Analysis of Imaging Configuration Effect on Surface Reconstruction of Complex Sites

Shalev MOR and Sagi FILIN, Israel

SUMMARY

Detailed three-dimensional mapping of small scale scenes sees increasing demand in recent years. This trend can be attributed to increasing computing power, graphic hardware, technological and algorithmic advances, and finally growing awareness of users to the value of scene documentation in 3D. A common method for acquiring 3D information of terrestrial scene is using photogrammetry. Generating 3D information from images requires solving the stereo correspondence between image pairs. In the past, stereo correspondence was predominantly approached using local computational methods. Such schemes generally focused on correspondence assignment per pixel using some variant of the 2D normalized cross-correlation. Recent studies of the stereo-correspondence problem view disparity maps between image-pairs as a global optimization problem. Such formulation has proved more effective in producing more accurate results compared to any local approach.

We present in this paper application of 3D surface reconstruction using close range image sets. Our aim is studying how imaging configurations characteristic to imaging campaigns of complex scene fair with such models. Models are tested on a prehistoric archeological site that offers limited accessibility, complex topography and features therein.

Analysis of Imaging Configuration Effect on Surface Reconstruction of Complex Sites

Shalev MOR and Sagi FILIN, Israel

1. INTRODUCTION

The commonly used strategies for terrestrial mapping consist of traditional land-surveying and more recently, terrestrial laser scanning. Conventional surveying is limited, however, in its ability to provide a detailed three-dimensional scene description within a reasonable amount of time, while terrestrial laser scanners, which have proven effective, suffer from high instrumentation costs and are still considered unaffordable and expensive. In this regards, availability of low-cost quality cameras motivates a greater use of photogrammetry as an efficient means for surface geometry 3D reconstruction. Mapping terrestrial scenes by a set of images offers both high-level of detail and accuracy. A main challenge still existing is the automation of 3D modeling techniques in order to reduce the large amount of time required to reconstruct a scene.

A large body of research has been dedicated to 3D surface reconstruction from image sets, with recent years seeing greater focus on generating models for mapping and surveying. As an example, Zheng et al. (2008) use images for reconstructing small objects based on close-range photogrammetry. Marinov (2008) proposes a method for 3D modeling of architectural scenes. Remondino and Menna (2008) review developments in terrestrial 3D surface reconstruction and object modeling using digital images. The authors demonstrate how low-cost digital cameras and commercial or in-house software can be sufficient for such modeling.

Advances in computational stereo raises the question of how suitable modern techniques are for reconstructing complex natural sites characterized by undulating surface topography, variation in surface texture, and break lines. Images will usually feature an oblique view of the scene with considerable variation in scale between the near and far ends of the image. Usually, the relative base-line between images will be much wider compared to the aerial case, and thus representation of the same surface elements may differ greatly between images. In this paper we study the reconstruction of a dense 3D surface model using graph-cut optimization. Graph-cut optimization strategies have been gaining popularity because of their ability to offer a unified approach for global optimization of different requirements which incorporate image content, general surface smoothness, and awareness of geometric features related to scene representation within an image. For this study, images describing a complex topography within an archeological site were acquired from varied vantage points, typical to imaging setups for such a scene. We test how the model handles reconstructing this scene, and consistency in the reconstruction.

The organization of the presentation is as follows: Section 2 provides a theoretical background on surface models from image pairs using graph-cuts. It first addresses 3D geometry from images and the correspondence problem, and then discusses differences between local and global methods for solving this problem. The section then presents a

solution to the correspondence problem using graph-cut optimization. Section 3 presents results of the graph-cuts-based technique for the reconstruction of a 3D scene, and provides analysis. Conclusions and avenues for further studies summarize the paper.

