HIGH-RESOLUTION DEM GENERATION USING SELF-CONSISTENCY

Frank Stolle [a], Howard Schultz [a], Dong-Min Woo [b]

[a] Department of Computer Science, University of Massachusetts, Amherst, MA, USA - {stolle, hschultz}@cs.umass.edu
[b] School of Electrical Engineering, Myongji University, Yongin, Korea - dmwoo@wh.myongji.ac.kr

Commission I, WG I/5

KEY WORDS: DEM Extraction, Matching, Satellite Imagery

ABSTRACT:

We present a practical method for producing a high-resolution digital elevation map (DEM) from a set of spatially overlapping images. We start with the principle that multiple low-resolution DEMs can be combined to form a high-resolution DEM provided that (i) the sample DEMs are registered to the same coordinate system; (ii) the elevation estimates at each posting are partially independent; and (iii) the elevation errors between postings within a DEM are partially independent. Under these conditions, each sample DEM contributes an incremental amount of new information to the composite DEM. As a result, increasing the number of sample DEMs will increase the spatial resolution as well as decrease its vertical error bounds of the composite DEM. The key to an effective implementation of this algorithm is the ability to produce sample DEMs which are free from large blunders.
We are able to separate valid elevation estimates from blunders by exploiting their uniquely statistical properties. In general elevation estimates (computed by image matching algorithms) are members of one of two populations: *Valid* estimates in which the image matching algorithm identified a pair of corresponding pixels (one in the reference image and one in the target image) that are projections of the same surface point; and *blunders* in which the image matching algorithm identified corresponding pixels that are not projections of the same surface point. Since blunders are not related to the surface, corresponding pixels can occur anywhere in the search range. As a result, valid elevation estimates have an error distribution characterized by an unbiased normal distribution with a small standard deviation, and (ii) blunders have an error distribution characterized by a uniform distribution spread out over a much larger range. The standard deviation of the valid error distribution corresponds to the elevation range seen by a single pixel; whereas, the width of the blunder distribution corresponds to the full search range.

1. INTRODUCTION

This paper is concerned with the process of fusing repeated samples of motion parallax captured by a moving camera to recover the 3D structure of a scene. In particular, we describe a DEM tiling approach that produces an accurate 3D representation by fusing a sequence of DEMs made from repeated samples of the motion parallax. The method generates a sequence of partially overlapping DEM tiles, which are averaged together to reduce noise and form a continuous model of the underlying terrain. For this method to be successful, a normally distributed, unbiased process must characterize the error distribution of the elevation data in individual tiles.

Unfortunately, computed DEMs often are contaminated by large errors (blunders) that result from errors during the image matching process. These errors produce spikes in the disparity and elevation maps. Although averaging reduces their magnitude, the spikes are not eliminated from the data. As a result, combining multiple DEMs has the undesirable effect of increasing the frequency of error spikes. Thus, identifying and removing large errors from the individual DEM tiles is key to fusing a collection of DEM tiles.

We use the principle of self-consistency developed by Leclerc, Luong and Fua (Leclerc et al.,1998a) to identify unreliable points in a distribution. The main focus of their work was to obtain a quality measure for correspondence algorithms without relying on ground truth. Their algorithm obtained a probability distribution by counting the number of corresponding image points for each object point that is consistent with the viewing geometry within a specified error limit. Consistency checking in stereo algorithms has previously been used (Chang et al.,1991; Faugeras et al.,1993; Fua,1993; Konolige,1997; Matthies et al.,1995). In a closely related application (Fua and Leclerc,1996; Leclerc et al.,1998b), extend their work to detect changes in terrain by applying the concept of self-consistency to elevations. We improve on consistency checking by first detecting unreliable elements in DEMs generated from two images and then excluding unreliable estimates from an information fusion process across more than two stereo pairs.

