Error Modeling Techniques for Improved Automatic DEM Estimation

Howard Schultz
University of Massachusetts
Department of Computer Science
Amherst, MA 01003
hschultz@cs.umass.edu

Abstract. The implementation of real-time Digital Elevation Model (DEM) estimation algorithms face several challenges. Perhaps the most difficult one is the detection of spurious elevation errors arising from errors in the image matching process. A close second is the multi-view stereo problem—finding methods that exploit the tremendous amount of topographic information captured by a sequence of spatially overlapping images. Paying attention to the appropriate error distribution can solve both challenges. We present a method for estimating the elevation error distribution and show that applying a threshold to the distribution can separate false and correct matches. In addition, a method is presented for estimating the vertical uncertainty and horizontal resolution of the DEM products. Two examples are presented. The first demonstrates the ability to recover a high-resolution DEM of a scene with highly repetitive patterns. In the second example recovers a DEM of Boston from nine Ikonos images.

1. Estimating vertical errors and horizontal resolution with pseudo ground truth

The lack of dense ground truth (one ground truth posting per image pixel) severely limits the ability to analyze the performance of DEM extraction methods. To overcome problems associated with limited ground truth a *pseudo ground truth* dataset is constructed. The process begins by creating a dense DEM and orthoimage from one or more pairs of spatially overlapping images. The computed DEM and orthoimage comprise the ground truth dataset. Because the ground truth was not surveyed, we refer to it as pseudo ground truth. In other words, we treat the first generation, computed DEM as a collection of measured elevation postings. Next, synthetic images are rendered using a photo-realistic ray-tracing program. The input to the ray-tracing program is the pseudo ground truth and camera orientation. The rendered images are treated as digitized photographs of the ground truth area. The performance of the DEM extraction algorithm is analyzed by comparing the second generation DEM (computed from the rendered images) to the first generation, pseudo ground truth DEM.

The pseudo ground truth method was used to analyze the quality of a digital elevation model of the USMC Air Ground Combat Center, Twenty-nine Palms CA. The DEM was created using the Terrest software. The input was two digitized aerial photographs taken at an altitude of approximately 5000 meters and digitized with a 30 micron spot size, resulting in approximately 7800 × 7800 pixel images. Two spatially overlapping 2000 × 2000 sub-images were selected for analysis. Figure 1 shows the two sub-images, and Figure 2 shows the estimated DEM. Next, two rendered images were generated using a photo-realistic ray-tracing program. Figure 3 shows a comparison between one of the original images and a rendered image, where the orientation of the rendered image was identical to the orientation of the original image.

A second generation DEM was generated from the rendered images and compared to the DEM generated from the original images. Figure 4 shows the second generation DEM, and the elevation errors found by taking the difference between the first and second generation DEMs. Figure 5 shows the histogram of the elevation errors.

The horizontal error decorrelation length was used as a measure of horizontal resolution. Where the decorrelation length is defined as the lag L, measured in pixels, where the square root error variogram $V(L)$ reaches 95% of its asymptotic value. The square root error variogram is defined by

$$V(L) = \sqrt{\frac{1}{2N(L)} \sum_{(i,j)} \left(\delta(i,j) - \delta(i+L,j) \right)^2}$$

$\delta(i,j) = Z(i,j) - Z_{TRUE}(i,j)$ is the vertical error at pixel (i,j), $Z(i,j)$ and $Z_{TRUE}(i,j)$ are the second generation and pseudo ground truth DEMs, and $N(L)$ is the number samples where both $\delta(i,j)$ and $\delta(i+L,j)$ are below the error detection threshold. Figure 6 shows the second generation DEM with false matches marked in blue. A false match threshold was 1 meter. Excluding the detected false matches the standard deviation of the remaining elevation errors was about 22 cm. At $L=0$ the vertical errors are completely correlated and $V(L)=0$, and when $L \rightarrow \infty$, the vertical errors become completely uncorrelated and the value of $V(L)$ is asymptotic to the standard deviation of $\delta(i,j)$. Figure 7 shows the square root error variogram with no false match detection (pink line) and with the false match detection threshold set to 0.5 meters (black line). With false match detection the square root variogram reaches 95% of its asymptotic value 0.22 meters at a lag of 4 pixels. Thus, the vertical error is approximately 22 cm and the horizontal resolution is approximately 4 pixels.

Without false match detection the square error root variogram does not approach an asymptotic value. This is expected if you consider that vertical errors are of two types: 1) Normally distributed errors associated with correct image matches, and 2) uniformly distributed errors associated with false image matches. Correct image matches have a normal distribution because the matched pixels in the left and right images triangulate to a point near the terrain surface. On the other hand, false matches occur randomly in the images and have no relationship to the terrain.

2. The Image Matching Error Model

The image matching problem is by far the most challenging aspect of computing a digital elevation model from a pair of images. The goal of image matching is to find a disparity map $\mathbf{D}$ such that the pixel (i,j) in the reference image and $(i+\mathbf{D}(i,j),j)$ in the target image view the same spot on the terrain. A match is found by searching for a pattern in the target image that resembles the pattern surrounding pixel (i,j) in the reference image. Usually the image matching algorithm makes a correct choice. However, occasionally a false match is found.

For correct matches vertical errors are normally distributed with a zero mean and a standard deviation of a few pixels of parallax. On the other hand, for false matches the pixels surrounding (i,j) in the reference image and $(i+\mathbf{D}(i,j),j)$ in the target image have no relationship to the terrain. As a result, the errors associated with false matches do not have a zero mean are not normally distributed. Because false matches can occur anywhere in the target image, false match elevation errors are approximately uniform.

Elevation errors, therefore, have a bimodal distribution comprised of a large number of normally distributed errors with a zero mean and small standard deviation, superimposed on a small number of uniformly distributed errors. The key to detecting false matches rests on the ability to find a threshold that efficiently separates the correct and false matches.

3. Estimating Elevation Errors

Starting with a pair of spatially overlapping images (labeled A and B) the image matching procedure finds a disparity map $\mathbf{D}_{AB}$ hat defines a mapping between the reference image A and the target image B such that the pixel (i,j) in A and the pixel $(i+\mathbf{D}_{AB}(i,j), j)$ in B are projections of the same surface element. The image matching process is not symmetric with respect to images A and B. Which is to say that the disparity map found by matching image B to A is not the inverse of the one found by matching image A to B. The asymmetry implies that the two DEMs, $\mathbf{Z}_{AB}$ and $\mathbf{Z}_{BA}$, computed by reversing the reference and target images, have partially uncorrelated errors. Remarkably, variations between $\mathbf{Z}_{AB}$ and $\mathbf{Z}_{BA}$ capture valuable information about the distribution of elevation errors.

