# Photogrammetric Area-Based Least Square Image Matching for Surface Reconstruction

Martinus Edwin Tjahjadi
Department of Geodesy and Geoinformatics
National Institute of Technology, Malang-Indonesia
edwin.tjahjadi@gmail.com

***Abstract*—In computer vision and robotics community, a normalized cross correlation image matching is widely known as a robust method to determine conjugate points between two overlapping images. In photogrammetric community, however, this method is less favor due to stringent requirements of precision. To achieve such a high standard, a least square adjustment is utilized to minimize a cost function of the image matching process, and then the sum of the residual errors of the cost function is employed to judge the precision and reliability of the match. This paper elaborates the least square image matching adjustment to match conjugate points for surface reconstructions in a highly convergent imaging network or in a wide baseline of stereo images.**

***Key-words*: *surface reconstruction, image matching, least-square, photogrammetry.***

## I. INTRODUCTION

Reconstructing three-dimensional surface models from one or more digital images has long been one of the central topics in photogrammetry [1, 2]. Surface reconstruction in industrial settings means determination of geometric models of three dimensional objects in arbitrary coordinate systems [3], and its ultimate goal is always to find a way to generate a computer model of the object surfaces which best fits reality [4].

In seeking this ultimate goal, the photogrammetric method nowadays requires a reliable mensuration system by means of digital images. This process consists of a few well-known steps [5], namely imaging network design, image measurement, geometric and texture modeling, and visualization of the results. As has been pointed out by some authors [6, 7], a final result of the measurement phase is usually three-dimensional coordinate data for the object points. To convert these finite points into a precisely meaningful surface, the geometrical condition of the object's surface per se must be taken into consideration [4]. Apart from the postulate that increasing sampling density of object points might increase the chance of the geometric modeling procedure recovering the unknown surface [2], the measured points must satisfy certain properties required by the algorithm to infer the correct geometry of the surface. For example, the image points must have little noise. This paper will provide a discussion of the fundamental mathematical concepts of the least squared image matching including a normalized cross correlation method, by which 3D object space determination is achieved. Its reliable photogrammetric point determination is central to accurate surface reconstruction.

## II. AREA –BASED IMAGE MATCHING

The term 'image matching' refers to the process of finding corresponding or conjugate points in digital images (or parts thereof) in the form of a matrix of reflectance levels [8]. Photogrammetric literature shows that there are three general methods of image matching, namely area-based matching, feature-based matching and relational matching [9]. Since this research focuses on the area-based matching, it gives an insight into the utilization of this technique only.

Area-based matching is based on the idea that grey values of pixels of conjugate points have similar radiometric characteristics [9]. The process generally requires a close approximation to the matched patches in order to ensure a successful match. In other words, having a point in one image, its conjugate in the other one is obtained by optimizing a certain similarity measure, defined over the pixel grey values within the image window. Two techniques are adopted to calculate the possible similarity measures: a normalized cross correlation method and a least square matching method.

### *A. Normalized Cross Correlation (NCC) Method*

The general procedure of the cross correlation method is to calculate a similarity measure between a patch f(x,y) on a reference image, and a matching or target patch g(x,y) on an overlapping or matching image (Fig. 1). The position of the best agreement is assumed to be the location of the reference patch on the matching image. The similarity measure is indicated by a cross correlation coefficient [9] which is computed by comparing every pixel in the reference patch with the corresponding pixel in the matching patch. A common similarity measure is the normalized cross-correlation coefficient [10, 11]:

$$\rho = \frac{\sum_{i=1}^{m}\sum_{j=1}^{n}\left[\left(f_{ij}-\bar{f}\right)\left(g_{ij}-\bar{g}\right)\right]}{\sqrt{\left[\sum_{i=1}^{m}\sum_{j=1}^{n}\left(f_{ij}-\bar{f}\right)^{2}\right]\left[\sum_{i=1}^{m}\sum_{j=1}^{n}\left(g_{ij}-\bar{g}\right)^{2}\right]}} \tag{1}$$

In Eq. (1), it is the normalized cross-correlation coefficient; $m$ and $n$ are the numbers of rows and columns of the patches respectively; $f_{ij}$ is the $i^{th}$ row and $j^{th}$ column of the grey value from the reference patch; $g_{ij}$ is the ith row and $j^{th}$ column of the grey value from the matching patch; and are the arithmetic means of the grey values in the reference patch and the matching patch, respectively.

