A Framework for Generation of High Quality Digital Elevation Models in Urban Environments from Aerial Images

Ahmed F. Elaksher
Department of Computer and Information Science, Community Collage
Taibah University, Almadinah Almunawwarah, Kingdom of Saudi Arabia
ahmedelaksher@yahoo.com

ABSTRACT

Digital elevation models (DEMs) are required for several applications, such as cartographic analysis, mobile communication, urban planning, and visualization. A conventional stereo pair is insufficient to create complete and accurate DEMs particularly in urban areas. Hence, robust techniques for creating DEMs are imperative. This paper presents a novel technique for simultaneously matching several overlapping images. The conventional cross correlation technique is used to produce a good approximation of the DEM. Least squares matching is employed to improve the DEM accuracy. Due to surface discontinuities, the DEM do not represent the exact 3D surfaces. Therefore, a refinement process is applied to enhance the quality of the DEM utilizing the geometrical properties of man-made features such as buildings. Points contributing to roof patches are delineated and fitted through 3D planes. The final results look appealing and suggest the use of the proposed algorithm in DEM generation and surface reconstruction in urban environments.

KEY WORDS: Correlation, Matching, Least squares estimation, Image segmentation

1. INTRODUCTION

DEM generation in urban areas is a complex problem in photogrammetry. High quality DEMs are essential for cartographic analysis, urban area planning, and visualization. Although one pair of images is adequate to find the 3D position of two visibly corresponding image features, it is inadequate to generate a high quality DEM in urban areas due to hidden parts that are obscured in the image pair. DEM can be extracted from multiple images to avoid this deficiency.

Kraub et. al. (2005) presented an algorithm to automatically improve current methods for creating DEMs from stereoscopic image pairs based on the epipolar geometry. First, correspondence lines were found and radiometric corrections were applied for each line using a shifting technique. The shifting technique was based on attempting to fit the actual image patch with the correlated patch in the other image shifted by a displacement value to obtain local brightness and contrast parameters. Second, a radiometricly corrected left image was generated from these parameters. Third, the standard cross correlation was applied.

Although the presented algorithm yielded more details in dense senses, more blunders were observed.

Gruen and Li (2002) derived Digital Surface Models (DSM) from Three-Line-Scanner (TLS) raw images through combining grind-point and feature-point matching procedures. Points from the three images were matched to compute pixel and object coordinates for grid points simultaneously. An additional feature-point matching procedure was performed to compensate for the disadvantage of modelling the terrain with grid points. This was performed via a modified Multi-image Geometrically Constrained (MPGC) matching algorithm.

In Msemakweli (1999) object space matching using the Vertical Line Locus (VLL) was presented to directly generate DEMs in object space from one pair of images. Starting with a horizontal grid at the mean terrain, a template window, for each post, was centred on the given elevation and projected to the images. Consequently, the cross-correlation was computed between the images, and the template was moved up and down. The elevation with the best correlation was selected. Dynamic programming algorithms are utilized in image matching (Sun, 1999; Apaphant, 1999). These algorithms construct a cost matrix between each pair of corresponding epipolar lines (Faugeras, 1999) in two images and then find the least cost path in the cost matrix. The cost function between two pixels depends on the similarity of their intensity and the order of the two pixels with respect to other pixels.

Wiman (1998) utilized a multi-image cross correlation matching function for object space matching. First an initial DEM was driven at the mean terrain height. Each post was then moved up and down and at each location the correlation coefficient was computed. The elevation with the largest correlation was recorded at this post. The idea presented is comparable with the VLL approach except that a new correlation function was implemented. In Gruen and Baltsavias (1985) and Rosenholm (1986) multi point geometrically constrained least squares matching was introduced. The basic idea was to have one condition equation for each pair, then introduce constraint equations including: epipolar constraint, colinearity constraint, and Z constraint. A complete study on the least squares matching could be found in Atkinson (1996).

