# Region growing stereo matching method for 3D building reconstruction

Gaurav Gupta* and M.S. Rawat

Department of Mathematics,
H.N.B. Garhwal University,
Srinagar, India
E-mail: guptagaurav.19821@gmail.com
E-mail: hnbrawat@gmail.com
*Corresponding author

R. Balasubramanian and Rama Bhargava

Department of Mathematics,
Indian Institute of Technology Roorkee,
Roorkee, India
E-mail: balaiitr@gmail.com
E-mail: rbharfma@iitr.ernet.in

B.G. Krishna

Space Applications Centre (SAC),
Indian Space Research Organisation (ISRO),
Ahemdabad, India
E-mail: bgk@sac.isro.gov

**Abstract:** This paper presents a new stereo matching algorithm based on region growing algorithm. To avoid visiting the entire disparity space, an algorithm has been proposed that greedily grow the corresponding patches from a given set of reliable seed correspondences. The proposed algorithm is using regions as matching primitives and defines the corresponding region energy functional for matching by utilising matching technique. Initially, a pre-matching technique i.e., Harris corner detector is used to obtain the initial matching points called, seed points. Secondly, a local window-based matching method is used to determine the disparity estimate from an initial set of seeds. Finally mode filter technique fills the gap of insufficiency of the sparse disparity for whole surface reconstruction. The results show that the proposed scheme is reliable, accurate and robust to high resolution aerial images. The proposed approach is very useful in 3D city model as buildings are the most important objects in producing a 3D city model.

**Keywords:** stereo matching; region growing; feature matching; building reconstruction; disparity.

Reference to this paper should be made as follows: Gupta, G., Rawat, M.S., Balasubramanian, R., Bhargava, R. and Krishna, B.G. (2011) 'Region growing stereo matching method for 3D building reconstruction', *Int. J. Computational Vision and Robotics*, Vol. 2, No. 1, pp.89–98.

**Biographical notes:** Gaurav Gupta received his MSc from Gurukul Kangri Vishwavidyalaya, Haridwar in 2002. He is pursuing his PhD in the Department of Mathematics, H.N.B. Garhwal University, Srinagar. His area of research is computer vision and image processing.

M.S. Rawat is working as an Associate Professor in the Department of Mathematics, at H.N.B. Garhwal University, Srinagar. He has 21 years teaching and research experience. His area of research is fluid dynamics and computer vision.

R. Balasubramanian is an Assistant Professor in the Department of Mathematics at the Indian Institute of Technology Roorkee since 2006. He received his PhD in Mathematics from the Indian Institute of Technology, Madras, India. He has worked as a Postdoctoral Associate at VIZ Lab, Electrical and Computer Engineering Department, the State University of New Jersey, USA from 2002 to 2003. He has more than 100 research papers in various reputed international journals and conference. His areas of research include computer vision, graphics, satellite image analysis, scientific visualization and watermarking.

Rama Bhargava is working as a Professor in the Department of Mathematics, at Indian Institute of Technology Roorkee since 1979. She has been working in the field of computational fluid dynamics, computer graphics and mobile computing. An Australian Endeavor award winner in 2008. She has been recipient of DAAD fellowship also. She has an expertise in finite element and has shifted herself now in computer graphics and image processing.

B.G. Krishna received his MScTech (Electronics) from Andhra University, Visakhapatnam, India, in 1981 and his MTech in Electronics from the Indian Institute of Technology, Kharagpur, in 1984, with a specialisation in satellite communications and remote sensing. He joined the Space Applications Centre, Indian Space Research Organisation, Ahmedabad, in 1984. Since then, he has been involved in satellite data processing, image processing, and data analysis forIRS-1A/1B, IRS-1C/1D, Cartosat-1, and Cartosat-2. He is currently heading the Satellite Photogrammetry and Digital Cartography Division. His research interests include digital photogrammetry, geometrical data processing for remotely sensed data, image mosaicing, stereo image analysis, and pattern matching.

