A HYBRID METHOD FOR STEREO IMAGE MATCHING

M. T.. Silveira [a]*, R. Q. Feitosa [b], K. Jacobsen [c], J. L. N. S. Brito [b], Y.Heckel [a]

[a] Pontifícia Universidade Católica do Rio de Janeiro, Departamento de Engenharia Elétrica, Rua Marques de São Vicente 225, Rio de Janeiro, Brasil - (marts, raul)@ele.puc-rio.br, heckel@gmail.com
[b] Universidade do Estado do Rio de Janeiro, Programa de Pós Graduação em Engenharia de Computação - Geomática, Rua São Francisco Xavier 524, Rio de Janeiro, Brasil - jsilvabr@gmail.com
[c] Leibniz Universität Hannover - Institut für Photogrammetrie und GeoInformation - Nienburger Str. 1D -30167, Hannover, Germany - jacobsen@ipi.uni-hannover.de

Commission I, WG I/5

KEY WORDS: Photogrammetry, matching, feature, extraction, DEM/DTM

ABSTRACT:

The accuracy of the 3D models of the Earth surface based on stereoscopic image pairs depends on the accuracy of corresponding points located in both images. Area-based automatic image matching combined using a region-growing technique are able to provide a dense and accurate grid of corresponding points. However the region-growing process may stop at image patches where the x-parallax is suddenly changing. In such a case new seed points must be provided, usually by human operator. From the additional seed points the region-growing procedure may continue. Depending upon the type of image and the 3D-structure of the mapped area, the human intervention may be considerable. A fully automatic alternative that combines scale invariant feature transform, least square matching and region-growing technique is presented. Experiments conducted on a stereo pair of IKONOS images covering different terrain types have shown the effectiveness of the proposed method especially in locations with abrupt height changes, such as façades of high buildings.

1 INTRODUCTION

After the high-resolution satellite images became commercially available, 3D surface models generated from space-born stereo images turned into an attractive alternative for applications such as telecommunication planning, disaster monitoring and urban planning.

Generally the quality of height models generated from stereo pairs depends essentially on how accurate corresponding points are determined. The matching approaches to perform this task can be classified into two main categories: area-based methods and feature-based methods.

Area-based methods (Trucco and Verri, 1998), may be comparatively more accurate, because they take into account a whole neighborhood around the points being analyzed to establish correspondences. The simple cross correlation may be considerably decrease over steep slopes, what is not so much the case for least squares matching, respecting the object inclination. Moreover, these methods require a good start solution to produce satisfactory results. This can be achieved by combining an area-based method with a region-growing strategy. Starting from a single pair of so-called seed points, usually provided by human operator, these methods may cover the whole area if no sudden height changes are available. During the process, the operator is required to place new seed points on some not reached regions and the process has to restart from there on. Depending upon the imaged terrain, the operator may be asked to repeatedly provide new seeds. As a result, the whole process may involve a considerable amount of human intervention.

Generally speaking, feature-based methods in most cases are less accurate than area-based methods and produce not a satisfying point density, however they are fully automatic.

A hybrid method to locate corresponding points in stereo pairs is shown. This hybrid method combines area-based and feature-based methods, forming a single strategy that combines the advantages of both approaches. The developed method generates an accurate and dense set of corresponding points with a minimum of human intervention.

2 RELATED TECHNIQUES

The method proposed in this paper essentially combines three well known techniques that are briefly described in the next sections.

2.1 Least Squares Matching

Assuming that g_1 and g_2 are two images of the same scene, area-based methods try to find neighborhoods centered in the positions (x_1,y_1) and (x_2,y_2) respectively of g_1 e g_2, in which the relation

$$g_2(x_2, y_2) = \alpha\, g_1(x_1, y_1) + \beta \qquad (1)$$

is valid in least square sense for some value of α and $\beta \in \mathbb{R}$.

This model is invariant to the brightness (β) and the contrast (α) between the images. The normalized cross correlation assumes that (x_1,y_1) and (x_2,y_2) are related to a mere translation, as shown in the left side of table 1. However, this model is not correct

when the related object part is not horizontal or the images are rotated against each other.
An improvement of the normalized cross correlation is the least squares matching, taking the geometric differences between the image sub-areas into account, which is modeled by an affine transformation applied to the second image sub-area, as shown in the left side of table 1.

