf{I}'$), such that, $\langle \mathbf{x}, \mathbf{x}' \rangle$ verify $\mathbf{x}'^{\mathsf{T}} \mathbf{F}_{\mathrm{A}} \mathbf{x} = 0$ and, the image patches at $\mathbf{x}_i$ and $\mathbf{x}'_i$ are as much "similar" as possible. Thus, we set the point matching as a process that minimizes the sum of square differences (SSD) of two pixel neighbourhoods centered at $\mathbf{x}_i$ and $\mathbf{H}_{\mathrm{A}}\mathbf{x}_i$. Formally, this can be expressed as:

$$\min_{\{\mathbf{H}_{\mathrm{A}}\}} \sum_{x=x_i-\omega_x}^{x_i+\omega_x} \sum_{y=y_i-\omega_y}^{y_i+\omega_y} \left[\mathbf{I}(x,y) - \mathbf{I}'\left(a_{11}x + a_{12}y + t_x, a_{21}x + a_{22}y + t_y\right) \right]^2 \tag{6}$$

where $(2\omega_x + 1) \times (2\omega_y + 1)$ is the size of the pixel neighbourhood considered. If no information is available *a priori*, $\mathbf{H}_{\mathrm{A}}$ can be initially taken as the identity matrix.

Finally, in order to compensate for the radiometric differences between both images, the image patches are conveniently normalized, and (6) can be rewritten as:

$$\min_{\{\mathbf{H}_{\mathrm{A}}\}} \sum_{x=x_i-\omega_x}^{x_i+\omega_x} \sum_{y=y_i-\omega_y}^{y_i+\omega_y} \left[\mathbf{I}(x,y) - \alpha \mathbf{I}'\left(a_{11}x + a_{12}y + t_x, a_{21}x + a_{22}y + t_y\right) + \beta \right]^2 \tag{7}$$

where α and β are the local contrast and brightness normalization coefficients, respectively, estimated from the values of both pixel neighbourhoods (Benke & Hedger, 1996).

3. Description of the proposed method

For a registration process to be successful, especially with high-resolution images, it is necessary that the CP pairs are distributed on the images according to their geometric differences or, failing that, the CP density must be high enough to account for the most geometrically distorted area. This paper is concerned with the second choice, that is, since we assumed that no *a priori* knowledge of the geometric distortions of images is available; we present a technique to robustly extract high-densities of CP pairs uniformly distributed over the image. As depicted in figure 3, the proposed method consists basically of two steps: 1) an initial set of CP pairs is identified in both images based on a coarse-to-fine matching strategy, and 2) this CP set is processed in order to discard outliers as well as to refine the position of the inliers according to the recovered affine epipolar geometry.

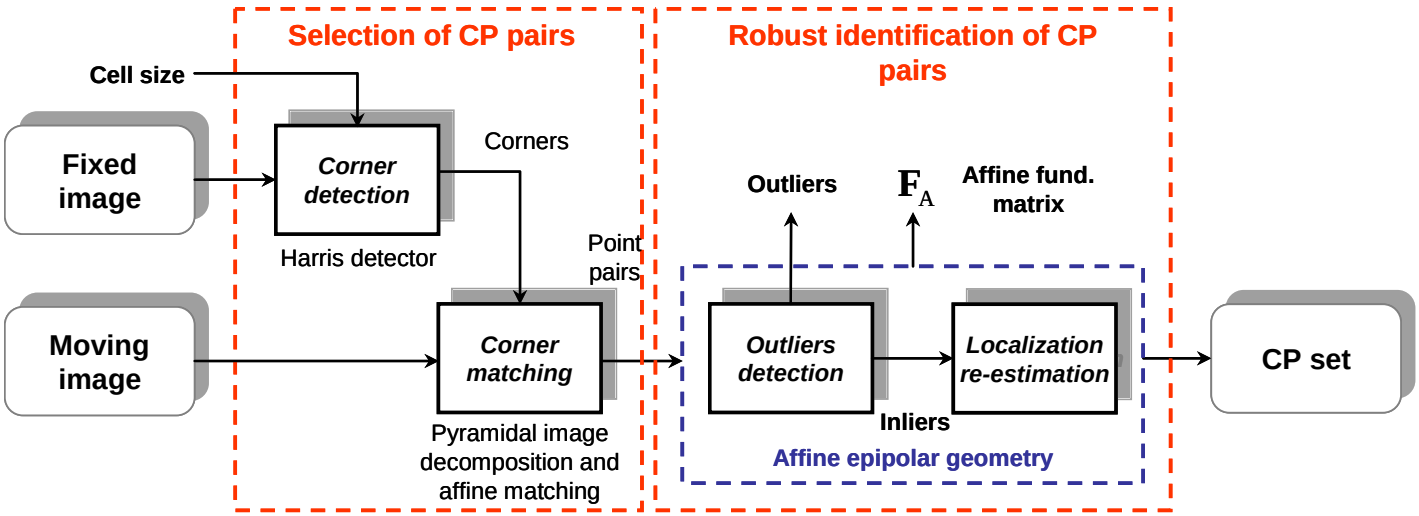

Figure 3: Structure of the proposed method.