2. METHODOLOGY

Our algorithm uses the self-consistency difference to separate valid elevation estimates from blunders. The self-consistency difference is defined as the elevation difference arising from changing roles of the reference and target images. From a pair of spatially overlapping images *A* and *B*, two DEMs can be produced $\mathbf{Z}_{AB}$ and $\mathbf{Z}_{BA}$ with a self-consistency difference ($\mathbf{Z}_{AB}$ - $\mathbf{Z}_{BA}$). We show through an analysis using simulated and real data that the self-consistency difference ($\mathbf{Z}_{AB}$ - $\mathbf{Z}_{BA}$) in conjunction with the match scores can be used to separate valid elevation estimates from blunders.

This model predicts that the blunders will have a disproportionate influence on the tails of the distribution of elevation errors. Thus, to accurately represent the elevation error distribution, the histogram of ($\mathbf{Z}_{AB}$ - $\mathbf{Z}_{BA}$) is fit to a model that includes a Gaussian distribution plus a constant,

$$h_{\max} \exp - \left[\frac{(h_i - z_0)^2}{2s^2} \right] + h_{\min} \quad (1)$$

where h_i are the histogram entries from ($\mathbf{Z}_{AB}$ - $\mathbf{Z}_{BA}$) , z_0 and s are the mean and standard deviation of the distribution, respectively, $h_{\min}$ is the asymptotic value of the tails and $h_{\max}$ is the value at the peak of the distribution. The parameters (z_0, s, $h_{\min}$, $h_{\max}$) are found by fitting the model expressed in Equation 1 to the histogram of ($\mathbf{Z}_{AB}$ - $\mathbf{Z}_{BA}$).

Based on this noise model, three methods for detecting blunders are explored:

1. Set a blunder threshold to a multiple of s (the width the distribution in Equation 1). The elevation estimates $\mathbf{Z}_{AB}(x,y)$ and $\mathbf{Z}_{BA}(x,y)$ are considered reliable if $|\mathbf{Z}_{AB}(x,y) - \mathbf{Z}_{BA}(x,y)| < n \cdot s$, (where n is a user defined parameter).

2. Set a blunder threshold to a user defined absolute value. Since the self-consistency difference can be interpreted as an uncertainty, the operator can reject elevation estimates where the self-consistency difference is above a desired value. For example, the operator may wish to reject any posting where the uncertainty is greater that 1 meter.

3. Set a blunder threshold to accept a percentage of the postings.

If there are N overlapping images, each from a unique viewpoint, then there are a total of $N(N-1)$ ordered image pairs. Thus, there are a maximum of $N(N-1)$ samples at every posting in the DEM. Applying one of the blunder detection criteria may reduce the number of elevation estimates. However, as long as there are at least two elevation estimates at a posting, an improved mean and estimated error bounds can be found.

3. EXPERIMENTS

The method is tested using two data sets, a synthetic data set consisting of a dense array of elevation postings and a set of photorealistic images of a desert scene, and a set of nine Ikonos images of central Boston. The synthetic data set was used to evaluate the spatial error characteristics of the DEM fusion method. For example, to examine the spatial decorrelation characteristics requires ground truth that is denser than the decorrelation length scale of the data. The use of synthetic data is based on the assumption that a previously acquired DEM and ortho-image can serve as pseudo ground truth over an extended area, and that a photo-realistic synthetic image can act as an image source. Starting with a high resolution DEM and ortho-image, the method generates synthetic images from arbitrary viewpoints using a ray-tracing algorithm. The quality of the DEM fusion algorithm can be tested by re-computing a DEM from the synthetic images and comparing it to the pseudo ground truth.

3.1 Constructing synthetic pseudo ground truth

We have constructed a pseudo ground truth DEM and ortho-images from a set of four overlapping aerial images. We then used a ray tracing program to create synthetic aerial images. These synthetic images were used as input to reconstruct a DEM. The degree of similarity between the true and reconstructed DEM was analyzed using error variograms.

The following is a summary of the procedures used to create the pseudo ground truth. We started with a set of four 9 x 9 inch aerial images (digitized to 7800 x 7800 pixels) of a high desert, taken from an altitude of about 5000 feet. To make the tests more tractable, we selected four 2000 × 2000 pixel sub-images that were almost completely mutually overlapping. Figure 1 and Figure 2 show two of these sub-images.

