

SEMIAUTOMATIC 3D MAPPING OF BUILDINGS FROM MEDIUM SCALE (1:40,000) AERIAL PHOTOGRAPHS

Yair Avrahami ^{a*}, Yuri Raizman ^b, Yerach Doytsher ^a

Department of Transportation and Geo-Information Engineering, Faculty of Civil and Environmental Engineering, Technion – Israel Institute of Technology.

Technion City, Haifa 32000, Israel. (yaira, doytsher)@tx.technion.ac.il

^b Survey of Israel, 1 Lincoln St., Tel-Aviv, 65220 Israel .yuring@mapi.gov.il

ThS12

KEY WORDS: digital photogrammetry, semi-automation, building extraction, mapping, small scale, aerial images, GIS

ABSTRACT:

In the last decade, much of the research dealing with 3D building extraction from aerial photographs has been carried out by the photogrammetry and computer vision communities. The increased usage of 3D City Spatial Information Systems and National GISs for control, management and planning, necessitated development of fully or semi-automatic methods for establishing and updating these systems. Most research tries to reconstruct the building space from large (~1:4,000) scale photographs, mainly for establishing or updating the 3D city model systems. However, the research presented in this paper focuses on 3D mapping from medium scale (~1:40,000) aerial photographs, specifically for establishing and updating the building layer in nationwide GIS databases. The algorithm developed for semi-automatic building mapping is based on a 2D approach to solving the 3D reality. The algorithm consists of five consecutive stages: pre-processing, left image operations, height extraction, right image operations and final mapping of the buildings. The first stage is the only stage performed manually in order to achieve specific goals: model solution, photograph preparation and designating the desired building roof. From the second stage onwards, the process is fully automatic. This algorithm can be employed in two ways, either as part of a fully automatic mapping of all buildings in the overlapping area, or as stand alone, enabling a new technology for semiautomatic mapping within a non-stereoscopic environment, without using 3D spectacles. Based on the algorithm, a system for semi-automatic mapping of buildings was developed in order to test its efficiency and accuracy. The results are satisfactory with an accuracy of 0.5m for planimetric measurement and 1m for altimetric measurement.

1. INTRODUCTION

1.1 General Background

When a human eye looks at an aerial image it is usually easy to detect what is the building and location of its corners. This ability consists of the reticular reception stage and information processing by the brain. In automatic mapping procedures, human vision is replaced by the computer. There is an analogy between human vision and computer vision. The reception stage parallels the scanning and saving of data in the computer memory. The information processing stage parallels a set of topology, geometry and radiometry rules, which enable the computer to detect and extract the building. In 3D mapping, these rules are required to simultaneously detect and extract from both images. Formulating these rules is a complex procedure. This complexity stems from the difficulty in formulating stable rules that remain valid for a large variety of images taken under diverse conditions (different orientation, camera, time photographed, season, etc) and suitable for all buildings. In order to surmount this difficulty, it is possible to divide the problem into two independent sequential stages. The first includes an automatic process to find votes (pointers) for the buildings in a single image. The second includes an automatic process to extract the 3D contour building for every vote. The current research focuses on developing a detailed algorithm for the second process and deals with three issues: Identifying the contour building in a single image, extracting categories (i.e., specific areas, scales, objects etc.). In the last decade, much of the research dealt with 3D building extraction

the building contours simultaneously in both images and finally, 3D mapping of the building. Implementation of this algorithm enables mapping a building, while relying on a 2D building vote. This process can be referred to in two ways: either as part of a fully automatic mapping of all the buildings in the overlapping area, or as a separate part that facilitates a new technology for semiautomatic mapping within a non-stereoscopic environment and without using 3D spectacles. Based on this algorithm, a system for semiautomatic mapping of buildings was developed in order to test its efficiency. The experiments were conducted on two residential buildings areas in Tel-Aviv using medium scale images (~1:40,000) scanned at a pixel size of 14 μm . This article presents the algorithm, the experiments and the results.