Similar approaches have been proposed, for example, in disciplines such as computer vision for feature tracking (Kang et. *al.*, 1997; Bouguet, 1999), stereo matching (Guoqing et. *al.*, 1996; Zhang et. *al.*, 1996), motion estimation (Tomasi & Kanade, 1991; Feng & Perona, 1998), etc. Most of these methods cope with video sequences or pairs of images where the relative geometric differences between two consecutive frames can be approximated, in most of cases, by a similarity transformation (translation, rotation, and scale change). However, these approaches may not perform well when applied to typical remote sensing images which are usually acquired from very different observation angles, lighting conditions, and with significant scene changes. On the other hand, none of them include specific mechanisms to assure the geometric consistency and/or the correct localization of the detected pairs of control points.

Next, we describe in detail each of above-mentioned steps.

3.1. Selection of control point pairs

The selection of a CP set is addressed in two phases. First, we select distinctive points (i.e. pixels of maximum curvature or corners) in the fixed image. In order to get them uniformly distributed over the image, we pick a point from every cell of a rectangular grid arranged on it (the cell size is adjustable). The corner detector used in this work is the well-known Harris operator (Harris & Stephens, 1988), which presents two interesting characteristics for this work: it generates a big number of reliable corners and provides sub-pixel accuracy. Secondly, the matches of the fixed-image corners are identified in the moving one. To accomplish such a task, we employ a pyramidal implementation (Bouguet, 1999) of the classical Lucas-Kanade feature tracker (Lucas & Kanade, 1981) that we have extended for working with affine matching (instead of similarity) as well as for

compensating image radiometric differences (as described in section 2). More precisely, this approach addresses the minimization of (7) in a Newton-Raphson style for each fixed-image point. The minimization is applied throughout the n levels of the Gaussian pyramid decompositions* of the two images with a fixed size of the pixel neighbourhood (Burt & Adelson, 1983). More specifically, the matching process starts from the top level of the pyramid, and the results in the higher levels are used as approximations in the subsequent lower levels. This continues until the entire pyramid has been processed, or certain degree of similarity between the image patches has been achieved. In the latter case, the control point localizations are conveniently scaled (in the same proportion) to compensate the pyramidal decomposition. Because of fixed pixel size during this iterative process, we are able to cope with large displacements between corresponding points (typical in satellite image registration). Figure 4 shows graphically the selection of control point pairs described above.

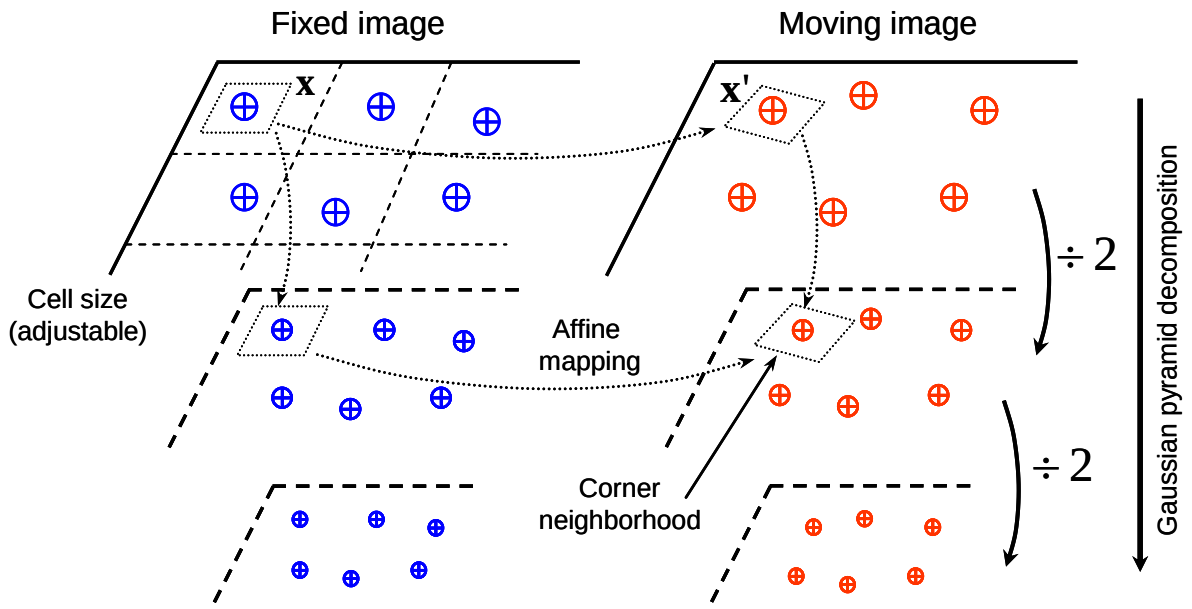

Figure 4: Selection of control point pairs: corner selection and affine matching. The selection step is addressed in a Gaussian pyramid decomposition of both images, which allows us to deal with significant geometric differences between the images.