Normalized cross correlation	least squares matching
$\begin{cases} x_2 = x_1 + a \\ y_2 = y_1 + b \end{cases}$	$\begin{cases} x_2 = a + bx_1 + cy_1 \\ y_2 = d + ex_1 + fy_1 \end{cases}$

Table 1. Comparison of normalized cross correlation with least square matching

In addition the least squares matching respects the linear grey value changes in the x- and y-direction. The 8 parameters are determined by adjustment.

2.2 Region Growing

The region-growing (Otto and Chau, 1989) procedure begins with a pair of corresponding points, called seed points, usually identified by a human operator. An area-based method is then applied to determine more accurately its position in the second image. Depending on the similarity value, the matched corresponding points are kept or disregarded. Once the accurate position has been determined by the area-based method, up to four new pairs of points are generated *d* pixels up, down, left and right from the last localized point, where *d* is a chosen parameter, which establishes the steps in pixels for the next point location. These corresponding points are now new seed points and their positions are refined by an area-based method. The procedure is recursively repeated spreading new points over both images, and providing a dense grid of corresponding points.
Figure 1 shows an example of the result produced by least squares matching with region growing, starting from four seed points (red numbered crosses) selected manually. The encountered points are represented by white dots.
On the lower right region of figure 1 an important limitation of this approach can be observed, here is a high building. Note that no point was found around it. Generally speaking, the region growing stops over areas having larger height variation, with large differences of corresponding images caused by occlusion and different view to facades.
This problem can be solved with additional seed points inside the non-reached regions, and the region growing process restarts from there on. Depending on the size of the image and the terrain a manual measurement of seed points may be time consuming.

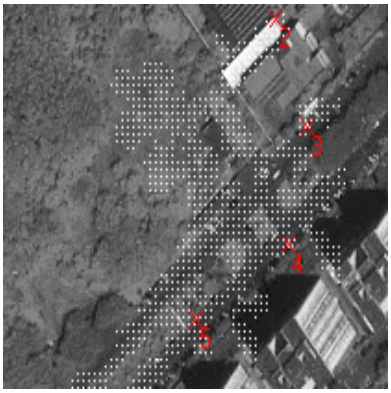


Figure 1. corresponding points determined by least squares matching with region growing

2.3 Scale Invariant Feature Transform

A feature-based method, recently proposed by David Lowe, known by the acronym SIFT (*Scale Invariant Feature Transform*), has been successfully applied in robotics. It is invariant to scale and rotation, and partially invariant to illumination and 3D view point. A complete description of this method can be found in (David Lowe, 2004). Following the three main steps that compound this method are briefly described.

Step 1: Scale-space extrema detection

Scale invariance is achieved by building a pyramid of images, as depicted in figure 2. A number of new images is generated from the input image, by a smoothing Gaussian filter, making up image octaves. One of these images is reduced in size, generating a new image with half the size of the images in the previous octave. The successive smoothing process followed by reducing in size may be repeated several times, thus generating new octaves. The images in the pyramid have different scales and hence different levels of details.

Figure 2. Pyramid with 3 octaves with 5 levels each

Key points are then selected across this pyramid. Some methods have been proposed in the literature to find the best key points (Mikolajczyk and Schmid, 2004; Mikolajczyk and Schmid, 2005). Here we have tested the method based on scale-space extrema detection. For each octave, the difference between images in adjacent levels is computed. This operation produces the so called difference of Gaussians (DoG) pyramid, as illustrated in the figure 3. The extrema values of the DoG pyramid computed on scale and space are selected as key points.

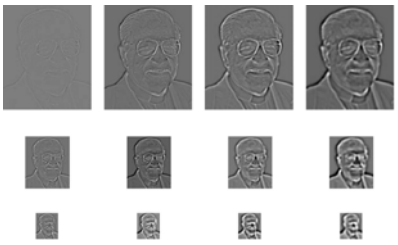


Figure 3. Difference of Gaussian Pyramid (DoG)

In a further step, key points on low contrast neighborhoods or along edges are discarded. Figure 4 shows in yellow key point groups obtained with an IKONOS image of Rio de Janeiro city.

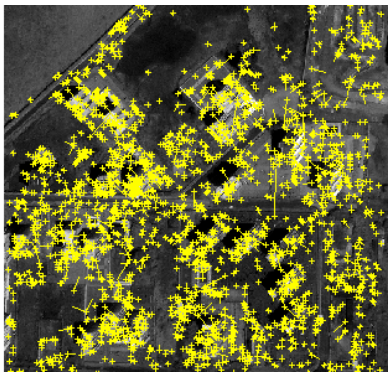


Figure 4. Example of key points corresponding to extrema on a DoG pyramid.

