C. Sun, A Fast Stereo Matching Method, In Digital Image Computing: Techniques and Applications, Massey University, Auckland, New Zealand, 10-12 December 1997, pp.95-100.

A Fast Stereo Matching Method

Changming Sun

CSIRO Mathematical and Information Sciences
Locked Bag 17, North Ryde, NSW 2113
AUSTRALIA
changming.sun@cmis.csiro.au

Abstract

Stereo matching is important in the area of computer vision and photogrammetry. This paper presents a fast stereo matching algorithm which produces a dense disparity map by using a pyramid structure, fast correlation and dynamic programming techniques. Fast correlation is achieved by using the box filtering technique which is invariant to the size of the correlation window. The disparity for each scan line is found in the correlation matrix by finding the best path using dynamic programming rather than simply choosing the position that gives the maximum correlation coefficient. Both synthetic and real image tests have been performed, and good results have been obtained.

Keywords: *Image matching, Stereo vision, Pyramid, Coarse-to-fine, Fast correlation, Dynamic programming, Box filtering, Similarity measure.*

1 Introduction

The correspondence problem in stereo vision concerns the matching of points or other kinds of primitives in two images such that the matched points are the projections of the same point in the scene. The disparity map obtained from the matching stage may then be used to compute the 3D position of the scene points given knowledge about the transformation between the two cameras.

Similarity is the guiding principle for solving the correspondence problem. Corresponding features or areas should be similar in the two images. Because of factors such as noise and occlusion, the appearances of the corresponding points will differ in the two images. For a particular feature in one image, there are usually several matching candidates in the other image. It is usually necessary to use additional information or constraints to assist in obtaining the correct match. Some of the commonly used constraints are:

1. Epipolar constraint: Under this constraint, the matching points must lie on the corresponding epipolar lines of the two images;
2. Uniqueness constraint: Matching should be unique between the two images;
3. Disparity gradient constraint: For certain kinds of 3D surfaces, the disparity gradient should be within a certain limit.

Lotti and Giraudon [1, 2] used a correlation based algorithm with an adaptive window-size that is constrained by an edge map extracted from the image. They presented results on real aerial images. Intille and Bobick [3] presented a stereo algorithm that incorporates the detection of the occlusion regions directly into the matching process. They developed a dynamic programming solution that obeys the occlusion and ordering constraints to find a best path through the disparity-space image. They also used ground control points to eliminate sensitivity to occlusion cost. Xiong *et al* [4] presented a stereo matching approach which integrates area-based and feature-based processes.

In this paper we address some of the efficient and robust implementation aspects of the stereo matching algorithms by using fast correlation and dynamic programming techniques in a multi-resolution scheme.

The rest of the paper is organised as follows: Section 2 reviews the box filtering techniques and derives the fast correlation method. The detailed matching method is described in Section 3. Section 4 shows the experimental results obtained using our stereo matching method. Section 5 contains concluding remarks.

2 Fast Correlation

Barnea and Silverman [5] introduced a class of sequential algorithms for fast image registration. They were designed to reduce computation in matching procedures using minimum dissimilarity measures like the sum of the absolute differences (SAD). Konecny and Pape [6] reviewed image correlation techniques according to photogrammetric and mathematical fundamentals.

Different similarity measures have been used in the literature [7, 8], and their performance and computation cost vary. It has also been shown that the zero mean normalized cross correlation and the zero mean sum of squared differences tend to give better results [9]. We will use the zero mean normalized cross-correlation (ZNCC) coefficient as the similarity measure of the candidate matching areas. The estimate is independent of differences in brightness and contrast due to the normalization with respect to mean and standard deviation.

Let f_{mn} be the intensity value of an $M \times N$ image f at position (m, n), where f is to be box filtered into $\bar{f}$, i.e. obtaining the mean of the original image within the box. We also have similar definition for a second image g. The normalized cross-correlation of two windows can be written as follows:

$$c_{ij,d} = \frac{cov_{ij,d}(f,g)}{var_{ij}(f) \times var_{ij,d}(g)} \tag{1}$$

where

$$cov_{ij,d}(f,g) = \sum_{m=i-K}^{i+K} \sum_{n=j-L}^{j+L} (f_{m,n} - \bar{f})(g_{m+d,n} - \bar{g}) \tag{2}$$

$$var_{ij}^2(f) = \sum_{m=i-K}^{i+K} \sum_{n=j-L}^{j+L} (f_{m,n} - \bar{f})^2 \tag{3}$$

$$var_{ij,d}^2(g) = \sum_{m=i-K}^{i+K} \sum_{n=j-L}^{j+L} (g_{m+d,n} - \bar{g})^2 \tag{4}$$

and d is the shift along epipolar lines; K and L define the correlation window size. It can be seen from this equation that the co-variance between f and g and the variances of f and g at different positions in the image need to be evaluated.

