# A Self-consistency Technique for Fusing 3D Information

**Howard Schultz, Allen R. Hanson,**
**Edward M. Riseman, Frank R. Stolle, Zhigang Zhu**
Computer Science Department
University of Massachusetts
Amherst, MA  01003, USA
{hschultz,hanson,riseman,stolle,zhu}@cs.umass.edu

**Woo Dong-Min**
School of Electrical Engineering
Myongji University
South Korea
dmwoo@wh.myongji.ac.kr

**Abstract -** *This paper describes a robust method for recovering an optimal DEM and its variance from multiple, randomly oriented views of a surface. The method generates a set of DEM tiles in a common coordinate system from multiple overlapping images, and then employs the concept of self-consistency to detect and remove errors from the tiles. The clean tiles are averaged together to form a low noise composite DEM. The method is tested on real and photo realistic simulated data. Results show that the method is capable of producing a virtually error free composite DEM.*

**Keywords:** Stereo, multi-view stereo, photogrammetry, DEM extraction, resource management.

## 1    Introduction

We have all noticed that viewing a scene from a moving car enhances our perception of the 3D structure of the surrounding environment.  From a photogrammetric point of view, the induced motion parallax allows our vision system to form a robust 3D model of our surroundings.  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 will use the principle of self-consistency developed by Leclerc, Luong and Fua [10] 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.  In a closely related application [4][9], they extend their work to detect changes in terrain by applying the concept of self-consistency to elevations.  We extend the idea of this work to detect unreliable elements in a DEM generated from stereoscopic image pairs.

For the work described in this paper, DEM tiles are generated in a three-step process [6][7] using the UMass Terrest system [12].  First, pairs of overlapping images are resampled such that scan lines are coincident with epipolar lines.  This simplifies the problem by ensuring that corresponding pixels lie on the same scan line [13].  Next, the image matching procedure is invoked to produce a disparity map $\mathbf{D}_{AB}$, which defines a one-to-one 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.  In the final step, the disparity map and the orientation information are combined to form a DEM $\mathbf{Z}(x,y)$, which is a dense array of elevation estimates in a world coordinate system.

The Terrest image matching procedure employs a nonlinear, hierarchical algorithm to generate disparity maps [8][12][14].  A useful characteristic of the algorithm is that when the reference and target images are reversed,

the resulting errors are partially uncorrelated. In other words, the disparity map $\mathbf{D}_{AB}$ is not completely determined by $\mathbf{D}_{BA}$. We will show that this property can be exploited to successfully identify and remove blunders in a DEM.

# 2 Self-consistency

At the heart the method is an expectation that reversing the reference and target images will lead to similar (or self-consistent) results when the image matching algorithm finds correct correspondence, and dissimilar results when the image matching algorithm fails. The measure of similarity/consistency will be the signed difference between the two disparity maps or two DEMs (derived from the disparity maps).

Two self-consistency methods will be explored: (1) object space self-consistency in which the analysis takes place after the disparity maps and orientation information have been converted to an array of elevations in object space, and (2) image space self-consistency in which comparisons are made between disparity maps. The majority of the work presented here will focus on object space self-consistency applications.

## 2.1 Object Space Self-Consistency

We begin by denote the DEM recovered from the disparity maps $\mathbf{D}_{AB}(i,j)$ and $\mathbf{D}_{BA}(i,j)$ by $\mathbf{Z}_{AB}(x,y)$ and $\mathbf{Z}_{BA}(x,y)$, where $(x,y)$ are the coordinates of the DEM nodes in a common world coordinate system. Without loss of generality, $\mathbf{Z}_{AB}$ may be written as the sum of the actual surface shape $\hat{\mathbf{Z}}$ and an error term $\delta_{AB}$. Thus, for any overlapping image pair $(A,B)$ the two recovered surface models are

$$\mathbf{Z}_{AB} = \hat{\mathbf{Z}} + \delta_{AB}$$

$$\mathbf{Z}_{BA} = \hat{\mathbf{Z}} + \delta_{BA}$$

Taking the difference of the two DEMs gives an expression that is independent of the surface shape $\hat{\mathbf{Z}}$

$$\mathbf{Z}_{AB} - \mathbf{Z}_{BA} = \delta_{AB} - \delta_{BA} \qquad (1)$$

The left-hand side of Equation 1, which is very similar to self-consistency described in [10], depends only on images $A$ and $B$. The right hand side is the difference between the geospatial errors associated with the DEM generation process. Taking the standard deviation of both sides of Equation 1,