Step 2: Key point Descriptors

Once the key points are found, their descriptors are calculated as follows. The gradients on a neighborhood around each key point are computed. The selected neighborhood is divided into sub-regions, as depicted in figure 5.
For each sub-region the histogram of gradient directions is computed. In the setting up of those histograms the accumulated values are weighted by the corresponding magnitude of the gradient. In the example showed in figure 5, there are four histograms comprising 8 main directions. The histogram counts are stacked, forming a descriptor vector of that particular key point. The use of relative directions instead of absolute ones makes the descriptors rotation invariant.

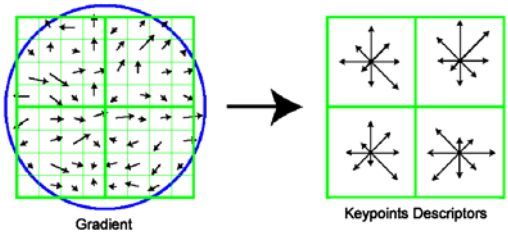


Figure 5. The SIFT descriptor

Step 3: Matching

The process described until now is applied to both images of the stereo pair thus producing two sets of key points, each one represented by its descriptor. The correspondence level between key points of both sets is given by the Euclidian distance between their descriptors. A pair of points will be considered as being correspondent if:

- the Euclidian distance between descriptors is less than a given threshold, and
- the Euclidian distance to the second nearest descriptor is less than a second given threshold.

Further requirements such as the epipolar constraint, can be imposed so as to reduce the number of false matches.

3 PROPOSED METHOD

The point matching algorithm proposed here adds to the least squares matching with region growing method an automatic mechanism that replaces the human operator in providing seed points whenever the matching procedure stops advancing over the image for any reason.
Basically, we use a feature-based method to generate a kind of seed repository from which we pick up a new seed whenever a barrier in the image is met, typically due to large differences of corresponding image caused by sudden height changes in the scene or other reasons. Many feature-based methods could be used for that purpose (i.e. Harris &Stevens 1988). Here we use SIFT. The whole method can be described by the following five sequential steps:

i) apply SIFT to produce a seed point repository,
ii) randomly draw a seed point from the repository,
iii) execute the least squares matching with region growing starting from the seed drawn in the previous step until it stops,
iv) delete from the repository alls seed points located in the area covered in the previous step,
v) repeat steps ii) to iv) until the repository becomes empty.

4 EXPERIMENTAL ANALYSIS

Experiments have been conducted to determine the effectiveness of the proposed method for automatic determination of corresponding points on stereoscopic images.
The experiments were based on a pair of IKONOS images, covering an area of approximately 100 km^2 in Rio de Janeiro city with heights ranging from 0 to 500 m.
Four types of regions were cropped from the image for analysis:

a) forest: characterized by a predominance of high and dense trees,
b) rural: dominated by grassland and savannah,
c) residential: with predominance of man-made features, specially single floor residential units, and
d) high–buildings: urban regions with multi-floor buildings.

Five sub-images of 500x500 pixels representing each region type were manually selected from the input images. Two experiments were conducted on each sub-image:

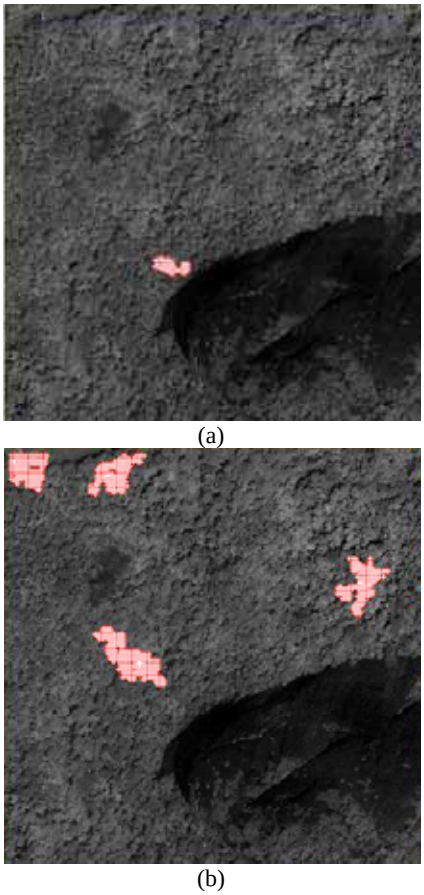
i) a single pair of seed points was placed manually by a human operator and least squares matching with region growing was executed from there on. This procedure was repeated 5 times for each sub-image; the initial seed points were selected so as to start from quite different image regions in each run; and
ii) the proposed hybrid method (SIFT followed by least squares matching with region growing) was applied.