Dragos et. al. (2004) demonstrated the benefits of using break lines in generating and filtering DSMs. The raw data was a point cloud acquired by a laser-ranging scanner. The DSM was produced through a moving plane interpolation algorithm in a hierarchical approach. First, a rough approximation of the surface was computed. Then, the oriented distances from the surface to the original cloud of points were computed. Each measured elevation was given a weight according to its distance value. The surface was then recomputed using the moving plane under the consideration of weights. However, the moving plane was trimmed when a break line is crossed.

Skarlatos and Georgopoulos (2004) utilized a pair of stereo images to refine a coarse DEM. Two orthoimages were generated, one for each image. The parallax between the two orthoimages was computed by image correlation methods based on reference template and search template taken from the two orthoimages. The displacement of the best match point was used to compute elevation errors in DEM and refine the DEM.

Xu et. al. (2002) presented techniques for producing hierarchical DEMs from descent and rover imagery for Mars mapping and rover localization. During a descending process of a Mars spacecraft, ten descent images were taken at approximately every half of the altitude. An initial DEM of the landing site was generated from these images. The DEM was then refined both in accuracy and resolution to form a five-layer hierarchical DEM, with resolution ranging from one centimeter in the immediate area of the landing center to one meter in the boundary region about 1 km away from the center. The technique was based on area-based matching assisted by epipolar constraints and the least-squares matching.

The approach presented in this paper depends on utilizing a large number of images with large overlap and sidelap percentages. The images used have about 80% forward overlap and 60% sidelap. This guarantees that each point in the DEM appears in more than two images. The algorithm starts by generating an initial DEM from the images through conventional correlation methods. Least squares matching is applied to refine the generated DEM. Geometric constraints assists in the 3D modelling of planer surfaces. The remaining of the paper is organized as follows. The next section explains the DEM generation. Then, the refining step is illustrated. Conclusions are then stated.

2. DEM GENERATION

This section presents the algorithm implemented to generate the DEM from a large number of highly redundant aerial photos. The first part is like the VLL approach except that a large number of images are handled rather than in the case of one image pair. After that, a spike detection algorithm is applied to detect spikes and outliers. Least-squares matching is finally employed to refine the generated DEM.

2.1. Correlation Matching

The first step is to construct a horizontal grid with elevations near the mean elevation in the test area. At each location in the grid a vertical line is constructed. The vertical line is divided into a number of elevations and at each elevation a window is built for each DEM post. Window points are projected back to all images and their intensities are recorded. The value of the cross-correlation between each image pair, equation (1), is computed.

Since several images are used, at each assumed ground elevation a number of correlation coefficients are defined, figure 1. These coefficients are used to generate one value as follows. First, the minimum and maximum correlations are ignored. Then a weighted mean of the remaining coefficients is computed through equation (2). The weighted value for each image pair is computed as the inverse of the distance between the exposure stations of the image pair, equation (3). If the weighted standard deviation of the 13 values is large, the weighted average correlation is ignored. We then have to generate a number of subsets from the 13 correlations. The weighted average correlation for each four images is computed and the one with the smallest weighted standard deviation is chosen. In addition, the average gray value for each window is computed and windows with low intensities are excluded. The reason for this test is to prevent matching windows at shadow areas were the average intensity is small.

$$C_{ij} = \frac{\sum_{k=1}^{n}(G_{ki} - G_i)(G_{kj} - G_j)}{\sqrt{\sum_{k=1}^{n}(G_{ki} - G_i)^2 \bullet \sum_{k=1}^{n}(G_{kj} - G_j)^2}} \qquad (1)$$

where C_{ij} is the correlation coefficient between images *i* and *j*,

n is the total number of points in the window,

G_i and G_j are the mean intensity values for windows in the i[th] and j[th] images respectively,

G_{ki} and G_{kj} are the intensity value at position *k* in image *i* window and image *j* window respectively.

$$C = \frac{\sum_{i=i}^{N} w_i C_i}{\sum_{i=i}^{N} w_i} \qquad (2)$$

where C is the weighted correlation value,

C_i is the correlation value between image pair *i*,

w_i is the distance inverse between image pair *i*,

N is the total number of image pairs.

$$w_{ij} = \frac{1}{\sqrt{(X_i - X_j)^2 (Y_i - Y_j)^2}} \qquad (3)$$

where w_{ij} is the weight between the two images i and j

X_i, Y_i, X_j, and Y_j are the exposure station coordinates for the two images i and j.