$$\sigma(\mathbf{Z}_{AB} - \mathbf{Z}_{BA}) = \sigma(\delta_{AB} - \delta_{BA}) \qquad (2)$$

gives a relationship between the distribution of the observed object-space self-consistency difference and the distribution of the elevation errors. Assuming the $\delta_{AB}$ and $\delta_{BA}$ have identical distributions, Equation (2) can be rewritten as

$$\sigma(\mathbf{Z}_{AB} - \mathbf{Z}_{BA}) = C\sigma(\delta), \qquad (3)$$

where the constant C is a measure of the degree of correlation between the two random variables $\delta_{AB}$ and $\delta_{BA}$. Equation 3 represents a simple relationship between the standard deviation of the elevation errors $\sigma(\delta)$ and the self-consistency distribution $\sigma(\mathbf{Z}_{AB} - \mathbf{Z}_{BA})$, which can be inferred directly from the computed elevation data.

Implicit in the formulation of Equation 3 is the assumption that the elevation estimates $\mathbf{Z}_{AB}$ and $\mathbf{Z}_{BA}$ are samples of a normally distributed random variable. In reality, however, $\mathbf{Z}_{AB}$ and $\mathbf{Z}_{BA}$ are samples of two distinct populations. The first population is attributed to cases where the image matching algorithm returns two sets of corresponding pixels that project to a point in space near the same surface element, i.e., $\mathbf{Z}_{AB}(x,y) \approx \mathbf{Z}_{BA}(x,y)$, which implies that $\mathbf{Z}_{AB}(x,y) - \mathbf{Z}_{BA}(x,y)$ is sampled from a normally distributed, zero mean process. In the second case, the matching algorithm fails and $\mathbf{Z}_{AB}(x,y) - \mathbf{Z}_{BA}(x,y)$ is sampled from a much broader distribution. We use this natural separation of scale to detect false correspondences.

Assuming that the matching algorithm returns more true correspondences than false ones, the elevation errors may be modeled by a mixture of a large number of samples from a normally distributed, zero mean random variable and a smaller number of samples from a uniformly distributed population of outliers. This model predicts that the outliers will have a disproportionate influence on the tails of the distribution. To minimize the influence of the tails when computing the moments of the distribution, the histogram of data $(\mathbf{Z}_{AB} - \mathbf{Z}_{BA})$ is fit to a model that included a Gaussian distribution plus a constant,

$$h_{max} \exp-\left[\frac{(h_i - z_0)^2}{2s^2}\right] + h_{min}, \qquad (4)$$

where $h_i$ are the histogram entries from $(\mathbf{Z}_{AB} - \mathbf{Z}_{BA})$, $(z_0, s)$ are the mean and standard deviation of the distribution, $h_{min}$ 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 4 to the histogram of $(\mathbf{Z}_{AB} - \mathbf{Z}_{BA})$.

2

The value of the false correspondence threshold is set to a multiple of the width of the central peak of the distribution. Thus, a pair of elevation estimates $\mathbf{Z}_{AB}(x,y)$ and $\mathbf{Z}_{BA}(x,y)$ are considered reliable if the absolute value of the self-consistency difference falls below the threshold, e.g., if $|\mathbf{Z}_{AB}(x,y) - \mathbf{Z}_{BA}(x,y)| < n \cdot s$, (where $n$ is a user defined parameter), otherwise both are considered unreliable.

If there are $N$ overlapping images of a surface region, each from a unique viewpoint, then there are a total of $N(N-1)$ ordered image pairs, e.g., three views $(A,B,C)$ would result in 6 pairs $(A,B)$, $(B,A)$, $(A,C)$, $(C,A)$, $(B,C)$, $(C,B)$. Thus, there are a maximum of $N(N-1)$ samples at every node in the DEM. Once all of the elevation estimates are labeled reliable or unreliable, the optimal DEM and its variance are simply set to the mean $\overline{Z}(x,y)$ and variance of the reliable elevations.

# 3 Results

The self-consistency fusion algorithm was tested on two data sets: (1) A set of four overlapping high altitude images taken over a desert environment, and (2) a set of photo-realistic simulated images generated from a know DEM and ortho-image.