In the following analysis the coverage will be used as a performance metric. This is given by the proportion of the image area covering the neighbourhoods of $(2d+1)\times(2d+1)$ pixels surrounding each pair of corresponding points found by the matching procedure, where d is the step size used in least squares matching with region growing phase.

As a matter of fact, aspects other than the coverage alone must be considered in this performance assessment. These aspects will be introduced below.

4.1 Forested Regions

Forested regions come out as high textured regions whose local appearance is significantly affected by different view directions. The caused different imaging can hardly can be compensated by an affine transformation as in the least squares matching. This makes it difficult to find conjugate image pairs in forested regions.



(a)

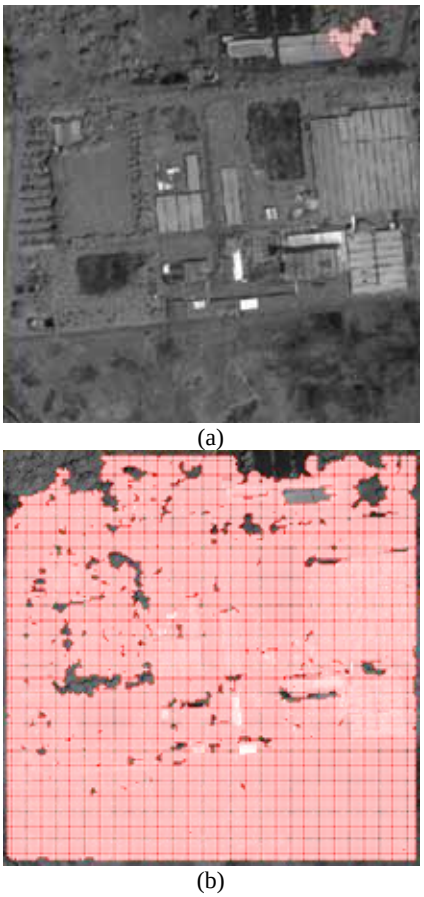
(b)

Figure 6. Typical identified corresponding points over a forested region (a) starting from a single manually selected seed point and (b) by the hybrid method

Figure 6 shows typical results. In all cases the attained coverage by corresponding points was very small. Note that the single group of corresponding points shown in figure 6a, which started from a single seed, does not appear in figure 6b, that was produced by the proposed hybrid method. This is because no key point has been found by SIFT over there. In fact, the SIFT algorithm also performs poorly over forested regions. Generally, in relation to the single-seed point approach the proposed hybrid method does not bring a significant gain in terms of coverage over forest regions.

4.2 Rural Regions

Figure 7a and 7b show results obtained on a rural region. In both cases the algorithm started on a single, but different initial seed point. The results are remarkable different. In the case of figure 7a the seed point was placed within a region surrounded by trees, water bodies and big homogeneous image objects that confined the region growing process to a small area. In the case of figure 7b most of the image was reached just based on one seed point. The barrier that stopped the process in the previous case was overwhelmed on some point, and even the spot shown in figure 7a has been reached. This is a typical example showing that the result of least squares matching with region growing is highly dependent on the selected seed point.
The result achieved by the hybrid method on that sub-image was visually indistinguishable to the case shown in figure 7b. Yet, the algorithm failed to find corresponding points over clusters of trees, water bodies and large homogeneous objects, due to intrinsic limitations of least squares matching.



(a)

(b)

Figure 7. Typical identified corresponding points over a rural region starting from a single manually selected seed point in the worst (a) and in the best (b) case; results from the hybrid model are quite similar to (b)

4.3 Residential Regions

Figure 8 shows results obtained over a residential region. Figures 8a and 8b show quite different results, although both are based on a single, but different manually selected seed point. As before, this shows again that the results of least squares matching with region growing is strongly depending upon the selected initial seed point.
Figure 8c corresponds to the hybrid method, which is considerably better than the result shown in figure 8b.
Again, in this case the least squares matching with region growing did not advance over groups of trees and water bodies.