1.2 Related Works

Nowadays 3D mapping of buildings is carried out manually employing a digital photogrammetric workstation (DPW) or an analytical stereoplotter. The advantage of the DPW environment is the ability to develop automation for photogrammetric assignments. However, full automation of object space mapping is still far from being implementable. Various methods are available in the scientific community, for mapping object space at different automation levels. These methods are limited and designed for specific mapping

from aerial images executed by the photogrammetry and computer vision communities. The increased use of 3D City

Spatial Information Systems and National GIS for control, management and planning, necessitated the development of full or partial automation methods for establishing and updating these systems. In the first system, the aim is to achieve a complete 3D visual presentation of the buildings. This is generally accomplished by using a large scale ($\sim 1:4000$) at a high resolution. The methods of automation for this system are usually based on model fittings (Building Model Schemes) divided into four categories (Tseng and Wang, 2003): polyhedral models, prismatic models, parameterized polyhedral models and CSG models. In contrast, the second system aims to achieve mapping of the 3D building contour (outline) as viewed from the top. Such mapping is mostly accomplished by using a medium scale ($\sim 1:40,000$) and medium resolution. The methods of automation for this system vary and differ at the automation level offered. Usually, the automation level is determined by the point of origin. In the automatic methods the initial votes or the rough location of the buildings are automatically extracted. In some methods, the initial votes or the initial rough locations are 3D when extracted by exploiting the DSM or DEM (Weidner and Forstner, 1995; Cord and Declercq, 2001; Ruther et al., 2002). In other methods, they are 2D when using classification or texture analysis (Kokubu et al., 2001), shadow analysis (Irvin and McKeown, 1989) or finding local maximums in a cumulative matrix of possible votes (Croitoru and Doytsher, 2003). In the semi-automatic method the voting is accomplished manually. Relying on the initial votes or the initial rough location the building contours are extracted in the image space. Michel et al. (1999) suggest that the initial vote would be 2D (i.e., on the left image) and performed manually. The rough location would be spotted by using Region Growing operations on the intensity and disparity images. The exact location and the matching of the images would be carried out using Hough Transform or Snake, according to the shape and the operator's decision. Ruther et al. (2002) focus on building mapping in informal settlement areas and suggest extracting the rough location from the DSM. The exact location is extracted from an orthophoto using the Snake method. We suggest a new approach to mapping the buildings layer for GIS systems. This approach will facilitate a semi-automatic 3D building extraction from medium scale images within a nonstereoscopic environment and without using 3D spectacles. This is done by relying on an initial manual 2D vote.

1.3 Uniqueness of the Research

Usually it is accepted to divide the aerial images into small, medium or large image scales ranging from 1:70,000 to 1:4,000 (Mayer, 1999). This research is unique because it focuses on medium scale ($\sim 1:40,000$) panchromatic aerial images. There is a significant difference between large ($\sim 1:4,000$) and medium ($\sim 1:40,000$) scale images. In the first, the buildings appear clearly, the DSM extracted from those images is detailed, and describes the surface in a credible manner. Therefore, the accuracy of the resulting 3D mapping of the building space is high. In the second, the buildings do not appear clearly, the DSM is less detailed, and describes the surface in an unreliable manner. In this case, the accuracy of the 3D building space mapping is low and it is necessary to map the building contour (outline) only as it appears in a top view.

The considerations for focusing on medium scale images were as follows: at this scale, the images include a vast area which is useful for mapping and updating a large area, and the National GIS is constructed on this scale (Israel-1:40,000, France-1:30,000, etc) and moreover, this scale will soon be available in commercial satellite images. In addition, while there are many studies dealing with large scale automatic mapping, there are fewer studies dealing with the medium scale.

2. METHODOLOGY

2.1 The Algorithms

The inputs of a semi-automatic system for building mapping are two aerial images with an overlapping area. The algorithm consists of five stages (Figure 1): pre-processing, left image operations, height extraction, right image operations and mapping the object space. The first stage is performed manually in order to achieve three purposes, namely, a mathematical solution of the model, image processing and a manual vote on the desired building roof. From the second stage onwards, the process is fully automatic. The second stage includes extraction of the building contour in the left image space. The height is calculated at the third stage by finding the homologue points (in the right image) for each of the left contour vertexes. Now the initial vote can be transferred to the right image space. The right image-building contour is extracted at the fourth stage in the same way as the left contour was extracted in the second stage. At the last stage, the final 3D building contour is calculated using the information achieved in the previous stages.