In our tests, we have employed a grid with square cells of 50 pixels of side (for panchromatic QuickBird images this width corresponds to 30 meters). This size is small enough for precisely registering images acquired from different observation angles and covering diverse terrain profiles, as analyzed in (Arévalo & González,

* A Gaussian pyramid is a technique broadly used in image processing, which involves creating a series of images that are weighted down using Gaussian kernel (a low-pass filtering) and scaled down. It creates a stack of successively smaller images, with each pixel contains the local average that corresponds to a pixel neighbourhood on a higher level of the pyramid.

2006). Regarding the parameters of the matching procedure, we have considered 4 levels for the Gaussian pyramid and 25×25 pixels for the size of the pixel neighbourhood. Figure 5 shows the control point pairs identified in a pair of QuickBird images of a typical urban scene.

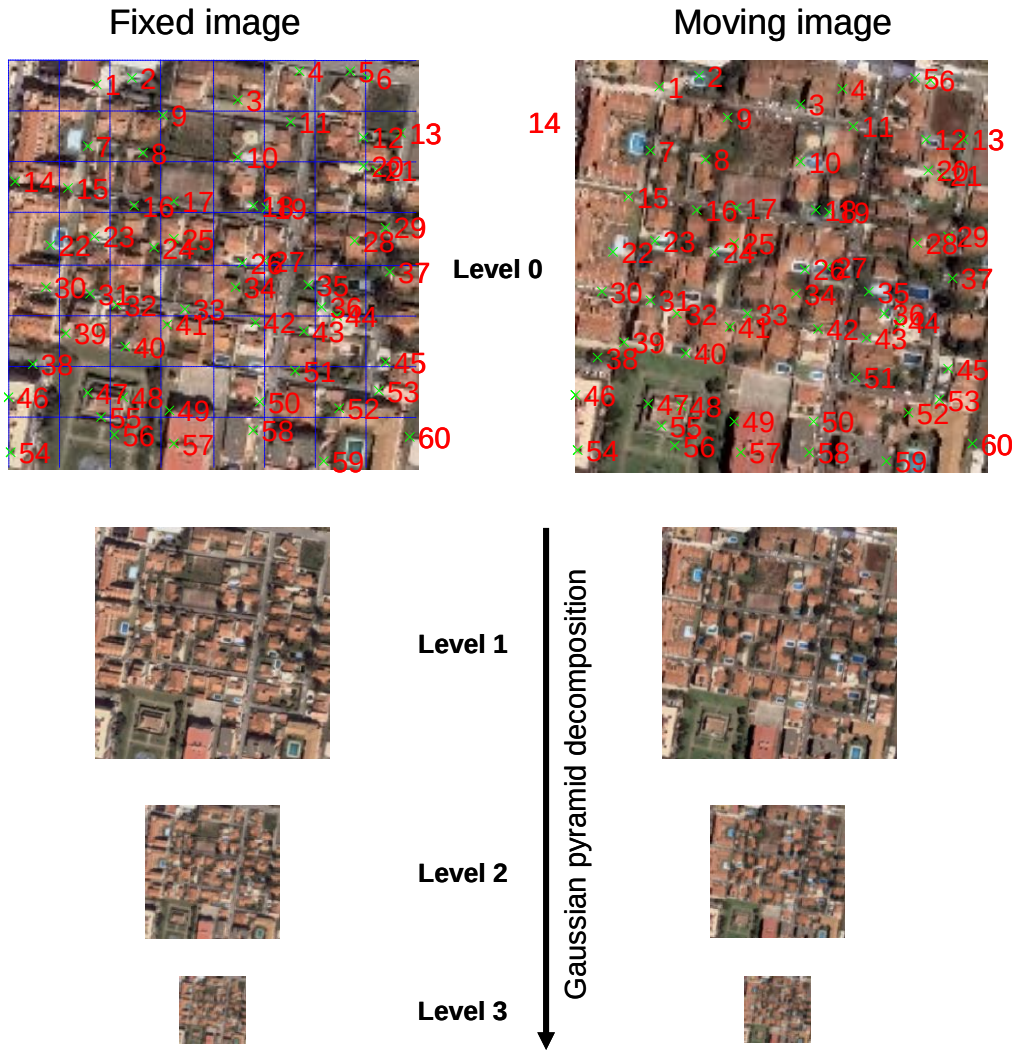

Figure 5: Selection of control point pairs illustrated on two QuickBird images of a typical urban scene. Gaussian pyramids of both images and set of CP pairs provided by the matching procedure proposed in section 3.1.