2. GRAPH CUTS BASED RECONSTRUCTION OF 3D SURFACE FROM IMAGES

2.1 Reconstruction of 3D Surfaces from Image Pairs

Formation of a 3D surface model from images requires understanding image geometry and relations between points in an image and object space. Given two images $\mathcal{L}$ and $\mathcal{R}$, and their exterior orientation parameters, the coordinates of any given object-space point that is visible in both images can be determined by coordinates of the left and right image points, $\boldsymbol{u} = (u,v)$ and $\boldsymbol{u}' = (u',v')$ respectively. Assuming use of distortion free cameras, the collinearity equations linking an image and object space point can be written as:

$$\boldsymbol{x} = \boldsymbol{c} + \lambda \cdot \mathbf{R} \cdot \begin{pmatrix} u \\ v \\ -f \end{pmatrix} = \boldsymbol{c}' + \lambda' \cdot \mathbf{R}' \cdot \begin{pmatrix} u' \\ v' \\ -f \end{pmatrix} \tag{1}$$

with c and c', the exposure coordinates of the left and right camera respectively; **R** and **R'**, the rotation matrices; f the focal length; λ and λ' are scalar scale multipliers; and $\boldsymbol{x} = (x, y, z)^T$ the object space coordinate of the given point. Eq. (1) provides six equations and five unknowns, ($\boldsymbol{x}$, λ and λ'), which can be estimated via least-squares adjustment.

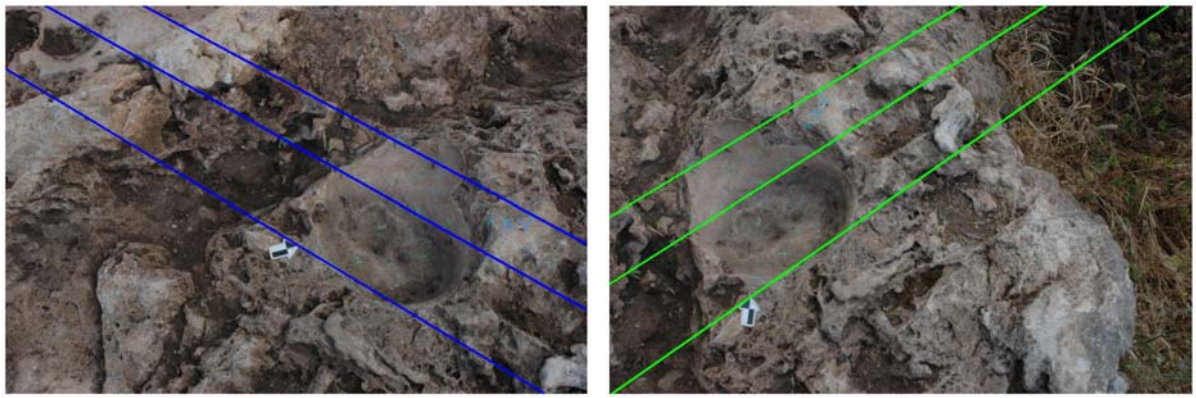
In order to create a detailed 3D model of a given scene, it is essential to establish a dense set of corresponding image points. The correspondence problem is a prominent research field in computer-vision and photogrammetry. A simple formulation of the correspondence problem can be expressed as follows: given images $\mathcal{L}$ and $\mathcal{R}$ of the same scene, find for any point, $\boldsymbol{u}$, in image $\mathcal{L}$, the matched point in image $\mathcal{R}$. Using epipolar geometry that constrains the geometric locus of corresponding points from both images, this problem is reduced from the complete image as search space into a single line in $\mathcal{R}$. For computational efficiency this problem can be simplified if images are converted such that the epipolar lines in both images are parallel to the u-axis. This is achieved by performing an epipolar image rectification in a way that constrains any two corresponding epipolar lines to share the same v ordinate (Fig. 1 demonstrates results of such rectification). Rectification of the two images is implemented by applying a planar projective transformation to each of the two images, defined by 3×3 homographic matrices, **H** and **H′**, for the left and right images respectively. The back and forth transformation from image coordinates $\boldsymbol{u} = (u,v)$ to rectified coordinates $\bar{\boldsymbol{u}} = (\bar{u},\bar{v})$ is given in Eq. (2) by the use of homogenous coordinates.

$$\begin{pmatrix} \bar{u} \\ \bar{v} \\ 1 \end{pmatrix} \square \mathbf{H} \cdot \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \qquad \Leftrightarrow \qquad \begin{pmatrix} u \\ v \\ 1 \end{pmatrix} \square \mathbf{H}^{-1} \cdot \begin{pmatrix} \bar{u} \\ \bar{v} \\ 1 \end{pmatrix} \tag{2}$$