To understand how to exploit the asymmetry, we begin by modeling the computed DEMs, $\mathbf{Z}_{AB}$ and $\mathbf{Z}_{BA}$, as the sum of the true surface $\mathbf{Z}_{\text{TRUE}}$ plus an error term δ_{AB},

$$\mathbf{Z}_{AB} = \mathbf{Z}_{\text{TRUE}} + \delta_{AB}$$

Similarly, $\mathbf{Z}_{BA}$ can be modeled by

$$\mathbf{Z}_{BA} = \mathbf{Z}_{\text{TRUE}} + \delta_{BA}$$

Subtracting the two equations eliminates $\mathbf{Z}_{\text{TRUE}}$

$$\mathbf{Z}_{AB} - \mathbf{Z}_{BA} = \delta_{AB} - \delta_{BA}$$

This gives a relationship between the measurable *self-consistency difference* $(\mathbf{Z}_{AB} - \mathbf{Z}_{BA})$ and the unknown elevation uncertainties δ_{AB} and δ_{BA}. Taking the variance of both sides, and making the assumption that δ_{AB} and δ_{BA} have the same variance, gives a relationship between the vertical uncertainty VAR(δ), the variance of the self-consistency difference $(\mathbf{Z}_{AB} - \mathbf{Z}_{BA})$, and the correlation between δ_{AB} and δ_{BA}

$$\text{VAR}(\delta) = \tfrac{1}{2}\,\text{VAR}(\mathbf{Z}_{AB} - \mathbf{Z}_{BA})/(1-\text{CORR}(\delta_{AB},\delta_{BA})) \qquad (1)$$

If $\mathbf{Z}_{AB} = \mathbf{Z}_{BA}$ the errors are uncorrelated and the right side of Equation (1) is undefined, which

means that Equation (1) cannot be used to estimate VAR(δ). If δ_{AB} and δ_{BA} are partially uncorrelated, Equation (1) says that the elevation uncertainty VAR(δ) is a function of the self consistency difference ($\mathbf{Z}_{AB}$ - $\mathbf{Z}_{BA}$) and the correlation between δ_{AB} and δ_{BA}.

The degree of correlation between the elevation errors can be estimated by measuring the decrease in error variance arising from the summation of $\mathbf{Z}_{AB}$ and $\mathbf{Z}_{BA}$. For the two random variables δ_{AB} and δ_{BA} with the same distribution, the reduction in error variance of the average ($\mathbf{Z}_{AB}$ + $\mathbf{Z}_{BA}$)/2 is related to the correlation between δ_{AB} and δ_{BA} by

$$\frac{VAR\left(\frac{1}{2}\left(\delta_{AB}+\delta_{BA}\right)\right)}{VAR\left(\delta_{AB}\right)}=\frac{1}{2}\left(1+CORR\left(\delta_{AB},\delta_{BA}\right)\right) \qquad (2)$$

If δ_{AB} and δ_{BA} are highly correlated, the right-hand side of Equation (2) becomes unity and there will be no reduction of variance from averaging. On the other hand, if δ_{AB} and δ_{BA} are uncorrelated, the error variance is reduced by 1/2.

For the pseudo ground truth example, described in Appendix A, the reduction in variance from averaging can be estimated from the square root variograms of $\mathbf{Z}_{AB}$ and ($\mathbf{Z}_{AB}$ + $\mathbf{Z}_{BA}$)/2 shown in Figure 8. The asymptotic standard deviations of $\mathbf{Z}_{AB}$ and ($\mathbf{Z}_{AB}$ + $\mathbf{Z}_{BA}$)/2 are 0.224 and 0.204 meters respectively. Using Equation (2) CORR(δ_{AB},δ_{BA}) = 0.659 then using Equation (1)

$$VAR(\delta) = 1.47 \cdot VAR(\mathbf{Z}_{AB} - \mathbf{Z}_{BA})$$

4. The Characteristics of Automatic False Match Detection

Considering the bimodal distribution of the self-consistency difference, it is reasonable to assume that large values of ($\mathbf{Z}_{AB}$ - $\mathbf{Z}_{BA}$) are associated with false matches. Thus, false-matches can be detected by applying a threshold to the self-consistency difference. One method for setting the threshold is to fit a bimodal model consisting of a Gaussian plus a uniform distribution to the histogram of ($\mathbf{Z}_{AB}$ - $\mathbf{Z}_{BA}$), and set the threshold to a value that approximately separates the two distributions, e.g., 2σ.

The ability of this algorithm to detect false matches was tested using pseudo ground truth. False matches were defined as any posting where the absolute difference between the second generation DEM and the pseudo ground truth DEM was greater than 1 meter (a relatively large threshold). A detected false match was defined as any case where the self-consistency difference was above variable threshold. The two plots in Figure 9 shows the percent of false negatives (cases where a good match was classified as a false match) plotted against the percent of false positives (cases where a false match was classified as a good match), and the false-match detection threshold δZ as a function of the false positive rate. For example, to achieve a false positive rate of 10%, the self-consistency threshold was 0.1 meters. With a self-consistency threshold of 0.25 meters, the false positive and false negative rates would be approximately 20% and 10%.

5. Fusing DEMs from multiple views

DEM fusion is straightforward. Two DEMs are computed from each pair of spatially overlapping images. The two DEMs are averaged together. The average DEMs are cleaned up by removing false matches with the self-consistency difference method. The cleaned and averaged DEMs are then overlaid on a master grid that covers the full observation area. For an image sequence, a clean, average DEM is computed for all spatially overlapping image pairs. Once all images pairs are processed, the fused DEM is found by averaging the elevation entries in the master grid.

6. Example 1: Multispectral images of a highly repetitive textured terrain

The data were a sequence of fifty 4-band (blue, green, red, near infrared) 1920 x 1080 x 10bit multispectral images of a dense forest. Forest images present a difficult example because the image texture patterns are highly redundant—one patch of forest looks like any other. The redundant texture problem is often referred to as the 'green carpet effect.' The image data were collected using a digital camera mounted to a Cessna 172, fitted with an 8-inch belly port. The images frames were collected with 80% along-track overlap. The spatial overlap allowed for as many as four overlapping images pairs for a given section of forest. DEM estimates were generated for each spectral band with a 20%, 40%, 60% and 80% overlap. Figure 10 shows a typical raw image, and Figure 11 shows the fused DEM, the number of reliable samples used to compute each fused DEM posting, and the corresponding ortho-image. The number of samples ranged between 0 where no reliable DEM estimates were found to 66. The greatest number of spatially overlapping samples were found near the center of the test strip where there we more spatially overlapping images. Notice that near the edges of the strip, where numbers of samples were small, errors in the form a seam can be scene.

7. Example 2: Oblique Satellite Views

A digital elevation model of central Boston was constructed from nine Ikonos images, collected from three orbital passes. From each 10K × 10K Ikonos images, a 2048 × 2048 sub-image of central Boston were extracted. One of the 2048 × 2048 sub-images is shown in Figure 12 and a 400 × 400 region from each of the nine sub-images is shown in Figure 13 (each image is labeled with two digits taken from its original imaging label). 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.

We applied an epipolar rectification algorithm to the set of sub-images and created nine 3072 × 3072 epipolar rectified image pairs. A larger output size was chosen to compensate for geometric distortion introduced by epipolar rectification. The image labels for the nine epipolar rectified image pairs are: Orbit 1 – image pairs 43-44, 43-45, 44-45; orbit 2 – image pairs 46-47, 46-48, 47-48; and orbit 3 – image pairs 49-50, 49-51, 50-51. A sample of one epipolar rectified image pairs (orbit 2 image pair 46-48) is shown in Figure 14. Notice that all stereo parallax is horizontal.

