RETURN

MULTI-IMAGE MATCHING FOR ARCHITECTURAL AND ARCHAEOLOGICAL ORTHOIMAGE PRODUCTION

Petros PATIAS, Vassilios TSIOUKAS
The Aristotle University of Thessaloniki
Department of Cadastre Photogrammetry and Cartography
Univ. Box 473, GR-54006, Thessaloniki, Greece
e-mail: patias@topo.auth.gr, tsioukas@egnatia.ee.auth.gr

CIPA-W G 4

KEY WORDS: multi images, matching, DSM, orthoimage, Architecture, Archaeology

ABSTRACT:
In close-range architectural and archaeological photogrammetric applications, very often, amateur cameras are used. While images are suffering from excessive distortions, the objects themselves exhibit remarkable anaglyph (walls, excavation holes, 3D objects, etc.), prohibiting thus the use of simple planar rectification. The 3D reconstruction of such objects unavoidably requires the production of 3D Digital Surface Models and the subsequent production of orthoimages. But then, the main problems we are faced with, are the big number of images, their arbitrary convergence, the need for automation in object reconstruction, in often cases repeated patterns (eg. stones, bricks, soil), unfavorably low image texture, etc. All these factors create a very unfavorable situation for the application of automatic image matching techniques. Moreover, the matching algorithms in commercial photogrammetric systems are mainly referring to the aerial case, where most of the above problems do not exist. For obvious reasons, also, they do not provide for multiple image matching.

This paper deals with the development of a flexible multi-image matching technique under the above mentioned situations. Its novelty is placed in the development of the mathematical models used but mainly on its ability to cope with harsh situations, often met in architectural and archaeological applications. In order to show the algorithm's performance extreme but real examples are shown, where subpixel accuracy is achieved.

1. INTRODUCTION

The orthoimages are very important in Photogrammetry as they can offer reliable metric information of objects especially in aerial cases. Their use in architectural photogrammetry is very restricted and they are barely used for documentation. In most cases of architectural photogrammetry the reliable metric information of the object is provided by the rectified images of the surfaces of the facades. But in some cases it is impossible to use rectified images due to the complexity of the object's facades. So the use of ortho images is unavoidable.

The most difficult part for the creation of an orthoimage is the extraction of the DTM that will be used for its creation. Although the automatic extraction of the DTM is not a problem at all for the aerial case, it is still a problem in the terrestrial case. The problem is mainly due to the inability of the creation of epipolar images in close range cases. The big rotation angles of terrestrial images are prohibiting the creation of epipolar images in Digital Photogrammetric Stations. The terrestrial stereo images have to be just like in the aerial case, as vertical as possible, in order to be inserted in the DPS. The situation however, rarely is achieved in terrestrial cases. In order to be able to take ideal images as required by a DPS, we must take a picture from long distance. But then the image scale would be very small and the expected accuracy would fall very much. Besides, lack of space or presence of obstacles prevent such setup, as well. It is, thus, unavoidable to take pictures of terrestrial objects that are very much inclined. Then DPS's are unable to produce epipolar images and proceed thereafter with image matching, for the DTM production.

The solution of the above problem would be the external formation of epipolar images and the subsequent insertion of them into the DPS for further processing. This way, one would take advantage of the full software support of a DPS, while externally could solve the problem of big inclination angles. Of course, The problem of the formation of epipolar images still needs a solution, and our proposal follows the next steps:

1. Interior orientation
2. Relative orientation through Gruber points measurements
3. Formulation of epipolar images, considering big rotation angles.

However, in terrestrial case, we usually have more than two photographs, another side-effect of big rotation angles.

These multi-images should certainly be taken into consideration since they offer valuable degrees of freedom, leading to blunder detection especially in heights (relief). Thus the formation of epipolar images should be bundled with a multi-image approach.

The additional steps should be:

4. Repetition of step 3 for any additional image
5. Repetitive exterior orientation computation of the formed stereo pairs
6. Multiple intersection for each computed DTM point for blunder detection.

Another drawback in producing DTM from terrestrial architectural and archaeological objects is the appearance of regularly repeating patterns all over the object's facades. These patterns could very easily lead to incorrect matching of points and the generated DTM suffers from the existence of many blunders. In this case, the present proposal suggests the use of an affine transformation that defines the appearance of the repeating patterns all over the facade of the object in both images of the stereo pair. That way template matching of the specific pattern is performed at those positions of the images that are defined by an affine transformation given by the user.

After that, the DTM is extracted and the creation of the orthoimage is an easy process. The orthoimage can be created and printed at any scale depending on the aim of the documentation. It is also possible to create 3D drawings of the object using the orthoimage and the underlaying DTM (used for the creation of the orthoimage). The planimetric information of the image features are obtained from the image coordinates of the orthoimage, while their height value is easily calculated through interpolation of heights from the DTM.