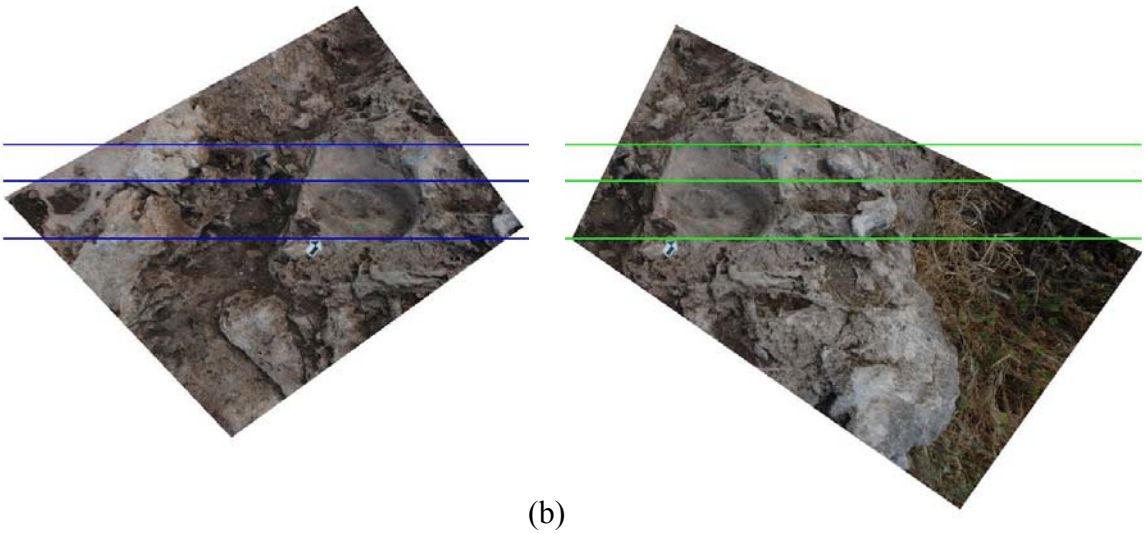
With the rectification of the images, the correspondence problem for a pair can be defined as follows: for any pixel, $\bar{\boldsymbol{u}} = (\bar{u},\bar{v})$, in $\bar{\mathcal{L}}$, the rectified left image, find the matching pixel, $\bar{\boldsymbol{u}}' = (\bar{u}+d,\bar{v})$, in $\bar{\mathcal{R}}$, the rectified right image, with d as the disparity value between the given pixels. As there is a disparity value for each pixel in the left image $\bar{\mathcal{L}}$, it is possible to create a disparity-map $\mathcal{D}$, which has the same dimensions as those of $\bar{\mathcal{L}}$. The main challenge is to establish these disparity maps autonomously, achieving results that weight both image content, effect of the perspective projection, and the general requirement of achieving a continuous and relatively smooth reconstructed surface.

2.2 Local Methods for Correspondence Problem

Until a decade ago, most algorithms for automatic generation of disparity-maps from image-pairs were based on local methods. For a given pixel in the rectified image $\bar{\mathcal{L}}$, these methods search for the most probable matching in $\bar{\mathcal{R}}$ by making use of the radiometric content in the image only. This is usually implemented by comparing two local windows centered on the left image pixel and candidate counterparts along the epipolar line in the right one. Comparison is calculated in several ways, with the most common one using cross-correlation, or alternatively, the sum or average of square- or absolute-differences. The fundamental algorithm for calculating disparity-maps using local methods is relatively straightforward. The process scans each line in both right and left images, and calculates a score value that echoes the level of similarity between the two pixels and their surroundings. These algorithms then select the pixel whose score is the highest, which ignores suitable matches with lower scores. They also ignore the continuity of most real-world objects, and thus do not consider smoothness or continuity of surfaces and objects as a constraint. Finally, occlusions are not treated by these methods. Since the left and right images are taken from different viewing directions, parts of the 3D scene may be seen in one image but occluded in the other. This fact may lead to a case where some pixels have no actual matchings may exist whatsoever.



(a)



(b)

Figure 1: Epipolar geometry and image epipolar rectification (a) original images (b) rectified images.