For each epipolar rectified image pair two DEMs were generated using the Terrest DEM extraction software, resulted in 18 DEM estimates. A sample DEM (from image pair 46-47) is shown in Figure 15. Notice the apparent lack of detail in the DEM. The eighteen individual DEMs were fused with a self-consistency threshold of 3 meters. The fused DEM also is shown in Figures 15. The fused DEM shows considerably more detail. Less than 5% of the fused DEM had no reliable samples and 90% had at least 14 reliable samples. Figure 16 shows the fused DEM re-projected back into the reference frame of image 47.

By re-projecting the fused DEM to one of the original images we were able to use an additional shadow constraint to label unreliable postings. In this case, if the image pixel had a value less than 127 the underlying DEM posting was labeled as unreliable. Figure 17 shows a composite color image of Boston with the original image encoded in grayscale value, elevation encoded in hue (color), and the standard deviation of the elevation errors encoded in saturation. In the color composite image taller objects are blue and areas below ground level (e.g., construction sites) are red. The brighter the color the small the elevation uncertainty (the absence of color indicates an unknown elevation). The large white object in the center of the image is a bloom caused by sunlight reflecting directly into the Ikonos image sensor.

8. Future Directions

Currently, the fusion algorithm uses an un-weighted average. As a result, the fused DEM does not have the minimum possible elevation errors. In addition, it is difficult to detect and exclude unreliable DEMs from the fusion process. To minimize the elevation error a method is needed for estimating the error covariance matrix. The pseudo ground truth method, described above, can be used to estimate the error covariance matrix. Unfortunately, the process is extremely slow and cannot be implemented in real-time. Recently, Corrada-Emmanuel and Schultz (2008)[1] described a method for estimating the precision error covariance (PEC) matrix without any ground truth. From the PEC matrix, averaging weights can be derived that will minimum error variance of the fused DEM. In the future we plan to implement the PEC matrix methods in the DEM fusion process.

9. Summary

Estimating the distribution of elevation errors is essential to computing a DEM from a sequence of spatially overlapping images. This report outlined methods, based on error estimation, for detecting large, spurious elevation errors, for estimating vertical uncertainty and horizontal resolution, and for fusing multiple DEMs. The methods are computationally efficient and can be implemented in real-time. Two examples were shown that demonstrated the ability of the Terrest system to operate in difficult circumstances.

[1] Corrada-Emmanuel and Schultz, "Autonomous geometric precision error estimation in low-level computer vision tasks," Proceeding of the 25th International Conference on Machine Learning (ICML), Helsinki, Finland, pages: 168-175, 2008

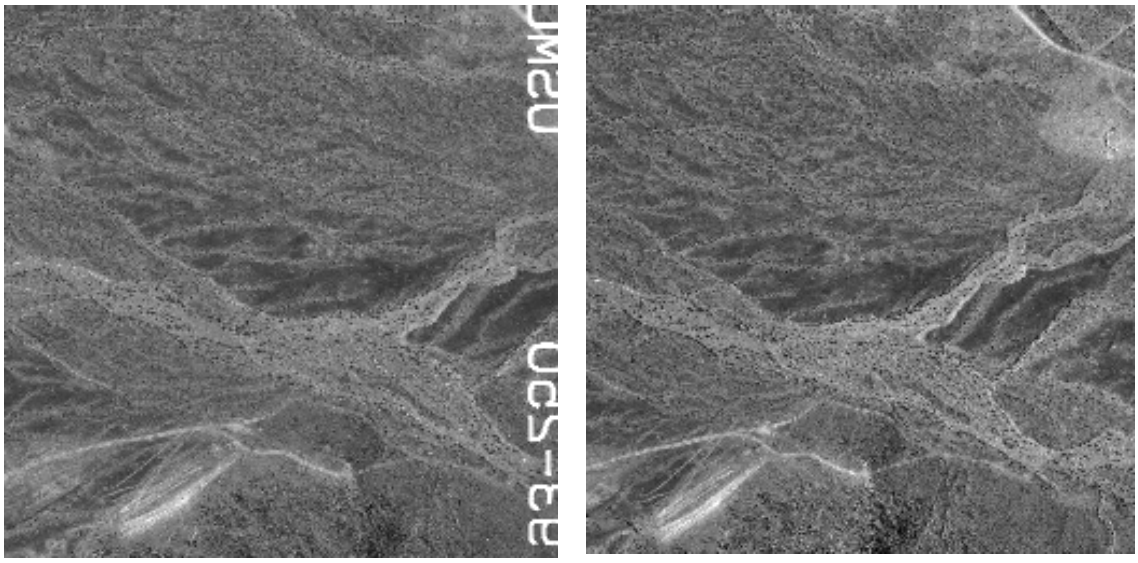

Figure 1. Two 2000 × 2000 spatially overlapping sub-images.

Figure 2. The pseudo ground truth DEM.

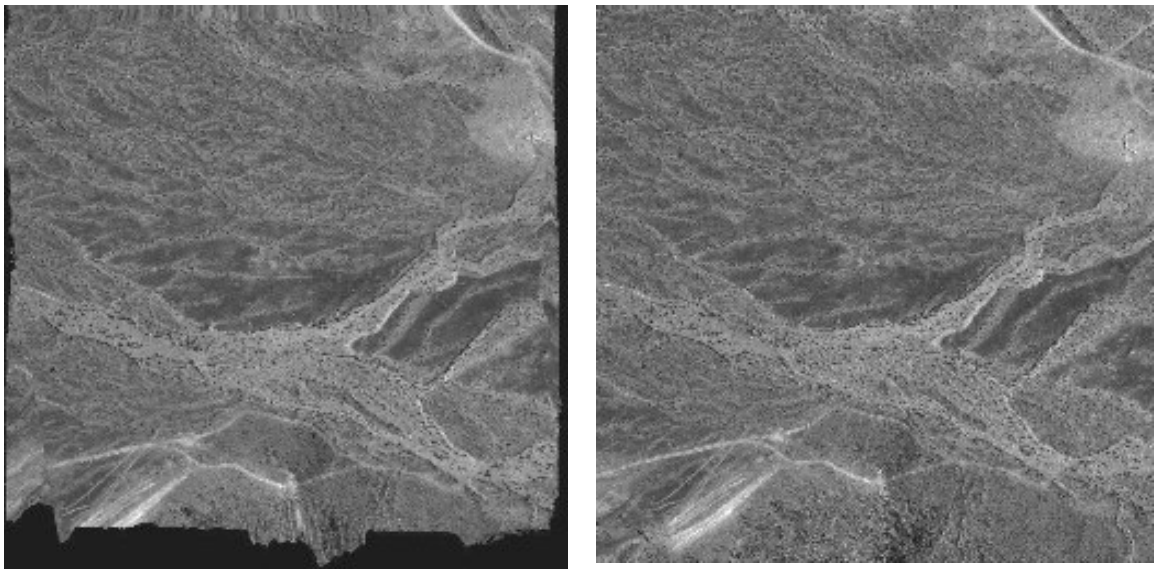

Figure 3. Synthetic image (left) and original image (right).

