MULTI-IMAGE MATCHING USING NEURAL NETWORKS AND PHOTOGRAMMETRIC CONDITIONS

Ahmed F. Elaksher

Faculty of Engineering, Cairo University, Egypt, -ahmedelaksher@yahoo.com

Commission III, WG III/1

KEYWORDS: Image matching, Neural networks, coplanarity condition, collinearity condition, correlation.

ABSTRACT:

Automatic determination of three dimensional information from digital images is a fundamental problem in digital photogrammetry and computer vision. The hardest part of the problem is finding conjugate points in two images. Despite the wealth of information contained in digital images, factors such as occlusion and discontinuity weaken several matching algorithms. However, image matching using more than a pair of stereo images enhance the reliability of the image matching process. This paper presents an alternative approach to match image points across several views. For each pair of images, the coplanarity condition and the correlation coefficient of image intensities are computed for each pair of image points. These two measures are feed into a feed-forward neural network used to solve the multi-image correspondent. The collinearity condition is then used to validate the outputs of the neural network and to compute the 3D coordinates of the matched points. The detection rate of the neural network is about 95% to 98% and the false alarm rate is about 7% to 4%. In addition, the collinearity condition eliminated several of the incorrect matches and reduced the false alarm rate to less than 2%. The RMS errors of the ground coordinates are seven to eight centimetres.

1. INTRODUCTION

Retrieving 3D spatial information using photogrammetric models depends extremely on solving the correspondence problem between image features (Croitoru and Vincent, 2003). However, feature correspondence is not a trivial task and several algorithms are presented in the literature to solve this problem (Lane and Thacker, 2007). Several factors affect this task such as: lack of texture, feature topology, condition of surrounding features, and noise.

Researchers have proposed a variety of techniques to solve and improve the performance of matching techniques. These techniques are categorized into stereo matching and multi view matching. In either case, a variety of constraints, approaches, criteria are used to find conjugate features depending upon the properties of the image (Vincent and Laganière, 2001). Techniques that uses only a pair of stereo images, usually suffer from missing or hidden information. Although there has been several attempts to solve this problem (Mordohai and Medioni, 2004; Zitnick and Kanada, 2000; Sun et al., 2005), missing features continue to challenge stereo matching. On the other hand, multi view matching can overcome such problem. This paper presents an alternative approach to solve the multi view feature correspondence problem using neural network and photogrammetric condition.

The approach take advantage of the fact that neural network can be used to infer a function from observations (Haykin, 1994). Neural networks are being used in several scientific applications to solve a variety of problems in pattern recognition, prediction, optimisation associative memory and control (Russell and Norvig, 2002). None of the conventional approaches to these problems is flexible enough to perform well outside their domain.

The image matching technique presented in this paper consists of three major steps. In the first step, feature vectors are computed for all image points in all image pairs. The elements of the feature vectors are computed, for each image point, using both image coordinates and local intensities. These two quantities are computed for all pairs of image points. The second step includes the use of a feed-back neural network to find conjugate points. The algorithm assumes that all the images are triangulated. In the last step, the collinearity condition is used to check the outputs of the neural network and compute the ground coordinates of the image points. The remaining of the paper is organized in the following order. First, a background on recent research in image matching is summarized. The proposed approach is then illustrated. Experimental results are then presented and discussed. The last section then states the research conclusions.

2. BACKGROUND

Multi view matching has been addressed by several researchers in the photogrammetric and computer vision communities. The technique presented in Maas (1996) shows a multi image matching algorithm using discrete points extracted by an interest operator. Image matching is then carried out using epipolar line intersection. The technique presented in Brown et al. (2005) starts by locating interest points using Harris corner detection (Harris and Stephens, 1998). Matching is then performed using a fast nearest neighbour algorithm that indexes features based on their low frequency Haar wavelet coefficients. Moreover, an outlier rejection procedure is also introduced that verifies a pairwise feature match based on a background distribution of incorrect feature matches. Feature matches are then refined using the Random Sample Consensus RANSAC (Fischler and Bolles, 1981).