2.3 Global Methods for Correspondence Problem

In contrast to local methods, global ones seek optimal disparity-maps for the complete image rather than for a single pixel. In most cases, the global methods cast this problem as a global optimization problem, and minimize the cost (or energy) of a given objective function for which the parameters are the disparity values of all pixels. The general form of this function is shown in Eq. (3).

$$E(\mathcal{D}) = \underset{\text{data}}{E}(\mathcal{D}) + \underset{\text{smooth}}{E}(\mathcal{D}) \tag{3}$$

The data term refers to the radiometric differences between matched pixels, which should be as small as possible. The main difference of global from local methods is the smoothness term. The purpose of minimizing this term lies in the realization that neighboring pixels should have the same or close disparity values as long as their radiometric values are similar. This is because discontinuity in image radiometry is usually related to discontinuity in the scene's surface.

Most global algorithms use methods such as graph-cut optimization (Kolmogorov and Zabih, 2001) or belief propagation (Sun et al., 2002). Research in recent years has dealt mainly with improvement of run-time and refinements; such improvements in run-time can be achieved by replacing the graph structure with a tree structure, and solving the optimization problem using dynamic programming (Veksler, 2005). Improving the resulted disparity-map can also be achieved by performing image segmentation as a preliminary stage (Yang et al., 2006).

2.4 Graph-cuts-based implementation

Although the function given in Eq. (3) is a useful objective function for computing a disparity map, it does not account for the existence of occlusions (Section 2.2). As noted, for some pixels, no matching can be found in the other image, as the viewed entity is occluded from that vantage point. In order to find the best disparity-map, occlusion-related considerations should be taken into account within the optimization framework. Thus, Eq. (3) is appended with another term leading to:

$$E(f) = \underset{\text{data}}{E}(f) + \underset{\text{occ}}{E}(f) + \underset{\text{smooth}}{E}(f) \tag{4}$$

with *f*, a configuration that assigns to any pixel in the rectified left image a pixel in the rectified right one, where some pixels in both images can be considered by *f* as occluded.

For any two pixels, $\bar{u} \in \bar{\mathcal{L}}$ and $\bar{u}' \in \bar{\mathcal{R}}$, which were assigned to each other, the data term is calculated as follows:

$$\underset{\text{data}}{E}(f) = \sum \left(I(\bar{u}) - I(\bar{u}') \right)^2 \tag{5}$$

with $I(\boldsymbol{u})$, the image intensity value at pixel $\boldsymbol{u}$. The occlusion term in Eq. (4) is defined as a sum of penalties for all pixels in $\boldsymbol{O}$, the set of all occluded pixels as assigned by *f*

$$\underset{\text{occ}}{E}(f) = \sum_{\bar{u} \in \boldsymbol{O}} \lambda_{\bar{u}} \tag{6}$$

The value for $\lambda_{\bar{u}}$ should be set such that this assignment will be chosen when no reasonable matching is found. By doing so, the minimization should result in a configuration that holds very few occluded pixels. Occlusion will be assigned only when it is the better choice; namely, any matching will increase the value of the objective function in Eq. (4).

The smoothness term is also implemented as a sum of penalties, as shown in Eq. (7)

$$\underset{\text{smooth}}{E}(f) = \sum_{\{a_1, a_2\} \in \mathcal{N}} \lambda_{a_1, a_2} \tag{7}$$

with a_1, pixel assignment ($\bar{\boldsymbol{u}}_1$, $\bar{\boldsymbol{u}}'_1$); a_2 pixel assignment ($\bar{\boldsymbol{u}}_2$, $\bar{\boldsymbol{u}}'_2$); and $\mathcal{N}$ is the set of assignment pairs, where either $\bar{\boldsymbol{u}}_1$, $\bar{\boldsymbol{u}}_2$ are neighbors in the left image, or $\bar{\boldsymbol{u}}'_1$, $\bar{\boldsymbol{u}}'_2$ are neighbors in the right image. For any such pair of neighboring assignments, a penalty value λ_{a_1, a_2} is assigned. If the disparity value for $\bar{\boldsymbol{u}}_1$ is the same as that of $\bar{\boldsymbol{u}}_2$, λ_{a_1, a_2} is set to zero. In any other case, λ_{a_1, a_2} depends on the radiometric difference for both neighboring pixels. If this difference is small, the assumption is both are lying on the same surface; a noticeable difference will likely suggest a stepped surface. Thus, in terms of smoothness, if the difference $\max\left(\left| I(\bar{\boldsymbol{u}}_1) - I(\bar{\boldsymbol{u}}_2) \right| , \left| I(\bar{\boldsymbol{u}}'_1) - I(\bar{\boldsymbol{u}}'_2) \right| \right)$ is small, the expected disparity values should be equal or

close to one another, for both assignments. If that is not the case, a penalty is assigned (Kolmogorov and Zabih, 2001).