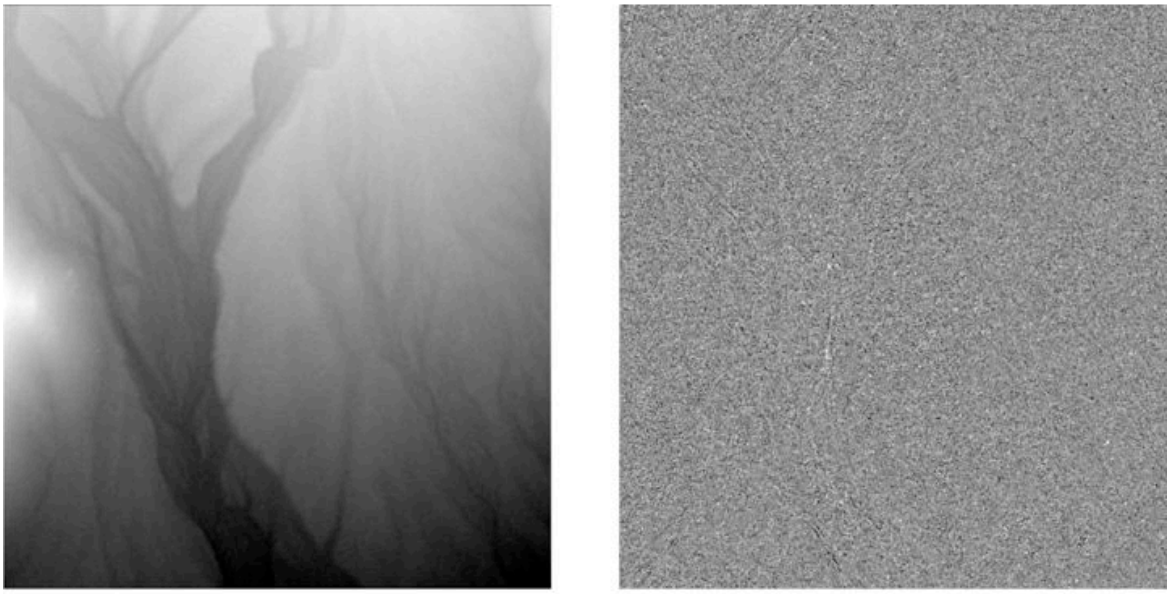

Figure 4. (Left) The second generation DEM created from the synthetic images, (Right) The second generation errors.

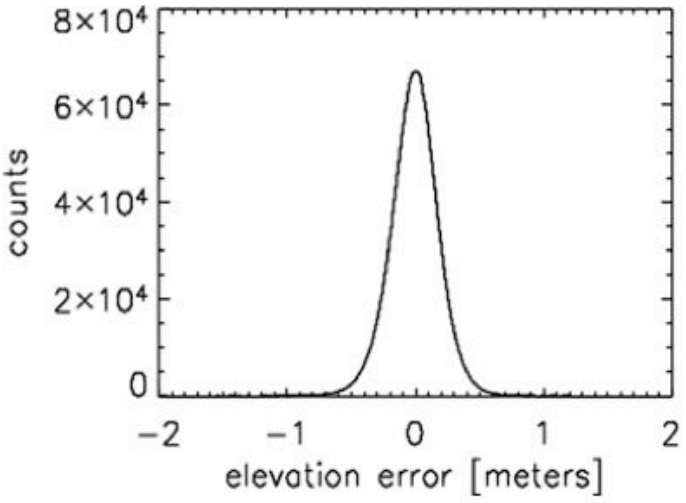

Figure 5. A histogram of the second generation errors.

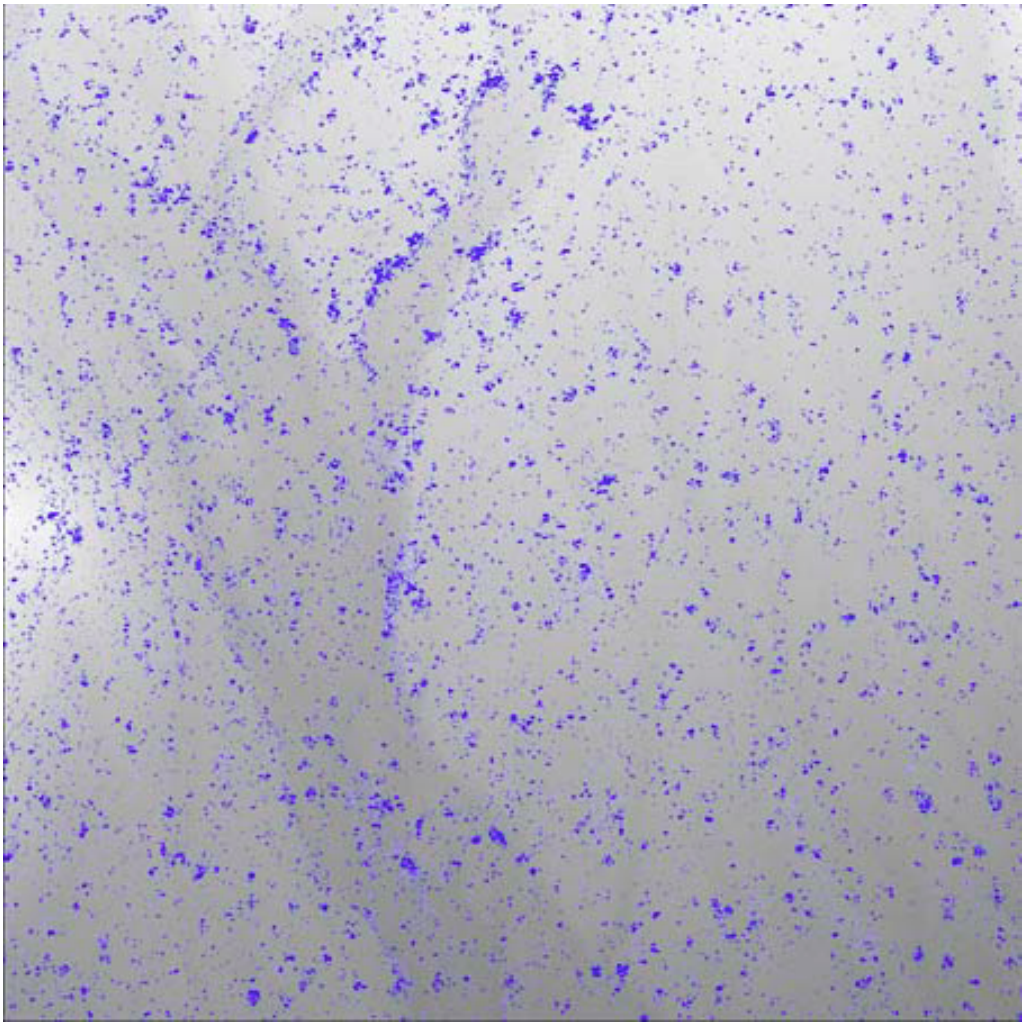

Figure 6. An image of the detected false matches marked in blue overlaid on the second generation DEM. The false match threshold was +/- 1 meter. Excluding the detected false matches, the standard deviation of the elevation errors was about 20 cm.