2.2 Stage 1: Pre-Processing

In order to facilitate a semi-automatic process it is necessary to prepare the environment. This includes scanning and saving the images, finding the model's solution and performing operations on the images to emphasize the mapping object in relation to its background and to achieve radiometric proximity (calibration) between images. At this point, the operator can vote (point on) the building roof. Pointing on the building is performed on the left image within a nonstereoscopic environment and without using spectacles. The input consists of image coordinates of any point on the building roof $n_{cL} = (x_{cL}, y_{cL})$. This manual operation defines the level of automation as semi-automatic. From this point on, the process is fully automatic.

2.3 Stage 2: Left Image Operations

At this stage, the 2D building contour on the left image is extracted by using the Region Growing method (Eq.1). Where: $\bar{I}$ is the average of radiometric values around the vote. σ is the standard deviation of radiometric values around the vote and ρ is the factor of the standard deviation.

$$\begin{aligned} n_i &= (x_i, y_i); \\ \text{if } |I(n_i) - \bar{I}| &< \rho \cdot \sigma \\ \text{then } n_i &\in \text{roof_building} \end{aligned} \quad (1)$$

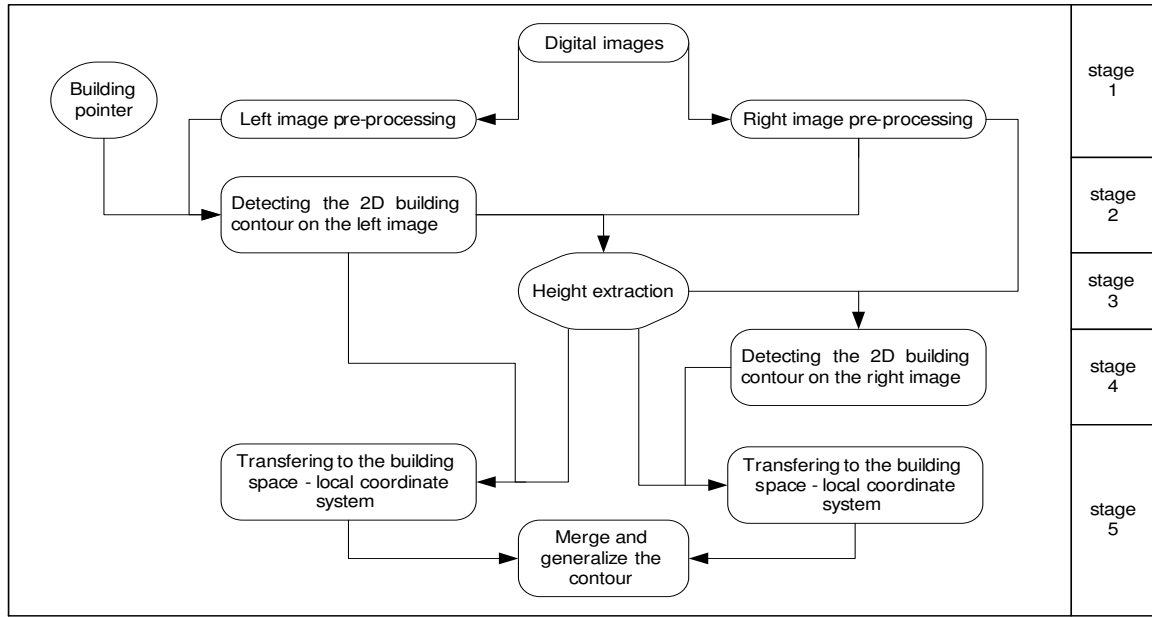


Figure 1: Flow chart of the semi-automatic algorithm.

The initial vote area is expanded to its neighbors (4 or 8) provided that they fulfill the criterion that connects them to the building roof. The building roof is extracted by an iterative process. The values $\bar{I}$ and σ are calculated at every iteration. This process is carried simultaneously out in all directions. A procedure is applied to the extracted building roof by filling the holes and Sobel edge detector. The created roof edge is converted from raster to vector.