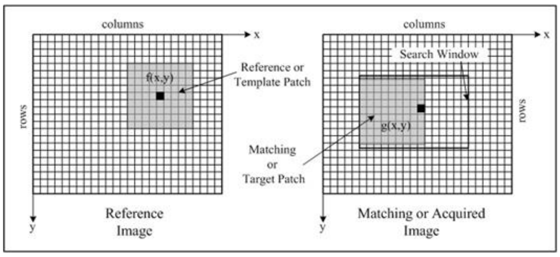


Fig. 1. The concept of area-based image matching

Despite its computational simplicity, it is computationally expensive considering that the correlation coefficient is calculated at every pixel in two directions over a given patch in the search window. Another major disadvantage of this method is that it neither takes into account the fact that there may be geometric and radiometric differences between the two patches being matched, nor adapts to distortions caused by scale and perspective differences between images, different lighting condition, and high frequency noise contaminations. Consequently, the match determined by this technique is error prone, and can often produce a misleading match. These disadvantages underpin a concept of solution which needs to account for geometric and radiometric differences between patches in order to seek a better match. This concept is set up in the context of least squares estimation.

*B. Least-Squared Image Matching (LSM) Method*

The concept of least squares matching is to minimize the grey level differences between the reference patch and the matching patch, whilst computing the position and the shaping parameters of the matching patch during the least squares estimation process. Therefore, the position and the shape of the matching patch are both varied until the grey level differences between the deformed matching patch and the reference patch reaches a minimum. The method of least squares matching (LSM) employs iterative radiometric and geometric transformations between the reference patch and the matching patch. As illustrated in Fig. 1, if f(x,y) is to be a reference patch of *n* x *n* pixels and g(x,y) is to be a matching patch of an equal size, the objective of LSM is to estimate a new location of g(x,y) such that the grey value differences between f(x,y) and g(x,y) are minimized. In an ideal situation where noise free patches exist, matching is established if the following condition is met [12]:

$$f(x, y) = g(x, y) \tag{2}$$

In a real situation, however, either one or both images are affected by noise. Thus, Eq. (2) becomes inconsistent. Therefore, assuming the reference image is noise free, a noise vector e(x,y) is added to the matching patch resulting in (3)

$$f(x, y) - e(x, y) = g(x, y) \tag{3}$$

e(x,y) is the true error vector of a goal function, which measures the differences of grey values between the reference patch and the matching patch. The goal function to be minimized is the quadratic form of the residuals of the least squares estimation. The Eq. (3) is a non-linear least squares observation equation in terms of g(x,y), which models the reference patch function of f(x,y) with the matching patch function of g(x,y). The position of the matching patch g(x,y) in the matching image has to be estimated to a positional tolerance of a pixel or so with respect to an approximate position of the matching patch, $g^{\circ}(x, y)$. The location is described by shift parameters $\Delta x$ and $\Delta y$, which are applied to the patch $g^{\circ}(x, y)$ to yield the best estimate for the position of g(x,y). In order to account for a variety of systematic image deformations and to obtain a better match, geometric corrections such as image shaping parameters as well as the shift parameters, and radiometric corrections are introduced [1] [12]. The image shaping parameters are determined by a resampling of $g^{\circ}(x, y)$ over the transformed grid points. The geometric correction parameters need to be estimated from Equation 3, and in order to conform with the least squares approach, the function g(x,y) must be linearized as follows:

$$\begin{aligned} & f(x, y) - e(x, y) = \\ & g^{\circ}(x, y) + \frac{\partial g^{\circ}(x, y)}{\partial x} dx + \frac{\partial g^{\circ}(x, y)}{\partial y} dy \end{aligned} \tag{4}$$

and

$$\begin{aligned} dx &= da_1 + x_o \, da_2 + y_o \, da_3 \\ dy &= db_1 + x_o \, db_2 + y_o \, db_3 \end{aligned} \tag{5}$$

And the Eq. (4) is then modified to become

$$f(x, y) = g^{\circ}(x, y) + g_x \, dx + g_y \, dy + e(x, y) \tag{6a}$$

$$g_x = \frac{\partial g^{\circ}(x, y)}{\partial x} \quad ; \quad g_y = \frac{\partial g^{\circ}(x, y)}{\partial y} \tag{6b}$$