Once the initial set of matches has been identified, the following step of our method consists of discarding those pairs non-consistently matched and refining the remaining matches' coordinates in order to compensate their localization errors.

3.2. Refinement of the control point pairs

At this stage, we make use of the affine epipolar geometry both to remove point pairs not consistent with that estimated fundamental matrix, that is, outliers that do not verify the equation (1), and to refine the localization of those non-discarded (inliers).

For recovering the affine fundamental matrix, we employ a robust methodology based on the RANdom SAmple Consensus (RANSAC) algorithm (Fischler & Bolles, 1981). This technique exploits the information redundancy to provide a robust estimate of the parameters of a model and to discard those samples do not consistent with it. In our particular case, the model is given by the affine fundamental matrix $\mathbf{F}_{\mathrm{A}}$ and the error function that measures the consistency of a pair of points $\langle \mathbf{x}, \mathbf{x}' \rangle$ is the first-order geometric error (also called Sampson distance), which is expressed as follow:

$$d\left(\mathbf{x}', \mathbf{F}_{\mathrm{A}}\mathbf{x}\right) = \frac{\left(\mathbf{x}'^{\mathrm{T}}\mathbf{F}_{\mathrm{A}}\mathbf{x}\right)^2}{\left(\mathbf{F}_{\mathrm{A}}\mathbf{x}\right)_1^2 + \left(\mathbf{F}_{\mathrm{A}}\mathbf{x}\right)_2^2 + \left(\mathbf{F}_{\mathrm{A}}^{\mathrm{T}}\mathbf{x}'\right)_1^2 + \left(\mathbf{F}_{\mathrm{A}}^{\mathrm{T}}\mathbf{x}'\right)_2^2} \tag{8}$$

where the terms of the form $\left(\mathbf{z}\right)_j^2$ represent the square of the *j*-th entry of the vector $\mathbf{z}$.

In practice, in order to achieve a better estimate, we employ the symmetric epipolar error (the distance of a point from its projected epipolar line in both images) which is derived from (8) and whose geometric interpretation is graphically shown in figure 6. The symmetric epipolar error is defined as:

$$d\left(\mathbf{x}', \mathbf{F}_{\mathrm{A}}\mathbf{x}\right)^2 + d\left(\mathbf{x}, \mathbf{F}_{\mathrm{A}}^{\mathrm{T}}\mathbf{x}'\right)^2 = \left(\mathbf{x}'^{\mathrm{T}}\mathbf{F}_{\mathrm{A}}\mathbf{x}\right)^2 \left(\frac{1}{\left(\mathbf{F}_{\mathrm{A}}\mathbf{x}\right)_1^2 + \left(\mathbf{F}_{\mathrm{A}}\mathbf{x}\right)_2^2} + \frac{1}{\left(\mathbf{F}_{\mathrm{A}}^{\mathrm{T}}\mathbf{x}'\right)_1^2 + \left(\mathbf{F}_{\mathrm{A}}^{\mathrm{T}}\mathbf{x}'\right)_2^2} \right) \tag{9}$$

The final step of the RANSAC algorithm is re-computing the model (i.e. the epipolar geometry) using all the inliers. We make this re-estimation by minimizing a cost function from which we deliver the Maximum Likelihood (ML) estimate of the affine epipolar geometry and, taking advantage of the simplifications derived from the paraperspective assumption, a better estimate of the coordinates of the matches (for a comprehensive description of this procedure, please, refer to (Hartley & Zisserman, 2003)). The ML estimate of the affine fundamental matrix assumes that the image point localizations are affected by Gaussian noise. In that case, the ML estimate is the one which minimizes the geometric distance (which is the re-projection error):

$$\min_{\{\mathbf{F}_{\mathrm{A}}, \hat{\mathbf{x}}_i, \hat{\mathbf{x}}'_i\}} \sum_i d\left(\mathbf{x}_i, \hat{\mathbf{x}}_i\right)^2 + d\left(\mathbf{x}'_i, \hat{\mathbf{x}}'_i\right)^2 \tag{10}$$

where $\langle \mathbf{x}_i, \mathbf{x}'_i \rangle$ are the initial CP pairs, and $\langle \hat{\mathbf{x}}_i, \hat{\mathbf{x}}'_i \rangle$ are the refined ones.