2.4 Stage 3: Height Extraction

A procedure for finding the homolog point is performed for every point on the left contour. The search for every point is performed on a specific area limited by geometric conditions: the epipolar-line equation and the possible maximum and minimum building height.

Following is the calculation of the epipolar-line equation (Eq.2) in the right image for each point (x_L, y_L) on the contour in the right image (Thompson, 1968) where A is the correlation matrix:

$$\begin{bmatrix} x_{iL} & y_{iL} & f \end{bmatrix} \begin{bmatrix} A \\ y_R \\ f \end{bmatrix} = 0 \quad (2)$$

All points on the left contour represent edge points and are defined as “interest points”. It would therefore be efficient to use the ABM (Area Based Matching) methods for finding their homologue points. Cross-Correlation Matching (Eq.3) calculates the matching at pixel level, where A is a template from the left image, B is a template from the right image and

$\bar{A}, \bar{B}$ are the averages of A and B . For greater accuracy, the matching at the sub pixel level can be calculated according to Eq.4 which uses Least Square Matching (Ackermann, 1984), where $G_T(X_i, Y_i)$ is the grayscale value in the target window, $G_s(x_i, y_i)$ is the grayscale value in the search window, A is an affine transformation model accommodating the geometric

difference between two images, and h_1, h_2 are used for the radiometric difference modeling.

$$r = \frac{\sum_m \sum_n (A_{mn} - \bar{A})(B_{mn} - \bar{B})}{\sqrt{\left(\sum_m \sum_n (A_{mn} - \bar{A})^2 \right) \times \left(\sum_m \sum_n (B_{mn} - \bar{B})^2 \right)}} \quad (3)$$

$$G_T(X_i, Y_i) = h_1 + h_2 \cdot G_s(x_i, y_i; A) \quad (4)$$

In calculating the building height one homologue point on the contour is sufficient (assuming that the building roof is flat). However, many observations will give a more accurate result. Therefore, the height (Z) is calculated by using the MEDIAN of the heights which has the highest value (over 0.9%) in the criterion for determining the optimal match.

2.5 Stage 4: Right Image Operations

After calculating building height (Z), it is possible to transfer the first vote $n_{cL} = (x_{cL}, y_{cL})$ to the right image in two steps: transferring to a local coordinate by using Eq.5 $(x_{cL}, y_{cL}) + Z \rightarrow X, Y, Z$ and transferring to a right image coordinate $X, Y, Z \rightarrow (x_{cR}, y_{cR})$. It is performed by using the co-linear equation (Eq. 6). From here on, Region Growing can be applied in the same way as in the left image.

2.6 Stage 5: Mapping the Object Space

At this stage the building is mapped according to the information extracted from the images in the previous stages. The mapping is performed in three parts:

The first is transferring the extracted contours to the object space using the extracted height. The second part consists of merging the contours. The advantage is that the information is derived from both images and many hidden details from one image can be complemented by the second.

The difference between the two contours is caused by different photographic conditions and inaccuracy in the model. There are three basic approaches for merging two polygons: polynomial transformation, triangulation transformation and polyline projection (Filin and Doytsher, 1999). The third part consists of generalization operations such as simplification and smoothing, which can be implemented by using a knowledge base, in order to facilitate the final mapping of the building. Application of this part is neither presented nor detailed in this article.

$$\begin{aligned} S &= (Z - Z_0) / (r_{13} \times (x_L - x_0) + r_{23} \times (y_L - y_0) - r_{33} \times f) \\ X &= S \times (r_{11} \times (x_L - x_0) + r_{21} \times (y_L - y_0) - r_{31} \times f) + X_0 \\ Y &= S \times (r_{12} \times (x_L - x_0) + r_{22} \times (y_L - y_0) - r_{32} \times f) + Y_0 \\ x &= x_0 - f \frac{r_{11}(X - X_0) + r_{21}(Y - Y_0) + r_{31}(Z - Z_0)}{r_{13}(X - X_0) + r_{23}(Y - Y_0) + r_{33}(Z - Z_0)} \\ y &= y_0 - f \frac{r_{12}(X - X_0) + r_{22}(Y - Y_0) + r_{32}(Z - Z_0)}{r_{13}(X - X_0) + r_{23}(Y - Y_0) + r_{33}(Z - Z_0)} \end{aligned} \quad (6)$$

