Cai, Jinhai (2007) Fast Stereo Matching: Coarser to Finer with Selective Updating. In *Proceedings Image and Vision Computing New Zealand 2007*, pages pp. 266-270, Hamilton, New Zealand.

Fast Stereo Matching: Coarser to Finer with Selective Updating

Jinhai Cai

Faculty of Information Technology
Queensland University of Technology
QLD 4001, Brisbane, Australia
Email: j.cai@qut.edu.au

Abstract

A coarse to fine approach for fast stereo matching is presented. In this approach, a dynamic programming (DP) based algorithm at the top of the pyramid is applied to obtain an initial coarse dense disparity map of high quality and to reduce computational cost. In each finer layer, a new dense disparity map is inherited from the coarser layer by interpolation. The new dense disparity map will be updated only in selected areas, instead of the whole map, according to the local matching cost and the depth difference among neighbouring areas. In this way, the proposed approach is able to obtain a smooth dense disparity map and to preserve discontinuity as well. The approach is evaluated using rectified stereo images and good results are achieved in terms of quality and running speed.

Keywords: Stereo matching, coarse to fine approach, dynamic programming, selective updating, fast algorithm, pyramid

1 Introduction

The dense correspondence problem in stereo vision has been actively studied for two decades. The problem is mainly to find a unique mapping of points belonging to two or more images of the same scene. As stereo techniques are able to convert 2D images to 3D structures, they have been applied in many industries for building 3D models of objects in computer graphics and virtual reality or finding relative locations of objects in understanding semantic relationships among objects. One of my current projects needs to conduct a survey on vegetation conditions (mainly the location and height of trees) in a very large scale for Ergon Energy in Queensland, Australia. Ergon Energy supplies electricity and related products and services to regional Queensland with about 10% populations but it covers about 90% areas of Queensland. Therefore, we will have to process a large amount of aerial videos/images within given time, and a fast stereo matching algorithm is desirable.

Many stereo matching algorithms have been developed in the last decade. Scharstein and Szeliski [1] presented taxonomy of dense, two-frame stereo matching methods. They provided an exhaustive comparison of the recently best-performing dense stereo correspondence algorithms. Usually, stereo matching algorithms can be divided into two categories: local (area/window-based) algorithms [2, 3, 4, 5, 6] and global (feature-based) algorithms [7, 8, 9, 10, 11, 12]. For local (area-based) algorithms, a smaller window is desirable to avoid unwanted smoothing. In areas of low texture, however, a larger window is needed so that the window contains enough intensity variation to achieve reliable matching [7]. Usually, area-based algorithms focus on the computation and the aggregation of the matching cost. Alternatively, global algorithms make explicit smoothness assumptions and then solve an optimization problem. In order to prevent over-smoothing in global algorithms, a discontinuity-preserving cost (energy) function is necessary. To minimise the global cost, many methods have been proposed. The framework based on graph cuts [8] can produce excellent dense disparity maps but it involves iterated optimisation processing which needs much higher running times than most other methods. Recently, global optimisation methods based on dynamic programming (DP) [1] are becoming increasing popular in stereo vision due to the highly efficiency of DP. However, it is difficult to enforce the inter-scanline consistency using DP. After all, most global correspondence methods are computationally expensive and sometimes need many parameters that are hard to determine [13].

Despite the demand for fast and reliable stereo matching techniques, there are only few fast algorithms available [2, 3, 4], and almost all of them are area-based. Some of them can produce smooth dense depth maps but without much details, and some produce results with too much noise. Usually, we face the trade-off between smoothness and discontinuity preservation of the estimated dense depth maps. The motivation of this research is to develop an algorithm for fast stereo matching that is able to produce smooth dense depth maps and preserve enough depth discontinuity.

2 Area-Based Matching

The basic concept of area-based stereo matching is to estimate the similarity between two blocks for obtaining a dense disparity map from binocular images. Ideally, the block is large enough to cover sufficient intensity variation so that the similarity estimation is robust to noise. The design of the cost/similarity function is vital to fast stereo matching. The cost/similarity functions should be robust to noise and illumination, and use as little computation as possible. Researchers have designed several cost/similarity functions. Among them, the most popular functions are the sum of absolute differences (SAD) and the sum of squared differences (SSD) due to their simplicity in implementation. However, they are sensitive to differences in illumination and camera gain or bias as illustrated in Figure 1.

Figure 1: Difference caused by camera bias.

In order to deal with camera bias, the zero mean normalised cross correlation (ZNCC) is adopted by many researchers. However, ZNCC is computationally expensive. In this paper, the zero mean sum of absolute differences (ZSAD) is adopted in my experiments as it is insensitive to differences in illumination and camera gain and cheap in computation. For the comparison, these four cost/similarity functions are given in Table 1.

3 Coarse to Fine Scheme

For area-based matching, it is vital to choose a suitable size of window. In general, the window should be large enough to include sufficient intensity variation and small enough to avoid oversmoothing problem. However, a suitable size of the window is application dependent. Even in the same image, the optimal size of the window is different from area to area. Kanade and Okutomi [14] developed a method to select a window size adaptively according to local intensity variation and dense disparity. However, it is not cheap to adaptively select the window size in terms of computational cost. Instead of working on choosing a suitable window size, we propose to use a large window size in the top layer of the pyramid to obtain a smooth initial estimation of the dense disparity map, and then refine the disparity map in the finer layer of the pyramid to recover the depth discontinuity.