## 1 Introduction

Reliable and accurate 3D reconstruction of buildings is important for many applications such as digital 3D city models etc. Large efforts are being directed towards the automation of building reconstruction because manual reconstruction of buildings from aerial images is time consuming and requires skilled personnel. The use of aerial imagery (taken from an aero plane or a satellite) is a special case of architectural reconstruction. Over the last few years, reconstruction of cartographic features from aerial images has become a subject of intensive research. The stereo matching techniques are used for automatically extracting the height of the buildings from aerial stereo images. The reconstruction of 3D buildings has a processing chain of many steps for the automatic extraction of 3D buildings directly from high-resolution stereo aerial images. Stereo matching techniques are playing a key role in this process. Many authors have used different local stereo matching techniques such as normalised cross correlation (NCC); feature matching and area-based matching (Noronha and Nevatia, 2001; Cornou et al., 2003; Yom et al., 2004) as well as global matching techniques to obtain the disparity map. Much work has been done on automatic stereoscopic matching, and two distinct matching methods have emerged: feature-based and area-based approaches (Woo et al., 2008; Baillard and Dissard, 2000). Feature-based matching consists of matching primitive sets extracted from each image. Common features in an urban environment are points of interest, segments, and linear structures (Baillard and Dissard, 2000). The feature-based approach is appropriate for discontinuity, because depth discontinuities commonly appear as intensity discontinuities in the images. In area-based matching, each pixel matches from one image with their corresponding pixels in other image by measuring the similarity of grey-level value. The accuracy of feature-based approach relies on quality of edge segmentation.

Stereo matching is the most challenging problem in computer vision. The goal of stereo matching is to determine the disparity map that can be turned into depth map from two or more images taken from distinct view points. Stereo matching is an ill-posed problem. Hence, the recovery of an accurate disparity map still remains challenging, generally due to poor texture regions, disparity discontinuous boundaries and occluded area. A broad overview on stereo matching can be found in Scharstein et al. (2001) and Brown et al. (2003). All the methods on stereo matching attempt to match pixels in one image with their corresponding pixels in the other image. These methods can be classified into local (window-based) and global methods. Local methods perform matching at each pixel, using intensity values within finite window whereas global methods incorporate explicit smoothness assumptions and determine disparity simultaneously by using various minimisation techniques such as dynamic programming, intrinsic curves, graph cuts, belief propagation, etc. The main disadvantages of the traditional area-based dense stereoscopic matching methods (local methods) are that they only reconstruct smoothed layers of disparities and they are very expensive in terms of time and memory. The problems encountered when these techniques are applied to stereo matching for urban imagery, and other forms of imagery in which disparity discontinuities, shadows and occlusions occurs. To avoid visiting the entire disparity space, an algorithm has been proposed that greedily grow the corresponding patches from a given set of reliable seed correspondences. Such algorithms assume that neighbouring pixels have similar disparity, not exceeding disparity gradient limit as in Pollard et al. (1985) or a similar constraint.

In 3D reconstruction, the region growing algorithm has been used in segmentation as in Haralick and Shapiro (1985) and stereo matching as discussed by Otto and Chau (1989), O'Neill and Denos (1992) and Alagoz (2008). In this paper, we have developed a 3D reconstruction techniques based on region growing algorithm. This work proposes a stereo matching approach, which has been designed to work well for 3D surface model acquisition from high resolution stereo aerial imagery urban areas. Initially, we obtain initial matching points using a pre-matching technique i.e., Harris corner detector. It gives a high density of reconstructed 3D points, called seed points, on which a surface representation can be reconstructed. Secondly, an unconstrained growth of disparity components from an initial set of seeds and an optimal matching that works with the set of components found in the first step have been done. Finally, mode filter technique fills the gap of insufficiency of the sparse disparity for surface reconstruction.