From these four sub-images (labelled A, B, C, and D) we constructed the pseudo ground truth as follows: First, each image pair (A-B, A-C, A-D, B-C, B-D, C-D) was converted to a pair of epipolar aligned images. Then the disparity maps between the epipolar aligned image pairs were found using the image matching algorithm developed in (Schultz,1995). For each epipolar pair (for example, A and B), two disparity maps $\mathbf{d}_{AB}$ and $\mathbf{d}_{BA}$ were computed. The first subscript denotes the reference image and the second subscript denotes the target image. The reference and target images are defined such that the pixels (i,j) in the reference image and (i+$\mathbf{d}$(i,j), j) in the target image are projections of the same surface facet.

Figure 1: Sub-image A of the high desert imagery (2000x2000 pixels).

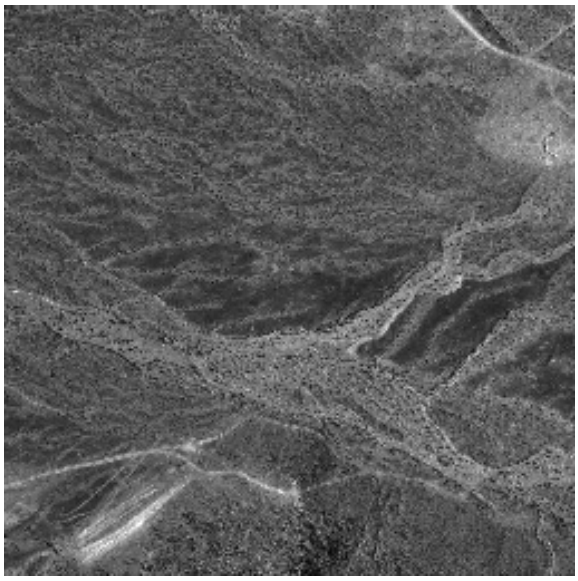

Figure 2: Sub-image B of the high desert imagery (2000x2000 pixels).

From an image processing point of view, the disparity maps define a function that warps the target image into the reference image. The function **d**(i,j) defines a displacement for the pixels in the target image that brings it into alignment with the reference image. From a pair of images, two warping functions (disparity maps) can be defined – $\mathbf{d}_{AB}$ which warps B into A, and $\mathbf{d}_{BA}$ which warps A into B.

In the absence of image matching errors the DEMs ($\mathbf{Z}_{AB}$ and $\mathbf{Z}_{BA}$) derived from the disparity maps ($\mathbf{d}_{AB}$ and $\mathbf{d}_{BA}$) would be similar. Matching algorithms, however, are notorious for finding false matches which produce large errors in the recovered DEMs. Our system was specifically designed to reduce the occurrence of false matches and to detect and remove false matches (or the elevation errors resulting from false matches). After image matching, the next step was ortho-rectification which converted each disparity map into a DEM. In this case, the four test images yielded 12 disparity maps ($\mathbf{d}_{AB}$, $\mathbf{d}_{BA}$, $\mathbf{d}_{AC}$, $\mathbf{d}_{CA}$, etc.) and twelve sample DEMs ($\mathbf{Z}_{AB}$, $\mathbf{Z}_{BA}$, $\mathbf{Z}_{AC}$, $\mathbf{Z}_{CA}$, etc.).

The final step in creating a pseudo ground truth data set was to fuse the 12 sample DEMs into the single high quality DEM and ortho-image. This was done by computing six self-consistency errors maps by taking the difference between the DEMs created from a pair of epipolar aligned images, $\boldsymbol{\delta}_{AB} = (\mathbf{Z}_{AB} - \mathbf{Z}_{BA})$, $\boldsymbol{\delta}_{AC} = (\mathbf{Z}_{AC} - \mathbf{Z}_{CA})$, etc. The self-consistency errors are compared to a threshold (e.g., 1.0 meter). If the self-consistency difference exceeds this threshold, the elevations estimates are labelled unreliable. Next, the fused DEM was formed by taking a weighted average of the reliable elevation estimates. The weights are the inverse of the self-consistency differences (i.e., $1/\boldsymbol{\delta}$). This maximum likelihood approach assigns a higher value to elevation estimates where self-consistency difference is small. The fused DEM is shown in
Figure 3. Further details of the fusing process are described in (Schultz et al.,2002).