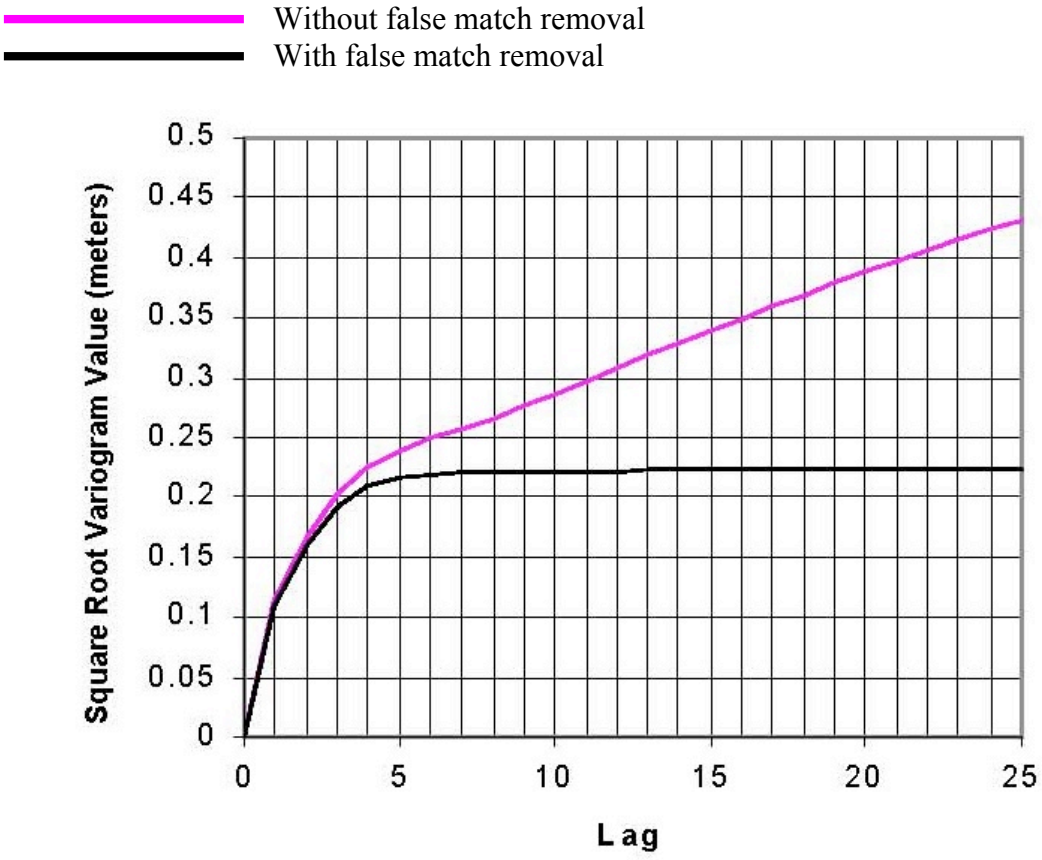

Figure 7. The square root variogram with and without false match removal.

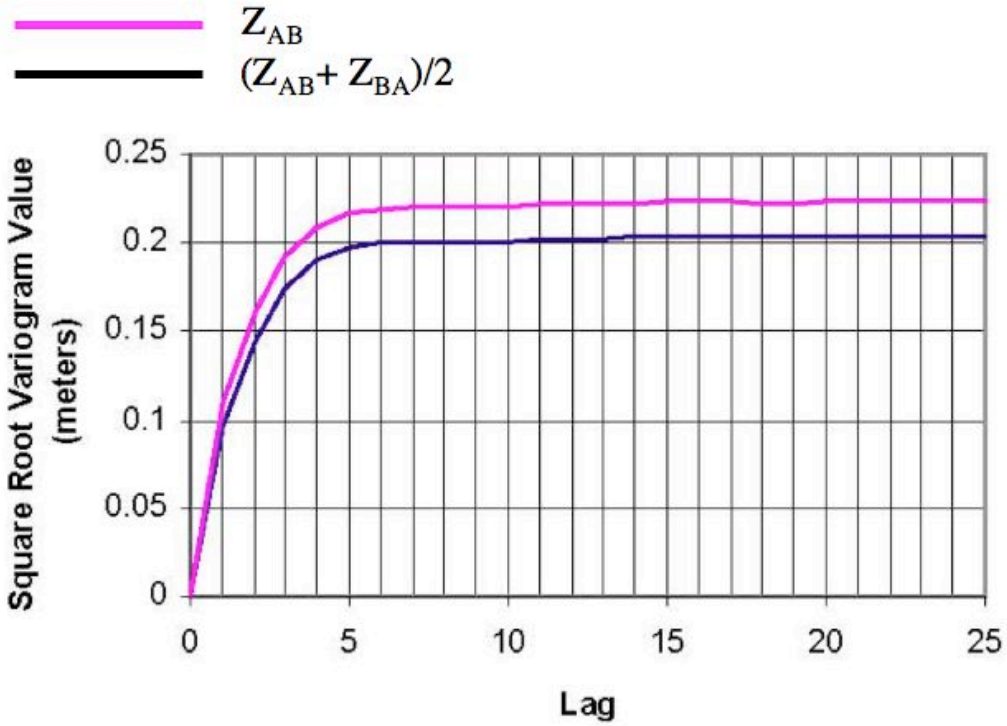

Figure 8. Comparison between either of the second generation DEMs (Z_{AB} or Z_{BA}) and average of the second generation DEMs (Z_{AB}+Z_{BA})/2

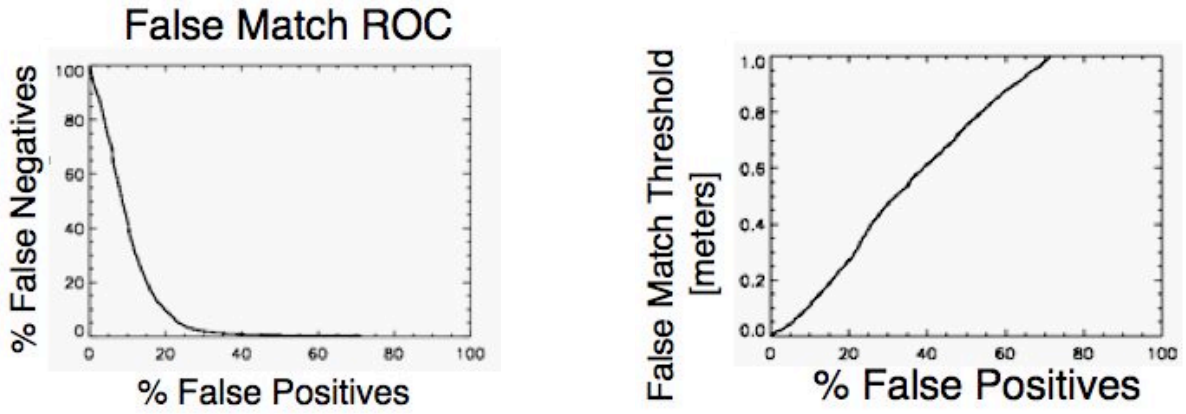

Figure 9. (Left) The ROC curve showing the percent of false negatives (false matches labeled as good matches) as a function of false positives (good matches labeled as false matches), (Right) A graph showing the relationship between the false match detection threshold and the percent of false positives.