2.1 Box Filtering

McDonnell [10] described a box-filtering procedure for mean calculation. The main advantage of box filtering is its speed, which approaches four operations for each output pixel and is independent of box size. For detailed description of the technique please see [10].

2.2 Fast Calculation of Variance

Rearranging Equation (3), the following equation (Eq. 5) is obtained. It can be seen that the pixel variance within the box can also be obtained during the same pass as when calculating the mean. This is achieved by accumulating the square of the intensity values while accumulating original pixel values for mean calculation. The variance of points within the box is calculated using Equation (5).

$$\begin{aligned} var_{ij}^2(f) &= \sum_{m=i-K}^{i+K} \sum_{n=j-L}^{j+L} (f_{m,n} - \bar{f})^2 \\ &= \sum_{m=i-K}^{i+K} \sum_{n=j-L}^{j+L} f_{m,n}^2 \\ &\quad -(2K+1)(2L+1)\bar{f}^2 \end{aligned} \tag{5}$$

Therefore we have a fast way to obtain the mean and variance of the input images for the calculation of the cross-correlation that is to be used as a measure of similarity between matching candidates from the left and right images.

2.3 Fast Cross Correlation

Here again we will use the technique described in Section 2.1 to achieve fast calculation of the cross correlation. Rewriting Equation (2), we have:

$$\begin{aligned} cov_{ij,d}(f,g) &= \sum_{m=i-K}^{i+K} \sum_{n=j-L}^{j+L} (f_{m,n} - \bar{f}) \times \\ &\quad (g_{m+d,n} - \bar{g}) \\ &= \sum_{m=i-K}^{i+K} \sum_{n=j-L}^{j+L} f_{m,n} \times g_{m+d,n} \\ &\quad -(2K+1)(2L+1)\bar{f} \times \bar{g} \end{aligned} \tag{6}$$

Equation (6) is the numerator of Equation (1). Most, if not all, of the image correlation in the literature is performed using direct calculation of Eq. (6). Direct calculation of equation (6) has $(2K+1)(2L+1)$ redundancies. Similar to the calculation for the variance, cross correlation of two images can be obtained using only a few multiplications.

In our case, the correlation is performed along the epipolar line. If for any point in the left image, the search window is assumed to be within $[-w, +w]$ in the right image, then the value of d in equation (6) varies from $-w$ to $+w$. The traditional way of obtaining the correlation is to fix a point in the left image and vary d in the range of $[-w, +w]$ to calculate the correlation coefficients.

In our new algorithm for fast correlation, we first fix on one particular d all the points in the left image and calculate the cross correlation between the left image and the shifted right image of the amount d. Then we increase the number of d by 1, and repeat the process of correlation calculation until the value of d has gone through $[-w, +w]$.

The complexity of the algorithm is $O(MND)$, where M, N are the image row and column numbers and D is the maximum disparity search range.

3 Matching Strategy

3.1 Correlation Cube

The result of the above correlation is a cube containing the correlation coefficients as shown in Fig. 1(a). The size of the cube depends upon the image size and the disparity range $(2w + 1)$.

3.2 Best Path in the Matrix

Most researchers [11] choose the position that gives the maximum correlation coefficient as the disparity value. We choose a slice of the correlation coefficient cube as a 2D Correlation Matrix for each scan line of the input image and use this matrix to obtain more reliable disparities. The width of the matrix is the same as the length of the scan line, and the height of the matrix equals the correlation search range, $2w$+1. A typical Correlation Matrix is shown in Fig. 1(b). This image/matrix is actually one slice of the correlation cube obtained in Section 2.3. We will use the correlation matrix to find the disparity for any one scan line. Rather than choosing the maximum correlation coefficient, we find a best path through the correlation matrix. The position of the path indicates the best disparity for this scan line.