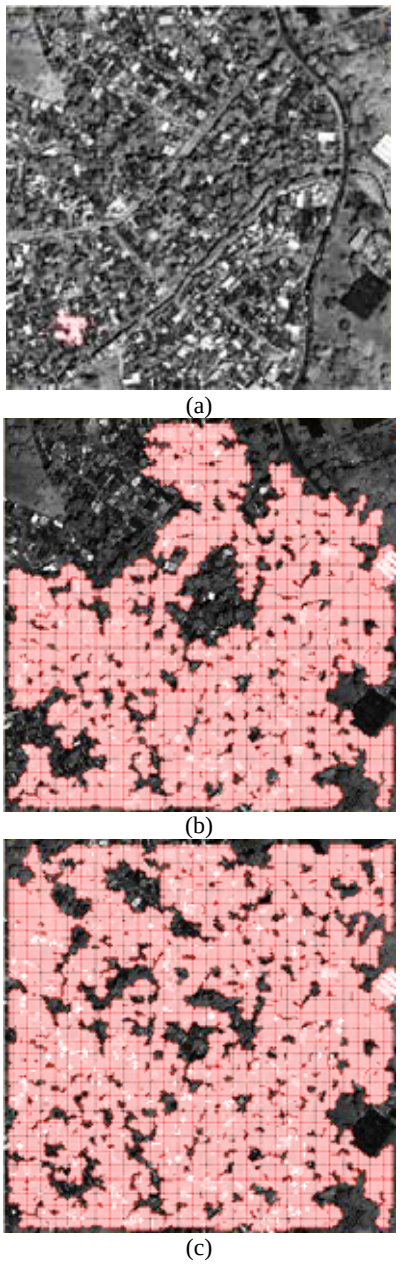


Figure 8. Typical results over a residential region starting from a single manually selected seed in the worst (a) and in the best (b) case; results from the hybrid model (c)

4.4 Regions with High Buildings

As mentioned before, the large height changes may cause significant differences of corresponding images. Finally the performance of the proposed method over areas with high buildings has been analyzed.

Among the four region types considered in this analysis so far, these are the regions that are expected to benefit most from the proposed method. One of the five sample sub-images used in our experiments for that purpose is shown in figure 9a. In total 24 high buildings are included. Figure 9b shows the overlay of corresponding points obtained from a single initial seed point to the image. It can be observed that the least squares matching with region growing method fails at façades, and seldom reaches the top of buildings.

Figure 9c shows the result obtained by the hybrid method. On the first view the overall coverage seems not to be quite different to figure 9b, but nearly all building tops are included. This leads to a quite better description of the achieved digital surface model. Of course also in figure 9c gaps are included, but mainly caused by the fact, that the different images show different facades of the same building.

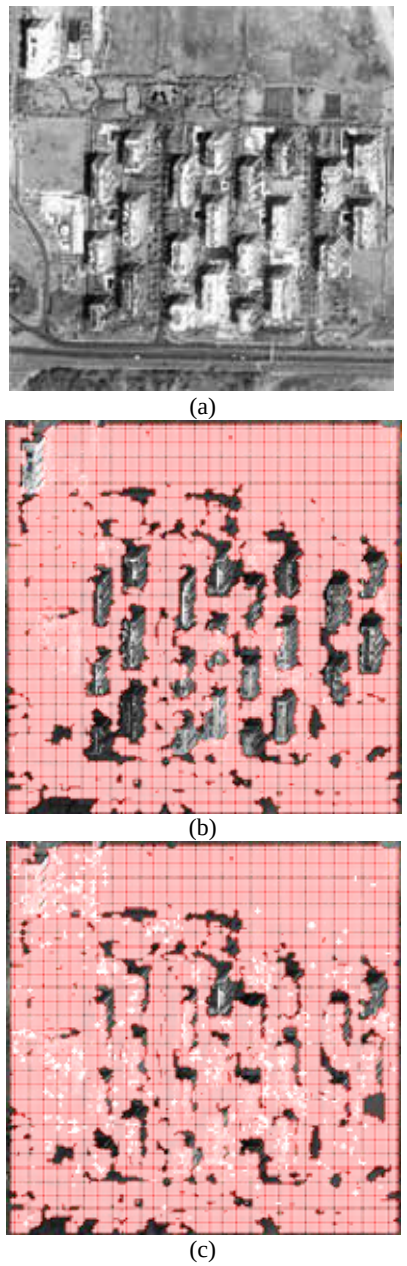


Figure 9. Sub-image containing high buildings (a); results obtained starting from a single manually selected seed (b); results from the hybrid model (c)

Table 2 presents the number of buildings in each of the five sub-images used in this series of experiments, as well as the number of building tops that have been reached by the manual and by the hybrid method.

