Terrain Reconstruction from Multiple Views

Georgy L. Gimel'farb[1] and Robert M. Haralick[2]

[1] International Educational and Scientific Center
for Information Technologies and Systems
Kiev-22, 252022 Ukraine
[2] Intelligent Systems Laboratory, University of Washington
Seattle, WA 98195, U.S.A

Abstract. A two-stage approach is discussed for reconstructing a dense digital elevation model (DEM) of the terrain from multiple pre-calibrated images taken by distinct cameras at different time under various illumination. First, the terrain DEM and orthoimage are obtained by independent voxel-based reconstruction of the terrain points using simple relations between the corresponding image gray values. As distinct from other approaches, possible occlusions and changing shadow layouts are taken into account implicitly by evaluating a confidence of every reconstructed terrain point. Then, the reconstructed DEM is refined by excluding occlusions of more confident points by less confident ones and smoothed with due account of the confidence values. Experiments with RADIUS model-board images show that the final refined and smoothed DEM gives a feasible approximation to the desired terrain.

1 Introduction

Photogrammetric image processing has a significant place in today's robotics, cartography, and remote sensing [2,4]. It includes, in particular, the calibration of imaging cameras and the DEM reconstruction from the calibrated images. The calibration estimates, by using visually or automatically detected ground control points (GCP) with known world 3D coordinates, cameras model parameters that relate to where any 3D point will project on each imaging plane (see, for instance, [4,5,11]). Here, we address the problem of multi-view DEM reconstruction using a set of the pre-calibrated images.

The DEM reconstruction is most extensively studied in binocular stereo. As does the majority of other inverse photometric problems, stereo belongs to the class of ill-posed mathematical problems [7] because, even without a noise, there always exist several 3D surfaces that produce the same stereo pair. Adequate regularizing heuristics sometimes permit making the DEMs reconstructed by stereo close enough to the desired surface [2,3,6]. One way to help decrease the ill-conditionedness is to use multiple views [1].

In a few known works on the multiple-view reconstruction of dense terrain DEMs a prior knowledge about or restrictions on illumination, reflectance (albedo), and smoothness of the surface are involved to simplify the problem

[9,10]. But, generally, terrains have arbitrary shapes with discontinuities and varying albedo. The images are sensed by several cameras with various resolutions, positions, and orientations, at different times when positions of some mobile objects may change, and under distinct illuminations giving changing shadow layouts. This results in a wide scatter of gray values representing the same surface point in the images. Our goal is to judge how to compute, under these conditions, a rough but plausible approximation to the dense DEM of arbitrary terrain if we presume no prior knowledge about the terrain features but can use simultaneously all the sensed image signals.

2 Metodology

We exploit a voxel representation of a 3D surface $\mathbf{Z} = \{Z(X,Y) : (X,Y) \in \mathbf{Q}\}$ over a supporting domain $\mathbf{Q}$ in the plane OXY in the world coordinate system $OXYZ$. Let the voxels $\langle (X_i, Y_j, Z_{ij}) : (X_i, Y_j) \in \mathbf{Q}_{IJ},\ Z_{ij} \in \mathbf{H} \rangle$ represent the digital surface $\mathbf{Z}$ over an equi-spaced lattice $\mathbf{Q}_{IJ} = \{(X_i, Y_j) : i = 0, \ldots, I-1;\ j = 0, \ldots, J-1\}$. The set $\mathbf{H}$ of heights is a set of K equi-spaced values, $\mathbf{H} = \{Z_k : k = 0, \ldots, K-1\}$. For simplicity, we restrict the consideration to cubic voxels whose faces are aligned normal to the axes of the world coordinate system. Figure 1 shows a X- or Y-section of the 3D space where each voxel is represented by three sides of a square depicted by boldface lines with "bullet" ends. Black arrows show viewing directions, and "H", "VR", and "VL" denote, respectively, a horizontal face, visible to cameras with higher Z-positions, and vertical faces, visible to cameras with greater or smaller X- or Y-positions (that is, placed to the right or to the left of the voxel). Generally, the actual visibilty of these faces as well as admissible X- or Y-transitions, depicted by thin lines in Figure 1, between the visible neighboring faces have to be taken into account.