2. RECENT RESEARCH

In most cases of architectural objects the objects' facades are plane surfaces and with the use of a single image it is possible to create rectified images with a minimum number of 4 control points. The mosaicing of many rectified images of a single facade is possible with the use of many commercial photogrammetric packages. In other cases the objects' facades form an analytical surface of a cylinder or a cone (Karras-Patias-Petsa, 1996). In such a case we can take advantage of the equation of the analytical surface that is satisfied from all the points that lie on it and the creation of the unwrapping of the cylindrical surface from a single exterior oriented image is possible. In cases where the surface of the building forms a spherical surface (which can be generalized as an ellipsoid) the creation of a special projections (e.g. mercatorian projection) can be used to obtain metric information for the object.

However in a general approach of a building the object's facade does not form a regular shape such as a plane surface, a cylindrical or conical shape, or a part of sphere. Complex 3D structures can be automatically extracted from stereo pairs, using image matching techniques. It is even possible to use more than 2 images of a terrestrial object to detect and eliminate possible blunders of the generated DTM grid points.

Image matching for DTM generation of terrestrial architectural and archaeological objects is not a usual practice. The main reason is that there are many drawbacks when using terrestrial images for that purpose, as described in the previous paragraph.

The recent research is based on multi image matching and the extraction of DTM from objects in industrial applications. In industrial applications, however, Shao and Fraser (1998) introduced an automatic reconstruction of a digital multi-ray photogrammetric network of images. During this reconstruction a multi-image feature based matching technique was used.

Maas (1996) proposed the use of many images and the use of each one of them as a fixed reference in the search of conjugate points. The proposed methodology extracts interest points from the reference image and uses the intersection of epipolar lines in the other images to extract their exact matching. Range constraints of the ground coordinates of matched points are used to eliminate blunders and an accumulator is used to count the hits of successful matching of each point.

3. PREPROCESING OF IMAGES

Consider the case of a terrestrial application which is comprised of a stereo pair. The relative orientation is achieved through Gruber points' measurement. The right image is considered as reference and using their known relative orientation the epipolar image of the left is formed.

The main purpose for the creation of the epipolar image is the convenient search of conjugate point from the one of the images onto the other. The conjugate of a point on the reference image lies along the same scanline of the epipolar image. Below is shown a pair of images of a church in a suburb of Thessaloniki. (Fig.3 and Fig 4)

The images were taken with an uncalibrated Sony Mavica digital camera, and the calculation of the exterior orientation was produced using the 3D Builder software. The calculation of the exterior orientation was rather rough, because only one measured distance and several parallelity and perpendicularity conditions were used. There were no restrictions of the attitude of the camera during the procedure of photographing. The only restriction was the full coverage of the object which leads to an almost 100% overlap of the images.

4. INTEREST OPERATOR

The success of the matching between parts of the epipolar images depends very much on the texture of the image window that will be used as a prototype from the original image. This window will be called template window and its central point will be used as an intersection point. Its conjugate will be searched along the same scanline of the epipolar image and this pair of observations will be used by an intersection process to calculate the ground coordinates of the imaged point.

Since richness of texture is important, the points to be used as potentially matched points have to be extracted from one of the images through an interest operator evaluation process. The most commonly used interest operator is the Foerstner operator.

The advantages of the Foerstner interest operator is its fast speed and great accuracy in finding interest points that are centers of circular features or intersections of lines. Considering an image window the roundness (q) and weighting (W) can be found:

$$q = \frac{4DetN}{(TrN)^2} \qquad (1)$$

$$W = N^{-1} = \begin{bmatrix} gu^2 & gu \cdot gv \\ gu \cdot gv & gv^2 \end{bmatrix} \qquad (2)$$

where gu and gv are the derivatives of the grey values of image pixels across the rows (u) and columns (v) of the window. If the values of W and q are greater than some thresholds and are extreme maxima in the neighbourhood, then the window center is an interest point (center of circular feature or intersection of lines).

In many cases of architectural or archaeological objects there exist no points that are centers of circular features or intersections of lines namely Foerstner points. We then have to use other characteristic points along the epipolar lines. These are generally edges. These edges should be perpendicular to the scanlines of the images and their gradient is parallel to the scanlines. We could also accept points whose gradient angle is a few degrees more or less than the angle of the epipolar lines (after the creation of the epipolar image the angle of all epipolar lines is known and is equal to 0°).

With this operator, points are extracted that are centers of window images (which dimensions are defined by the user) with rich texture in the center or across or even diagonally to the direction of the window. The grey value of gradient angle at that point have to be parallel the epipolar line or with a little variation from parallelity.