The algorithm for finding the best path through the correlation matrix is performed by using a dynamic programming technique [12]. The best path gives the minimum cost when certain constraints are imposed.

Sub-pixel accuracy can be obtained by fitting a second degree curve to the correlation coefficients in the neighbourhood of the disparity and the extrema of the curve can be obtained analytically. This second degree curve can be a parabola.

3.3 Coarse-to-fine Scheme

It has been shown that a multi-resolution or pyramid data structure approach to stereo matching is faster than one without multi-resolution [13], as the search range in each level is small. Besides fast computation, a more accurate disparity map can be obtained by exploiting multi-resolution. The upper levels of the pyramids are ideal to get an overview of the image scene. The details can be found down the pyramid at higher resolution [14].

During the process of projecting the disparity map from the current level of the pyramid to the next (if current level is not level 0), the image size was scaled up by the value of r (reduction ratio), and the disparity value by the same r. The disparity value where the position (i, j) of the new image is not a multiple of r was obtained by linear interpolation.

Our proposed algorithm for stereo matching is:

1. Build pyramids with k levels, and the reduction ratio of r; The upper or coarse resolution levels are obtained by averaging the corresponding $r \times r$ pixels in the previous level;
2. Initialize the disparity map as zero for level k;
3. Perform image matching using the method described in sections 3.1-3.3;
4. If $k \neq 0$, propagate the disparity map to the next level using linear interpolation, set $k = k - 1$ and then go back to step 3; otherwise go to step 5;
5. Smooth the disparity map using a median filter.

4 Experiment Results

This section shows some of the results obtained using the described method. A variety of images have been tested, including Random Dot Stereograms, synthetic images, and different types of real images. The image size does not have to be a power of 2.

Fig. 3 shows the result obtained by applying the algorithm to a pair of Random Dot Stereograms (RDS). Fig. 3(a)(b) show the original left and right RDS, and Fig. 3(c) is the 3D surface obtained showing the different levels of “cakes”. Fig. 4 gives the result of the algorithm running on a synthetic image. Fig. 5 shows the results for two close range real images. Fig. 6 shows the results for two aerial photo images.

The computer used is a Sun SPARC10 running Solaris 2.5. The typical running time for the algorithm on a 256$\times$256 image is in the order of half minute rather than hours. Table 1 gives some of the typical running times of the algorithm on different size of images with different disparities.

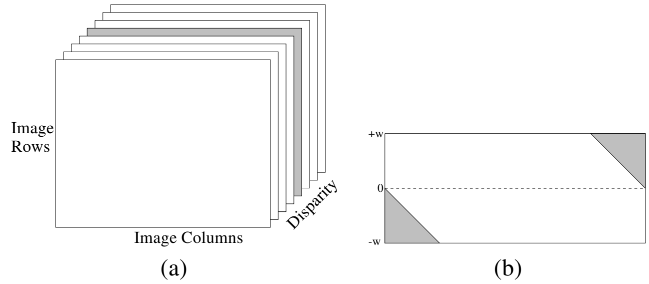


Fig. 1: *(a) An illustration of the correlation matrix. (b) One horizontal slice of the correlation matrix in (a).*

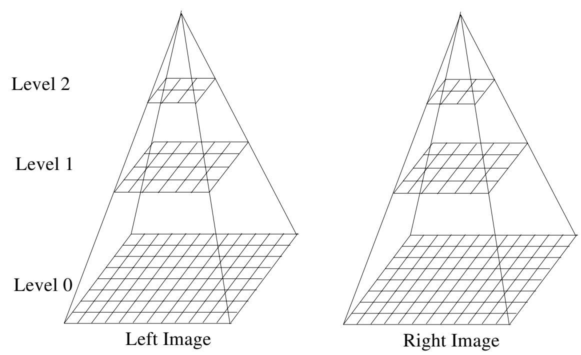
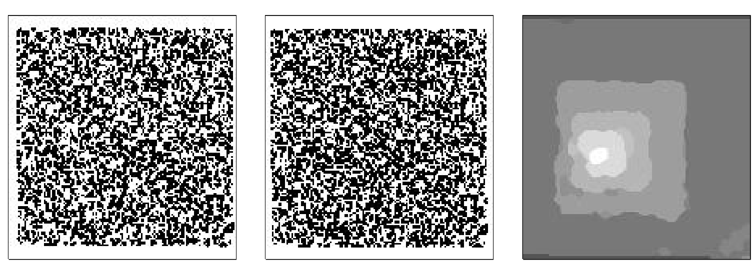
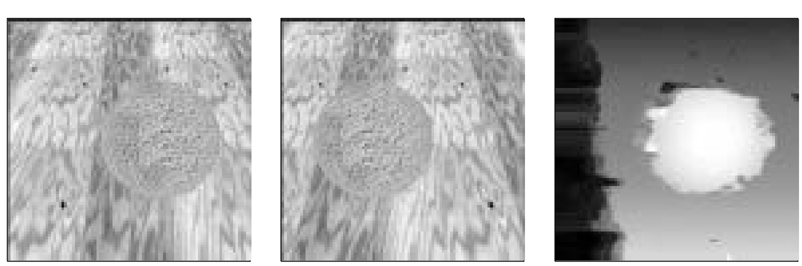