## 2 Pre-matching

We use a pair of rectified stereo images. By rectifying the images, the corresponding epipolar lines which lie along horizontal scan-lines and the two-dimensional correspondence search problem is reduced to a scan-line search, greatly reducing both computational complexity and the likelihood of false matches. Our goal is to find the correspondence between rectified stereo images i.e., disparity space in which each element (*x*, *x*′, *y*) denotes a possible correspondence between (*x*, *y*) and (*x*′, *y*), where (x, y) and (*x*′, *y*) are the elements of left and right images. First, we have generated automatic seed points by using Harris corner detector (feature-based stereo matcher). We have found all correspondences of Harris interest points as initial seeds. Then we have to find out a large number of disparity components from a small set of seeds.

The feature selection process is highly dependent on the kind of application for which the vision is employed. The work presented here evolves from the analysis of point projections and their correspondence between image frames. In order to improve the correspondence finding, the number of points is selected corresponding to image corners or highly textured patches. The selection of image points is based on the well-known corner detector algorithm proposed by Harris. This detector finds corners in step edges by using only first order image derivative approximations. Given an image *I*(*x*, *y*), following steps are used to detect whether a given pixel (*x*, *y*) is a corner feature or not:

- Set a window *w* of fixed size, and compute the image gradient ($I_x$, $I_y$).
- At all pixels in the window *w* around (*x*, *y*), compute the matrix:

$$G(x, y) = \begin{pmatrix} I_x^2 & I_x I_y \\ I_x I_y & I_y^2 \end{pmatrix} \tag{1}$$

- A corner strength signal is given by:

$$\varnothing(x, y) = \left(\left|G(x, y)\right|\right) + kTrace^2\left(G(x, y)\right) \tag{2}$$

where $x \in R$ is a scalar, and different choices of *k* may result in favouring gradient variation in one or more directions, or maybe both. Local maxima is also extracted by performing a grey scale morphological dilation and then finding points in the corner strength image that match the dilated image which are also greater than the threshold. Thereby, a non-maximal suppression is done to determine the strongest corners. The corners detected, using the above mentioned scheme, are superimposed on the two consecutive frames.

## 3 Feature matching

From the two successive images, each has their own list of features. It must be verified which features from the first image match with that of the second. The next stage involves matching these features between both frames that are maximally correlated with each other. That is, for every feature at ($x$, $y$) in the first image, the match with the highest neighbourhood cross-correlation in the second image selected within a window centred at some ($x'$, $y$). For each feature in the second image, the best match is similarly sought in the first image. There will be a case in which a feature in one image is deemed the best neighbour by more than one feature in other image. In such cases, only the match with the highest cross-correlation is kept. Algebraically, if $I_k(x, y)$ and $I_{k+1}(x, y)$ are the two frames and $X_k$ and $X_{k+1}$ represent the detected features in $I_k$, then the correlation measure is represented by the following:

$$C_m\left(X_{k,k+1}\right)=\frac{1}{\sqrt{\sigma^2\left(X_k\right)\sigma^2\left(X_{k+1}\right)}}\sum_{i=-w}^{w}\sum_{i=-w}^{w}\genfrac{}{}{0pt}{}{\left[I_k(x+i,y+j)-I_k\right]}{\left[I_{k+1}(x+i,y+j)-\bar{I}_{k+1}\right]} \tag{3}$$

where $\sigma^2$ is the variance of patch intensity and $\bar{I}$ is the average intensity of the patch, $w$ is the size of window. To reduce the effects of ambient lighting, each pixel intensity is normalised by the average intensity of the window. It has been found that normalised correlation measure gives significantly better matches than the un-normalised ones. The assumption of small disparity between the two image frames can be used to significantly reduce the computational burden. The consecutive image frames are subtracted with an averaging filter. This compensates for the brightness differences in each frame and allows faster correlation calculation.