This algorithm is even more faster than the Foerstner operator and extract points that have good texture across the epipolar lines, and can be applied to the whole or part of the original or epipolar image.

5. CROSS CORRELATION

After the extraction of all good points with rich texture in the original image we are ready to find all their matches in the scanlines of the epipolar image. This is done with the application of the least squares image matching technique, which is described in the following paragraph. The success of the algorithm of image matching depends on the prior good knowledge of the initial values of the correct position of the interesting points of the first image (extracted by the previously described algorithm of interest operator) in the other image. In order to succeed in the matching of the template window of the original image and a neighboring window of the epipolar image (called patch window), conjugate of a point must be known to an accuracy of at least 2-3 pixels.

This leads to the use of the algorithm of cross correlation. For each template window of interest points we produce the cross correlation coefficient between the grey values of the window and the grey values of the windows of all the points in the epipolar line, using the following formulas:

$$\boldsymbol{r}_{12} = \frac{\boldsymbol{s}_{g1g2}(u,v)}{\boldsymbol{s}_{g1} \cdot \boldsymbol{s}_{g2}} \qquad (3)$$

$$\boldsymbol{s}_{g1g2}(u,v) = \frac{1}{m-1}\Big[\sum_{i=1}^{m} g_1^2(r_i - u, c_i - v) -$$

$$g_2(r_i, c_i) - \frac{1}{m}\sum_{i=1}^{m} g_1(r_i - u, c_i - v)\sum_{i=1}^{m} g_2(r_i, c_i)\Big] \qquad (4)$$

$$\boldsymbol{s}_{g1}^2(u,v) = \frac{1}{m}\Big[\sum_{i=1}^{m} g_1(r_i - u, c_i - v) -$$

$$\frac{1}{m}\big(\sum^{m} g_1(r_i - u, c_i - v)\big)^2\Big] \qquad (5)$$

The point of the epipolar line that gives the largest cross correlation factor $\boldsymbol{\tilde{n}}_{12}$ is closer to a candidate of the matching point in the epipolar image. This point should lie in the same scaline of the epipolar image. However in some cases where the relief of the object is big the conjugate of a point could lie in a band of scanlines. This is the reason why the search is done, not only on the same scanline but in a band of lines of the epipolar image.

The cross correlation algorithm gives good results only when the images do not suffer from difference in scale and rotation. This is almost our case because any scale and rotation effects between the images have been removed after the creation of the epipolar image.

6. LEAST SQUARES MATCHING

The accuracy of the cross correlation is within the pixel size. Such an accuracy may be adequate in some cases, but in the majority of the cases, this is only an approximate solution, which serves as initial values to the next step: least squares image matching. Least squares image matching can be considered as a generalization of cross correlation.

The algorithm is performed with the use of a geometric transformation between pixels of a window in the original image (template) and a window from the epipolar image (patch). The geometric transformation used is a generalized affine transformation and the minimization of the difference of the grey values of patch and template windows lead to the determination of its parameters. Each pixel of the template window with coordinates (x,y) is associated with its corresponding pixel in the patch window with the following equations:

$$x = a_1 + a_2 \cdot X + a_3 \cdot Y$$

$$y = b_1 + b_2 \cdot X + b_3 \cdot Y \qquad (6)$$

The difference of the grey values of the pixels of patch and template windows is:

$$G(X,Y) - e(x,y) = g(x,y) \qquad (7)$$

where G(X,Y) is the grey value of pixels in the patch window, g(x,y) is the grey value of pixels in the template window and å(x,y) their remaining differences to be minimized with the use of a least squares procedure.

After linearization and substitution we derive the observation equations for the adjustment of the geometric parameters ($a_1, a_2, a_3, b_1, b_2, b_3$).

$$G(x,y) - e(x,y) = g^0(x,y) + g_x(x,y) \cdot dx + g_y(x,y) \cdot dy \qquad (8)$$

$$\begin{aligned} G(x,y) - e(x,y) = g^0(x,y) + g_x(x,y) \cdot a_1 + g_x(x,y) \cdot a_2 \cdot X + \\ g_x(x,y) \cdot a_3 \cdot Y + g_y(x,y) \cdot b_1 + g_y(x,y) \cdot b_2 \cdot X + \\ g_y(x,y) \cdot b_3 \cdot Y \end{aligned} \qquad (9)$$

and in matrix notation we have:

$$-e = A \cdot x - l \qquad (10)$$

where

A is the design matrix and x the vector of unknowns:

$$x^T = [a1 \quad a2 \quad a3 \quad b1 \quad b2 \quad b3]. \qquad (11)$$

The corrections of initial values are:

$$x = \left(A^T \cdot P \cdot A + P_x\right) A^T P l \qquad (12)$$

The P matrix is the matrix of the weights of observations, and P_x is the matrix of weights, which defines our prior knowledge of the unknown parameters. Since the matched windows are referring to epipolar images the scale and rotation parameters are very small. Therefore our choice of the weight matrix P_x is:

$$Px = diag[0 \quad 0 \quad p \quad p \quad p \quad p] \qquad (13)$$

where p is a big number e.g. 10^{13}, and the off-diagonal elements are all zeros. The initial values of unknowns are:

$$x^T = [0 \quad 0 \quad 1 \quad 0 \quad 0 \quad 1] \qquad (14)$$

7. CALCULATION OF GROUND COORDINATES

After the determination of the exact location of the conjugate point in the epipolar image we back project it onto its corresponding original image. Then by intersection the ground coordinates of the imaged point are computed.

There is always a possibility of a false match and therefore an automatic blunder detection is indispensable. To this end the use of multi-image matching is quite desirable.

8. MULTI IMAGE

Having more than two images of an object is possible to construct epipolar images for each one of them. In our algorithm the insertion of more than two epipolar images and the application of the previously described steps is possible. So having first extracted the interest points, calculation of conjugates points in each epipolar image is achieved, leading to the computation of ground coordinates through multiple intersection, while a blunder detection process takes care of the mismatches.

Possible mismatches are recovered only through consistency checks on individual intersections. If exterior orientation is known, or if other external constraints can be imposed (e.g. parallelity, coplanarity, etc), those possible mismatches can be actually evaluated. Otherwise the matched points are marked in different color, for the user to interactively make this evaluation.

9. REPEATING PATTERNS

In many architectural and archeological objects repeating patterns may occur. A typical example of such a pattern is easily observed in the images presented in fig. 1. The joint of the bricks create a T-junction shape which is repeating along the whole building. This pattern is very easily detected in the lighted and nearly flat areas of the building and can be used as interest points for the determination of DTM. It is rather difficult to extract the features in the shaded or inclined areas, and the multiple occurrence of the pattern along the epipolar image can lead to mismatches of intersection points. However, the information of the repetition frequency of the pattern can be used in order to define conjugate points more easily along the epipolar lines of images.

The proposed algorithm estimates an affine transformation that connects corresponding points of repeating patterns in the images and applies a geometrically constrained least square image matching between them.

The estimation of the affine transformation that relates a point to its conjugate onto the other image can be easily inserted with the manual collection of at least 4 pairs of conjugate points in those images. With the manual collection of another pair of conjugate points it can also be determined the step of repetition. Fig. 1 is showing the collection of 4 pairs of conjugate points for the determination of the affine transformation between the images. Another point (numbered 999) gives the step of repetition.