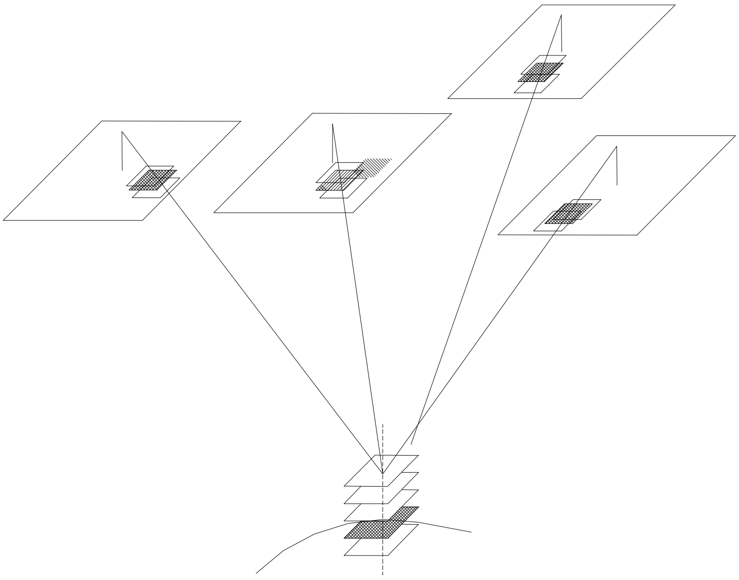


Figure 1. Object Space Matching from Multiple Images

The shape and size of the window are important factors to provide high quality matching results. In this research, only horizontal windows were tested; hence, if a large window size is chosen the horizontal plane assumption will be violated. Conversely, small window correlation will are unreliable. Therefore, different window sizes are tested, and the optimum size was seven. The spacing between the points in the window is also a vital factor that affects the matching outcomes. The vertical distance between the elevations of the windows is another significant factor in the process. It is related to the window size, the image resolution, and the terrain shape. Figure 2 shows the aerial images used for one site.

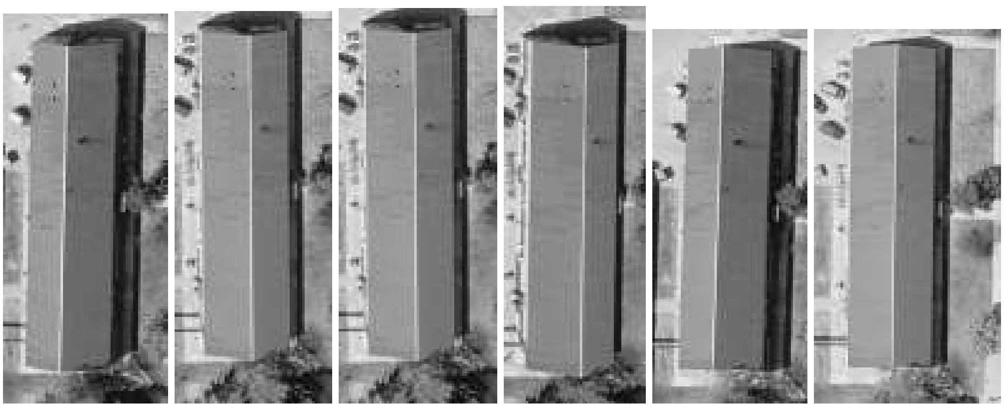


Figure 2. Six Images for the Creative Art Building, Purdue University

2.2. Spike Detection and Removal

The outputs of the object space correlation contain elevation spikes due to false correlation or incorrect modelling. The weighted median filter is used to remove the spikes in the DEM. The weights depend on the final correlation value of each post. The median filter window size is three by three posts and at each post the elevation is recomputed using the elevations of the points in the window.

In order to prevent smoothing the DEM at building edges, the window at each post is examined. This is achieved by investigating the nine elevations of the window using the leave-one-out method. The mean is computed using all 8 combinations of the nine elements and if the difference between the mean and the remaining post is large and the standard deviation on the 8 posts is small the point is a spike and its elevation has to be corrected.