3. IMPLEMENTATION AND EXPERIMENTS

In the course of research, a semi-automatic system for building mapping from a medium image scale (~1:40,000) was developed in order to examine the algorithm efficiency. The system enables opening a pair of aerial images, in order to perform a manual vote on the wanted building roof in the left image, mapping the 3D building contour and transferring it to a Geographic Information System in a local coordinate system. Since the aim was to develop a semi-automatic approach for constructing and updating the buildings layer of the GIS, the Israeli national GIS was chosen as the pilot environment and the same conditions used for its construction and updating were retained. The Israeli national GIS has been characterized by Peled and Raizman (1997) as follows: (a) mapping is based on photogrammetric mapping of 1:40,000-scale air photographs by 1st and 2nd class photogrammetric stereoplotters; (b) the planimetric and altimetry accuracies of the mapping are ± 2 meters, suitable to the 1:5,000-scale traditional mapping; (c) the level of mapped details is according to regional mapping at 1:10,000 scale; (d) the DEM is measured at 50 meter resolution. The experiments were conducted on two residential building areas in Tel-Aviv using two medium scale (~1:40,000) aerial images scanned at a pixel size of $14 \mu m$. The first test area was in north Tel-Aviv and included 80 residential buildings, while the second was in central Tel-Aviv and included 97 residential buildings. The chosen test areas were large enough and represented buildings in a flat crowded urban area. The buildings had 4-24 corners and most had few floors and flat roofs. Figures 2, 3 present the 2D building extraction in the left and right images of both areas. Figure 4 presents the semi-automatic building mapping (dark) upon the manual mapping (bright) in both areas.

4. ANALYSIS AND DISCUSSION OF THE RESULTS

Results are analyzed separately, as qualitative results and quantitative results.

4.1 Qualitative Results

In the qualitative analysis, the aim is to evaluate whether the approach is practical, i.e., what percentage of buildings can be mapped using this approach. For this evaluation Eq. 7 was

employed, where BSM is the number of buildings successfully mapped, BPM is the number of Buildings partially mapped, BNM is the number of buildings not mapped and K is a weight for evaluating success in the BPM category ($k=0.5$). Table 1 presents the success percentage in each test area. These results show that a significant percentage (76%) of the buildings was mapped. However, the major innovation is that the operator can see at a glance all buildings that can be mapped using this approach. Therefore, even if the success rate was smaller, it would still be efficient to use this approach initially and complete the mapping by using the traditional method.

$$\text{Buildings_mapping_rate} = \frac{BSM + k \cdot BPM}{BSM + BPM + BNM} \times 100 \quad (7)$$

Test Area	North	Center
BSM	62	66
BPM	6	5
BNM	12	26
Mapping Rate (%)	81%	71%
		76%

Table 1: Success percentage in each test area

4.2 Quantitative Results

The quantitative analysis was based on comparison between the 3D buildings contours extracted using the semi-automatic approach and 3D manual mapping of these buildings made by a professional operator using the same images and same solution model. The deviation vector $d = [dx \ dy \ dz]$ of each building corner from the manual mapping and the closest point on the semi-automatic building contour were measured. Altogether 1444 deviation vectors belonging to 139 buildings in the test areas were measured.

	Test area	dX (m)	dY (m)	dZ (m)
Mean	North	0.22	-0.12	0.44
RMS	728 vertexes	0.72	0.59	1.06
Mean	Center	0.27	-0.19	0.12
RMS	716 vertexes	0.70	0.67	1.35
Relative Accuracy	1444 vertexes	0.71	0.63	1.21

Table 2: mean and RMS of the deviation vectors.

Based on 1444 deviation vectors between the manual and the semi-automatic mapping, the relative accuracy (semi-automatic to manual) was calculated. In Table 2, the mean and the RMS of the deviation vectors in each area are presented.