An adaptive multi image matching technique was presented in Pateraki and Baltsavias (2002) using the 3 panchromatic channels of the ADS40 digital camera. Edge pixel matching is performed based on cross-correlation and similarity measures to provide pixel approximate positions. These positions are subsequently refined using sub-pixel matching techniques. The geometry of the sensor is used to apply matching constraints via a modified epipolar geometry designed for the pushbroom sensor. In addition a modified image pyramid approach is used for derivation of approximations. Another multi image matching method is presented in Gruen and Li (2002) to generate DSM using Three-Line-Scanner (TLS) raw images. The proposed method combines matching procedures based on grid point matching and feature point matching. The three images are matched to provide pixel and object coordinates for grid points simultaneously. An additional feature-point matching procedure is then performed to compensate for the disadvantage of modelling the terrain using grid points. This was performed via a modified Multi-photo Geometrically Constrained (MPGC) matching algorithm.

The process presented in D'Apuzzo (2002) is used to model human faces in a multi image environment. The multi image matching process is based on the geometrically constrained least squares matching algorithm presented in Gruen (1985). The process produces a dense set of corresponding points in the five images. Neighborhood filters are then applied on the matching results to remove outliers. After filtering the data, the three dimensional coordinates of the matched points are computed by forward intersection using the camera interior and exterior parameters. The researchers in Kang et al. (2001) provided two techniques to assist in solving the multi image matching problem. First, they implemented a combination of movable windows and a dynamically selected subset of the neighboring images to perform the matches. Secondly, they explicitly labeled occluded pixels within a global energy minimization framework, and reasoned about visibility within this framework so that only truly visible pixels are matched.

The researchers in Wang and Hsiao (1999) used neural networks in stereo matching. Two different types of neural networks were used. The first network utilizes intensity, variation, orientation, and position of each image pixel to facilitate self-development in network growing. The network then classify the input image into several clusters, and results are then used by a second network to achieve accurate matching. The second network is used to generate an initial disparity map. With the clustering results and the initial map, a matching algorithm that incorporates a back propagation network is then applied to recursively refine the disparity map. In the matching process, useful constraints, such as epipolar, ordering, geometry and continuity, are employed to reduce the occurrence of mismatching. Researchers in Memony and Khanz (2001) and Mendonça (2002) proposed a method to establish the relationship between 3D coordinates and the image coordinates of a point using neural networks. A three layer neural network was used. The input layer receive the image coordinates in both images. One hidden layer was used in both researches. The out layer presents the ground coordinates of each point. The networks were trained using several control points with known image and ground coordinates and then the network was used to compute the ground coordinates of a number of test points from there image coordinates. The computed ground coordinates of the test points where compared with those computed using the traditional image triangulation technique and the differences in the coordinates were insignificant.

3. METHODOLOGY

The image matching technique presented in this paper depends on generating all matching hypotheses between each pair of images. For each hypothesis, a feature vector is computed using image coordinates and intensities. A neural network is then used to find the appropriate correspondent points in all images. Once the image point matching is solved the collinearity equation is used to compute the 3D ground coordinates of the point and to discard any false matching. Image intensities is also used to assist in removing false matches. Figure 1 shows a flow chart of the proposed framework. The next three sections provide a detailed explanation of the algorithm.