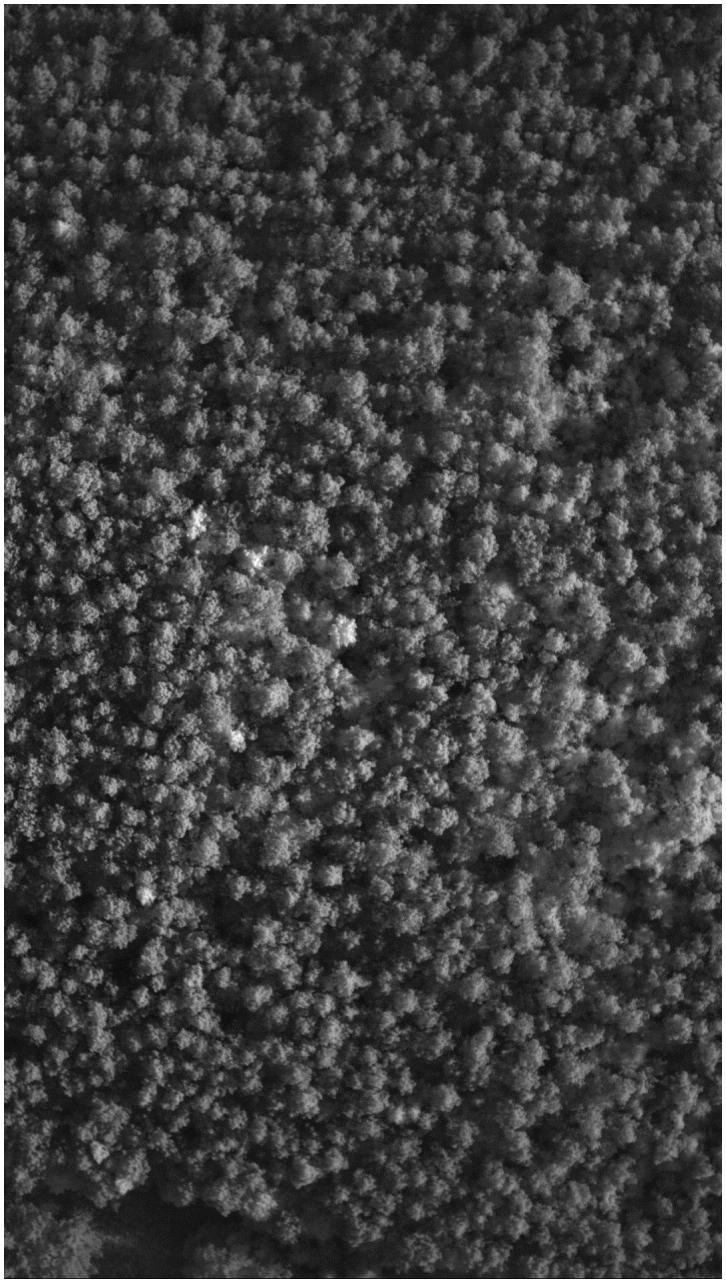

Figure 10. A typical 1920 x 1080 x 10bit image from the test strip.

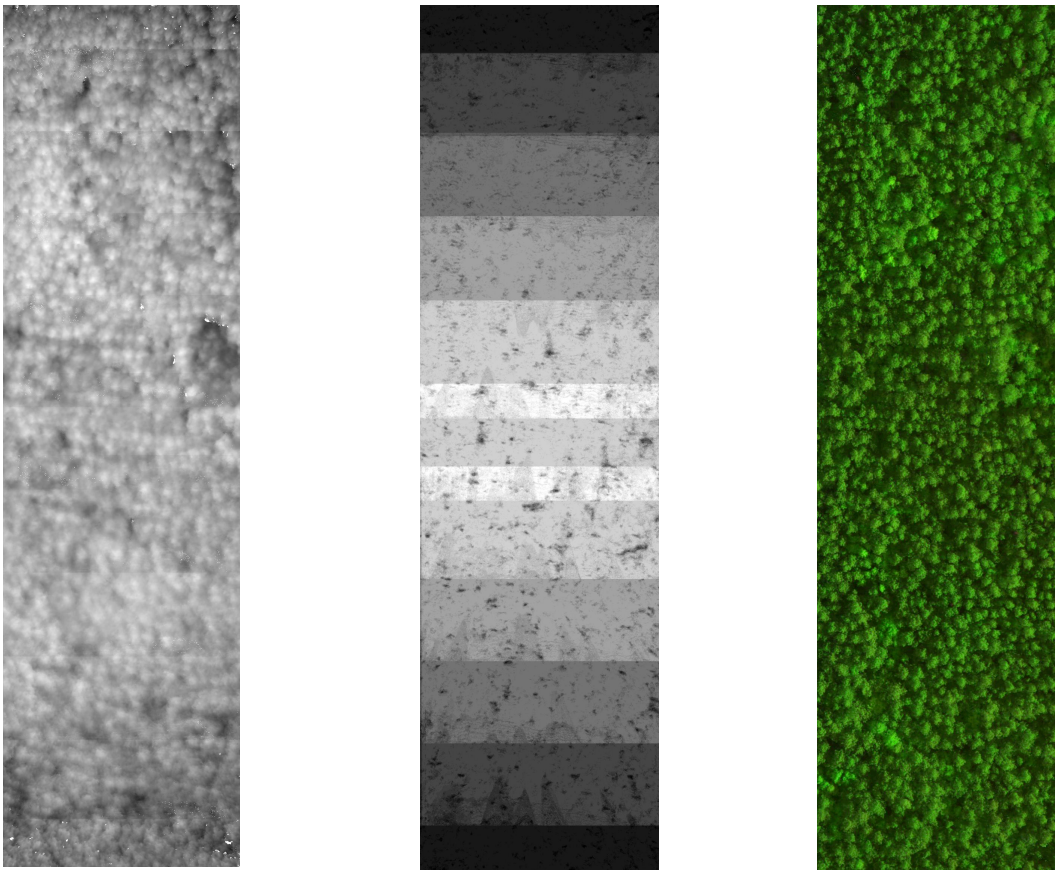

Figure 11. (Left) The fused DEM, elevations are indicated by grey level with dark shades indicating lower elevations and lighter shades indicating higher elevations. (Middle) The image shows the number of reliable elevation estimates used to compute the fused DEM postings, the values range between 0 and 66. (Right) The ortho-image created by projecting the raw images on to the fused DEM.