## 3.1 Real images

The object space self-consistency DEM fusion method was tested on a set of four 9 inch × 9 inch aerial images of a desert terrain near 29 Palms, California. The image scale was 1:10,000 and the four-way overlap covered a region of approximately 2k × 2k pixels. One of the overlap regions is shown in Figure 1.



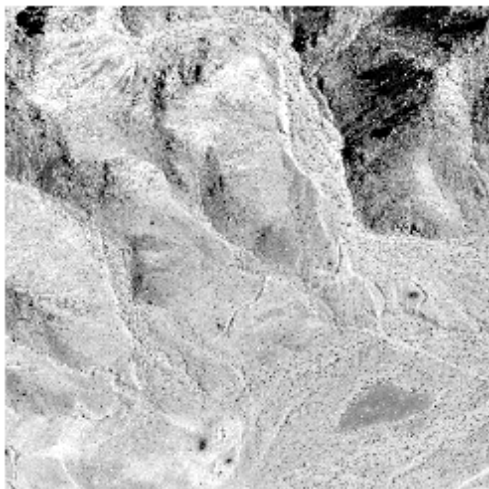**Figure 1.** One of the four 2k × 2k overlap regions.

From the four images $(A,B,C,D)$ a total of twelve DEMs $(\mathbf{Z}_{AB}, \mathbf{Z}_{BA}, \ldots, \mathbf{Z}_{CD}, \mathbf{Z}_{DC})$ and six self-consistency difference arrays $(\mathbf{Z}_{AB} - \mathbf{Z}_{BA}, \ldots, \mathbf{Z}_{CD} - \mathbf{Z}_{DC})$ were generated. The histogram made from all twelve arrays is shown in Figure 2. A reliability threshold of $2\sigma$ was selected, which corresponded to $\pm 0.4$m.

Figure 3 shows the fused DEM, a map of the number of reliable elevation estimates, and a rendered view of the ortho-image draped over the fused DEM. 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.
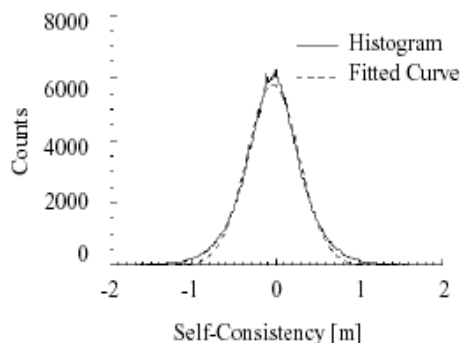


**Figure 2.** The histogram of the self-consistency difference arrays and the fitted model.

In addition, there were no apparent blunders or spikes in the fused DEM, and the rendered surface appeared realistic. The examination procedure consisted of checking the relationship between objects such as rocks and ditches to their shadows, and checking to see if streams (or in this case dry stream beds) flow downhill.

## 3.2 Photo realistic simulation analysis

The methods described in this paper produce a dense array of elevation estimates, which typically involves millions of samples. Consequently, a comprehensive evaluation of the self-consistency method requires an equally dense array of ground truth. A few manually gathered ground control points will not provide a sufficient number of samples to test the validity of the method. When available, LIDAR may provide large, dense arrays of ground truth. However, LIDAR data have several disadvantages, including (1) vertical errors that are similar to the self-consistency errors; (2) it is difficult to register LIDAR and image data; (3) LIDAR data are expensive to collect.