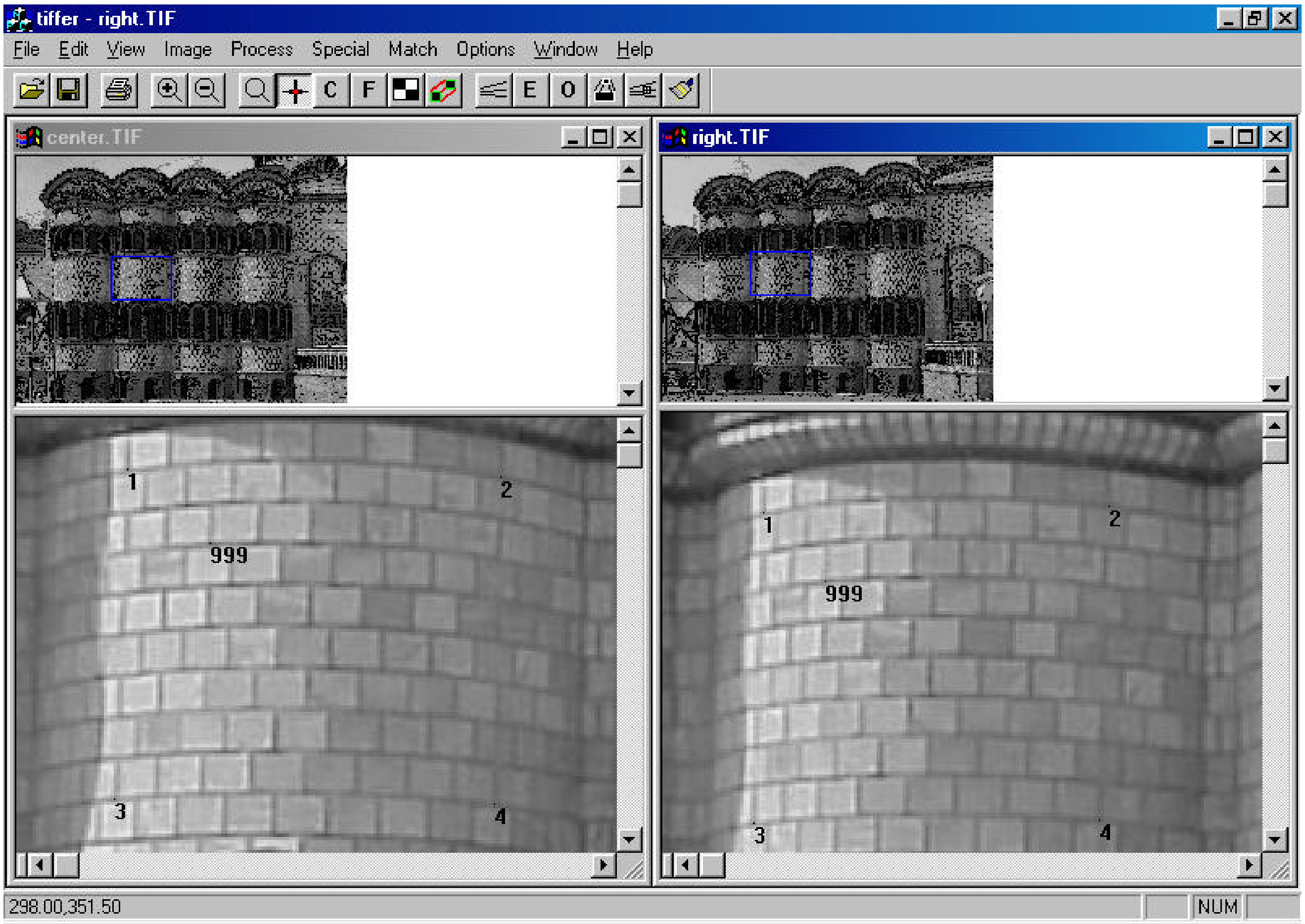

Fig. 1. *Example of manualy collected repeating pattern points*

The definition of the relative orientation of the images and the estimation of the affine transformation that connects the conjugate points in the images, constrains the search of these conjugate points in a certain area of the images. The algorithm is very fast and efficient especially in cases with low relief.

10. ORTHOIMAGE CREATION AND DIGITAL MONOPLOTTING

The orthoimage creation is an easy process after the extraction of the DTM using the one or the other method. In order to create the orthoimage a regular grid of DTM should be geberated. This has been done with the use of a commercial software package (Surfer) and the distance between the gird points was 2 cm. The same value has been used for the size of the pixel of the orthoimage. A small part of this orthoimage is shown in fig. 2. The distortions appering on the image is a side effect of the nearest neighbor algorithm that was used to create the regular grid of the DTM points.

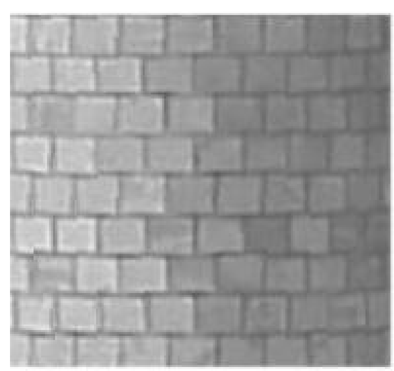

Fig. 2 *Orthoimage of the allmost cylindrical part of the facade*

The center image of our example has been used for the creation of the ortho image as it is less distorted than the other images. The created ortho image along with the generated DTM can be inserted in our program and a special measuring mode can be used in order to obtain metric information of the imaged points. The 3D coordinates of every point can be extracted and the tracing of the linear features can be saved in a DXF drawing file. For further processing it is possible to use the DXF with another commercial software package (AutoCAD) and produce a hardcopy output of the object's 3D model. In this way 3D drawings of architectural or archaeological objects can be created, with the use of just one image.

11. RESULTS

The DTM of a region of the almost cylindrical facade has been manually extracted (Fig. 8). A total number of 98 points were calculated. This reference DTM was compared with the DTM which is automatically extracted through matching of one (Fig. 9) or two (Fig. 10) pairs of epipolar images and with the DTM created using epipolar pattern matching (Fig. 11). The results from this comparison are presented in the table below. In figures 12,13 and 14 the difference of the reference DTM and the automatically created ones are shown.

	1 pair of images	**2 pairs of images**	**Pattern matching**
Mean error (cm)	6	3	5
Std Dev. (cm)	10	5	6
Max. error (cm)	34	12	13
Number of Matched pts	85	68	25

The precision of the DTM is greater in the front part of the surface. The left and right parts of this cylinder are nearly perpendicular to the image axis and are very much deformed between the most left and most right images. That is why only a small number of points could be matched in these parts of the object.