Fig. 2: *Image pyramids construction.*



(a) (b) (c)

Fig. 3: *The matching result for Random Dot Stereograms. (a) left image; (b) right image; and (c) the 3D surface recovered.*



(a) (b) (c)

Fig. 4: *The matching result for synthetic images. (a) left image; (b) right image; and (c) the 3D surface recovered.*

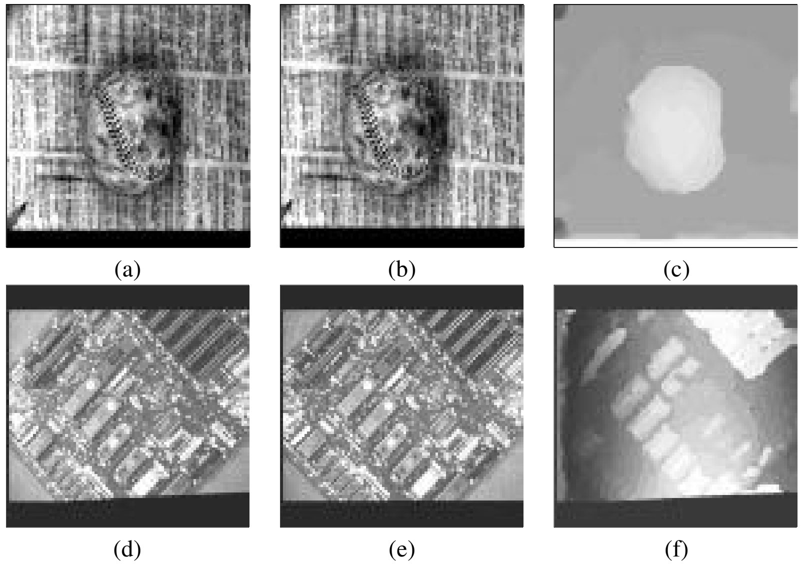


Fig. 5: *The matching result for some close range images. (a,d) left image; (b,e) right image; and (c,f) the 3D surface recovered.*

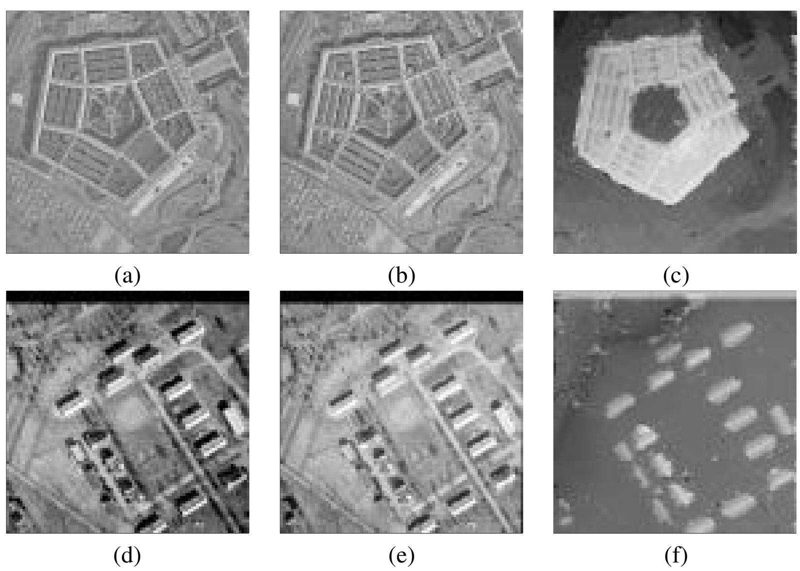


Fig. 6: *The matching result for some aerial photo images. (a,d) left image; (b,e) right image; and (c,f) the 3D surface recovered.*