To provide a detailed analysis of the self-consistency fusion technique, we created a pseudo ground truth data set. The process begins with an existing DEM and ortho-images. A photo realistic ray-tracing program is
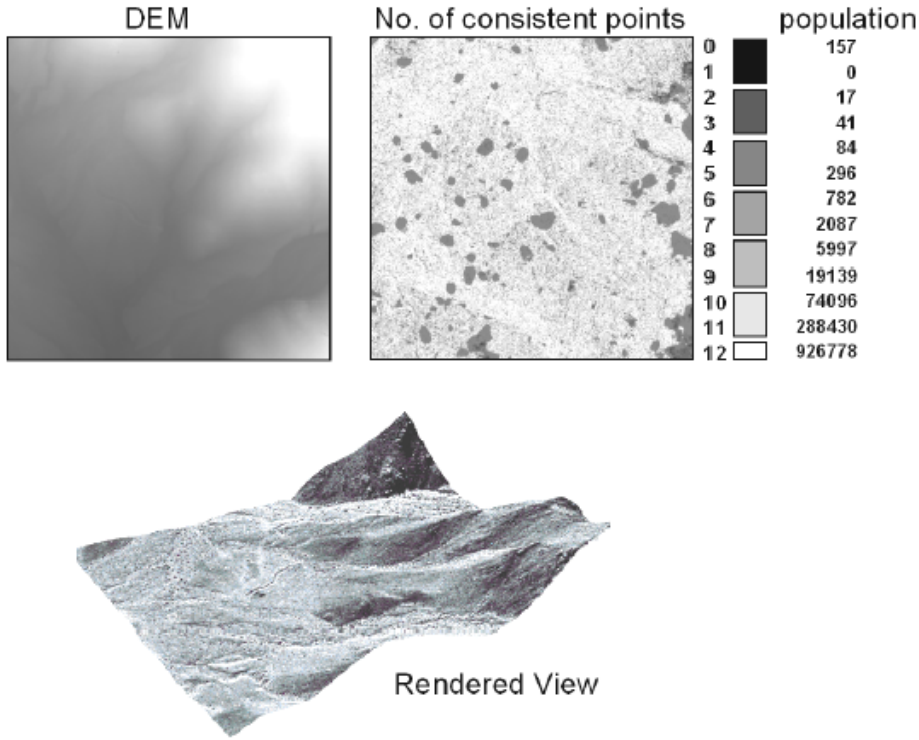
**Figure 3.** (Top Left) The fused DEM generated from four overlapping views of the terrain shown in Figure 1. The ground sampling distance is approximately 35cm, the average elevation variation is 17cm, and the elevation range varies from a low 762.7m to a maximum of 885.7m. (Top Right) A map showing the number of reliable elevation over the DEM. (Bottom) A rendered view of the ortho-image draped over the fused DEM.

used to synthesize images of the surface from arbitrary viewpoints. Next, the synthetic images are used to regenerate the DEM. The regenerated DEM can then be compared on a point-by-point basis to the pseudo ground truth. Clearly, pseudo ground truth and synthetic images are not a completely satisfactory substitute for real data. Nevertheless, the method does provide the means for generating otherwise unobtainable ground truth samples.

For this study, a small region in the center of the fused DEM covering an area of 157.5m × 368.4m was selected for the pseudo ground truth. Photo realistic synthetic images were then generated at incidence angles that ranged from -30° to +30° at 15° intervals. An example of an image chip from one of the real images and a synthetic image chip of the same terrain is shown in Figure 4. From these images, five stereo pairs with base-to-height ratios (*b/h*) that ranged from 0.277 to 1.23 were analyzed. The results are summarized in Table 1. The results were generally consistent with previous analyses of the Terrest algorithm [12]. The standard deviation of the self-consistency difference decreased with increasing *b/h*.

**Table 1.** Pseudo ground truth results for five synthetic image pairs with varying base-to-height (*b/h*) ratios and incidence angels ($\theta_A$, $\theta_B$). The dependent variable are the standard deviation of the self-consistency difference $\sigma(\mathbf{Z}_{AB}-\mathbf{Z}_{BA})$, the percent of reliable elevation estimates (using a $2\sigma$ threshold), and the standard deviation of the pseudo ground truth error $\sigma(\hat{\mathbf{Z}} - \overline{\mathbf{Z}})$, where $\hat{\mathbf{Z}}$ is the pseudo ground truth DEM and $\overline{\mathbf{Z}}$ are the average reliable elevations. All angles are in degrees and all standard deviations are in meters.