The calibration yields a projective correspondence between the 3D point coordinates (X, Y, Z) and the 2D image point coordinates $(x_{[t]}, y_{[t]})$ for every camera $t \in \mathbf{T} = \{1, \ldots, T\}$. If $G_{ij} \equiv G(X_i, Y_j, Z_{ij})$ and $g_{[t]} \equiv g_{[t]}(x_{[t]}, y_{[t]})$ are, respectively, the gray values in the 3D terrain point (X_i, Y_j, Z_{ij}) and in the corresponding 2D point $(x_{[t],ij}, y_{[t],ij})$ of the image $\mathbf{g}_{[t]}$ received by the camera t then $\mathbf{G_{IJ}} = \{G_{ij} : i = 0, \ldots, I-1;\ j = 0, \ldots, J-1\}$ is a terrain orthoimage.

Our methodology produces a simple two-stage DEM reconstruction. In the first stage, every position $(X, Y) \in \mathbf{Q_{IJ}}$ is examined. For each height $Z \in \mathbf{H}$, there is a corresponding 2D perspective projection of the 3D point (X, Y, Z) on each of the T images. For each image for which the 2D perspective projection of (X, Y, Z) lies on the image, there is an observed gray value. This produces the gray values $g_1, \ldots, g_S$. Let g_{min}, g_{max}, and g_{med} be, respectively, the minimum, the maximum, and the median of these gray values. We define the dissimilarity of the S gray values by $\hat{d} = (\max\{0, \varepsilon_{min} \cdot g_{max} - \varepsilon_{max} \cdot g_{min}\})^2$, where ε_{min} and ε_{max} are given numbers which bound the admissible variations in the surface albedo and transfer factors for the cameras. Some other tested measures, say, $\hat{d} = \max\{0, \varepsilon_{min} \cdot g_{max} - g_{med},\ g_{med} - \varepsilon_{max} \cdot g_{min}\}$, gave worse results in our