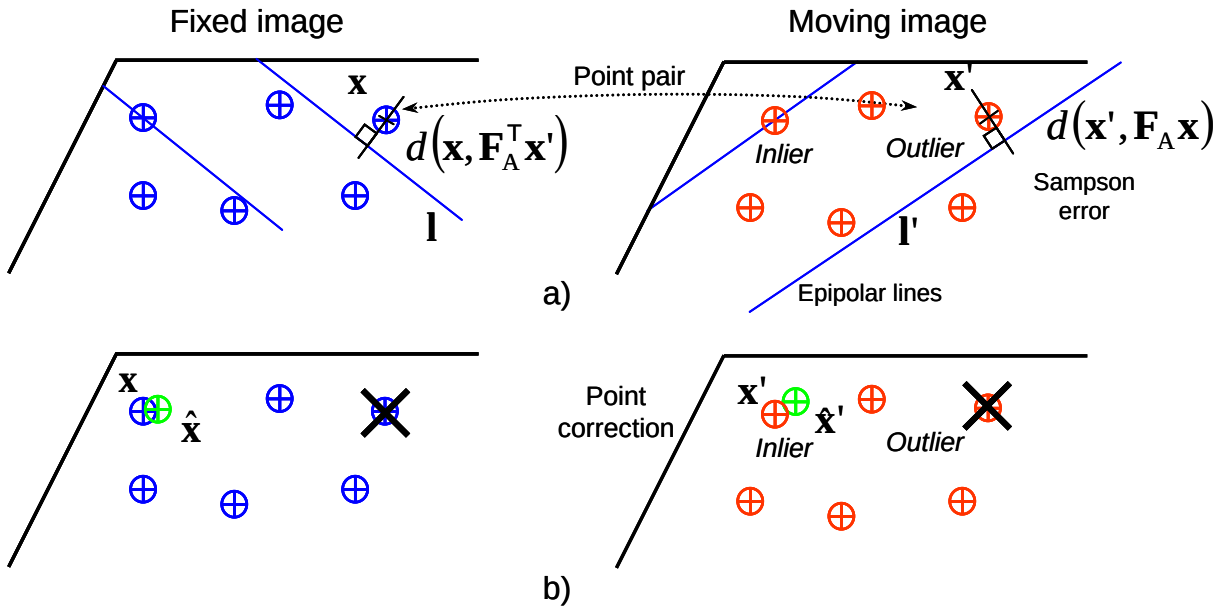

Figure 6: Robust identification of control point pairs. By forcing the pairs to satisfy the affine epipolar geometry that relates both images, we can (a) discard the outliers and (b) re-estimate the localization of the inliers.

Figure 7 illustrates the refinement of the control point pairs identified in the pair of QuickBird images shown in figure 5. In particular, figure 7(a) illustrates the case of a pair classified as outlier, since its symmetric epipolar error (equation (9)) is greater than a given threshold (0.75 pixels in our tests). Figure 7(b) shows the inliers once their localizations have been corrected. As observed in figure 7(b), the proposed method is able to provide a set of point pairs distributed throughout the images despite their difference in radiometry, angle of observation and content (e.g. cast shadows, different sides of the buildings, scene changes, etc.). Next, a variety of experimental tests to quantify and evaluate the method performance are presented.