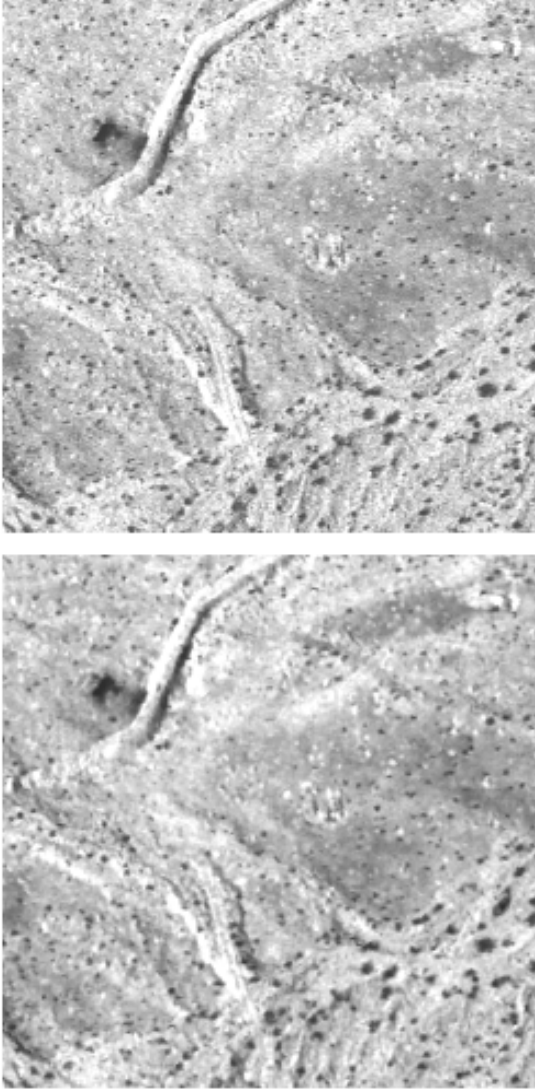| b/h | $\theta_A$ | $\theta_B$ | $\sigma(\mathbf{Z}_{AB}-\mathbf{Z}_{BA})$ | % Reliable ($2\sigma$ cutoff) | $\sigma(\hat{\mathbf{Z}} - \overline{\mathbf{Z}})$ |
|---|---|---|---|---|---|
| 0.277 | 0° | 15° | 0.451 | 91.90 | 0.214 |
| 0.293 | 15° | 30° | 0.486 | 92.50 | 0.266 |
| 0.575 | 15° | -15° | 0.312 | 91.36 | 0.131 |
| 0.868 | -15° | 30° | 0.204 | 89.40 | 0.152 |
| 1.230 | 30° | -30° | 0.168 | 84.24 | 0.156 |

4

**Figure 4.** A $400 \times 400$ pixel image chip (top) and a synthetic view of the same region (bottom).

In addition, the percentage of reliable elevation estimates also decreased with increasing $b/h$. This later result probably was caused by an increase in occluded surfaces resulting from steeper incidence angle. The slight increase in pseudo ground truth error $\sigma(\hat{\mathbf{Z}} - \overline{\mathbf{Z}})$ was likely cause by contamination from unreliable elevation estimates, which could be corrected by choosing a lower threshold.

## 4 Image space self-consistency

As discussed in Section 2, the self-consistency principle can be applied in object space and image space [2][11]. The primary advantage of an image space representation is that the disparity map naturally aligns with the reference image. In contrast, in object space the ortho-image is registered to the DEM, and elevation errors will cause distortions in the ortho-image. Consequently, an image space representation is especially useful when trying to establish the 3D location of straight lines, such as building edges.

Application of the self-consistency principle in image space requires resampling $\mathbf{D}_{AB}$ into the reference frame of image $B$, or $\mathbf{D}_{BA}$ into the reference frame of image $A$. Since the disparity map $\mathbf{D}_{BA}$ is defined such that the pixel pair $(i+\mathbf{D}_{BA}(i,j),j)$ in image $A$ and $(i,j)$ in image $B$ are projections of the same surface element, the disparity in image $B$ for the pixel $(i + \mathbf{D}_{BA}(i,j), j)$ in image $A$ is $-\mathbf{D}_{BA}(i,j)$. This relationship implies a simple algorithm for changing the reference image from $B$ to $A$. For every line in image $A$ ( $j = 0,1,2,\dots$ ) generate an array of irregularly spaced independent values $\mathbf{X} = i + \mathbf{D}_{BA}(i,j)$, $i = 0,1,2,\dots$ and a corresponding array of dependent values $\mathbf{Y} = -\mathbf{D}_{BA}(i,j)$, $i = 0,1,2,\dots$. Then, interpolate $\mathbf{Y}$ at equal steps along $\mathbf{X}$. The result is a line in the resampled disparity map $\hat{\mathbf{D}}_{AB}$, which contains the same information as $\mathbf{D}_{BA}$, but in the reference frame of image $A$. Self-consistency analysis can now be applied to $\mathbf{D}_{AB}$ and $\hat{\mathbf{D}}_{AB}$.