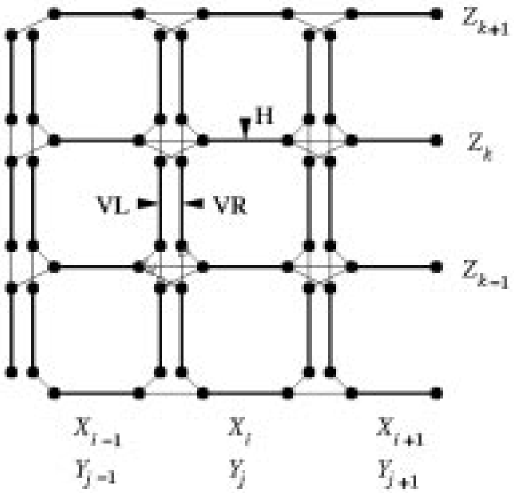


Fig. 1. Transitions between the voxel faces

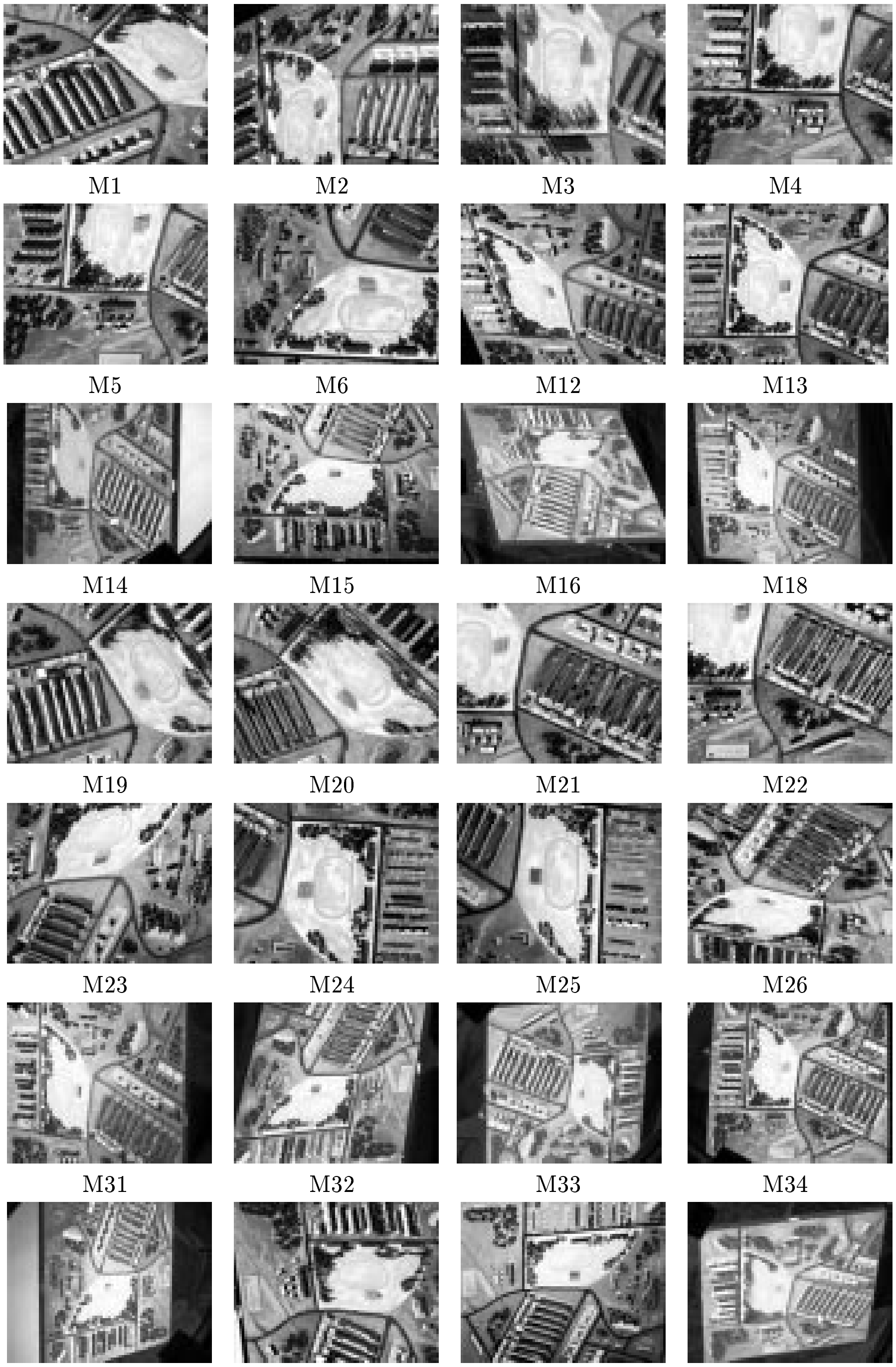
experiments. We choose the height Z giving the smallest value of dissimilarity. And we assign the gray value g_{med} as the gray value at position (X, Y) of the ortho-image. For a confidence measure we use the range $\hat{R} = g_{max} - g_{min}$.

At the second stage, the reconstructed DEM is refined by checking possible occlusions of its voxels. If any less confident voxel occludes the more confident one from the viewing camera then the height of the occluding voxel is cut so as to exclude the occlusion. The confidence values are used, also, for the adaptive moving-window median smoothing of the refined DEM. The window contains only the points that have the same or higher confidence as the central window point and form a continuous region around it. In spite of simplicity, the proposed approach gives promising results for real pre-calibrated images.

3 Basic Features of Multiple Terrain Views

These features are evident from the RADIUS model-board image sets [8]. Figure 2 show reduced examples from the set "M" containing 40 digital images, the size of 1350 pixels × 1035 lines. These images were taken with different resolution (compare, for example, M16 and M20 or M35 and M36), at different time, and under the distinct illumination.

The terrain smoothness varies arbitrarily and there are notable surface discontinuities, say, for the platform in the stadium or for the buildings. Only a central part of the model board is covered by all the views. Other parts are viewed only by different subsets of the cameras, down to two cameras per point. Due to occlusions, the image gray values collected for a 3D point which could be visible to several cameras, may correspond to different surface points. There are differently placed mobile objects such as cars in different parts of the images.