The crucial part of calculating the optimal disparity-map is finding the optimal configuration, f^*, which would minimize the objective function $E(f)$. This optimization process is implemented iteratively, beginning with an initial configuration, f_0, that assigns an initial disparity value (or occlusion) for each pixel. Within each iteration a disparity value k is chosen, from within a predefined disparity range, and the optimal configuration f_C is calculated, where any assignment is either from f_0 or one proposed by k. If $E(f_C) < E(f_0)$, f_C replaces f_0 as the current configuration. Repeating this process for all possible disparity values, the disparity-map can be enhanced in any iteration. If there are n_k possible disparity values, the process requires n_k iterations to complete a single cycle. To obtain the optimal disparity-map, such a cycle must be performed until no further enhancement can be achieved. This results in f^*, the optimal configuration, which corresponds to the optimal disparity-map.

In this formulation, the minimization problem becomes finding an optimal configuration, f_C, for which any assignment is either already in f_0, or relates to assignment of disparity k. One approach for solving this optimization is casting it as finding the minimal cut on a graph.

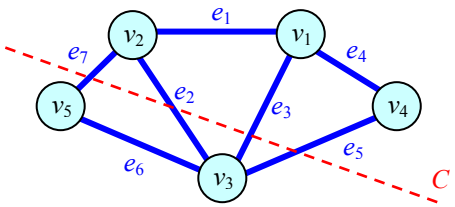


Figure 2: A cut on a graph.

Graph-cut – given a graph $\mathcal{G} = (\mathcal{V},\mathcal{E})$, with $\mathcal{V}$ as the set of nodes, and $\mathcal{E}$, the set of edges, a cut on $\mathcal{G}$ is defined as a partition of its vertices into two disjoined sets. For example, in Fig. (2), the cut C is a partition of the graph into $\{v_1, v_2, v_4\}$ and $\{v_3, v_5\}$. Another way of defining a cut on a graph is by the set of edges it crosses; therefore the cut in Fig. (2) can also be defined as the set of edges $C = \{e_2, e_3, e_5, e_7\}$. If the graph edges are weighted, the weight of the cut is the sum of weights for all the edges in C. For a directed weighted graph with a source node s, and a sink node t, several algorithms can be applied for finding the optimal cut (having a minimal weight) which partitions the graph into two sets, S and $\mathcal{T}$, where $s \in S$ and $t \in \mathcal{T}$. Such algorithms have been proposed by Ford and Fulkerson (1962), and Ravina et al. (1993).

The disparity-map minimization problem can be solved by finding the minimal cut. In each iteration, a graph is built so that each pair of pixels ($\bar{\boldsymbol{u}}$, $\bar{\boldsymbol{u}}'$) that is a feasible assignment in f, is then made a node in the graph, and linked to the source and sink nodes. The nodes are linked by edges such that the weight of any graph-cut separates the source node from the sink node, and so corresponds to the value of the objective function in Eq. (4). Thus, for any iteration, an

optimal configuration can be found. A detailed explanation for the creation of the graph is given in Kolmogorov and Zabih (2001).

3. EXPIRIMENTAL RESULTS

The proposed model is applied for surface reconstruction for an archeological excavation site. It is demonstrated on a documentation project in Raqefet cave in the Carmel mountain ridge, as part of ongoing research studying the origins and use of Manmade-bedrock-holes (MBH) in those sites. An excavation campaign performed in June 2008 serves as the case study here. Images were acquired during the excavation with a Nikon D70 camera, with 48mm Nikkon lens.

We focus on the reconstruction of a 3D model of MBH embedded within a rough topographic setting. For the experiment, four models were formed automatically using six images (Fig. 3). The sets of image pairs used for the experiment are {1,2}, {1,3}, {1,4} and {5,6}. Our aim is not only testing the ability to reconstruct the surface using the image pairs, but to also test consistency of the reconstruction based on different images or image pairs. All images were acquired from ~ 2-3m from the MBH and base line between images ranges between 1-3m. Set {1,2} features a horizontal displacement of the camera position between the two images, Set {1,3} features both offset and rotations, Set {1,4} features images that show a drastically different view of the same object with significant perspective-related changes. Finally, Set {5,6} features an image set of the same object, but using different images, and used for comparison of the reconstructed scene based on "independent" images. Comparison is performed by 3D registration of the models into a single reference frame.