The $g_x$ and $g_y$ are a discrete first derivative (or a gradient) in the x-direction and in the y-direction, respectively. Values of $g_x$ and $g_y$ are evaluated as the slopes of the reflectance levels in the x and y directions across the initial matching patch before performing iteration, and across the transformed matching patch thereafter. Schenk [9] reported that even if the position and image shaping model of the projected patch are correctly determined in the acquired image to obtain g(x,y), the grey values of f(x,y) and g(x,y) are generally going to differ due to other factors such as temporal differences of illumination source radiance, different distance and viewing angles of the cameras to the object, lens distortion, and errors in image acquisition. To compensate for these errors and acquire a better match, a set of radiometric transformation parameters for g(x, y) is incorporated. Two radiometric parameters, $r_o$ (grey value shift) and $r_1$ (grey value scale), are introduced into the system Equation 6 and it gives a result as follows

$$\begin{aligned} & f(x, y) - e(x, y) = \\ & g^{\circ}(x, y) + g_x \, dx + g_y \, dy + r_o + r_1 \, g^{\circ}(x, y) \end{aligned} \tag{7}$$

The radiometric transformation parameters introduced in Eq. (7) compensate for grey value differences in terms of brightness and contrast between the reference and matching patches. This transformation would perform a general brightness shift $r_o$ and contrast stretching $r_1$ to perform a radiometric adjustment of the image characterised by $g^{\circ}(x, y)$. And it gives the result:

$$\begin{aligned} f(x, y) - e(x, y) = \; & g^{\circ}(x, y) + g_x\, da_1 \\ & + g_x\, da_2 + g_x\, da_3 + g_y\, db_1 + g_y\, db_2 \\ & + g_y\, db_3 + r_o + r_1\, g^{\circ}(x, y) \end{aligned} \tag{8}$$

Thus, an equation of the following form conforms to the standard indirect least squares adjustment and it can be written for each pixel as follows:

$$\begin{aligned} \{ f(x, y) - g^{\circ}(x, y) \} - e(x, y) = \; & \\ g_x\, \delta a_1 + g_x\, x_o\, \delta a_2 + g_x\, y_o\, \delta a_3 + g_y\, \delta b_1 & \\ + g_y\, x_o\, \delta b_2 + g_y\, y_o\, \delta b_3 + r_o + r_1 g^{\circ}(x, y) & \end{aligned} \tag{9}$$

Combining the parameters in (9) in the parameter vector x gives:

$$x^T = \begin{bmatrix} da_1 & da_2 & da_3 & db_1 & db_2 & db_3 & r_o & r_1 \end{bmatrix} \tag{10}$$

their coefficients in the design matrix A, and the vector differencef(x,y) – $g^0(x,y)$ in ℓ, the observation equations are obtained in classical notation(with e = e(x,y)) as $\ell - e = Ax$. In the standard indirect model for n x n pixels of a reference patch, the Equation (8) can be rewritten as:

$$\underset{(n\cdot n,8)}{A} = \begin{bmatrix} g_x & g_x x_{o_1} & g_x y_{o_1} & g_y & g_y x_{o_1} & g_y y_{o_1} & 1 & g^{\circ}(x_1, y_1) \\ g_x & g_x x_{o_2} & g_x y_{o_1} & g_y & g_y x_{o_2} & g_y y_{o_1} & 1 & g^{\circ}(x_2, y_1) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ g_x & g_x x_{o_n} & g_x y_{o_1} & g_y & g_y x_{o_n} & g_y y_{o_1} & 1 & g^{\circ}(x_n, y_1) \\ g_x & g_x x_{o_1} & g_x y_{o_2} & g_y & g_y x_{o_1} & g_y y_{o_2} & 1 & g^{\circ}(x_1, y_2) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ g_x & g_x x_{o_n} & g_x y_{o_2} & g_y & g_y x_{o_n} & g_y y_{o_2} & 1 & g^{\circ}(x_n, y_2) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ g_x & g_x x_{o_{n-1}} & g_x y_{o_n} & g_y & g_y x_{o_{n-1}} & g_y y_{o_n} & 1 & g^{\circ}(x_{n-1}, y_n) \\ g_x & g_x x_{o_n} & g_x y_{o_n} & g_y & g_y x_{o_n} & g_y y_{o_n} & 1 & g^{\circ}(x_n, y_n) \end{bmatrix}$$