### *3.1 Outlier (mismatch points) rejection*

After the putative matches have been found, a set of displacement vectors relating the features between the image frames of the sequence is obtained. Due to error prone nature of the matching process, it is expected that a proportion of these putative correspondences obtained from the previous section will in fact mismatches. Thus, a robust estimation technique is needed for outlier rejection. The random sample consensus (RANSAC) algorithm is applied to the putatively matched set in order to estimate the homography between the frames and the correspondences which are consistent with that estimate, which is called as ‘inliers’. Each iteration of the RANSAC algorithm proceeds as follows:

*Algorithm*

```
1 Feature extraction: compute the points of interest in each image.
2 Putative correspondences: compute the set of interest point matches, based on proximity and
  similarity of their intensity neighbourhood.
3 RANSAC robust estimation: repeat for $N$ samples, where $N$ is determined adaptively.
    a Select a random sample of eight correspondences and compute the homography matrix
      F.
    b Calculate the distance $d$ for each putative correspondence.
    c Compute the number of inliers consistent with $H$ by the number of correspondences.
      Choose the matrix $H$ with the largest number of inliers. In the case of ties, choose the
      solution that has the lowest standard deviation of inliers.
```

First, a selection of a random eight set of matches is generated. We compute the homography of these eight points, where the homography is a matrix $H$ where $x' = Hx$, given $x$ is the set of features in the first frame matched to $x'$ in the second frame.

$$\begin{pmatrix} x' \\ y' \\ z' \end{pmatrix} = \begin{pmatrix} h_{11} & h_{12} & h_{13} \\ h_{21} & h_{22} & h_{23} \\ h_{31} & h_{32} & h_{33} \end{pmatrix} \begin{pmatrix} x \\ y \\ z \end{pmatrix} \tag{4}$$

Computing a normalised homography matrix (or fundamental matrix), requires eight points and this itself requires the normalisation of the matches, an singular value decomposition (SVD) calculation, and a denormalisation to find the matrix H. Then, the matrix is applied to the first set of points *x*, and the distance between the second set of points in the putative matches and the generated *x* is determined. Whatever distances lie below the user-defined threshold, *t*, are deemed inliers. Whatever differences lie above *t* are discarded, and after several iterations, the number which is calculated according to the procedure detailed on Hartley and Zisserman (2003), considered as the best inliers and are kept. This procedure yields an acceptable set of inliers and respective homography, and at this point the RANSAC stage is complete. For a given pair of images, the homography is discovered and that information is printed or stored.

## 4 Disparity element growing

Suppose we have a list of disparity seeds *K* and each seed is a point in disparity space, k = (*x*, *x*′, *y*). Now, the neighborhood is selected so as to limit the magnitude of disparity gradient to the unity and to improve the ability to follow a disparity component even if the image similarity peak falls in between pixels in the matching points. Its neighborhood *N*(*k*) in disparity space consists of 16 points and the sets are as follows:

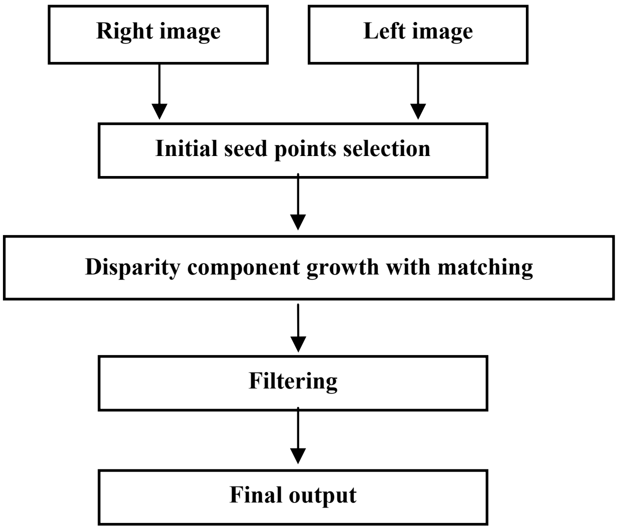
$$\begin{aligned} &N_1(k) = \{(x-1, x'-1, y),\ (x-2, x'-1, y),\ (x-1, x', x'-2, y)\}, \\ &N_2(k) = \{(x+1, x'+1, y),\ (x+2, x'+1, y),\ (x+1, x'+2, y)\}, \\ &N_3(k) = \{(x, x', y-1),\ (x \pm 1, x', y-1),\ (x, x' \pm 1, y-1)\}, \\ &N_4(k) = \{(x, x', y+1),\ (x \pm 1, x', y+1),\ (x, x' \pm 1, y+1)\}, \end{aligned}$$