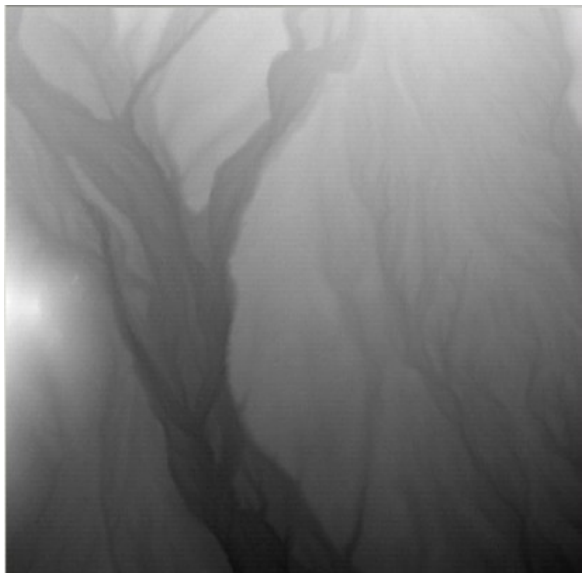

Figure 3 The fused pseudo ground truth DEM.

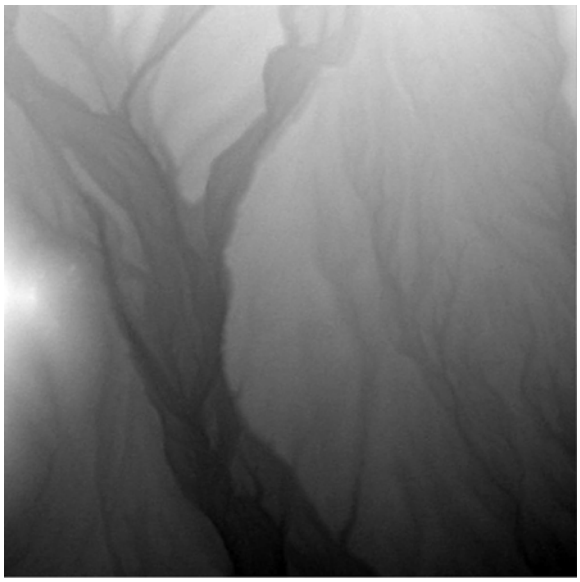

Figure 4: The DEM recovered by fusing the individual DEMs generated from the synthetic images.

3.2 Recovering the DEM from synthetic images

We then created a set of four synthetic images from the pseudo ground truth DEM using a photo-realistic ray tracing program. This allowed us to generate images of the pseudo ground truth from an arbitrary position. As an initial test we decided to place the synthetic cameras in the same position as the positions from which the original images were taken. This allowed us to double-check our viewpoint synthesis procedure by making sure that image features corresponded across the original and the synthesized images. Synthetic images were created at the same resolution as the ground truth DEM (2000 x 2000 pixels) so that we could test the ability to capture fine spatial detail.

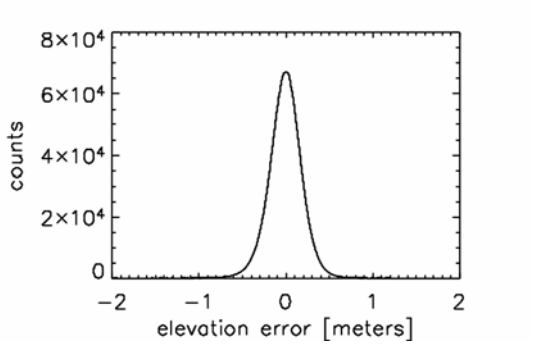

Figure 5: Distribution of differences of elevation estimates between Figure 3 and Figure 4.