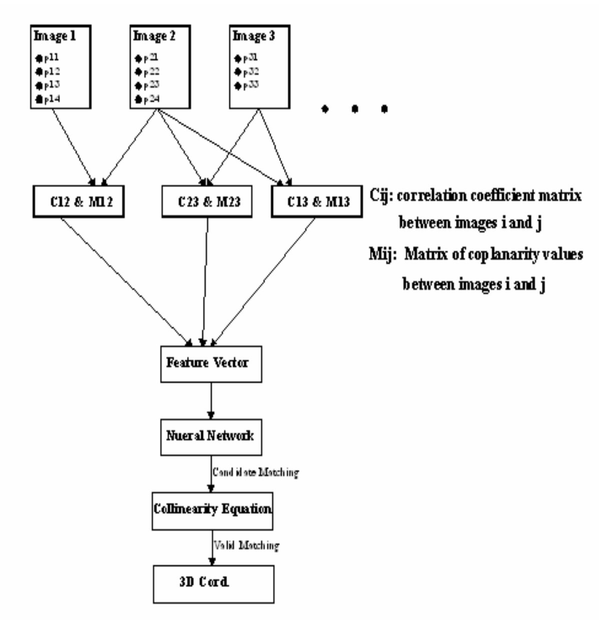

Figure 1. The proposed framework for multi image matching

3.1 Generating the pairwise matching hypotheses

For each pair of image points, there object space rays should intersect provided that they represent the same ground target. This geometric relation is represented by the coplanarity equation, equation 1. The determinate of (F), as presented in equation 1, should be zero provided that the second image point lay on the epipolar line of the first image (Mikhail et al., 2001). This fact have been used in recent computer vision and photogrammetric research to restrict the search space to a 1D space. In this research, the determinate of (F) is computed for all point pairs in each image pair. The computed values are stored in an *n*x*m* matrix (**F**), where *n* is the number of points in the first image and *m* is the number of points in the second image. For *N* images, !N/2 matrices are generated.

$$\mathrm{F} = \begin{vmatrix} b_x & b_y & b_z \\ u_1 & v_1 & w_1 \\ u_2 & v_2 & w_2 \end{vmatrix} = 0 \qquad (1)$$

where b_x, b_y, b_z = elements of the base vector representing the displacement between the perspective centers of two cameras,
u_1, v_1, w_1 = elements of the vector that representing the object space vector from the image points in the 1st image,
u_2, v_2, w_2 = elements of the vector representing the object space vector from the image points in the 2nd image.

The next step in this research is to compute the correlation coefficients for all point pairs in each image pair. The correlation coefficient is a measure used to quantify the similarity of the intensities between two image windows. It has been used widely to support pairwise image matching and to solve matching ambiguities (Helava, 1978). For each image pair, the values will be stored in another *n*x*m* matrix, i.e. the matrix of the correlation coefficients (**C**).

3.2 Computing the feature vector elements

The data input into the neural network is presented as a feature vector. The elements of the feature vectors affect the performance of any neural network significantly. Hence, they have to be carefully selected. The elements of the two matrices, i.e. the matrix of the correlation coefficients and the matrix of the coplanarity values are used to generate the feature vector as presented in equations (2a and 2b). Two different feature vectors (FV) are tested. In the first feature vector (equation 2a), both image intensities and point geometry, i.e. the values of (F), are utilized. The second feature vector (equation 2b) uses only the point geometry.

$$FV^{ijk\ldots}=[\,\mathbf{C}_{12}^{ij}\quad \mathbf{F}_{12}^{ij}\quad \mathbf{C}_{13}^{ik}\quad \mathbf{F}_{13}^{ik}\quad \mathbf{C}_{23}^{jk}\quad \mathbf{F}_{23}^{jk}\ \ldots]^t \qquad (2a)$$

$$FV^{ijk\ldots}=[\,\mathbf{F}_{12}^{ij}\quad \mathbf{F}_{13}^{ik}\quad \mathbf{F}_{23}^{jk}\ \ldots]^t \qquad (2b)$$

where $\mathbf{C}_{12}^{ij}$, $\mathbf{C}_{13}^{ik}$, $\mathbf{C}_{23}^{jk}$ = elements of the correlation matrices for points i, j, and k, in images 1, 2, and 3 respectively,
$\mathbf{F}_{12}^{ij}$, $\mathbf{F}_{13}^{ik}$, $\mathbf{F}_{23}^{jk}$ = elements of the coplanarity matrices for points i, j, and k, in images 1, 2, and 3 respectively,
dots in both formulas represent the ability for more images.

Several remarks are observed in the selected feature vector. The values of the coplanarity relation between each pair of images are dependent. However, neural networks overcome this problem and doesn't run into singularity. Moreover, for cases where one of the values is zero the others will only be zero if and only if the points represent the same ground object. In addition, since automatic point locating is not perfect, the quantities will be minimum, for matched points, and not equal to zero. The correlation coefficient values are effected by several factors such as the size of the local window, the orientation of the images, and the topology of the object. For this research the local window size is fixed to seven by seven pixels. On the other hand, the correlation coefficients could reach their maximum values even if the image points don't correspond to the same object point.