NCC is used to compute the similarity measure from small image window around pixels (*x*, *y*) and(*x*′, *y*).

$$\gamma_i = (x, x', y)_i = \frac{\sum_{x,y}\left(I_2(x,y) - \bar{I}_r\right),\ \left(I_2(x+d,y) - \bar{I}_r\right)}{\sqrt{\sum_{x,y}\left(I_2(x,y) - \bar{I}_r\right),\ \left(I_2(x+d,y) - \bar{I}_r\right)^2}} \tag{5}$$

Now, we prepare an empty matching table *d* and start growing disparity components by drawing an arbitrary seed *K* from *d*, adding it to *d*, individually selecting the best-similarity neighbours $\gamma_i$, over its four sub-neighbourhoods $N_i(k)$ and putting these neighbours $\gamma_i$, to the seed list if their inter-image similarity exceeds a given threshold. If we draw a seed from the list *S* that is already a member of the matching table, then we discard it. Hence, up to four new seeds are created. The growth must stop in a finite number of steps by exhausting the list *S*.

**Figure 1** Flow-chart of the proposed approach


Right image
Left image
Initial seed points selection
Disparity component growth with matching
Filtering
Final output

The output from the growth phase is a partially filled matching table whose connected regions in 3D represent disparity components grown around the initial seeds. We use an implementation that has been proposed for semi-dense stereo matching as in Sara (2002) and that does not require a dense matching. This final matching is computationally very efficient because of the sparsity.

Stereo pair of high resolution aerial colour images with 25-centimeter resolution obtained from http://isprs.ign.fr./packages/packages¡en.html (accessed on 21 September 2010).

**Figure 2** (a) and (b) stereo pair of aerial images, (c) and (d) pair of stereo aerial images marked with feature points, (e) obtained spase (semi-dense) disparity map and (f)–(g) different views of obtained 3D model of above 2D aerial images (see online version for colours)