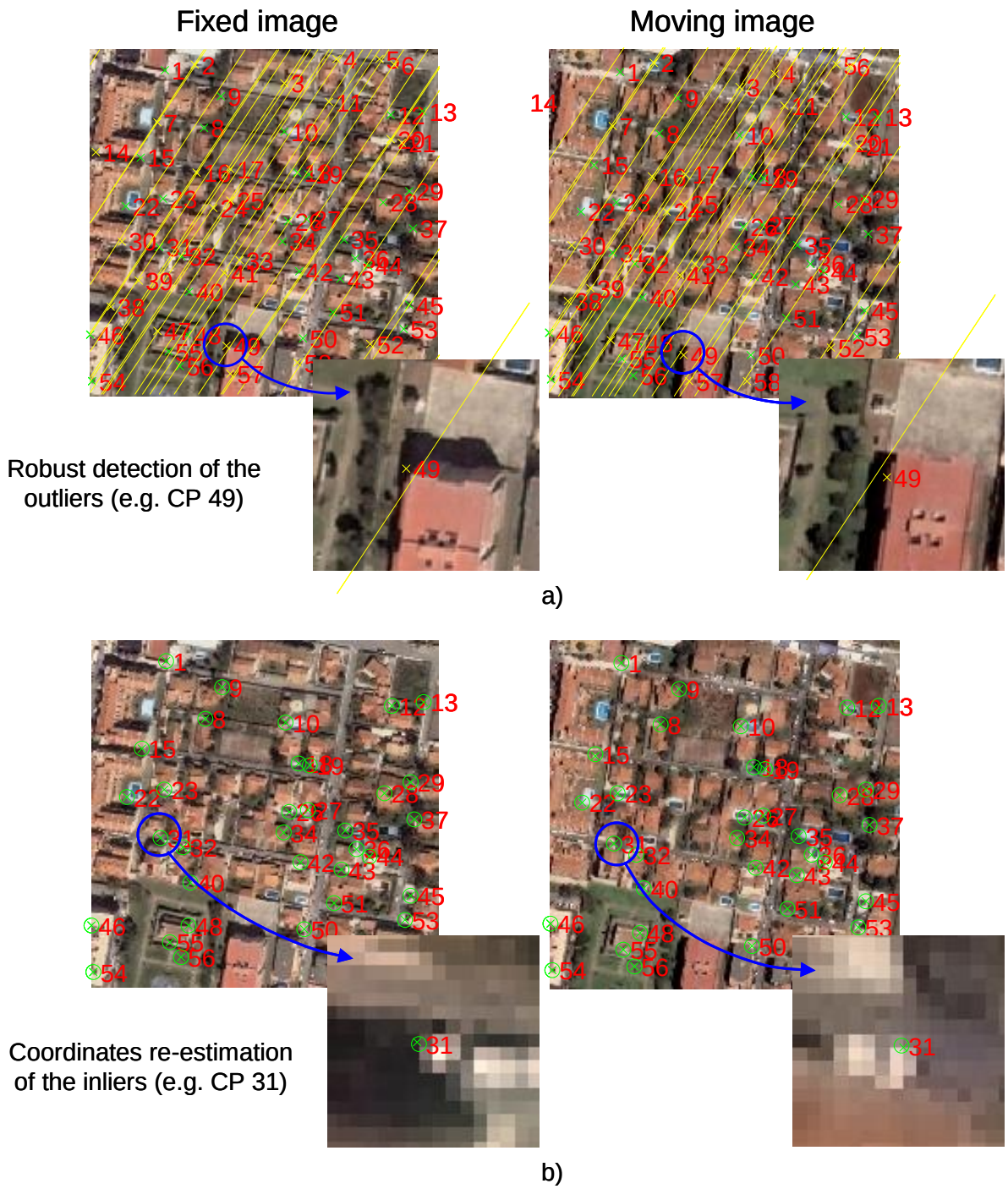

Figure 7: Refinement of the control point pairs detected in two QuickBird images of a typical urban scene: a) robust detection of the outlier matches and b) coordinates re-estimation of the inliers. The enlarged images show, in detail, the actuations described in section 3.2.

4. Experiment and results

In this section, we evaluate the accuracy of the matches identified by the proposed method. Because the large number of them and differences between both images, it is difficult to manually deal with this task. Thus, we propose the following automatic procedure: first, we register the two images with a powerful elastic method using the identified control point set as input; and second, we measure the registration errors (RMS and CE90) of an independent check point (ICP) set selected by hand. The registration method applied is a thin-plate-spline (TPS) function, which a part from its effectiveness, takes into account the contribution of all control points in the resulting accuracy. With regard to the test data used in this work, we have considered two QuickBird images of the city of Rincón de la Victoria (Málaga-Spain) of $1.8 \times 0.9\,\mathrm{Km}^2$ acquired on different seasons (which leads to significant illumination changes) and from two very different poses.

A first set of experiments is aimed to check the robustness of the matching process. To this extent, we analyze the accuracy and the number of control points robustly matched by the proposed method when the images are affected by different noise patterns. Concretely, five pairs of images are evaluated: the original one and four pairs affected by different rates of either uniform or Gaussian noise. Given the size of the images ($1.8 \times 0.9\,\mathrm{Km}^2$) and the size of the cells (30 meters), the number of corners initially extracted in the fixed image is 1800 points. After applying the proposed method, the percentage of inliers and outliers as well as the registration errors for the five identified CP sets are shown in table 1. Two facts can be highlighted from these results: a) the robustness of the method despite the noise pattern applied (all experiments exhibit similar rates); and b) the high percentage of inliers, close to 90% (1613 pairs) for the five experiments. Such robustness and accuracy of matching is also appreciated in the registration errors, which remain steady for most experiments and close to 1 meter in all of them.

Table 1: Some statistical results of the method when the images to register are affected by different noise patterns.