2.3. Least Squares Matching

The least squares matching, (Gruen and Baltsavias, 1985) , is used in the next step to refine the generated DEM. Least squares matching needs a good DEM approximation; this is achieved by the results of the algorithm presented in sections 2.1 and 2.2. For a window of *n* points, the intensity difference in each pair of images for each point should idyllically equal zero. The previous statement is imposed as the condition equation of a least squares estimation model, equation (4).

$$\begin{bmatrix} F_1^{ij} \\ F_2^{ij} \\ F_2^{ij} \\ . \\ . \\ F_n^{ij} \end{bmatrix} = \begin{bmatrix} g_1^i - g_1^j \\ g_2^i - g_2^j \\ g_2^i - g_2^j \\ . \\ . \\ g_n^i - g_n^j \end{bmatrix} \tag{4}$$

where F_k^{ij} is the condition equation for point k between photos i and j,

g_k^i is the intensity value at position k in image i.

Equation (4) is forced between all independent image pairs. Each condition equation is a function of three parameters: the intensity value in each image and the point elevation. The least squares estimation model is solved numerically. Figure 3 presents three parts of the generated DEM for an urban area using the technique presented in section 2.

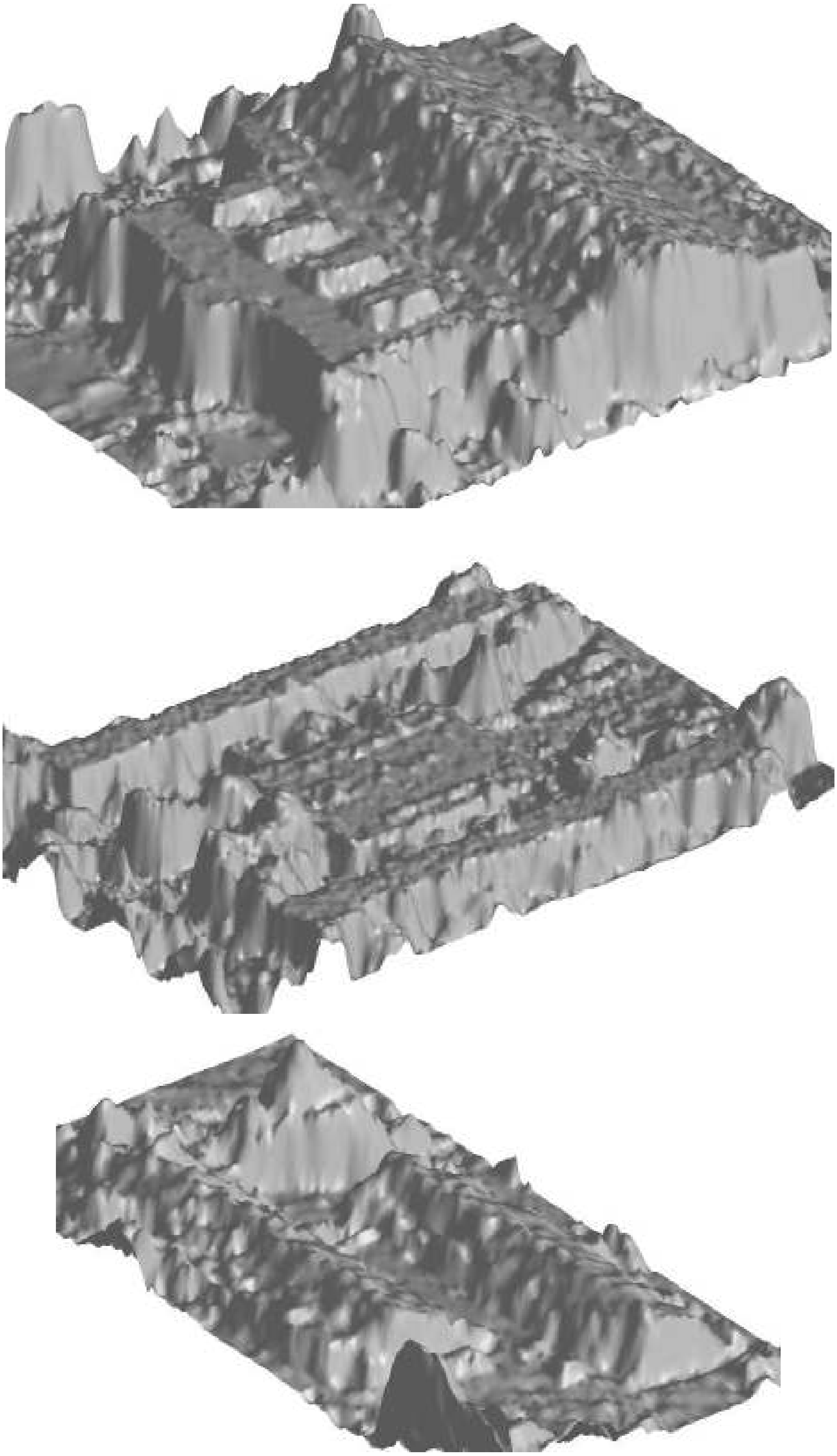


Figure 3. Parts of the Generated DEMs

3. DEM REFINEMENT

Figure 2 expresses that the generated DEMs still needs to be refined geometrically to represent the object surfaces correctly. Hence, a new technique is employed to refine the generated DEMs. The nearest image to the DEM is automatically picked, and the area covered by the DEM is extracted from the image. Image regions are extracted and projected on the DEM. For each region, all contributing DEM points are fitted through a plane. Outliers are detected and removed, then a least squares estimation model is carried out to compute the adjusted plane parameters. Subsequently, the adjusted parameters are used to compute the vertical coordinates of all points in the corresponding region.

3.1. Region Extraction

In order to extract the regions the split and merge image segmentation technique is used, (Samet, 1982). First splitting the image, the image is recursively divided into smaller regions until a homogeneity condition is satisfied. Then adjacent regions are connected to form larger regions based upon the previous criterion. In the last step, small regions are either eliminated or attached with bigger regions. The criterion used in the split and merge image segmentation method is that the difference between the minimum and maximum intensities is less than a certain threshold. The corresponding region for each DEM point is determined through the collinearity equation from the ground coordinates of the DEM point, the orientation parameters of the image, and the results of the image segmentation.

The plan fitting technique should be applied only to roof regions; hence, these regions need to be discriminated from other regions. A surface morphology filtering approach is used to extract the DSM. The surface morphology filtering starts by convolving the DEM with a minimum filter then the results is convolved with a maximum filter. The process is done iteratively with a small window size. The differences between the DEM and the DSM contain information about elevated points. These points are detected by thresholding the differences between the DEM and the DSM. If the majority of the region points are elevated above the ground then the region is classified as a roof region.