3.3 Implementing the neural network

The multi image matching problem is defined as finding the corresponding points in all images. This problem could be considered as minimizing an *L2* function of the values of the *F* and maximizing an *L2* function of the correlation values. The mathematical solution of such system is hard to implement. However, Hornik et al. (1989) showed that for any given ε>0 and any *L2* function, there exists a three-layer back-propagation neural network that can approximate the function within ε mean squared error accuracy. Thus neural networks provide exciting solution for the multi image matching problem.

The neural network implemented in this research is a feed-forward back-propagation network. The network consists of three layers; an input layer, one hidden layer, and an output layer. The number of neurons in the first layer is the same as the number of elements in the feature vectors. For the first case, i.e. using the image intensities and the point geometry, the number of neurons in the input layer is six, while for the second case the number of neurons is three figure 2. Several experiments were conducted to determine the optimum number of neurons in the second layer. Results showed no significant changes in the output of the networks. Hence, the number of neurons in this layer was selected to be ten. The number of neurons in the third, i.e. last, layer is constrained to one. The output of this neuron is either one in case the points in all image match or zero in case the points do not match. The activation functions for all neurons in the first and second layers, is the sigmoid functions (Haykin, 1994). For the output neuron, the step function is chosen as the activation function. The results of the neural network are evaluated using the Mean Square Error computed using equation 3.

$$MSE=\frac{\sum_i (Ti-Oi)^2}{n} \qquad (3)$$

where T_i = one for correct matches or zero for false matches,
O_i = output value from the neural network,
n = number of points used to evaluate the network.

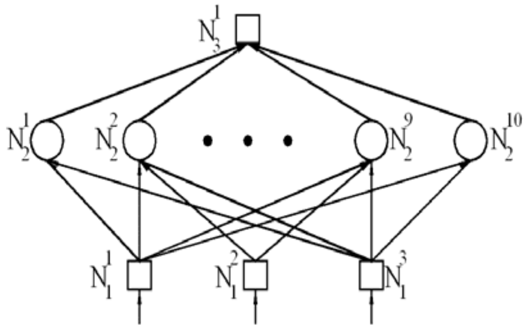

Figure 2. The implemented neural network (2nd case)

3.4 Computing ground coordinates and validating the neural network outcomes

The collinearity equation, equation 4, is used to compute the ground coordinates of any point given it's image coordinates in at least two images. For *N* images, 2*N* equations are written per point, however, only three equations are required to compute the ground coordinates of the point. This provides an over determinate system of equations. Hence, the least squares adjustment technique is used to compute the 3D coordinates.

$$F_{1\mathrm{i}}^{\mathrm{j}} = x_{\mathrm{i}}^{\mathrm{j}} - x_o + f\,\frac{U_{\mathrm{i}}^{\mathrm{j}}}{W_{\mathrm{i}}^{\mathrm{j}}}$$
$$F_{2\mathrm{i}}^{\mathrm{j}} = y_{\mathrm{i}}^{\mathrm{j}} - y_o + f\,\frac{V_{\mathrm{i}}^{\mathrm{j}}}{W_{\mathrm{i}}^{\mathrm{j}}} \qquad (5)$$

Where x_o, y_o, f = camera interior parameters,

x_i^j, y_i^j = coordinates of vertex (j) in image (i),

$$\begin{bmatrix} U_{\mathrm{i}}^{\mathrm{j}} \\ V_{\mathrm{i}}^{\mathrm{j}} \\ W_{\mathrm{i}}^{\mathrm{j}} \end{bmatrix} = R_i \begin{bmatrix} X^{\mathrm{j}} - X_{\mathrm{ci}} \\ Y^{\mathrm{j}} - Y_{\mathrm{ci}} \\ Z^{\mathrm{j}} - Z_{\mathrm{ci}} \end{bmatrix},$$

R_i = rotation matrix for image (i),

X_{ci}, Y_{ci}, and Z_{ci} = exposure station coordinates for image (i),

X^j, Y^j, and Z^j = object space coordinates of point (j).