Noise pattern	% inliers	% outliers	RMSE (m)	CE90% (m)
None	90.19	9.81	1.26	0.97
Gaussian (2.5%)	90.36	9.64	1.17	0.97
Gaussian (5%)	89.57	10.43	1.15	1.02
Uniform (2.5%)	89.80	10.20	1.24	1.06
Uniform (5%)	89.86	10.14	1.20	1.04

In a second experiment, we quantify the benefits of using the affine-based matching instead of the similarity-based one typically employed in other works. For this purpose, we compare the proposed method (affinity-

based) with a pyramidal implementation of the Lucas-Kanade feature tracker included in *The OpenCV Library*[†], which is based on similarity-based matching (Bouguet, 1999). Although the two methods start with the same set of fixed-image points, only those pairs matched by both methods are finally considered in the image comparison. As can appreciated in table 2, the our affinity-based approach outperforms the Lucas-Kanade feature tracker, because a matching process based on similarities is not able to accurately match features in images which exhibit local affine distortions.

Table 2: Results from the comparison between our affinity-based method and the Lucas-Kanade feature tracker.

Method	% inliers	% outliers	RMSE (m)	CE90% (m)
Proposed method	90.19	9.81	1.26	0.97
Lukas-Kanade feature tracker	91.40	8.60	3.57	3.15

Finally, the influence of the outlier detection and coordinates re-estimation on the method performance is analyzed with the comparative results shown in table 3. From that numbers, we can highlight the following points:

- The accuracy of the matching process is high even when no refinement is applied. We may confirm this fact comparing the results of the first test (initial CP set) with the registration errors provided by Lucas-Kanade feature tracker (see table 2). The differences are about 1 meter for both metrics.
- The identification of the outliers via affine epipolar geometry (second row) discards those pairs with extreme registration errors, as revealed by the RMSE value. The CE90 value is not affected by removing these pairs since the CE90 captures how good the 90% of the samples are, rejecting the remaining 10%. This actuation leads to a decrease of the RMSE and CE90 values up to 1.61 and 1.76 meters, respectively.
- The significant improvement in accuracy when the coordinate correction is applied, as revealed by the CE90 value (lower than 1 meter). This result reveals the importance of refining the CP localization.

[†] The Open Source Computer Vision Library is a collection of algorithms and sample code for various computer vision problems. The library is compatible with IPL and utilizes Intel® Integrated Performance Primitives for better performance. Visit http://www.sourceforge.net/projects/opencvlibrary for further information.

Table 3: Statistics of the proposed method after applying the following actuations: none, outlier detection and coordinate correction.

Actuation	% inliers	% outliers	RMSE (m)	CE90% (m)
Initial CP set	100	-	2.85	1.89
Outlier detection	90.19	9.81	1.61	1.76
Coordinate re-estimation	90.19	-	1.26	0.97

5. Conclusions and future work

High-resolution satellite images such as QuickBird are expected to play an important role in many remote sensing applications. For achieving that, tools commonly used for lower resolution images may not be appropriate, as it is the case of the methods for automatically detecting control point pairs.

This paper presents a technique to robustly extract pairs of control points for registering such images, and analyze their accuracy and robustness by means of a variety of experimental tests. We have described a coarse-to-fine approach for matching image points which exploits the affine epipolar geometry to discard the outliers (matches not consistent with the estimated 3D geometry) and to accurately estimate the correct localization of the inliers. In the experimental tests, we have used panchromatic QuickBird images (0.6 meter/pixel) acquired from very different poses and illumination conditions; and affected by different noise patterns.

This automatic procedure allows us to manage a large number of consistently-matched and uniformly distributed point pairs, as revealed in our tests. Nevertheless, it should be noted that, same as all methods for automatically detecting point pairs in high-resolution images (Zhang & Fraser, 2005; Paul et. *al*., 2002), the resulting set of matches could not be the most suitable ones to model the relative geometric differences in typical urban scenarios. For example, a perfectly-matched independent check point pair identified at a flat roof of a very high building is not appropriate for measuring how good the ground or low-height buildings have been registered (see figure 8). Thus, local geometric information should be considered to efficiently discard non-suitable (although correctly detected) pairs from the registration accuracy viewpoint. This is one of our concerns for the next future, as well as, to develop some strategy for recovering the points missed in the second step taking advantage of the recovered affine reconstruction.