The relative accuracy between the manual and the semi-automatic mapping is presented in the bottom row of the table. In order to compare the two mappings, the accuracy of the manual mapping should be calculated. This absolute accuracy (including orientation errors) can be evaluated according to Lobanov (1984) using Eqs. 8, 9, where: m is the image scale (1:41,892), $m\mu$ is the accuracy of the photogrammetric measurements (10 μ m), f is the focal length (153.941mm), and b is the base line (92mm). However, in the current research, both mappings were performed by using the same model solution, thus it is the measurement accuracy, rather than absolute accuracy, that was calculated using Eqs. 10, 11. The semi-automatic mapping accuracy can be calculated according to the manual accuracy and the relative accuracy using Eq. 12 and the results were $M_x = \pm 0.57$ m, $M_y = \pm 0.46$ m, $M_z = \pm 0.98$ m.

5. CONCLUSIONS

As can be seen, the planimetric accuracy of the semi-automatic extracted buildings was $M_x = \pm 0.57$ m and $M_y = \pm 0.46$ m. The altimetric accuracy was $M_z = \pm 0.98$ m.

The results show that using the developed semi-automatic approach has achieved numerous advantages, namely: 1) extracting 3D building contours from medium scale aerial images for mapping or updating the building GIS layer; 2) no limitations on specific building shapes or models; 3) this approach enables mapping a high percentage of the buildings; 4) mapping procedure is carried out within a nonstereoscopic environment and without using 3D spectacles; 5) using this approach increases efficiency, is more economical and reduces the building mapping work; 6) the operator can identify at a glance which buildings can be mapped by this approach.



Figure 2: North Tel Aviv - left and right image building extraction.



Figure 3: Central Tel Aviv - left and right image building extraction.

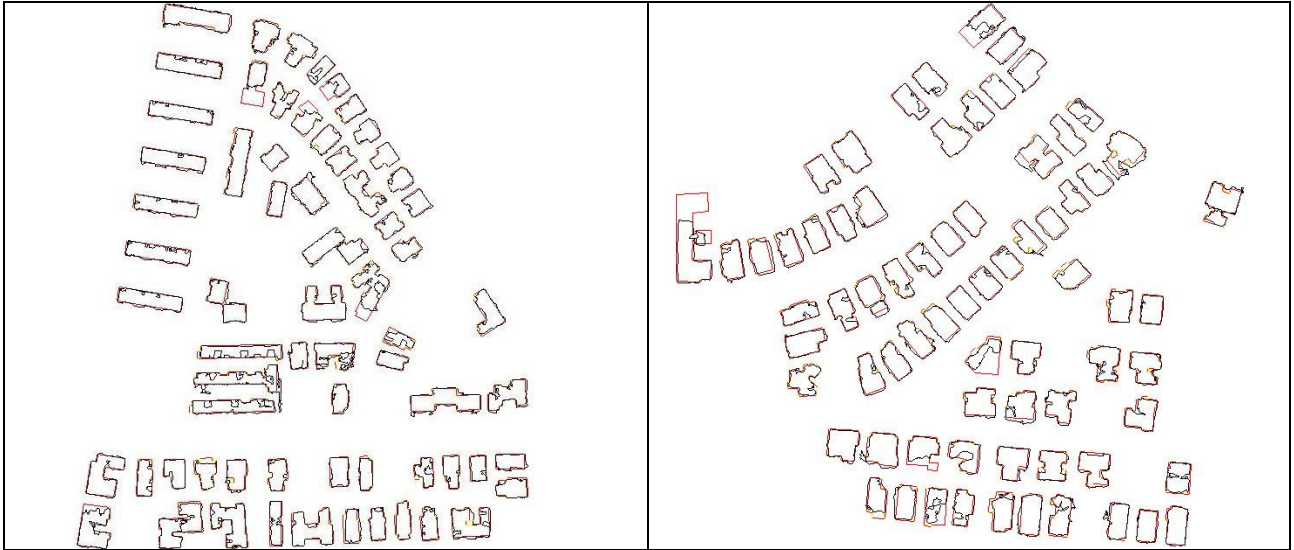


Figure 4: The manual mapping (bright) and the semi-automatic mapping (dark) are presented in local coordinates system.