These results indicate that the proposed method using the support by SIFT is very effective in providing a dense coverage of corresponding points in the presence of sudden height changes.

Sub-image	number of buildings		
	total	reached from a single seed point	reached by the hybrid method
1	25	4	18
2	21	5	16
3	14	0	7
4	17	6	10
5	24	0	22
average	100 %	15 %	72 %

Table 2 –building tops reached by each method

4.5 Overall Evaluation in Terms of Average Coverage

The graphic of figure 10 shows a summary of the average coverage obtained in 5 experiments conducted on each of the four types of sub-regions considered in the experimental analysis. In all cases the proposed method performed better than the single-seed alternative. For forested areas there was no important performance improvement, since finding conjugate points on such type of areas is difficult in any case. In terms of overall coverage the benefits of the proposed method become more evident in rural, residential and urban areas with high buildings.

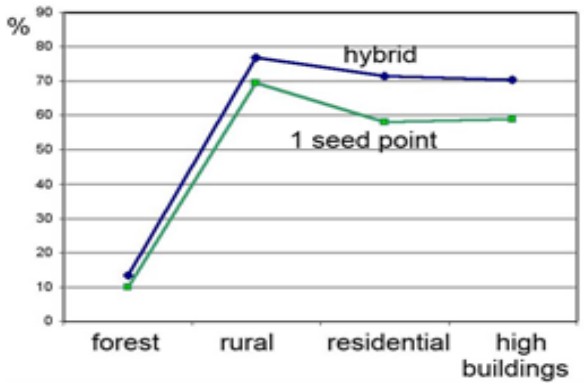


Figure 10. Average coverage obtained by using a single seed point and by the hybrid method

5 CONCLUSIONS

A fully automatic point matching algorithm that combines a feature and an area-based method is proposed. The algorithm first applies SIFT to provide a set of seed points used in a subsequent step by least square matching with region growing, so as to provide a dense and well distributed cloud of corresponding points on a pair of stereoscopic images with virtually no human intervention.

In tests performed on rural and residential areas as well as on areas with high buildings the proposed method always outperformed the least square matching with region growing initiated based on a single seed point. Particularly significant is the ability of the proposed method demonstrated in our tests to reach the tops of high buildings. Over forest areas the hybrid method was also better than the single-seed point approach but was still not able to produce a dense distribution of points, mainly due to the missing image similarity.

The tests have also suggested that the SIFT algorithm could be simplified for this kind of applications. Since the scale and the orientation of both stereo satellite images are quite constant over the images and usually known in advance, some steps of the basic SIFT algorithm designed to provide invariance to scale and rotation may be probably dropped or at least simplified without significant performance impact. An investigation about these possibilities is planned for the continuation of this research.

REFERENCES

David Lowe, 2004. Distinctive Image Features from Scale Invariant Key points. *International Journal of Computer Vision* pp. 91-110

Gruen, A.W., 1985. Adaptive Least Squares Correlation: A powerful Image Matching Technique. *South Africa Journal of Photogrammetry Remote Sensing and Cartography,* pp. 175-187.

Harris, C.; Stephens, M., 1988. A combined corner and edge detector. *In Fourth Alvey Vision Conference*, Manchester, UK, pp. 147-151.

Jacek Grodecki, 2001. Ikonos Stereo Feature Extraction – RPC Approach. *Proc. ASPRS Annual Conference*, St. Louis, pp. 23-27.

Jacobsen, K.; Büyüksalih, G., 2001. Determination and improvement of digital elevation models based on moms-2P imagery, *Turkish-German Geodetic Days, Berlin*.

Mikolajczyk, K.; Schmid, C., 2005. A performance Evaluation of local descriptors. *IEEE.*

Mikolajczyk, K.; Schmid, C., 2004. Scale & Affine Invariant Interest Point Detector. *International Journal of Computer Vision*, pp. 63-86.

Otto,G.P; Chau,T.K.W. 1989, Region-growing algorithm for matching of terrain images. *Image and Vision,* V. 7, N.2, pp. 83-94.

Trucco, E.; Verri, A., 1998, *Introductory Techniques for 3-D Computer Vision*, Prentice Hall.

ACKNOWLEDGEMENTS

The authors would like to thank the Brazilian National Council for Scientific Research (CNPq) and the German Aerospace Agency (DLR) for the financial support for this investigation.