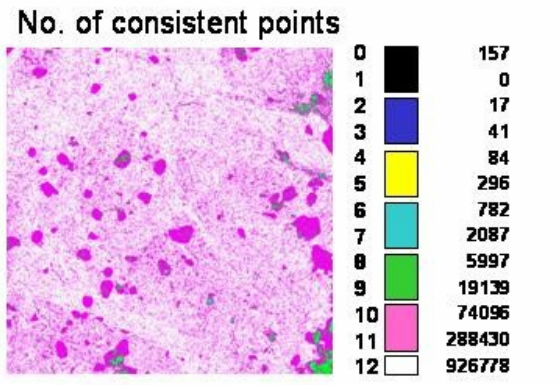

Figure 6: A map of the number of reliable elevation estimates contributing to the final result.

From the four synthetic images (*A,B,C,D*) a total of twelve DEMs ($\mathbf{Z}_{AB}$, $\mathbf{Z}_{BA}$, …, $\mathbf{Z}_{CD}$, $\mathbf{Z}_{DC}$) and six self-consistency difference arrays ($\mathbf{Z}_{AB}$,- $\mathbf{Z}_{BA}$, …, $\mathbf{Z}_{CD}$ - $\mathbf{Z}_{DC}$) were generated. Using the first method described above for selecting reliably postings, a threshold of 2*s* was selected, which corresponded to ±0.4m. The resulting DEMs were fused according to the above described procedure and the result is shown in Figure 4. Figure 5 shows the distribution of the differences between this DEM and the pseudo ground truth; Figure 6 shows a map of the number of reliable elevation estimates. About 98% of the fused DEM elements had 10 or more reliable values, 215 out of the 4 million had 3 or less reliable estimates, and only 157 had no reliable elevations estimates. In addition, there were no apparent blunders or spikes in the fused DEM, and the rendered surface appeared realistic.

3.3 Experiments with Ikonos data

In the second set of experiments, we constructed a fused DEM of central Boston from nine Ikonos images, collected from three orbital passes. From each of these 10K × 10K Ikonos images, a 2048 × 2048 sub-image that viewed the same area of Boston were extracted. Image 46, one of the 2048 × 2048 sub-images is shown in Figure 7 (each image was labelled by two unique digits taken from its original Space Imaging label).

To avoid possible epipolar alignment problems with cross-track image pairs, we generated DEMs only from image pairs taken along the same orbital track. The nine images resulted in eighteen pairs. We applied epipolar alignment to this set of images and created nine 3072 × 3072 epipolar aligned (rectified) image pairs (three image pairs from each of the three orbital passes). A larger output size was chosen to compensate for geometric distortions introduced by epipolar alignment.

Figure 7: Ikonos image 46 of central Boston.

Figure 8: A 400x400 sub-image of image 46, resampled in epipolar geometry, generated from image pair 46-48.

A sample of one epipolar aligned image pair (orbit 2 image pair 46-48) is shown in Figure 8 and Figure 9. Notice the large amount of perspective distortion associated with the tall buildings, the presence of confusing shadow patterns, and the large number of occluded surfaces. Also notice that all stereo parallax is horizontal. The epipolar-aligned pairs were examined manually to test for any vertical stereo disparity. Less than one pixel of vertical stereo disparity was detected. These results verified the quality of the RPC-epipolar alignment algorithm for images taken along a single orbit.

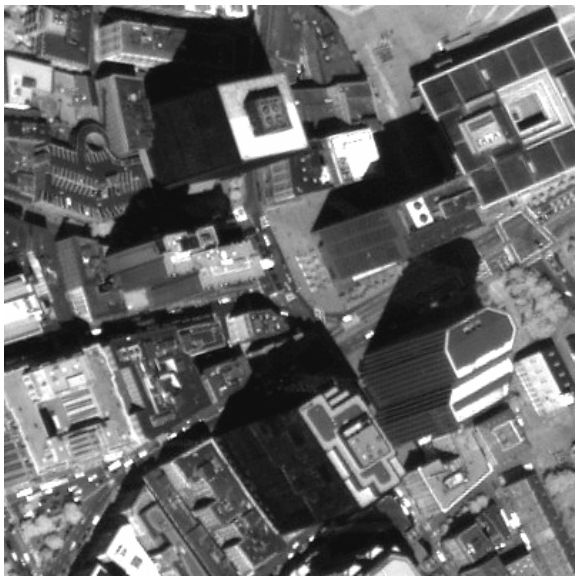

Figure 9: A 400x400 sub-image of image 48, resampled in epipolar geometry, generated from image pair 46-48.