Results of the surface reconstruction are presented in Fig. (4) with the radiometric content rendered on top of the surface (top) and with the reconstructed wireframe (bottom). For sets {1,2}, {1,3} and {5,6} the reconstruction results in a surface compatible with the actual 3D model of the MBH and the surrounding surface. The reconstruction failed with set {1,4}, where the results only roughly resemble the actual surface. This failure is attributed to the difference in viewing directions, and therefore the perspective distortions (foreshortening, occlusions), and to some changes related to illumination between both images.

In order to evaluate the surfaces created in each model, we compare surfaces of the same scene. Numerical comparison of the reconstructed surface models was implemented by evaluating z value differences between the surfaces, and calculated as the sum of square differences. We compare the models which share the same left image {1,2} with {1,3}, and models which were created from a different image pair {1,2} and {5,6}, with results summarized in Table 1. This shows, accuracy-wise, the average difference being ~10cm. Although all three models provide a true representation of the surface, differences can be attributed to the variation in the vantage points of images (leading to very different views) and mostly to the rough and undulating surfaces in the scene, which complicated the reconstruction. In terms of the surface reconstruction algorithm, some of the problems associated with it relate to the "layering" mode by which it operates (searching for a given disparity value in each iteration). Such layering may lead to strong sensitivity to image

orientation, and may be affected by the rectification of the images rather than depths within the scene. Additional cause for inaccuracies relates to the application of penalties in image-space rather than object-space. Thus, an offset of a single pixel will receive a penalty not related to its actual effect on depth in object space. Nonetheless, the fast and globally optimal solution by which the surface model was generated suggests that this approach can serve as a means to generate a coarse surface representation; this can then be refined while performing the optimization in object space by use of more aware considerations of the actual surface shape.

View #1

View #2

View #3

View #4

View #5

View #6

Figure 3: images of the scene taken from different views.

Models :		model (1,2) - model (1,3)	model (1,2) - model (5,6)
Number of mutual grid points in both models	n	38,104	28,990
Sum of Square Difference	SSD	355.87 m^2	391.86 m^2
Posterior Standard deviation of differences	$\hat{\sigma} = \sqrt{\frac{SSD}{n}}$	9.7 cm	11.6 cm

Table 1: comparison of two models.