M1 M2 M3 M4

M5 M6 M12 M13

M14 M15 M16 M18

M19 M20 M21 M22

M23 M24 M25 M26

M31 M32 M33 M34

M35 M36 M39 M40

Fig. 2. Model board images from the RADIUS "M" set

Along with changes of the albedo of the surface points, the overall illumination itself varies from one to another subset of the images so that these subsets have different layouts of shadows and distinct contrasts for the same objects (say, for the walls of the buildings or the stadium's platform). Also, the calibration errors result in matching neighboring but different surface points.

If the point is not occluded and sensed under the same illumination, the signals form, mostly, a cluster which depends only on variations in the surface albedo and cameras transfer factors. There can be several such clusters that correspond to different illuminations and changes of the shadow layouts. At the same time, the signals for the points occluding the current one from some cameras are more or less uniformly distributed over the gray range.

It is obvious that the less the signal range, the more plausible that there are no outliers, namely, signals for the occluded or shadow points. Thus, the signal range evaluates the confidence of the heights $\hat{\mathbf{Z}}$ found by minimizing the dissimilarities $\hat{d}$ for every model voxel over the supporting domain $\mathbf{Q}$.

4 Experimental Results and Conclusions

The experiments were carried out with the above-mentioned set "M" of the RADIUS images. The voxel lattice has the size $581(I) \times 581(J) \times 61(K)$ with the coordinate ranges $X_0 = -5$, $X_{I-1} = 53$, $Y_0 = -13$, $Y_{J-1} = 45$, $Z_0 = -1.5$, and $Z_{K-1} = 4.5$ units. Figure 3 shows the range image and orthoimage of the reconstructed DEM. By comparing the orthoimage with Figure 2 one can conclude that main features of this model-board scene are represented in the reconstructed DEM and its orthoimage. But, there are notable errors, mostly, in

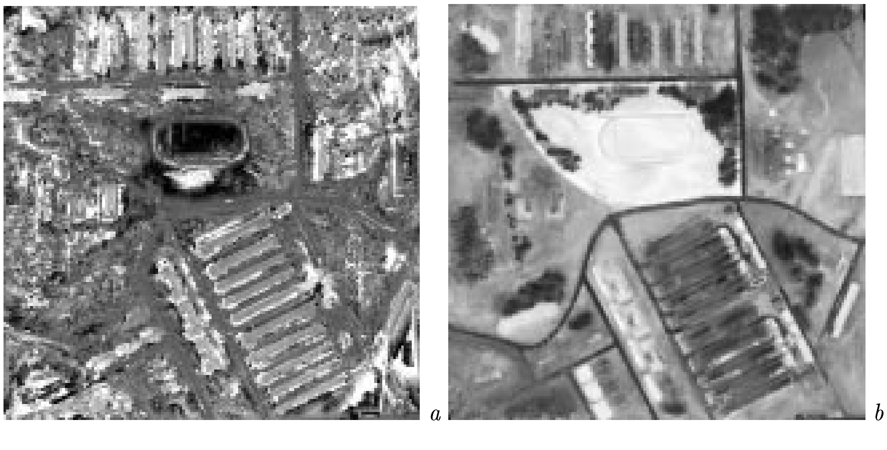


Fig. 3. Reconstructed DEM (*a*) and orthoimage (*b*)

the less confident areas (most them have large Z-values being white in the DEM image).

Figure 4,a displays the image of the confidence values: the darker the point, the higher the confidence, that is, the more narrow the signal scatter. As one might expect, less confident voxels are concentrated around buildings and vegetation, that is, in most occluded areas and areas where the shadow layouts are changing under different illumination. These errors are excluded by a subsequent refinement, as shown in Figures 4,b, and smoothing with the moving window 9×9 6,a. Figure 6,b, demonstrates the smoothed refined DEM with overlaid outlines of the real roofs of the buildings. It is easily seen that the resulting DEM has rather good correspondence with the ground truth.