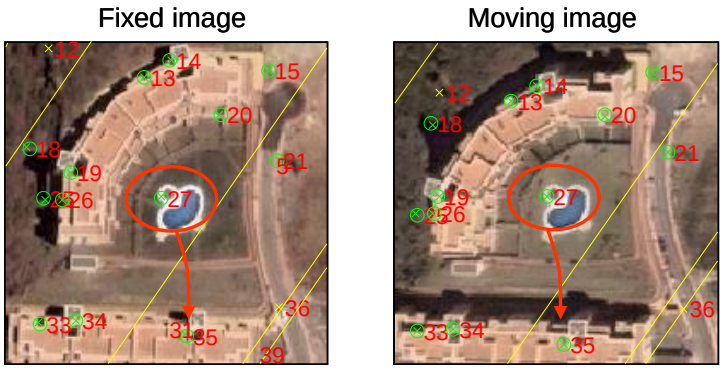

Figure 8: Control points pairs identified in different scene planes (e.g. ground and top of the buildings). This circumstance typically introduces additional registration errors in urban scenes.

Acknowledgments

The ©DigitalGlobe QuickBird images used in this study is distributed by Eurimage, SpA. (www.eurimage.com) and provided by Decasat Ingenieria S.L., Málaga, Spain. (www.decasat.com).

References

ARÉVALO, V. and GONZÁLEZ, J., 2006. A Comparison of Registration Techniques on QuickBird Satellite Images. *In Procs. of the ISPRS Mid-term Symposium 2006: Remote Sensing, From Pixels to Processes*, Enschede, The Netherlands, CD-ROM.

BENKE, K.K. and HEDGER, D.F., 1996. Normalization of Brightness and Contrast in Video Displays. *Eur. J. Phys.*, **17**, pp: 268–274.

BOUGUET, J.Y., 1999. Pyramidal Implementation of the Lucas-Kanade Feature Tracker: Description of the Algorithm. *Intel Corporation, Microprocessor Research Labs*. OpenCV Documents.

BURT, P.J. and ADELSON, E.H., 1983. The Laplacian Pyramid as a Compact Image Code. *IEEE Transactions on Communications*, **31**(4), pp: 532–540.

FENG, X. and PERONA, P., 1998. Real Time Motion Detection System and Scene Segmentation. *CDS Technical Report CDS 98-004*, California Institute of Technology.

FISCHLER, M.A. and BOLLES, R.C., 1981. Random Sample Consensus: A Paradigm for Model Fitting with Application to Image Analysis and Automated Cartography. *Communications of the ACM*, 24, pp. 381–395.

GONZALEZ, J.; AMBROSIO, G. and AREVALO, V., 2001. Automatic Urban Change Detection from the IRS-1D PAN. *In Procs. of IEEE/ISPRS Joint Workshop on Remote Sensing and Data Fusion over Urban Areas*, pp: 320-323, Rome (Italy).

GUOQING, Z., BAOZONG, Y. and XIAOFIANG, T., 1996. A Software Package of Stereo Vision. *In Procs. of 3rd International Conference on Signal Processing*, **2**(2), pp: 914–917.

HARRIS, C.J. and STEPHENS, M., 1988. A Combined Corner and Edge Detector. *In Proc. of 4th Alvey Vision Conference*, Manchester, UK, pp. 147–151.

HARTLEY, R.I. and ZISSERMAN, A., 2003. *Multiple view geometry in computer vision*. (Cambridge: Cambridge University Press), pp. 279–292.

KANG, S.B., SZELISKI, R., and SHUM, H.Y., 1997. A Parallel Feature Tracker for Extended Image Sequences. *Computer Vision and Image Understanding*, **67**(3), pp: 296–310.

LUCAS, B.D. and KANADE, T., 1981. An Iterative Image Registration Technique with an Application to Stereo Vision. *IJCAI81*, pp. 674–679.

PAUL, D., FRASER, C. and PENDLEBURY, N., 2002. Digital Orthomosaics as a Source of Control for Geometrically Correcting High Resolution Satellite Imagery. *In Procs. of the 23rd Asian Conference on Remote Sensing*, Kathmandu, Nepal, URL: http://www.gisdevelopment.net/aars/acrs/2002/vhr/108.pdf.

TOMASI, C. and KANADE, T., 1991. Detection and Tracking of Point Features, *Technical Report CMU-CS-91-132*.

XU, G. and ZHANG, Z., 1996, *Epipolar Geometry in Stereo, Motion, and Object Recognition: A Unified Approach*, (Norwell, SA, USA: Kluwer Academic Publishers), pp. 61–65.

ZHANG, C. and FRASER, C., 2005. Automated Registration of High Resolution Satellite Imagery for Change Detection. IAPRS, **36**, Hannover,Germany, CD-ROM.

ZHANG, Z., DERICHE, R., FAUGERAS, O.D. and LUONG, Q.T., 1995. A Robust Technique For Matching Two Uncalibrated Images Through The Recovery Of The Unknown Epipolar Geometry. *Artificial Intelligence*, **78**(1), pp: 87–119.

ZITOVÁ, B. and FLUSSER, J., 2003. Image Registration Methods: A Survey. *Image and Vision Computing*, 21, pp: 977–1000.