3.2. Region Refinement

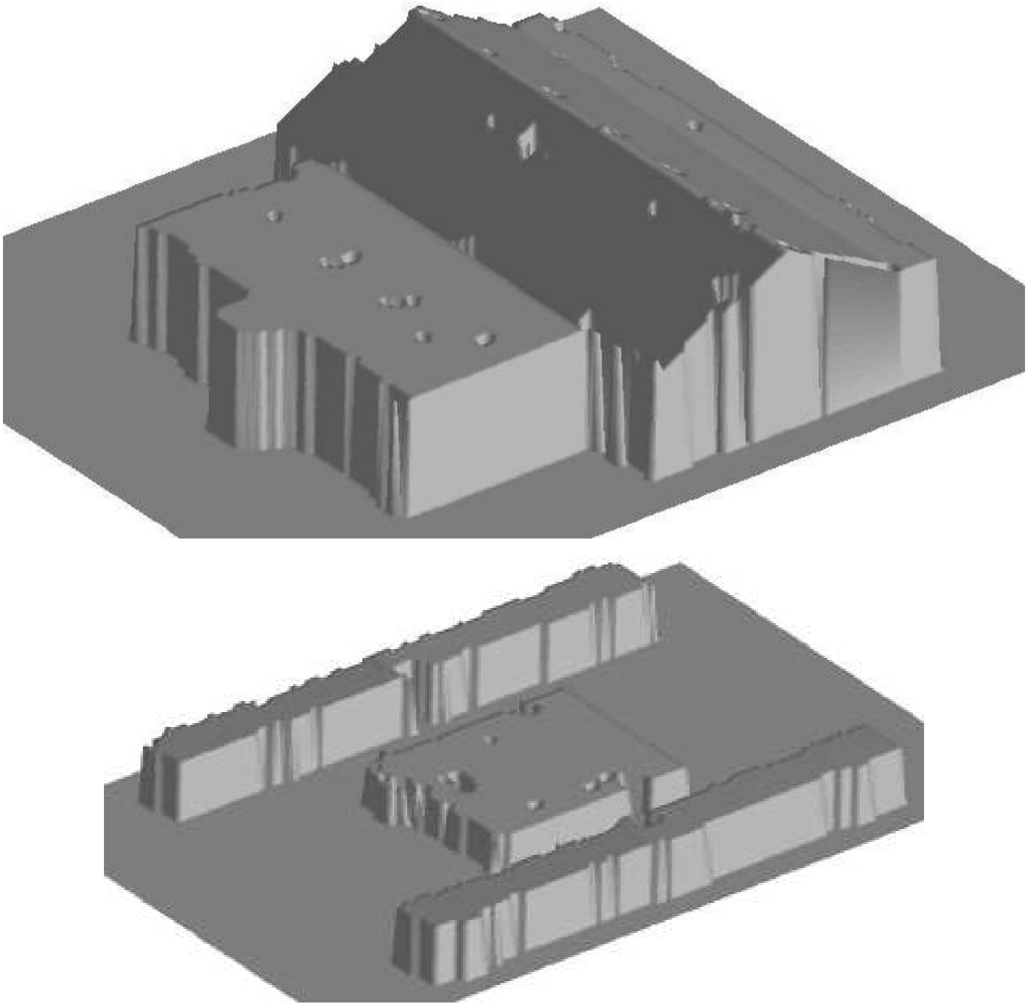
Roof regions are assumed to be planes. This assumption is true for most urban buildings except special buildings such as domes. For each plane its contributing points are used to find the plane parameters using the L1 norm (Marshall, 1998). L1 norm is a robust estimation technique that has the ability to detect outliers. First, each ground region is enlarged with two points from each side in order to include ground points that might have not been detected in the image region extraction process. This step is harmless and will not affect the final results, since the L1 norm has the ability to detect outliers. At the end of the process, only points that contribute the region will remain. Elevations of the points are taken as observations, in the L1 adjustment; the unknown parameters are the plane parameters. For each point in the region a plane equation, Equation (5), is written.

$$Z_i - aX_i + bY_i + c = 0 \tag{5}$$

where a,b,c are the parameters of the region plane,
X_i, Y_i and Z_i are the 3D coordinates of point i.

3.3. Least Squares Refinement

After removing the outlier points from each region, the remaining points are used to compute the plane parameters of the region via a least squares estimation model. The elevation of each DEM point is recomputed based on the adjusted plane parameters of the corresponding region. Figure 4 shows the refined DEMs produced by this algorithm.