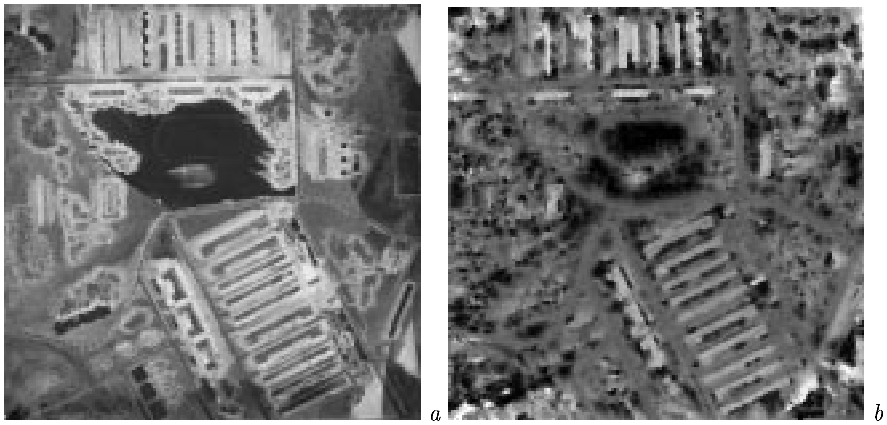


Fig. 4. Confidences (a) for the reconstructed DEM to get the refined DEM (b)

Reconstruction errors are estimated by comparing the DEM with the known 138 GCPs and 497 auxiliary passpoints used for the cameras calibration [11]. Figure 5, a gives positions of them in the reconstructed DEM. Here, cross sizes indicate relative error values. It should be noted thast most GCPs had been placed at the corners of the roofs and of the foundations of the buildings. These places are most difficult to our simplified reconstruction which searches for a single voxel per a planar position $(X_i, Y_j$ minimizing the dissimilarity between the corresponding signals so that chooses only one arbitrary voxel between the roof and the foundation along a visible wall of the building. The terrain discontinuities where all the voxels to be found have the same planar position need some other processing techniques taking into account all the visible voxel faces and admissible transitions between them (see Figure 1).

Bounds $[\varepsilon_{min}, \varepsilon_{max}]$ in the range $[1.0, 1.0] \ldots [0.7, 1.3]$ change the final error rate within 10-15% para to the best results obtained with the bounds $[0.9, 1.1]$. These latter results are summarized in Table 1 giving mean values ("mae") and

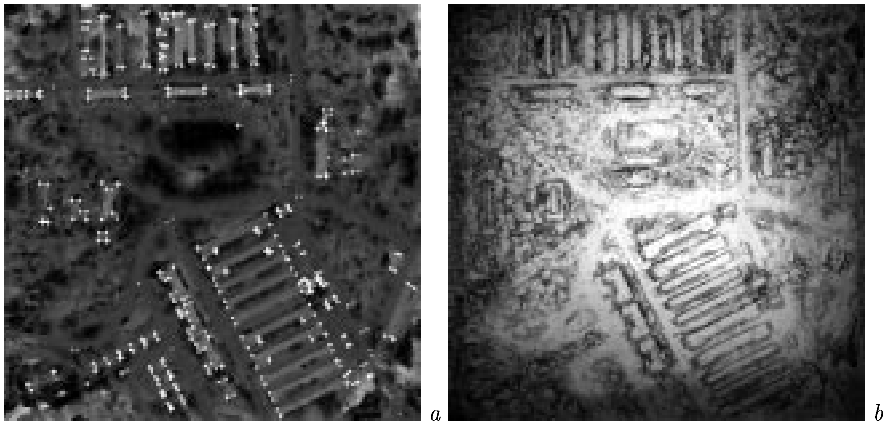


Fig. 5. Control points (a) and visibility pattern (b)

standard deviations ("std") of the absolute DEM height errors relative to the control points and their cumulative histograms. In total, 69.6% of the GCPs and

Table 1. Precision of the DEM reconstruction

DEM	138 GCPs						497 passpoints					
	mae	std	≤ 0.1	≤ 0.2	≤ 0.3	≤ 0.6	mae	std	≤ 0.1	≤ 0.2	≤ 0.3	≤ 0.6
reconstructed	0.49	0.64	47	73	82	96	0.60	0.85	200	255	283	329
refined	0.39	0.42	50	65	78	108	0.40	0.45	176	245	292	371
smoothed	0.28	0.34	67	81	96	118	0.31	0.37	211	293	330	398

66.4% of the passpoints have the absolute error less than 0.3, that is, less than 5% of the height range in our model. Thus, the proposed approach, in spite of its simplicity, yields rather good close approximation to the desired dense DEM. The overall quality of the resulting DEM can be checked qualitatively also by estimating the visibility of the terrain points in terms of numbers of the cameras that view every point. Such a "visibility" pattern of the final DEM is shown in Figure 5, *b*. Here, the signals are proportional to the numbers of the viewing cameras: the more black the point, the less the number in the range $2 \ldots 40$. It is apparent that the reconstructed DEM, in spite of some local errors, reflects most characteristic features of the observed scene. In total, this visibility pattern is consistent with the one expected by visual perception of the initial images.