For each epipolar-aligned image pair, two disparity maps were generated using the image matching procedure, resulting in eighteen disparity maps. Then a DEM was computed for each disparity map using ortho-rectification. We also ortho-rectified the match scores for each disparity estimate. Elevation estimates were labelled reliable if the self-consistency difference fell below a preset threshold and the match scores were above a preset threshold (in this case the thresholds were 3.0 meters and 0.8 respectively). The size of each DEM was 2454 postings in latitude by 3308 postings in longitude. The latitude range was from 42.3444° to 42.3632° North, and the longitude range was from -71.0736° to -71.0482° East, with a grid spacing in latitude and longitude of 7.66e-6°. The distribution of the elevations is shown in Figure 13. The elevation range was (-50m, 184m) relative to the WGS84.

The eighteen sample DEMs were fused to form (1) a single high quality DEM, (2) a map of the standard deviation of reliable elevation estimates, (3) map of the number of reliable elevation estimates used to compute each elevation estimate, and (4) a 9-bit code giving the image pairs that produced the elevation estimates (e.g., if bit-0 is set then the image pair 43-44 contributed to the elevation estimate, if bit-1 is set then image pair 43-45 contributed, etc.).

A typical sample DEM (from image pair 46-47) and it elevation distribution are shown in Figure 10 and Figure 11. Notice the apparent lack of detail in the DEM. The fused DEM and its elevation distribution are shown in Figure 12 and Figure 13. The fused DEM shows considerably more detail. Also note that small peaks can be seen in the tails of the elevation distribution which may correspond to tall buildings. To illustrate the relationship between the fused DEM and buildings, the fused DEM was projected back into one of the original Ikonos images. Figure 14 shows a 2k × 2k section from one of the original Ikonos images and the corresponding fused elevations. The power of the fusion method can be seen by comparing the sample DEM (Figure 10) with the fused DEM (Figure 12).

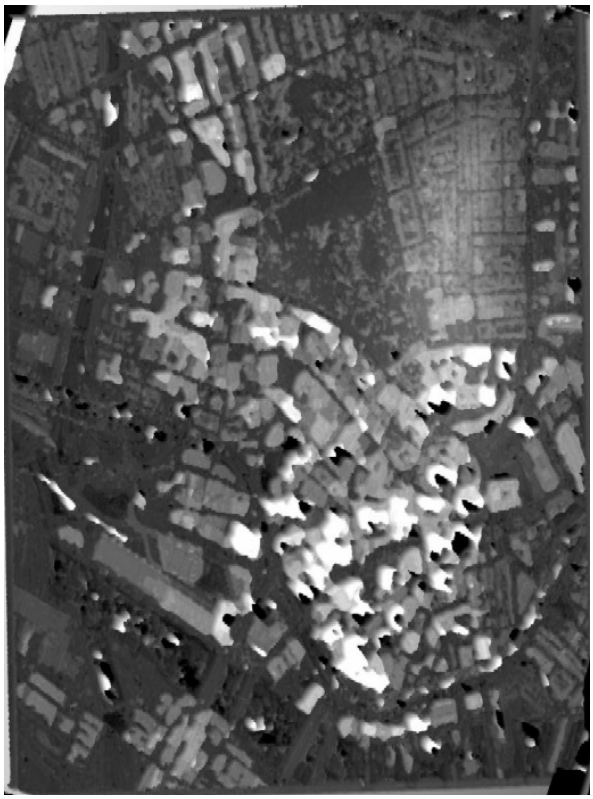

Figure 10: A DEM generated from a single pair (46-47) of Boston images.