Figure 12. One of nine Ikonos image of Boston (po_3947_pan_0000000)

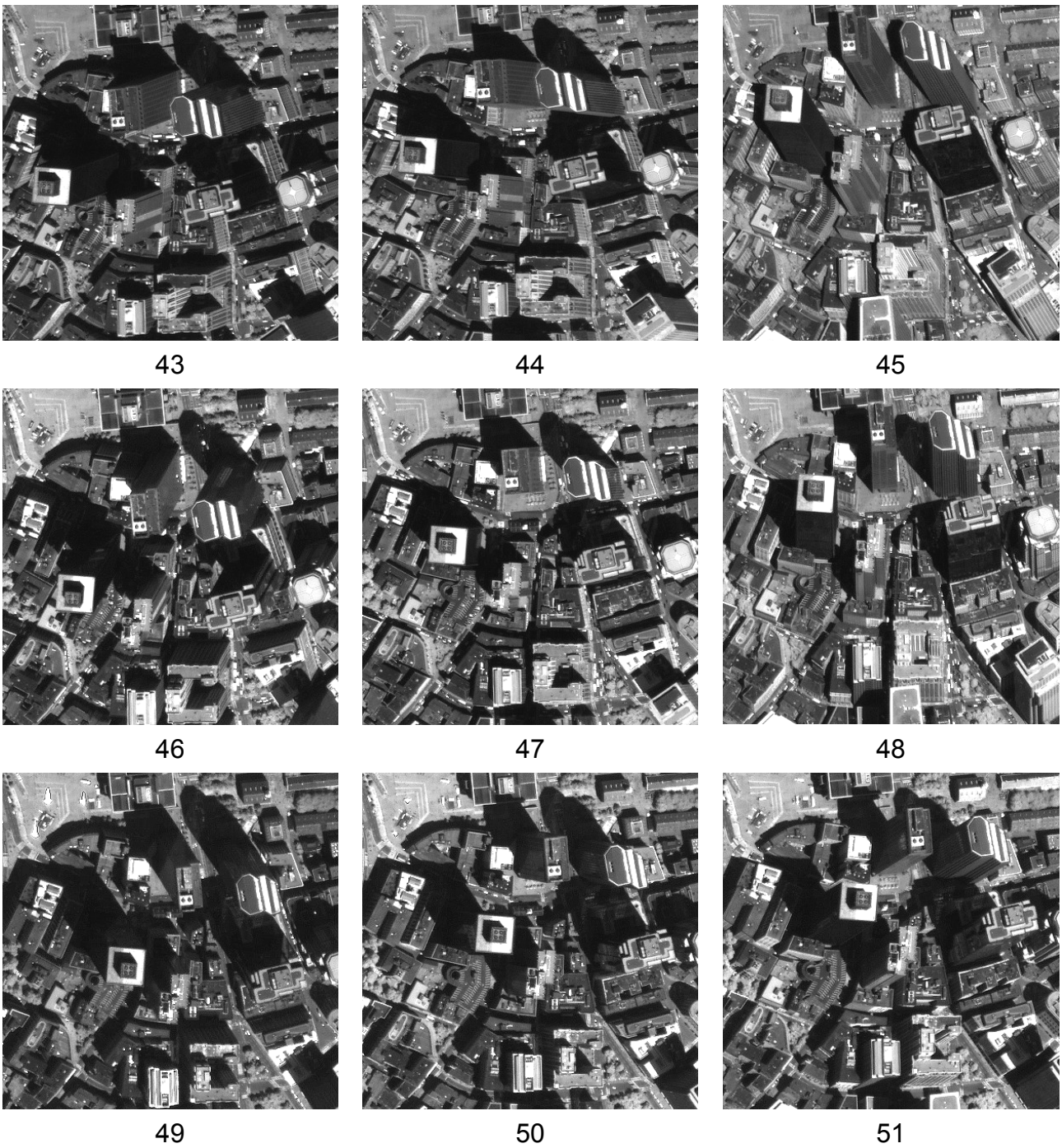

Figure 13. Nine 400 × 400 pixel regions taken from each Ikonos images. The image labels corresponding to the last two digits of the Ikonos image ID.

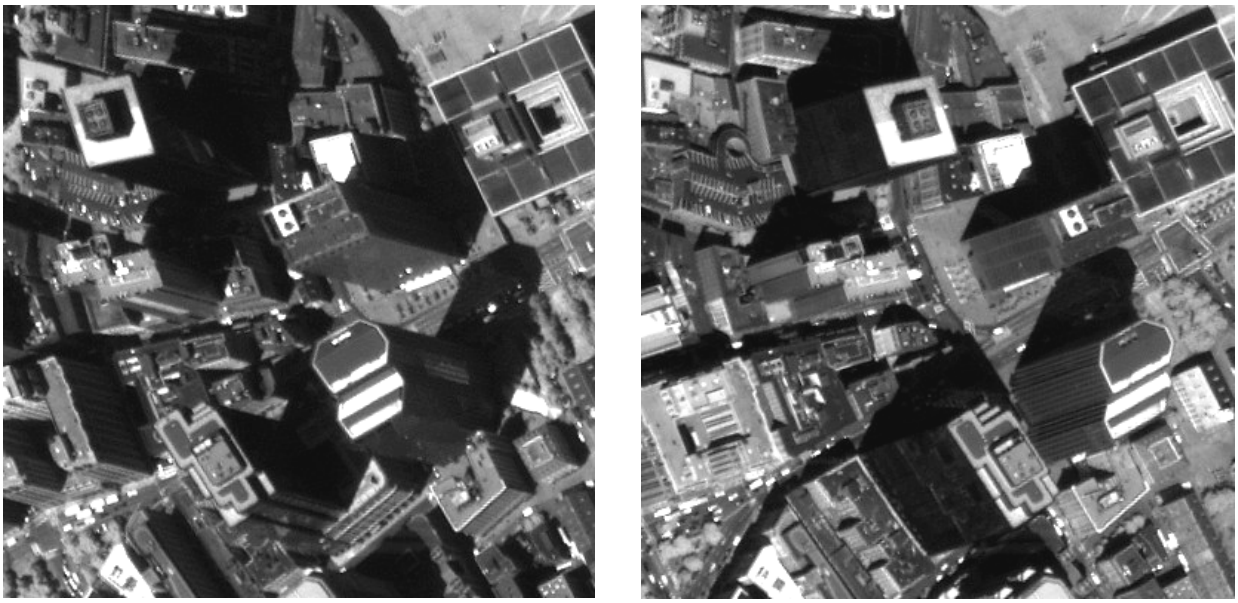

Figure 14. A 400 × 400 pixel region taken from the epipolar rectified image pair 46-48