Our experiments show that a feasible approximation to the dense DEM of the terrain viewed by a set of the calibrated cameras can be obtained by independent reconstruction of each terrain point. The confidence values for the chosen voxels are crucial in excluding most part of the errors from the reconstructed DEM.

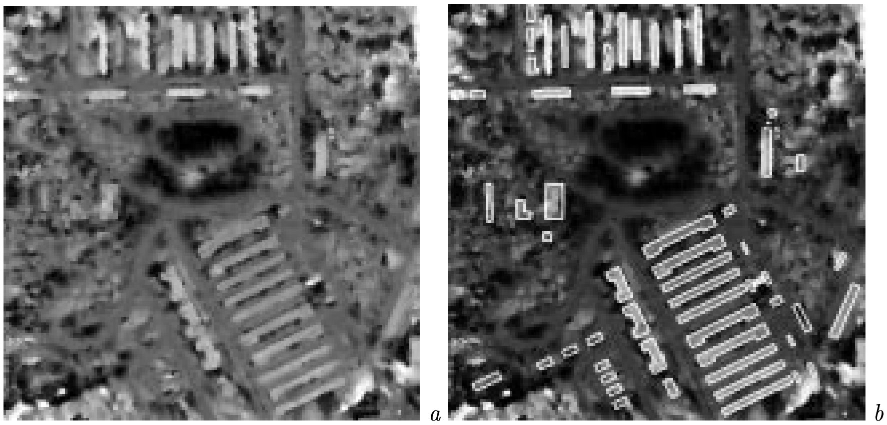


Fig. 6. Smoothed refined DEM (*a*) with overlaid roof outlines (*b*)

Of course, the obtained rough representation of the viewed terrain needs to be further refined by more elaborate techniques. But, it possesses basic features of the observed terrain and therefore can be useful in practice.

References

1. Agouris, P., Schenk, T.: Automated aerotriangulation using multiple image multipoint matching. Photogramm. Eng. Remote Sens. 62:6 (1996) 703-710
2. Baker, H. H.: Surfaces from mono and stereo images. Photogrammetria 39:4-6 (1984) 217-237
3. Gimel'farb, G. L.: Symmetric bi- and trinocular stereo: tradeoffs between theoretical foundations and heuristics. Computing Suppl. 11 (1995) 1-19
4. Haralick, R. M., Shapiro, L. G.: Computer and Robot Vision. Addison-Wesley Publ. (1993) Vol. 2
5. Haralick, R. M., Thornton, K. B.: On robust exterior orientation. Robust Computer Vision: Quality of Vision Algorithms (W. Förstner, S. Ruwiedel, Eds.). Herbert Wichmann Verlag: Karlsruhe (1992) 41-49
6. Jenkin, M. R. M., Jepson, A. D., Tsotsos, J. K.: Techniques for disparity measurement. CVGIP: Image Understanding 53:1 (1991) 14-30
7. Kireytov, V. R.: Inverse Problems of Photometry. Computing Center, Acad. Sci. USSR, Siberian Branch: Novosibirsk (1983) [*In Russian*].
8. RADIUS Model Board Imagery and Groundtruth. CD-ROM, Vol. 1 and 2. ISL, Univ. of Washington: Seattle, USA (1996)
9. Schultz, H.: Shape reconstruction from multiple images of the ocean surface. Photogramm. Eng. Remote Sens. 62:1 (1996) 93-99
10. Shekarforoush, H., Berthod, M., Zerubia, J., Werman, M.: Sub-pixel Bayesian estimation of albedo and height. Int. J. Computer Vision 19:3 (1996) 289-300
11. Thornton, K. B.: Accurate Image-Based 3D Object Registration and Reconstruction. Dissertation (Ph.D.), Univ. of Washington (1996)