$$\underset{(n\cdot n,1)}{v} = \begin{bmatrix} -e(x_1, y_1) \\ -e(x_2, y_1) \\ \vdots \\ -e(x_n, y_1) \\ -e(x_1, y_2) \\ \vdots \\ -e(x_n, y_2) \\ \vdots \\ \vdots \\ -e(x_n, y_n) \end{bmatrix} \quad \underset{(n\cdot n,1)}{\ell} = \begin{bmatrix} f(x_1, y_1) - g^{\circ}(x_1, y_1) \\ f(x_2, y_1) - g^{\circ}(x_2, y_1) \\ \vdots \\ f(x_n, y_1) - g^{\circ}(x_n, y_1) \\ f(x_1, y_2) - g^{\circ}(x_1, y_2) \\ \vdots \\ f(x_n, y_2) - g^{\circ}(x_n, y_2) \\ \vdots \\ \vdots \\ f(x_n, y_n) - g^{\circ}(x_n, y_n) \end{bmatrix} \tag{11}$$

The least squares estimation in model (9)-(11) leads to the unbiased, minimum variance estimators:

| | |
|---|---|
| $\hat{x} = \left(A^T P A\right)^{-1} A^T P \ell$ | solution vector |
| $\hat{\sigma}_0^2 = \frac{1}{r} v^T P v$ | variance factor |
| $v = A\hat{x} - \ell$ | residual vector |
| $r = n - u$ | |
| $C_x = \hat{\sigma}_0^2 \left(A^T P A\right)^{-1}$ | variance-covariance matrix |
| r… | redundancy |
| n… | number of observations |
| u… | number of transformation parameters |

Here **A** is the matrix of coefficients, x is the vector of corrections to approximate parameters values; ℓ is the discrepancy vector of constants between the reference patch and the initial measured matching patch; and **v** is the vector of noise values. The vector **v** can also be regarded as a measure of the quality of the mathematical model ([10]). The least squares solution minimises the sum of squares of the elements of **v,** which leads to the unbiased minimum variance estimators. **P** is the weight matrix which is usually approximated by the identity matrix by assuming an identical precision of all pixels; n is the number of pixel in row or column direction; $\hat{\mathbf{x}}$ is the solution vector; x is the correction vector which is applied for the geometric transformation parameters only; $r_o$and $r_1$ are linear apriori; and $\hat{\sigma}_o$ can be regarded as an a-posteriori estimator for the difference of the reference patch noise and the matching patch noise. $C_x$ is the variance-covariance matrix of the transformation parameters and it is used to judge the quality of parameter estimation. Since the function values of g(x, y) in Equation 3 are stochastic quantities, the design matrix **A** is not fixed. However Gruen[12] stated that ignoring the stochastic quantities does not significantly disturb the results.

## III. Computational Procedure

Schenk [9] observed that the adjustment procedure for least squares matching is somewhat different from the usual iteration cycle of a least squares adjustment. The first iteration commences with an approximate location of the matching patch. The coefficients of the design matrix A and discrepancy vector w are calculated using initial values of parameters to initiate the iteration. These initial values are often [12]: $r_0 = a_1 = b_1 = a_3 = b_2 = 0; r_1 = a_2 = b_3 = 1$.

Furthermore, since the matrix **A** includes digital numbers from the matching patch g(x,y), partial derivative terms must be obtained using discrete values to estimate the slope of the matching patch in both x and y directions. The slope gradients are calculated using the initial target patch $g^{\circ}(x, y)$ and formulated as follows:

$$\begin{aligned} g_x &= \frac{\partial g^{\circ}(x, y)}{\partial x} = \frac{g^{\circ}(x+1, y) - g^{\circ}(x-1, y)}{2} \\ g_y &= \frac{\partial g^{\circ}(x, y)}{\partial y} = \frac{g^{\circ}(x, y+1) - g^{\circ}(x, y-1)}{2} \end{aligned} \tag{12}$$

The Eq. (12) computes the estimate values for slopes both in the x and y directions, by taking the difference between the digital numbers of pixels to right and left, and above and below. Next, the transformation correction parameters and their estimated values can be determined by from solving Eq. (9). Before commencing the second iteration, the grey values for all positions $g^1(x, y)$ must first be determined.

This resampling process amounts to interpolating the grey values from the neighbouring pixels of the initial matching patch $g^{\circ}(x, y)$. According to the photogrammetric literature, it has been established that bi-linear interpolation is the best choice among the existing techniques available, such as the nearest-neighbour, bi-cubic, and distance weighted average interpolation methods [2][13]. In Wolf and Dewitt [11], the authors affirm that of the first three techniques, the simplest and fastest resampling method in terms of computation time is nearest-neighbour interpolation, which uses the value of the pixel closest to the transformed coordinates. However, since a continuous interpolation is not being performed, the resulting appearance can be very susceptible to aliasing. Bilinear interpolation, on the other hand, is slower than the nearest neighbour method, and has a smoother appearance effect due to partial elimination of high frequency detail. The bi-cubic technique is the slowest of the three with regard to computation time, but it is the most rigorous resampling method, and achieves a smooth appearance without sacrificing too much high frequency (edge) detail.