Two measures are used to validate the matching outcomes. The first measure is the quadratic form of the residuals, i.e. (***v***^{t}***v***) of the image coordinate. In addition, the sum of the pair-wise correlation coefficients is also used. If the quadratic value is larger than a given threshold and the sum of the correlation coefficients is small than another threshold, the candidate matching is removed. These two measures provide a tool for rejecting false matches that result from the neural network.

4. EXPERIMENTS AND RESULTS

4.1 Dataset description

The dataset used in this research is for the city hall of Zurich building. The data is available on the web page of the International Society of Photogrammetry and Remote Sensing (ISPRS). The dataset is provided with a complete description of the interior orientation of the used cameras, the images, and the coordinates and the description of the reference points measured on the facades of the building by geodetic means (Streilein et al., 1999). The images acquisition was performed using an Olympus C1400L camera of about ten millimeters focal length and a progressive scan CCD of 1280x1024 pixel resolution. Nine ground control points are used to compute the exterior orientation parameters. The Root Mean Square Errors (RMSE) of the check points in the three directions is about seven to eight centimeters. In order to test and evaluate the proposed technique, four images are used. Harris corner detector was applied to all four images. The image coordinates of the points are stored in an input file. In addition, the ground coordinates are computed and stored in another file to evaluate the results. The distribution of the points in one image is shown in figure 3.

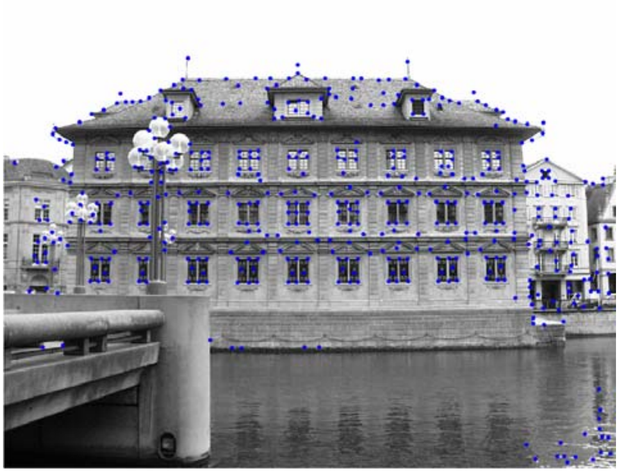

Figure 3. Points detected in one image

4.2 Data processing

The next step is to compute the pairwise matrices using both the coplanarity condition and the correlation coefficient values for each pair of points for each image pair. For the four images, used in this research, six coplanarity matrices and six correlation coefficient value matrices are generated. Figure 4 shows the coplanarity matrix and the correlation coefficient matrix between the first and second images. The next step is to generate the feature vector for each pair of image points in each pair of images. Training samples are then selected randomly from all available samples. Several experiments were conducting using a training sample size of 50 to 250 samples. The training dataset samples contain 10% samples of correctly matched points and 90% samples of false matching points. Two cases are tested, i.e. using both the correlation and coplanarity values and using only the coplanarity values. Figures 5 and 6 show the results for both cases.

The image coordinates of the correctly matched points are then feed to the least squares adjustment model. For each point eight equations are written. This provides a redundancy degree of five. The values of the quadratic form (***v***^{t}***v***) and the sum of the pair-wise correlation coefficients are computed and compared against two selected thresholds. The threshold for the quadratic form is selected to be 10e^{-3} and the threshold of the sum of the correlation coefficients is selected to 0.5. The final results showed that out of the 134 points only one points were missed. In addition, only two incorrect matching points passed the two thresholds and provided two final miss match points. The final set of matched points are shown in figure 7.

5. CONCLUSIONS

This paper presents an alternative approach to solve the multi image matching problem. Using any number of images, the process starts by computing the coplanarity condition between all pairs of images in each image pair. In addition, for each point pair the correlation coefficient is computed using local image intensities. The two measures are then used to discriminate between correct and incorrect multi image matches using a feed-forward neural network. Results of the neural network showed a detection rate of about 95% to 98% and a false alarm rate of about 4% to 7%. The collinearity condition was then used to compute 3D ground coordinates of all matched candidates. The residuals of the mage coordinate and the sum of the pairwise correlation coefficients are then used to eliminate false matches. This process reduces the false alarm rate to less than 2%.