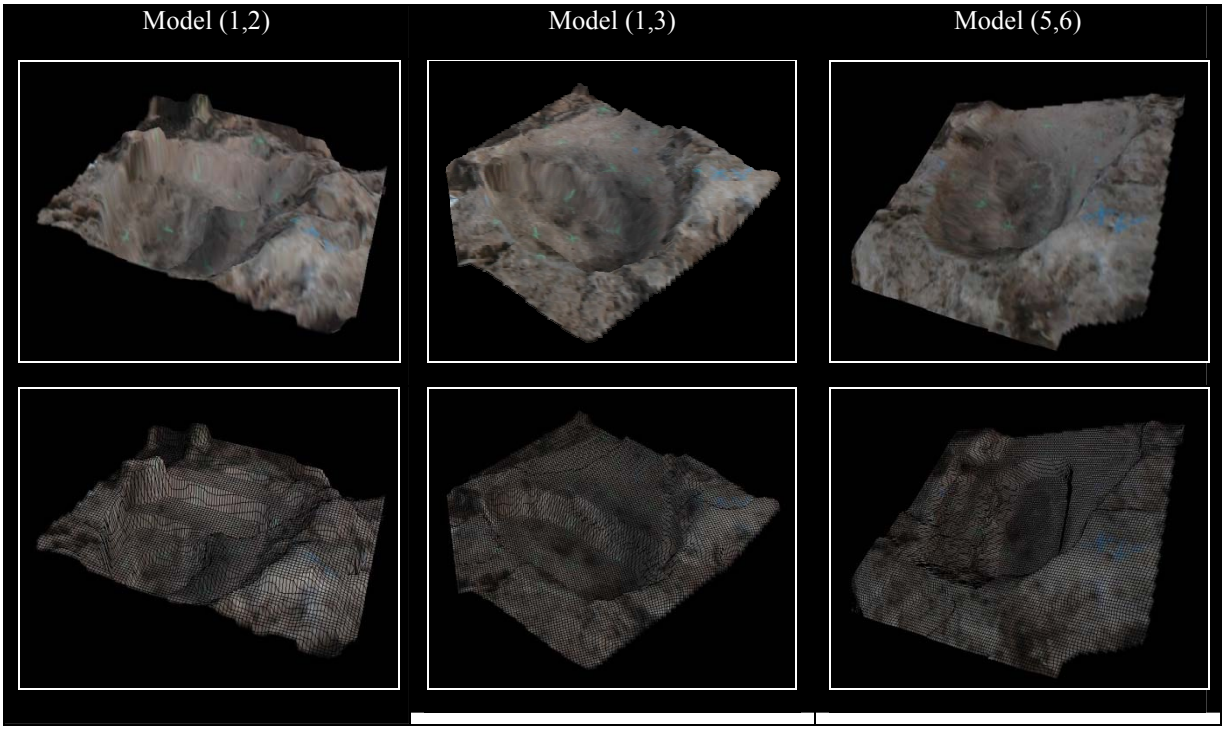


Figure 4: views of three-dimensional models.

4. CONCLUSIONS

This paper discusses formation of a dense 3D surface from a pair of images using graph-cuts. Considering the rough surface setup and the strong variation in depth, the graph-cut method performed better than conventional local photogrammetric or stereo-vision methods. Though differences in the reconstructions have been observed, results are adequate for an initial surface model featuring the main characteristics of the surface model. Results have also demonstrated the complexity of the reconstruction problem and its dependency on the image acquisition configuration.

Future work will deal with refinement of the initial results. Analysis will also look into the effect of image rectification, due to its potential influence on the resulting surface. Another avenue that will be explored is the generation of 3D data from multi-view sets of images. Such extension is only natural, considering the normative data acquisition procedure. Results from a multi-view set reconstruction are expected to provide a more accurate and reliable reconstruction of the surveyed scene.

REFERENCES

[1] Ford L., Fulkerson D., 1962. Flows in Networks. Princeton University Press, 1962.

[2] Kolmogorov V., Zabih R., 2001. Computing Visual Correspondence with Occlusions Using Graph Cuts., International Conference of Computer Vision. pp. 508-515.

[3] Marinov B. D., 2008, Hierarchical Modeling and Processing of Space Objects in Architectural Photogrammetry. International Archives of Photogrammetry and Remote Sensing, **37**(B5), pp. 401-408.

[4] Remondino F., Menna F., 2008, Image-Based Surface Measurement for Close-Range Heritage Documentation., International Archives of Photogrammetry and Remote Sensing, **37**(B5), pp. 199-206.

[5] Scharstein D., Szeliski R., 2002, A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms. International Journal of Computer Vision, **47**, pp. 7-42.

[6] Sun J., Shum H. Y., Zheng N. N., 2002, Stereo Matching Using Belief Propagation. ,European Conference on Computer Vision, Vol. **2**, LNCS 2351, pp. 450-452

[7] Ahuja R. K., Magnanti T. L., Orlin J. B., 1993, Network Flows: Theory, Algorithms, and Applications., Prentice-Hall, Englewood Cliffs, NJ, USA.

[8] Veksler O., 2005, Stereo Correspondence by Dynamic Programming on a Tree., International Conference on Computer Vision and Pattern Recognition. Vol. **2**, pp. 384-390.

[9] Zheng J., Yuan W., QingHong S., 2008, Automatic Reconstruction for Small Archeology Based on Close-Range Photogrammetry. International Archives of Photogrammetry and Remote Sensing, **37**(B5), pp. 165-168.

CONTACTS

Shalev Mor
Department of Transportation and Geo-Information Engineering
Faculty of Civil and Environmental Engineering
Technion – Israel Institute of Technology
Haifa 32000
ISRAEL
Tel. + 972 4 829 3065
Email: shalevm@technion.ac.il

Dr. Sagi Filin
Department of Transportation and Geo-Information Engineering
Faculty of Civil and Environmental Engineering
Technion – Israel Institute of Technology
Haifa 32000
ISRAEL
Tel. + 972 4 829 5855
Fax + 972 4 8295708
Email: filin@technion.ac.il