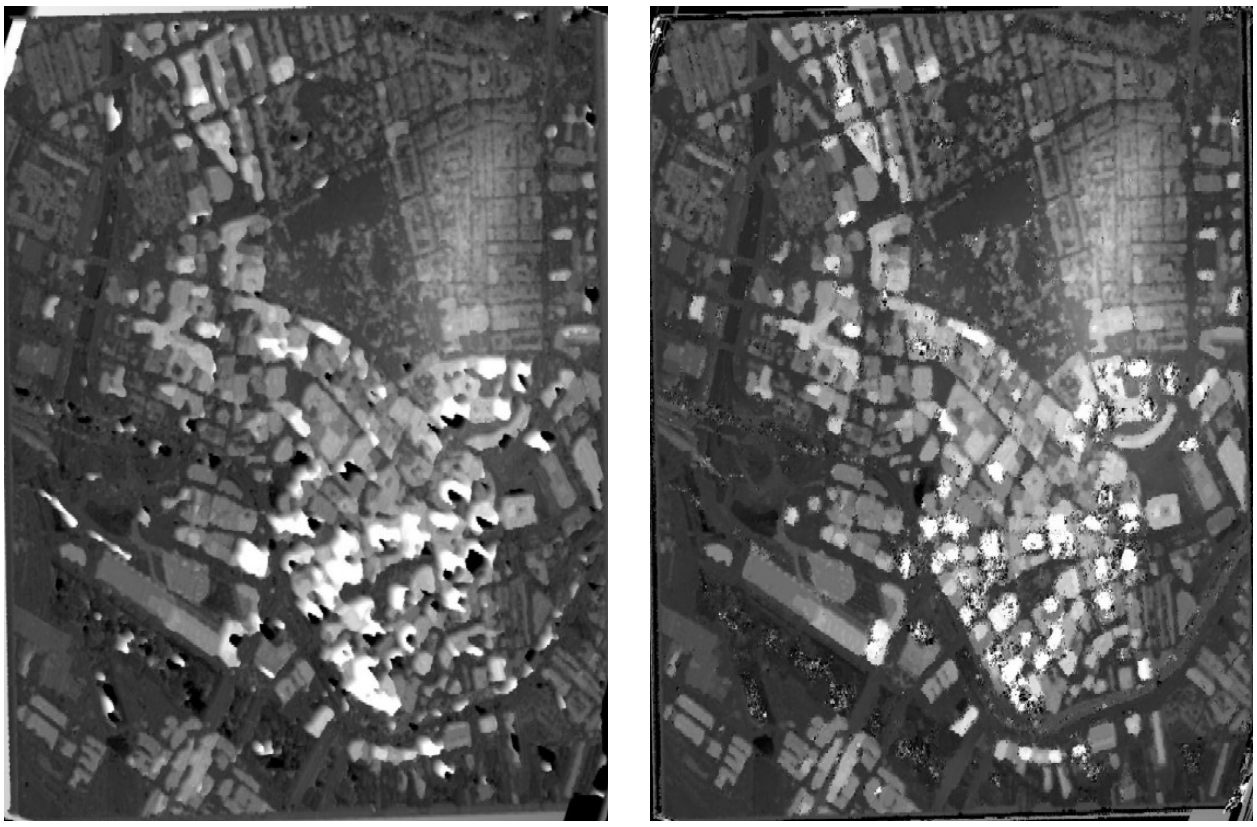

Figure 15. (Left) The DEM make from image pair 46-48, and (Right) The fused DEM create from nine Ikonos images.

Figure 16. The fused DEM projected into the Boston image shown in Figure 12

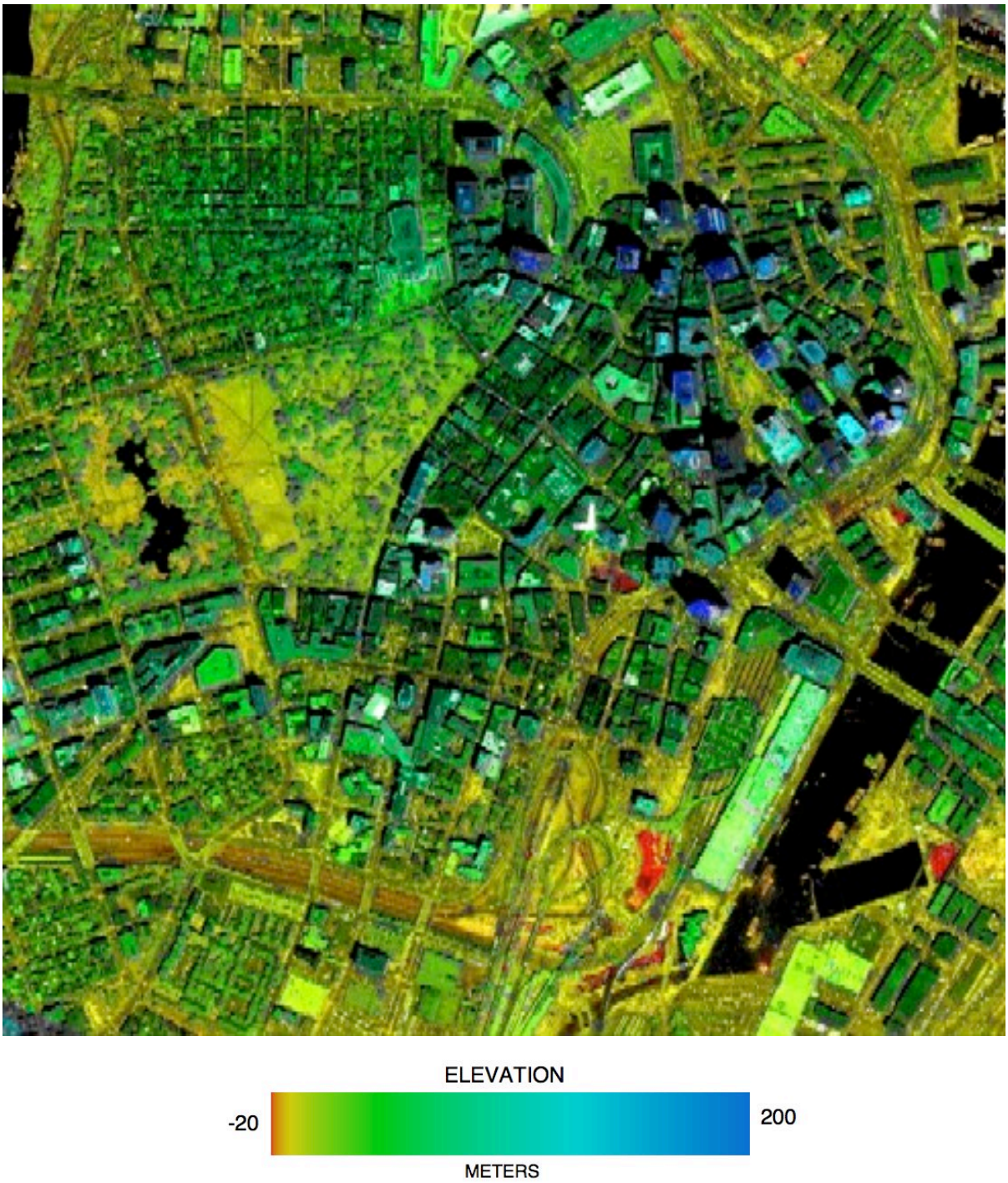

Figure 17. A composite color image of Boston with the original image (Figure 12) encodes in value, elevations (Figure 16) encoded in hue, and the standard deviation of the elevation errors encoded in saturation. Gray pixels indicate no reliable elevation estimates. The red areas are construction sites that extend below ground level.