12. CONCLUSIONS

The aim of this project is the creation of a handy tool for helping not only photogrammetrists but also architects and archaeologists, to automatically produce an orthoimage of a terrestrial object of archaeological or architectural interest.

During our research:

- We managed to create from stereo pairs (and multi images) epipolar images although rotation angles are large.
- We implemented two kinds of interest operators (the Foerstner operator and the one that extracts edges perpendicular to epipolar lines).
- We used least squares image matching for better accuracy in finding ground coordinates of conjugate points.
- We took advantage of multiple images to automatically reject blunders.
- We overpassed the limitations of DPS's and created DTM from archaeological or architectural close range objects.
- We have implemented a method to avoid pattern mismatch when there is a pattern repetition on an object's facade.
- We created orthoimages and used them for restitution using digital monoplotting techniques.

References

Heipke, C., 1992: A global approach for least-squares image matching and surface reconstruction in object space. *Photogrammetric Engineering and Remote Sensing*, 1992, 58(3), pp. 317-323.

Karras, G., Patias, P., Petsa, E., 1996: Digital monoplotting and photo-unwrapping of developable surfaces, *Intern. Archives of Photogrammetry & Remote Sensing*, Vol. XXXI, Part B5, pp. 290-294.

Maas, H.-G., 1996: Automatic DEM Generation by multi-image feature based matching. *International Archive of Photogrammetry and Remote Sensing*, 31(B3), pp. 484-489

Shao, J., Fraser, C., 1998: Multi image matching for automatic object reconstruction. *International Archive of Photogrammetry and Remote Sensing*, 32 (B5): pp. 148-154.

Tsioukas, V., Patias., P., 1999: A new approach on image matching techniques for the automatic production of DTM in terrestrial photogrammetry, *Intern. Archives of Photogrammetry & Remote Sensing*, Vol. XXXII, Part 5W11, pp. 89-94.