$$M_x = M_y = 2.5 \times m \times m_q = \pm 1.04m \quad (8,9)$$

$$M_z = 2.3 \times m \times m_q \times f / b = \pm 1.61m$$

$$M_x = M_y = \frac{Z}{f} \times m_q = m \times m_q = \pm 0.42m \quad (10)$$

$$M_z = m \times m_q \times \frac{f}{b} = \pm 0.70m \quad (11)$$

$$M_{semi-automatic}^2 + M_{manual}^2 = M_{relative}^2 = \frac{\sum d^2}{n} \quad (12)$$

6. REFERENCES

- Ackermann, F., 1984. "Digital Image Correlation: Performance and Potential Application in Photogrammetry". *Photogrammetria*, 11(64): 429-439.
- Cord, M. and Declercq, D., 2001. Three Dimensional Building Detection and Modeling Using a Statistical Approach. *IEEE Transactions on Image Processing*. V10, 715-723.
- Croitoru A., Doytsher Y., 2003. Monocular Right-Angle Building Hypothesis Generation in Regularized Urban Areas by Pose Clustering. *Photogrammetric Engineering & Remote Sensing*. 69(2): 151-169.
- Filin S., Doytsher Y., 1999. A Linear Mapping Approach for Map Conflation: Matching of Polyline. *Surveying and Land Information Systems*, Vol. 59(2), pp.107-114.
- Irvin R.B. and McKeown D.M., 1989. Method for Exploiting the Relationship between Buildings and their Shadows in Aerial Imagery, *IEEE Transactions on systems, man, and cybernetics*, 19(6):1564-1575.
- Kokubu K., Kohiyama M., Umemura F. and Yamazaki F., 2001. Automatic Detection of Building Properties from Aerial Photographs Using Color and 3D Configuration. Presented at the 22nd Asian Conference on Remote Sensing, November 2001, Singapore.
- Lobanov A.N., 1984. "Photogrammetry", Second edition, Moscow, "Nedra", p. 552.
- Mayer, H., 1999. Automatic Object Extraction from Aerial Imagery – a Survey Focusing on Buildings. *Computer Vision and Image Understanding* 74 (2), 138-149.
- McKeown, D.M., Bulwinkle, T.M., Cochran, S., Harvey, W., McGlone, C., Shufelt, J.A., 2000. Performance Evaluation for Automatic Feature Extraction. *International Archives of Photogrammetry and Remote Sensing* 33 (Part B2), 379-394.
- Michel, A. Goretta, O. and Lorin, L. 1999. 3D Building Extraction based on Aerial Stereoscopic Images. *Proceedings of SPIE - The International Society For Optical Engineering*; v.3808, p: 624-633.
- Peled, A., Raizman, Y. 1997. Toward Automatic Updating of the Israeli National GIS. *Proceedings, ISPRS WG III/2, III/3 and II/8 Joint Workshop on Theoretical and Practical Aspects of Surface Reconstruction and 3-D Object Extraction*, Haifa, ISRAEL, September 9-11, 1997. pp. 114-119.
- Ruther, H., Martine, H., and Mitalo, E. G. 2002. Application of snakes and dynamic programming optimisation technique in modeling of buildings in informal settlement areas. *Journal of Photogrammetry & Remote Sensing*, V56, p: 269-282.
- Shufelt, J.A., McKeown, D.M., 1993. Fusion of monocular cues to detect manmade structures in aerial imagery. *Computer Vision Graphics and Image Understanding* 57 (3), 307-330.
- Thompson E. H., 1968. The Projective Theory of Relative Orientation. *Photogrammetria* 23(2). pp. 67-75.
- Tseng Y. and Wang S., 2003. Semiautomated building extraction based on CSG model, *Photogrammetric Engineering & Remote Sensing*, 69(2):171-180.
- Weidner U. and Forstner W., 1995. Toward automatic building reconstruction from high resolution digital elevation model. *ISPRS Journal of Photogrammetry & Remote Sensing*, V50, No4, p: 38-49.