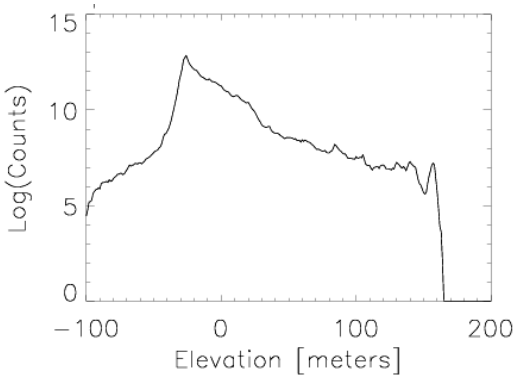

Figure 11: The elevation distribution for the sample DEM shown in Figure 10.

4. ACKNOWLEDGEMENTS

This research has been supported by the Army Research Office under grant number DAAD19-99-1-0016, the National Science Foundation under grants EIA-9726401, EIA-9726401 (KOSEF), EIA-0105272 (SGER), and IIS-0430742, and the Korean Agency for Defense Development (ADD), Seoul, South Korea.

Figure 12: The final DEM computed by fusing all DEMs generated from the set of nine Ikonos images. The elevations are encoded in greyscale such that lower elevations are dark and higher elevations are light. Unreliable elevations are encoded in black.

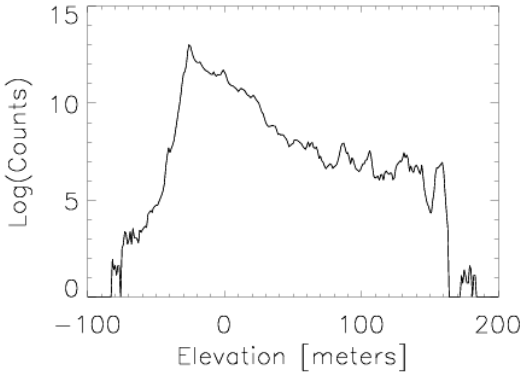

Figure 13: The elevation distribution for the fused DEM shown in Figure 12.

5. REFERENCES

Chang, C., Chatterjee, S. and Kube, P.R., 1991. On an Analysis of Static Occlusion in Stereo Vision, Computer Vision and Pattern Recognition, pp. 722-723.

Faugeras, O.D., Hotz, B., Matthieu, H., Vieville, T., Zhang, Z., Fua, P., Theron, E., Moll, L., Berry, G., Vuillemin, J., Bertin, P. and Proy, C., 1993. Real Time Correlation-Based Stereo: Algorithm, Implementations and Applications. INRIA Technical Report 2013.

Fua, P., 1993. A Parallel Stereo Algorithm that Produces Dense Depth Maps and Preserves Image Features. Machine Vision and Applications, 6: 35-49.

Fua, P. and Leclerc, Y., 1996. Taking Advantage of Image-Based and Geometry-Based Constraints to Revover 3-D Surfaces. Computer Vision and Image Understanding, 64(1): 111-127.

Konolige, K., 1997. Small Vision Systems: Hardware and Implementation, Eighth International Symposium on Robotics Research.

Leclerc, Y., Luong, Q.-T. and al., e., 1998a. Self-consistency: A novel approach to characterizing the accuracy and reliability of point correspondence algorithms, DARPA Image Understanding Workshop. Morgan Kauffman, Monterey, CA.

Leclerc, Y., Luong, Q.-T. and Fua, P., 1998b. A Framework for Detecting Changes in Terrain. IEEE Transactions on Pattern Analysis and Machine Intelligence, 20(11): 1143-1160.

Matthies, L., Kelly, A., Litwin, T. and Tharp, G., 1995. Obstacle Detection for Unmanned Ground Vehicles: A Progress Report, Intelligent Vehicles '95 Symposium, pp. 66-71.

Schultz, H., 1995. Terrain Reconstruction from Widely Separated Images., SPIE, Orlando, Florida, pp. 113-123.

Schultz, H., Hanson, A., Riseman, E., Stolle, F., Woo, D.-M. and Zhu, Z., 2002. A Self-consistency Technique for Fusing 3D Information, IEEE Fifth International Conference on Information Fusion, Annapolis Maryland.

Figure 14: A composite color image of Boston with the original image (Figure 7) encoded in value, elevations (Figure 12) encoded in hue, and the standard deviation of the elevation errors encoded in saturation.