## IV. Result and Discussion

Two photographs are taken by using camera Nikon D100. The imageries have resolution of 3000x2000 pixels. The left image (DSC_0037.tif) is assumed to be the reference image, whilst the right one (DSC_0044.tif) is of the matching image. In seeking the most precise of the matched points, the LSM is conducted. Since the LSM requires very close approximate values (small pull in range), the utilized image pair needs to be normalized first [9, 14]. The normalized images for the left and right images are of NM_DSC_0037.jpg and NM_DSC_0044.jpg accordingly. Figure 2 depicts this image pair and their normalized images counterparts.

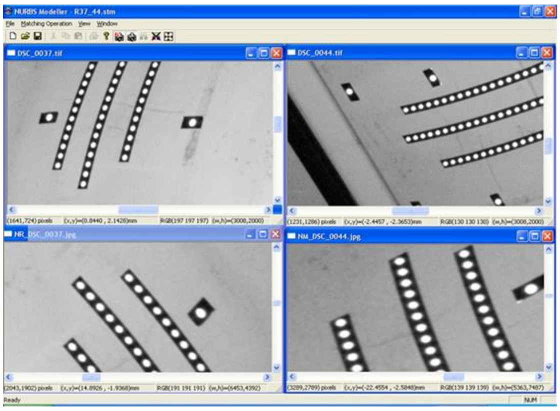

Fig. 2. The Stereo images (top) and their normalized pair (below)

When any feature is clicked on the reference image, the location of the selected feature is transformed to the normalized reference image (NM_DSC_0037.jpg). Then, the NCC is performed to compute the conjugate point on the normalized matching image (NM_DSC_0044.jpg), before it is transformed to the matching image (DSC_0044.tif). This transformed conjugate point acts as an approximate point to perform the LSM. As a result, a sub pixel conjugate point (the matched point) is obtained as shown on Figure 3.

Fig. 3. The Least Square Matching Process: a reference patch (top left), a normalized reference patch (bottom-left), a normalized matching patch (bottom-right), and a matching patch (top-right), as well as the matching result information (center window)

Fig. 3 depict a process of finding a matching entity in the stereo view using normalized images. The pixel in the center of the patch is transformed to the normalized counterpart (bottom-left). Then, the NCC process starts in seeking the best match on the normalized matching image along the epipolar line. Once found, the location of the best match is refined into the sub-pixel accuracy through the use of the LSM on the matching image. The LSM is an iterative process, during the iterations, the local patch is transformed into the reference patch. On the last iteration, therefore, the local patch resembles the reference patch; and its center pixel is to be the conjugate point of the center pixel of the matching patch, as shown in Fig. 4.

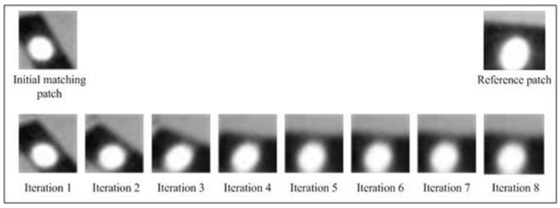


Fig. 4. Iterative transformations of the initial matching patch into the reference patch.

The adjustment equation for the LSM is usually very over determined. For example, a patch size of 21x21 pixels generates $n$ = 441 observations for only $u$ = 6 unknowns. Grey level gradients are used in the linearized correction equations. A solution exists only if enough image structures are available in the matched patch (Fig. 5f); while for homogeneous image patches, the normal equation system is singular, which is the situation illustrated in Fig. 5.e.

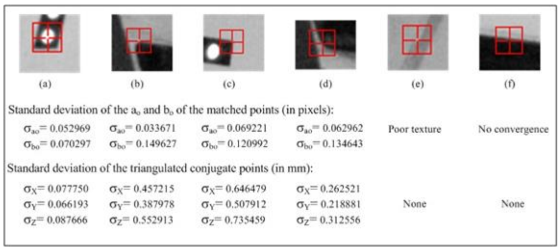


Fig. 5. The accuracies of the matched patches and the triangulated points

The standard deviations of shift parameters $a_0$ and $b_0$ are assessed to judge the accuracy of the matched points and the highest possible accuracies have been reported to be in the range of 0.01-0.04 pixels [10]. High accuracy assumes good similarity between the reference and the matching patch. A typical result of standard deviations of the $a_0$ and $b_0$ calculated by the software developed for this research is depicted in Figure 5(a)-(d).