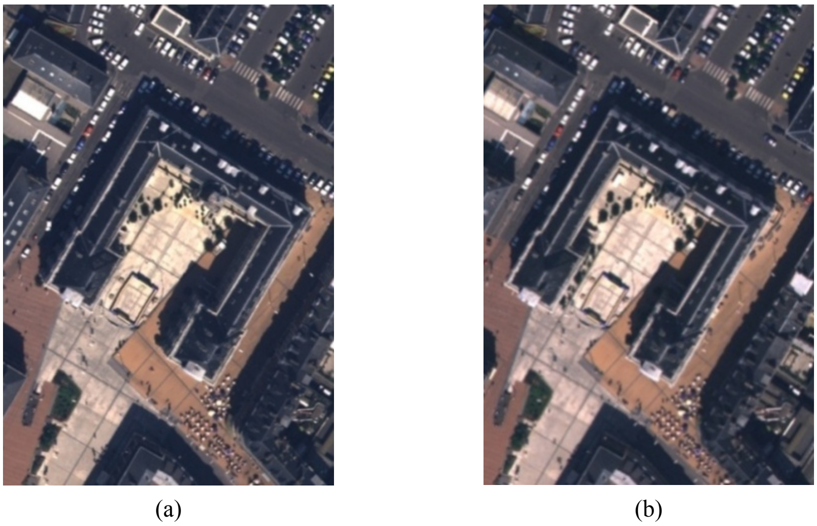
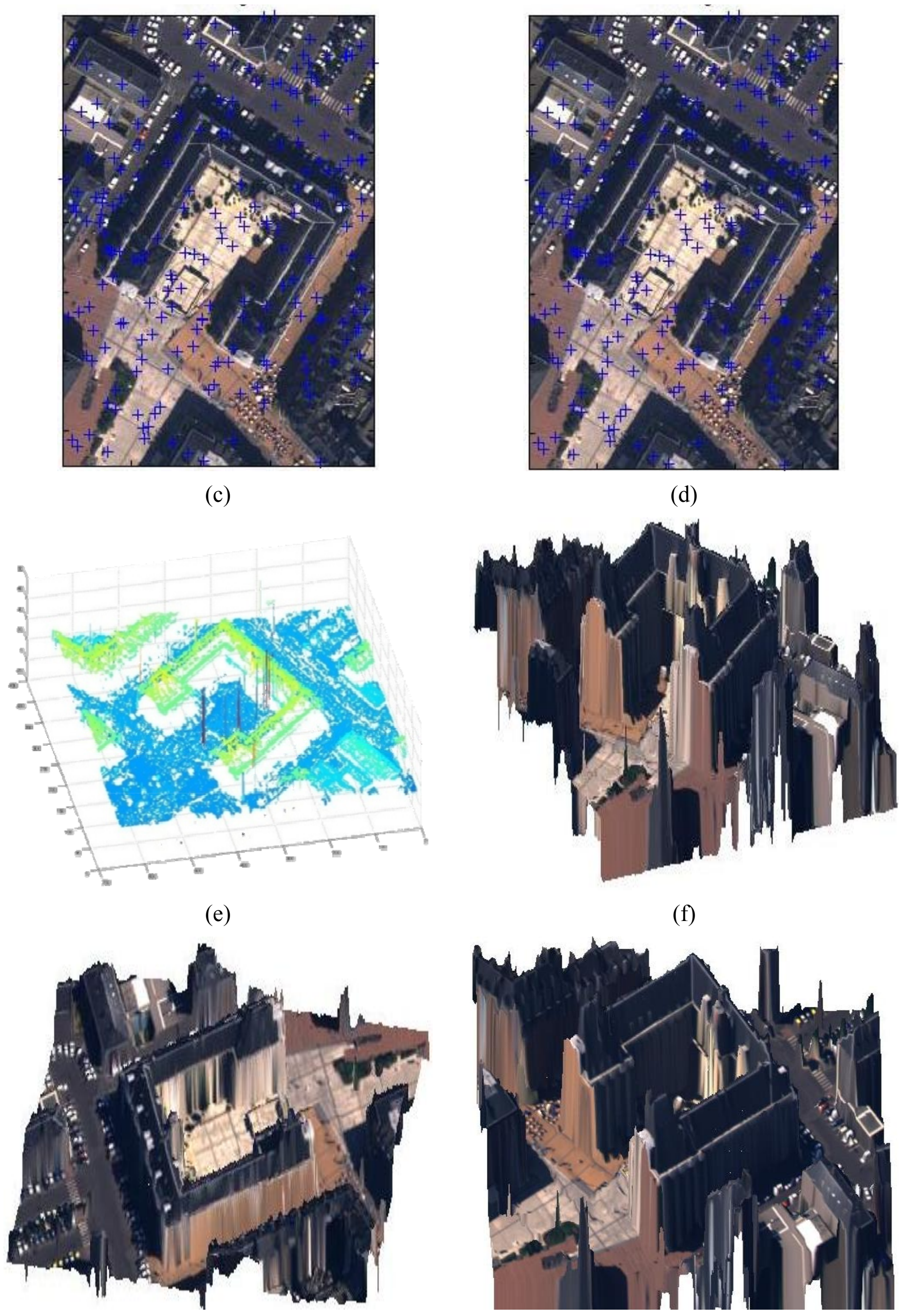


**Figure 2** (a) and (b) stereo pair of aerial images, (c) and (d) pair of stereo aerial images marked with feature points, (e) obtained spase (semi-dense) disparity map and (f)–(g) different views of obtained 3D model of above 2D aerial images (continued) (see online version for colours)



(c)

(d)

(e)

(f)

(g)

(h)