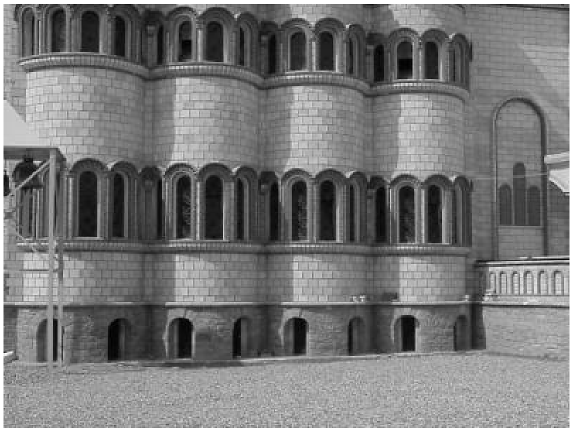

Fig. 3. *Left image of an architectural object*

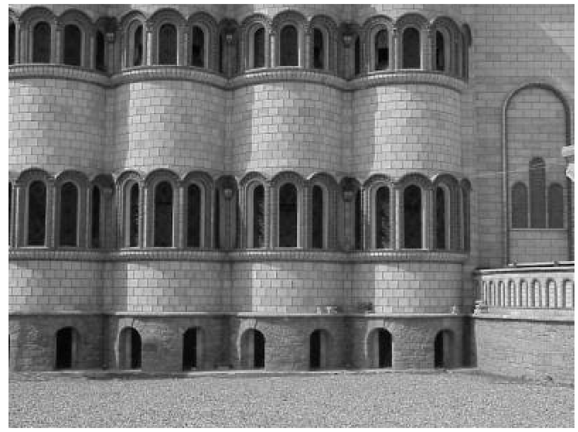

Fig. 4. *Central image of architectural object*

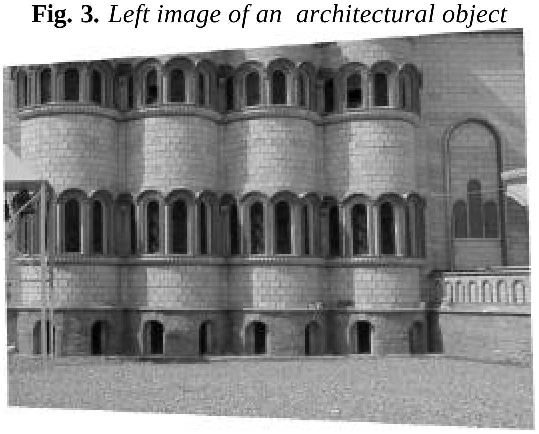

Fig. 5. *Epipolar image of the left original image*

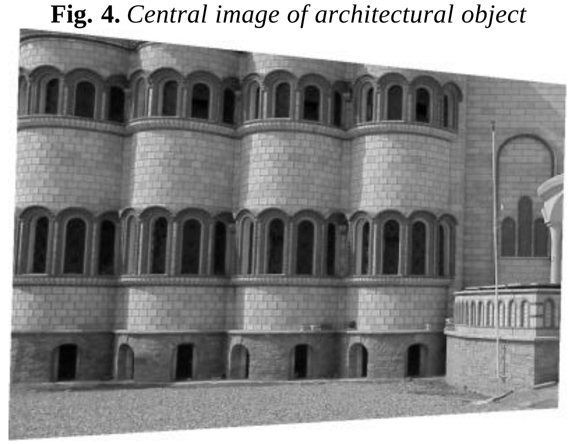

Fig. 6. *Epipolar image of a third photograph (right most original position)*

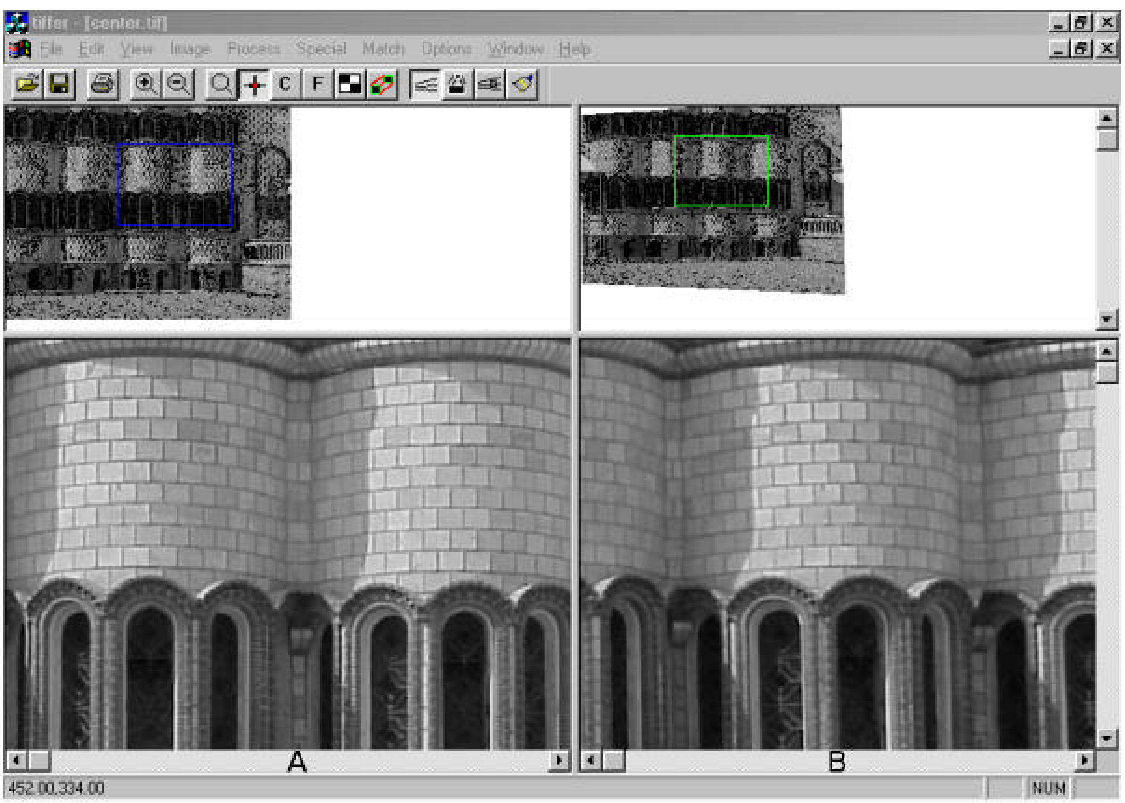

Fig. 7. *Projection of both reference(**A**) and epipolar (**B**) image on a single window of the program*