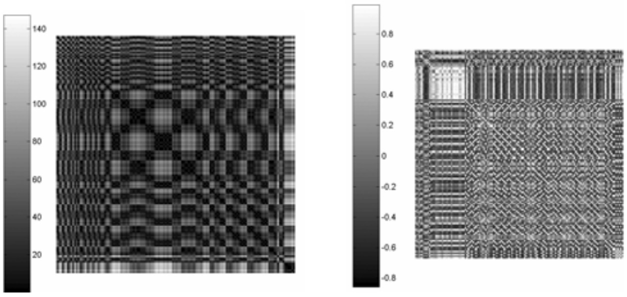

Figure 4. The coplanarity and the correlation coefficient matrices between the 1st and 2nd images

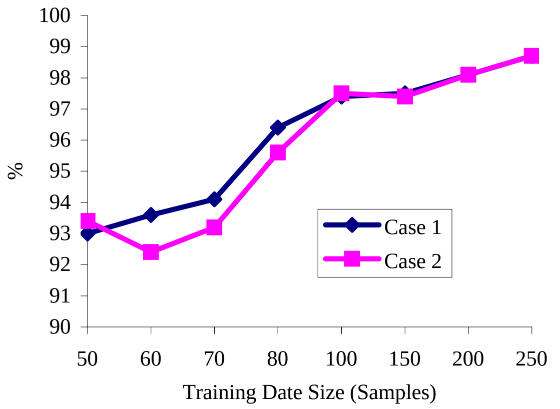

Figure 5. Percentage of correctly matched points using both cases

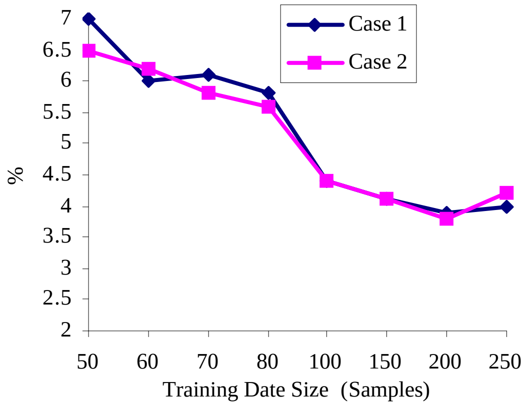

Figure 6. Percentage of false matched points using both cases

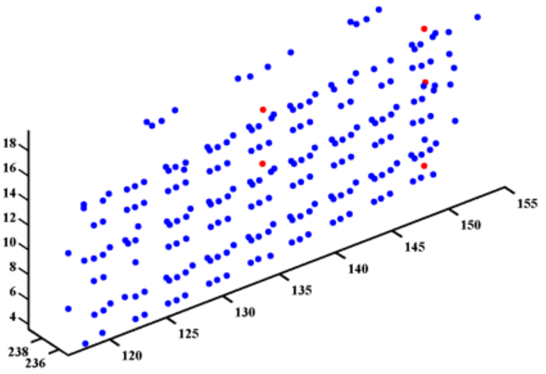

Figure 7. 3D coordinates (meters) for correctly matches (blue) and false (red) matches

REFERENCES

Brown, M., Szeliski, R., Winder, S., 2005. Multi-image matching using multi-scale oriented patches. *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, San Diego, Ca, USA, Vol. 2, pp. 510-517

Croitoru, A. and Vincent, T., 2003. An alternative approach to the point correspondence problem. *Proceedings of the ASPRS 2003 Annual Conference*, Anchorage, Alaska, USA, CD-ROM.

D'Apuzzo, N., 2002. Modelling human faces with multi-image photogrammetry. *Proceedings of SPIE*, San Jose, California, Vol. 4661, pp. 191-197.

Fischler, M.A. and Bolles, R.C., 1981. Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, Vol. 24(6), pp. 381-395.