## 5 Conclusions

We have presented an approach for automatic building reconstruction from stereo images which are robust with much more difficult cases (repetitive patterns, complex scene). The proposed approach is based on disparity component growing, which solves a global optimisation task. In this algorithm, the initial seed points are obtained using Harris corner detector technique and then we have formulated the dense stereoscopic matching task as a global discrete optimisation problem. The algorithm describes an efficient disparity component growth algorithm. The experimental results indicate that the proposed approach is able to reconstruct the buildings satisfactorily. As far as error analysis is concerned, pixel by pixel error estimation is not possible by way of ground data is not available. Visually, one can perceive the quality of the obtained results. The results from the proposed approach are encouraging. To evaluate the reconstruction results, we have applied this algorithm on different kind of stereo images.

## Acknowledgements

We gratefully acknowledge the financial support of Indian Space Research Organisation(ISRO), Ahemdabad, India through the Project scheme ISR-295-MTD to carry out this research work.

## References

Baillard, C. and Dissard, O. (2000) 'A stereo matching algorithm for urban digital elevation models', *Photogrammetric Engineering and Remote Sensing*, Vol. 66, No. 9, pp.1119–1128.

Alagoz, B.B., (2008) 'Obtaining depth maps from color images by region based stereo matching algorithms', *OncuBilim Algorithm and Systems Labs*, Vol. 8, No. 4.

Cornou, S., Dhome, M. and Sayd, P. (2008) 'Building reconstruction from N uncalibrated views', Paper presented at the *International Workshop Vision Techniques for Digital Architectural and Archaeological Archives*, 1–3 July 2003, Ancona, Italy.

Scharstein, D., Szeliski, R. and Zabih, R. (2001) 'A taxonomy and evaluation of dense two-frame stereo correspondence algorithms', Paper presented at the *IEEE Workshop on Stereo and Multi-Baseline Vision*, 9–10 December 2001, Kauai, Hawaii.

Haralick, R.M. and Shapiro, L.G. (1985) 'Image segmentation techniques', *Computer Vision, Graphics and Image Processing* (*CVGIP*), Vol. 29, pp.100–132.

Brown, M.Z., Burschka, D. and Hager, G.D. (2003) 'Advances in computational stereo', *IEEE Trans. on Pattern Analysis and Machine Intelligence* (*PAMI*), Vol. 25, No. 8, pp.993–1008.

Noronha, S. and Nevatia, R. (2001) 'Detection and modeling of buildings from multiple aerial images', *IEEE Trans. on Pattern Analysis and Machine Intelligence* (*PAMI*), Vol. 23, No. 5, pp.501–518.

O'Neill, M.A. and Denos, M.I. (1992) 'Practical approach to the stereo matching of urban imagery', *Image and Vision Computing* (*IVC*), Vol. 10, No. 2, pp.89–98.

Otto, G.P. and Chau, T.K.W. (1989) 'Region-growing algorithm for matching of terrain images', *Image and Vision Computing* (*IVC*), Vol. 7, No. 2, pp.83–94.

Pollard, S.B., Mayhew, J.E.W. and Frisby, J.P. (1985) 'A stereo correspondence algorithm using a disparity gradient constraint', *Perception*, Vol. 14, pp.449–470.

Sara, R. (2002) 'Finding the largest unambiguous component of stereo matching', Paper presented at the *European Conference on Computer Vision*, 27 May–2 June 2002, Copenhagen, Denmark.

Woo, D., Nguyen, Q., Nguyen, T.Q., Park, Q. and Jung, Y. (2008) 'Building detection and reconstruction from aerial images', Paper presented at the *International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 11–3 July 2008, Beijing, China.

Yom, J-H., Lee, D-C., Kim, J.W. and Lee, Y.W. (2004) 'Automatic recovery of building heights from aerial digital images', Paper presented at the *IEEE Proceedings of Geoscience and Remote Sensing Symposium*, 20–24 September 2004, Anchorage, AK.

Hartley, R. and Zisserman, A. (2003) *Multiple View Geometry in Computer Vision*, Cambridge University Press.