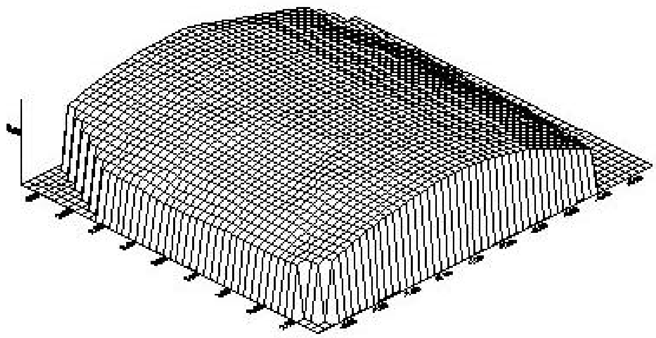

Fig. 8 *Reference DTM*

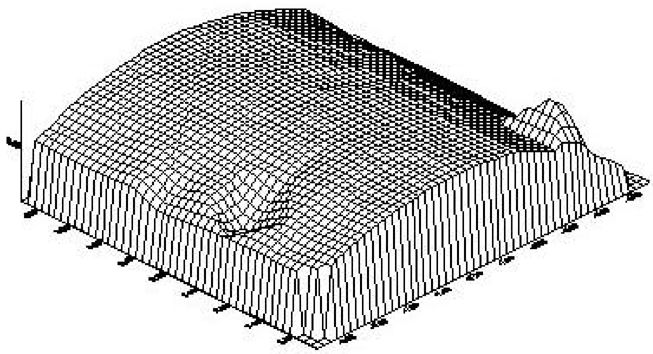

Fig. 9 *Automatically created DTM from a single pair of epipolar images*

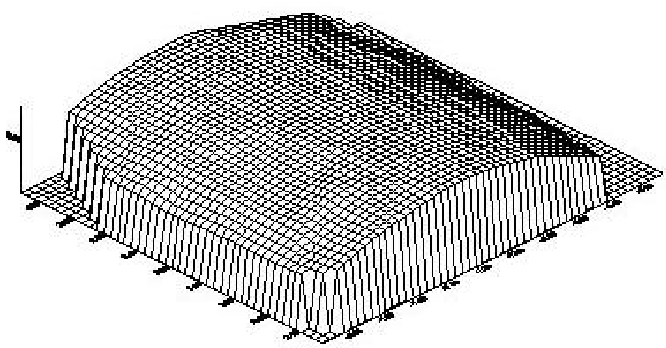

Fig. 10 *Automatically created DTM from two pairs of epipolar images*

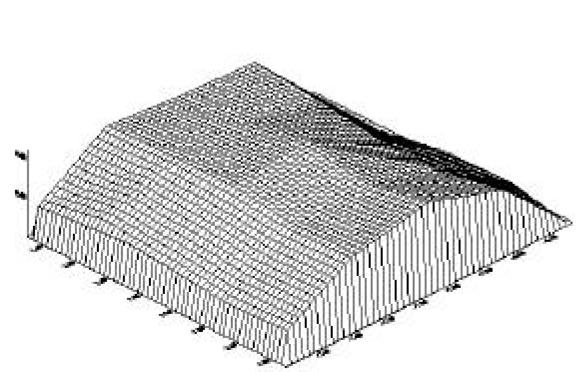

Fig. 11 *Automatically created DTM using repeating patterns*

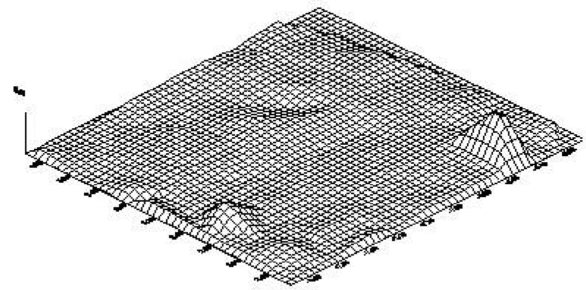

Fig. 12 *Difference of reference DTM from automatically created DTM from a single pair of epipolar images*

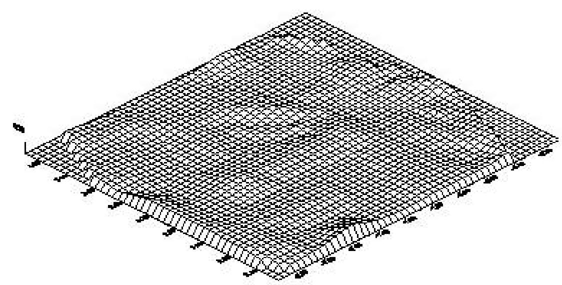

Fig. 13 *Difference of reference DTM from automatically created DTM from two pairs of epipolar images*

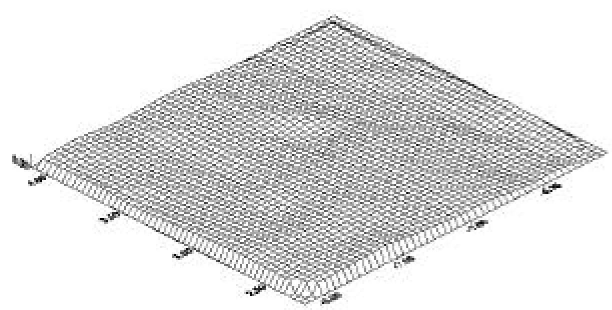

Fig. 14 *Difference of reference DTM from automatically created DTM using repeating pattern*