## V. CONCLUSION

This paper has presented the semi-automatic method used to produce object point coordinates from the image matching and spatial intersection process. The process starts by selecting a pixel on the reference image. Then, the program can be used to automatically find its conjugate point on the matching image, as well as the corresponding point in the object space. The automatic searching of the matched point is done in two processes. The first process is performed on the normalized image pair to compute the matched point using cross correlation matching. Using the matched point on the normalized matching image as the approximate value, the second process is performed on the original image pair to refine the matched point position on the matching image to obtain sub-pixel accuracy of the matched point through the LSM method.

## REFERENCES

[1] M. Pollefeys, R. Koch., M. Vergauwen, and L. V. Gool, "Automated reconstruction of 3D scenes from sequences of images," *ISPRS Journal of Photogrammetry and Remote Sensing,* vol. 55, pp. 251-267, 2000/11 2000.

[2] F. Rottensteiner, "Three Dimensional Object Reconstruction by Object Space Matching," *International Archives of Photogrammetry and Remote Sensing,* vol. 31, pp. 692-696, 9-19 July 1996 1996.

[3] W. Foerstner, B. Wrobel, F. Paderes, R. Craig, C. S. Fraser, and J. Dolloff, "Analytical Photogrammetric Operations," in *Manual Of Photogrammetry*. vol. 5th, C. J. McGlone, *et al.*, Eds., 5th ed Bethesda: American Society for Photogrammetry and Remote Sensing, 2004, pp. 763-948.

[4] F. Remondino. (2003, 24 September 2005). From Point Cloud to Surface: The Modeling and Vizualization Problem. *Proceedings of the ISPRS working group V/6: Workshop on Visualization and Animation of Reality-based 3D Models34(B5/W10)(34(B5/W10))*. Available: http://www.photogrammetry.ethz.ch/tarasp_workshop/papers/remondin.pdf

[5] A. W. Gruen. (2002, 13 September). *Return of the Buddha - New Paradigms in Photogrammetric Modeling, Keynotes at ISPRS Commission V Symposium in Corfu*. Available: http://www.photogrammetry.ethz.ch/general/persons/AG_pub/keynotes_corfu/index.htm

[6] S. F. El-Hakim, J. A. Beraldin, M. Picard, and G. Godin, "Detailed 3D Reconstruction of Large-Scale Heritage Sites with Integrated Techniques," *IEEE Computer Graphics and Applications,* vol. 24, pp. 21-29, 2004.

[7] A. W. Gruen, F. Remondino, and L. Zhang, "Photogrammetric Reconstruction of the Great Buddha of Bamiyan, Afghanistan," *Photogrammetric Record,* vol. 19, pp. 177-199, September 01, 2004 2004.

[8] A. W. Gruen, "Adaptive Least Square Correlation: A Powerful Image Matching Technique," *South African Journal of Photogrammetry, Remote Sensing and Cartography,* vol. 14, pp. 175-187, 1985.

[9] T. Schenk, *Digital Photogrammetry, Volume 1* vol. Volume 1. Laurelville: TerraScience, 1999.

[10] T. Luhmann, S. Robson, S. Kyle, and I. Harley, *Close Range Photogrammetry: Principles, Techniques and Applications*. Scotland, UK.: Whittles Publishing, 2006.

[11] P. R. Wolf and B. A. Dewitt, *Elements of Photogrammetry: with Applications in GIS*, 3rd ed. New York: McGraw-Hill Companies Inc., 2000.

[12] A. W. Gruen, "Least Square Matching: A Fundamental Measurment Algorithm," in *Close Range Photogrammetry and Machine Vision*, K. B. Atkinson, Ed., ed Scotland, UK: Whittles Publishing, 2001, pp. 217-255.

[13] G. H. Schut, "Review of Interpolation Methods for Digital Terrain Models," *The Canadian Surveyor,* vol. 30, pp. 389-412, December 1976 1976.

[14] W. Cho and T. Schenk, "Resampling Digital Imagery to Epipolar Geometry," *International Archives of Photogrammetry and Remote Sensing,* vol. 29, pp. 404-408, 1992.