An example of image space self-consistency analysis is shown in Figures 5-7. Figure 5 shows one of two $1000 \times 1000$ pixel sub-images (labeled 1 and 2) extracted from a pair of much larger Ikonos images of Hickam Air Force Base in Hawaii. The disparity maps $\mathbf{D}_{12}$ and $\mathbf{D}_{21}$ along with $\hat{\mathbf{D}}_{12}$ are shown in Figure 6. Finally, Figure 7 shows the average reliable disparity map $(\mathbf{D}_{12} + \hat{\mathbf{D}}_{12})/2$ for all cases where $|\mathbf{D}_{12} - \hat{\mathbf{D}}_{12}| < 1.5$ pixels, and the correlation match scores were above 0.98. An object detection analysis showed that all but one building and a smoke stack were found.



**Figure 5.** A $1000 \times 1000$ sub-image extracted from a larger Ikonos image of Hickam Air Force Base.
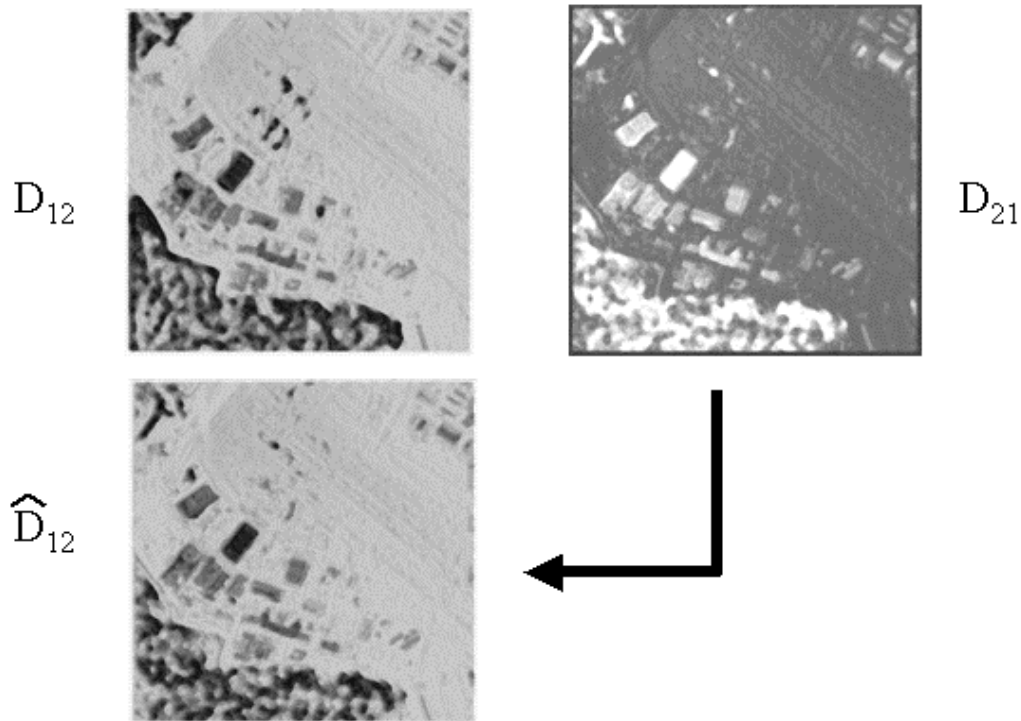
5

**Figure 6.** $\mathbf{D}_{12}$ is the disparity maps generated with image 1 as the reference and image 2 as the target; $\mathbf{D}_{21}$ is the disparity map generated with image 2 as the reference and image 1 as the target; and $\hat{\mathbf{D}}_{12}$ is $\mathbf{D}_{21}$ resampled into the reference frame of image 1.