Gruen, A., 1985. Adaptive least squares correlation: a powerful image matching technique. *South African Journal of Photogrammetry, Remote Sensing and Cartography*, Vol. 14(3), pp. 175-187.

Gruen, A. and Li, Z., 2002. Automatic DTM generation from three-line-scanner (TLS) images. *Proceedings of the ISPRS Commission III Symposium*, Vol. XXXIV, part 3A, pp. 131-137, Graz, Austria.

Harris, C. and Stephens, M., 1998. A combined corner and edge detector. *Proceedings of the 4th Alvey Vision Conference*, Manchester, England, pp. 147-151.

Haykin, S., 1994. *Neural networks: a comprehensive foundation*. 2nd edition, Prentice Hall, Upper Saddle River, New Jersey, pp. 842.

Helava, U.V., 1978. Digital correlation in photogrammetric instruments, Photogrammetria, Vol. 34, pp. 19-41.

Hornik, K., Stinchcombe, M., and White, H., 1989. Multilayer feed forward networks are universal approximators, *Neural Networks*, 2(5), pp. 359-366.

Kang, S.B., Szeliski, R., and Chai, J., 2001. Handling occlusions in dense multi-view stereo. *Proceedings of the IEE Conference on Computer Vision and Pattern Recognition*, Vancouver, Canada, Vol. 1, pp. 103-110.

Lane, R.A. and Thacker, N.A., 2007. Overview of stereo matching research. http://www.tina-vision.net/docs/memos_vision.php (accessed 21 Oct. 2007)

Maas, H.-G., 1996. Automatic DEM generation by multi-image feature based matching. *The International Archives of Photogrammetry, Remote Sensing, and Spatial Information Sciences*, Vol. 31, Part B3, pp. 484-489, Vienna, Austria.

Memony, Q. and Khanz, S., 2001. Camera calibration and three-dimensional world reconstruction of stereo-vision using neural networks. *International Journal of Systems Science*, Vol. 32(9), pp. 1155- 1159.

Mendonça, M., 2002. Camera calibration using neural networks. *Proceedings of the 10th International Conference in Central Europe on Computer Graphics, Visualization and Computer Vision*, Campus Bory, Plzen - Bory, Czech Republic.

Mikhail, E., Bethel, J., and McGlone, J., 2001. *Introduction to modern photogrammetry*, Join Wiley & Sons. Inc., New York, pp. .

Mordohai, P. and Medioni, G., 2004. Stereo using monocular cues within the tensor voting framework. *Proceedings of the 8th European Conference on Computer Vision*, Vol. 3024, pp. 588-601, Prague, Czech Republic.

Pateraki, M. and Baltsavias, E., 2002. Adaptive multi-image matching algorithm for the airborne digital sensor ADS40. *Proceedings of Map Asia 2002*, Bangkok, Thailand, (on CD-ROM).

Russell, S. and Norvig, P., 2002. *Artificial intelligence a modern approach*. 2nd edition, Prentice-Hall, Upper Saddle River, New Jersey, pp. 1132.

Streilein, A., Grussenmeyer, P., and Hanke, K., 1999. Zurich city hall: a reference data set for digital close-range photogrammetry. *Proceedings of the CIPA International Symposium*, Recife/Olinda-PE, Brazil.

Sun, J., Li, Y., Kang, S., and Shum, H., 2005. Symmetric stereo matching for occlusion handling. *Proceedings of the International Conference on Computer Vision and Pattern Recognition*, San Diego, CA, Vol. 2, pp. 399-406.

Vincent, E. and Laganière, R., 2001. Matching feature points in stereo pairs: a comparative study of some matching strategies. *Machine Graphics and Vision*, Vol. 10(3), pp. 237-259.

Wang, J. and Hsiao, C., 1999. On disparity matching in stereo vision via a neural network framework. *Proceedings of National Science Council*, Vol. 23(5), pp. 665-678.

Zitnick, C.L. and Kanada, T., 2000. A cooperative algorithm for stereo matching and occlusion detection. *Proceedings of the IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22(7), pp. 675-684.