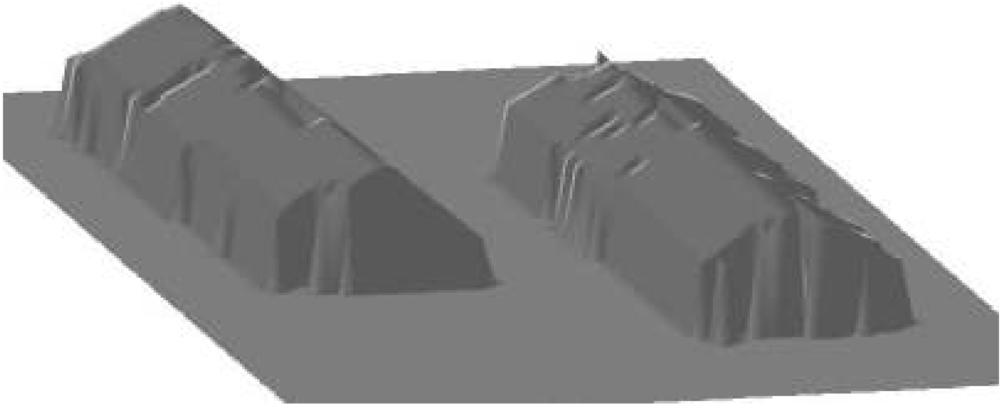


Figure 4. The Refined DEMs

4. CONCLUSIONS

Automatic terrain extraction techniques fail to provide accurate and reliable DEMs. However, high quality DEMs are essential for a variety of applications. This paper establishes a new algorithm to generate high quality DEMs in urban environments. The algorithm has the capability to use multiple images to provide solutions for factors like occlusion and discontinuity. A multi-image VLL algorithm is implemented to perform the multi-image matching. Correlation coefficients are computed for each image pair and a weighted sum of their values is computed. Subsequently, least squares matching is carried out to refine the generated DEM. A geometrically refinement process was applied to accurately represent object surfaces. Results demonstrate that the reliable and high quality DEMs could be achieved through the proposed algorithm.

REFERENCES

Apaphant, P., 1999. *Robust Image Matching for Object Space Reconstruction*. Ph.D. Thesis, Purdue University, West Lafayette, IN, 1999.

Atkinson, K.B., 1996. *Close Range Photogrammetry and Machine Vision*. Whittles Publishing, RoseLeigh House, Latheronwheel, Caithness, KW5 6DW, Scotland, UK.

Dragos B. , Acad I., Karsten J., 2004. Using break line information in filtering process of a Digital Surface Model. Proceedings of the 20th International Archives of Photogrammetry and Remote Sensing, July 12-23, 2004, Istanbul, Turkey, 35(3).

Faugeras, O., 1999. *Three Dimensional Computer Vision, a Geometric Viewpoint*. The MIT Press, Cambridge, Massachusetts, London, England.

Gruen, A. and E.P, Baltsavias, 1985. Adaptive Least Squares Correlation with Geometrical Constraints. *SPIE 595*, pages: 72-82.

Gruen, A. and Z. Li, " Automatic DTM Generation from Three-Line-Scanner (TLS) Images", Proceedings of the ISPRS Commission III Symposium, PCV'02, Graz, Austria, 9-13 September, 2002.

Kruab T., Reinarz P., Lehner M., Schroeder M., and Stilla U., 2005. DEM generation from very high resolution stereo data in urban areas. Proceedings of the ISPRS 3rd International Symposium on Remote Sensing and Data Fusion Over Urban Areas, 14 – 16 March, 2005, AZ, USA.

Marshall, J., 1998. *Gross error detection in survey networks by L1-norm minimization.* Ph.D. theses, Purdue University, 1998.

Msemakweli, C., 1999. *Design considerations in the generation of digital terrain models by image matching for GIS databases*. Ph.D. Thesis, Purdue University, West Lafayette, IN.

Rosenholm, D., 1986. Accuracy improvement of digital matching for evaluation of digital terrain models. *ISPRS*, Vol. 26(3/2), pages: 573-587.

Samet, H., 1982. Neighbor Finding Techniques for Images Represented by Quadtrees. *Computer Graphics and Image Processing*, Vol. 18, pages: 37-57.

Skarlatos D. and Georgopoulos A., 2004. Automating The Checking and Correcting of Dems without Reference Data. Proceedings of the 20th International Archives of Photogrammetry and Remote Sensing, July 12-23, 2004, Istanbul, Turkey, 35(2), pp. 553-558.

Sun, C., 1999. Multi Stereo Matching Using Maximum-Surface Techniques. *Digital Image Computing: Techniques and Applications*, Perth, Australia, pages: 195-200.

Wiman, H., 1998. Automatic Generation of Digital Surface Models through Matching in Object Space. *Photogrammtric Record*, 16.

Xu, F., K. Di, R. Li, L. Matthies and C. Olson 2002. Automatic Feature Extraction and DEM Generation for Mars Rover Mapping and Localization. Proceedings of the ISPRS Com. II Symposium, Xi'an, China, August 2002, pp.549-554.