## 5 Conclusions

The goals of self-consistency processing are to (1) detect and remove errors in disparity maps or DEM tiles that result from failures during the image matching process, and (2) to combine the individual tiles to form a large area, low noise model. The self-consistency principle is based on the expectation that reversing the reference and target images will lead to similar (or self-consistent) results when the image matching algorithm finds correct correspondence, and dissimilar results when the image matching algorithm fails. The measure of similarity/consistency is the signed difference between two disparity maps or DEMs (derived from the disparity maps) generated by reversing the reference and target images. A reliability threshold, based on the width of the central peak of the distribution of the self-consistence difference measures ($\mathbf{D}_{AB}$-$\mathbf{D}_{BA}$) or ($\mathbf{Z}_{AB}$-$\mathbf{Z}_{BA}$) for all image pairs (*A, B*), is used to separate reliable and unreliable disparity or elevation estimates. An element in a disparity map or DEM is considered reliable if and only if the corresponding self-consistency measure falls below the reliability threshold.
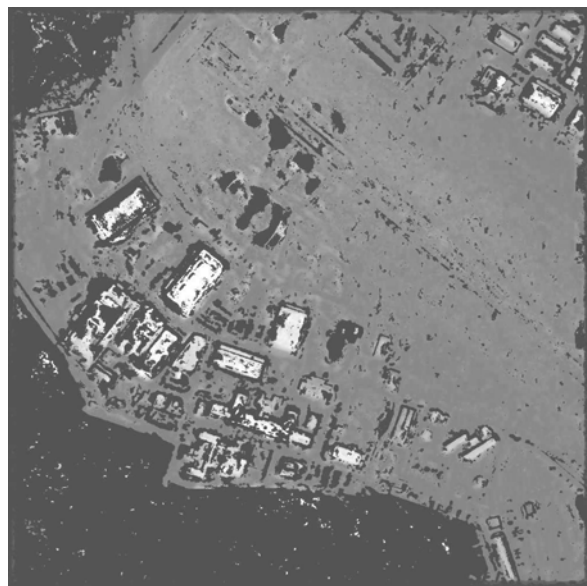


**Figure 7.** The Average reliable disparity map. Unreliable disparity values masked out with a black background color.

The concept of self-consistency was successfully applied to the task of forming a composite DEM from a collection of smaller DEM tiles for real and synthetic data. Based on the synthetic data, the technique was able to produce a virtually error free composite DEM. In addition, visual analysis of real high-altitude aerial and satellite data showed that the system is capable of consistently detecting and removing the effects of image matching failures.

## Acknowledgements

## References

[1] Agouris, Peggy, T. Schenk, Automated Aerotriangulation Using Multiple Image Multipoint Matching, Photogrammetric Engineering and Remote Sensing, Vol. LXII, No. 6, June 1996, pp. 703-710.

[2] Ayache, N., and B. Faverjon, "Efficient Registration of Stereo Images by Matching Graph Description of Edge Segments," Int'l J. Computer Vision, pp. 107-131, 1987.

[3] Aschwanden, P. and W. Guggenbuehl, "Experimental Results From a Comparative Study on Correlation-Type Registration Algorithms," Robust Computer Vision, Foerstner and Ruwiedel, eds., pp. 268-289, Wichmann, 1993.

[4] Fua, P. and Y.G. Leclerc, "Taking Advantage of Image-Based and Geometry-Based Constraints to Recover 3-D Surfaces," Computer Vision and Image Understanding, vol. 64, no. 1, pp. 111-127, 1996.

[5] Grimson, W.E.L., "Computational Experiments With a Feature Based Stereo Algorithm," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 7, no. 1, pp. 17-34, January, 1985.

[6] Hoff, W., and N. Ahuja, "Surface From Stereo: Integrating Feature Matching, Disparity Estimation and Contour Detection," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 11, pp. 121-136, 1989.

[7] Horn, Berthold K. P., Robot Vision, MIT Press, Cambridge, MA.., 1986.

[8] Kanade, T. and M. Okutomi, "A Stereo Matching Algorithm With an Adaptive Window: Theory and Experiment," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 16, no. 9, pp. 920-932, Sept. 1994.

[9] Leclerc, Y.G., Q.T. Luong, and P. Fua, "A Framework for Detecting Changes in Terrain," IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 20, no. 11, pp. 1143-1160, November 1998.

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

[11] Medioni, G. and R. Nevatia, "Segment-Based Stereo Matching," Computer Vision, Graphics, and Image Processing, vol. 31, pp. 2-18, 1985.

[12] Schultz, H., "Terrain Reconstruction from Widely Separated Images", Proc. SPIE, Volume 2486, pp. 113-123, Orlando, FL, April, 1995.

[13] Slama, Chester C. (Editor), Manual of Photogrammetry, 4ed., American Society of Photogrammetry, Falls Church, VA., 1980.

[14] Witkin, A., D. Terzopoulos, and M. Kass, "Signal Matching Through Scale Space," Int'l J. Computer Vision, pp. 133-144, 1987.