Table 1: *Running times of the algorithm on different images.*

Image name	**Image size**	**Disparity range**	**User time**	**CPU time**
`x_3d`	285x206	[-15,13]	22.650s	0.510s
`circuit`	512x512	[-29,29]	53.040s	3.900s
`ruts`	512x512	[-69,69]	96.680s	13.540s
`flat`	1000x1000	[-39,27]	206.100s	16.830s

5 Conclusions

We have developed a fast stereo matching method using fast correlation and dynamic programming techniques in the coarse-to-fine framework. The algorithm produces a reliable dense disparity map. The fast cross-correlation was realized by using the box-filtering technique. The time spent in the stage for obtaining the mean and standard deviation for the normalized cross-correlation is almost invariant to the search window size. The typical running time for a 512×512 image is in the order of minutes rather than hours.

By using the zero-mean normalized cross correlation (ZNCC) similarity measure rather than the simple SSD or SAD, the reliability of the algorithm was increased. The algorithm was shown to be reliable by testing on several different types of images.

References

[1] J.-L. Lotti and G. Giraudon, "Adaptive window algorithm for aerial image stereo," in *Proceedings of International Conference on Pattern Recognition*, 1994.

[2] J.-L. Lotti and G. Giraudon, "Correlation algorithm with adaptive window for aerial image in stereo vision," in *European Symposium on Satellite Remote Sensing (EUROPTO)*, (Rome, Italy), pp. 2315–10, September 1994.

[3] S. S. Intille and A. F. Bobick, "Disparity-space images and large occlusion stereo," in *Proceedings of European Conference on Computer Vision*, (Stockholm, Sweden), 1994.

[4] Y. Xiong, D. Wang, and G. Zhang, "Integrated method of stereo matching for computer vision," in *SPIE Proc. Applications of Digital Image Processing XIX* (A. G. Tescher, ed.), (Denver, Colorado), pp. 665–676, 7-9 August 1996.

[5] D. I. Barnea and H. F. Silverman, "A class of algorithms for fast digital image registration," *IEEE Transactions on Computers*, vol. C-21, pp. 179–186, 1972.

[6] C. Konecny and D. Pape, "Correlation techniques and devices," *Photogrammetric Engineering and Remote Sensing*, vol. 47, pp. 323–333, March 1981.

[7] M. Rechsteiner, B. Schneuwly, and G. Troester, "Dynamic workspace monitering," in *International Archives of Photogrammetry and Remote Sensing* (H. Ebner, C. Heipke, and K. Eder, eds.), vol. 30, (Munich, Germany), pp. 689–696, September 1994.

[8] P. Aschwanden and W. Guggenbühl, "Experimental results from a comparative study on correlation-type registration algorithms," in *Robust Computer Vision* (W. Förstner and S. Ruwiedel, eds.), pp. 268–289, Wichmann, 1992.

[9] Q. X. Wu, S. J. McNeill, and D. Pairman, "Fast algorithms for correlation-relaxation technique to determine cloud motion fields?," in *Digital Image Computing: Techniques and Applications*, (Brisbane, Australia), pp. 330–335, 6-8 December 1995.

[10] M. J. McDonnell, "Box-filtering techniques," *Computer Graphics and Image Processing*, vol. 17, pp. 65–70, 1981.

[11] S. D. Cochran and G. Medioni, "3-D surface description from binocular stereo," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 14, pp. 981–994, October 1992.

[12] M. Buckley and J. Yang, "Regularised shortest-path extraction," *Pattern Recognition Letters*, To appear 1997.

[13] K. S. Kumar and U. B. Desai, "New algorithms for 3D surface description from binocular stereo using integration," *Journal of the Franklin Institute*, vol. 331B, no. 5, pp. 531–554, 1994.

[14] F. Ackermann and M. Hahn, "Image pyramids for digital photogrammetry," in *Digital Photogrammetric Systems* (H. Ebner, D. Fritsch, and C. Heipke, eds.), pp. 43–58, Wichmann, 1991.