3.1 The initial estimation of the disparity map

It is reasonable to start a disparity map obtained from the top layer by winner-take-all. However, the disparity map obtained by winner-take-all may contain significant error in flat areas even the window size is large. Figure 2(a) gives such an example. As the top layer is much smaller than the original image, we are affordable to apply dynamic programming on the cost measure of the top layer to obtain a much better quality estimation of the disparity map. The cost measure is defined as

$$\begin{aligned} C(x,y,d) = \quad & \tfrac{1}{S}\textstyle\sum_{x,y} |(I_1(x,y) - \overline{I}_1(x,y)) \\ & -(I_2(x+d,y) - \overline{I}_2(x+d,y))|, \end{aligned} \tag{1}$$

where $C(x,y,d)$ is the cost measure at position (x,y), d is the disparity, S is the window size, I_1 and I_2 are the intensities of stereo images, and $\overline{I_1}$ and $\overline{I_2}$ are the mean values of intensities. We then apply interpolation on the disparity map of the top layer to obtain the initial estimation of the disparity map for the second top layer.

3.2 Selective updating

At each finer layer, the initial disparity map is obtained from the refined map of the coarser layer using interpolation. In the smooth area, the accuracy of the initial disparity map is usually excellent and updating is unnecessary, so that this

Table 1: Area-based matching metrics.

SAD	$\sum_{x,y} \lvert I_1(x,y) - I_2(x+d,y) \rvert$
SSD	$\sum_{x,y} (I_1(x,y) - I_2(x+d,y))^2$
ZNCC	$\dfrac{\sum_{x,y}(I_1(x,y)-\overline{I}_1(x,y)) * (I_2(x+d,y)-\overline{I}_2(x+d,y))}{\sqrt{\sum_{x,y}(I_1(x,y)-\overline{I}_1(x,y))^2 * \sum_{x,y}(I_2(x+d,y)-\overline{I}_2(x+d,y))^2}}$
ZSAD	$\sum_{x,y} \lvert (I_1(x,y)-\overline{I}_1(x,y)) - (I_2(x+d,y)-\overline{I}_2(x+d,y)) \rvert$

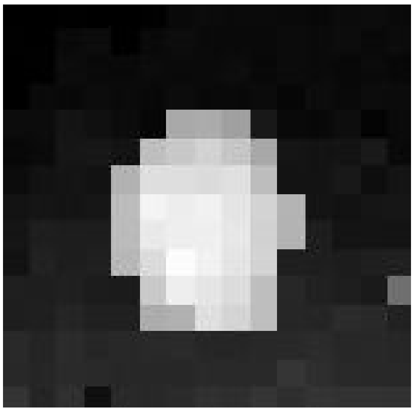
scheme is faster than many other stereo matching algorithms. If the disparity difference among neighbouring positions exceeds a threshold value or the cost measure at the position for the estimated disparity is too large, the disparity at that position will be updated. The criteria for updating disparity are given as follows.

$$|d(x,y) - d(x-1,y)| > \mu, \tag{2}$$
$$|d(x,y) - d(x,y-1)| > \mu, \tag{3}$$
$$C(x,y,d) > \nu, \tag{4}$$

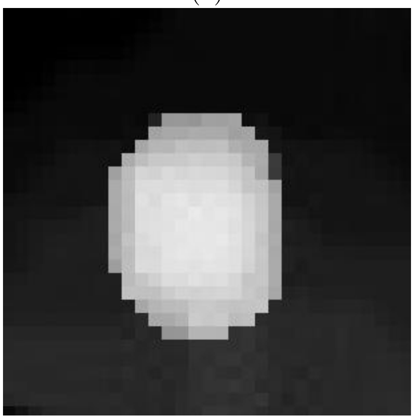
where μ and ν are thresholds, and $d(x,y)$ is the disparity at position (x,y). The updating is a local process in which the continuity with two causal neighbours only is under the consideration. The updating process is to minimise the $E_{smooth}(x,y,d)$:

$$E_{smooth}(x,y,d) = C(x,y,d) * min(|d(x,y) - d(x-1,y)| + |d(x,y) - d(x,y-1)|). \tag{5}$$

Figure 2(b) gives an example of the refined disparity map of the second top layer. The updating process will repeat few times utill the bottom layer is reached.



(a)



(b)

Figure 2: The disparity map: coarse to fine. (a) The disparity map obtained from the top layer by winner-take-all. (b) The disparity map of the second layer obtained by interpolation and refinement.

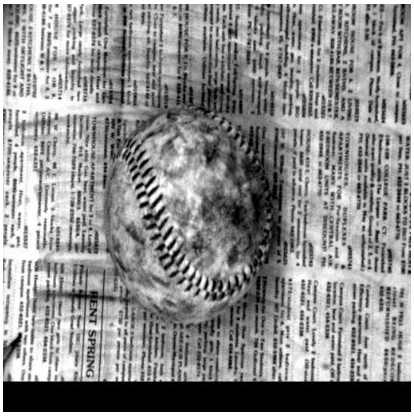
4 Experimental Results

The proposed fast stereo matching algorithm has been evaluated using several real stereo image pairs.

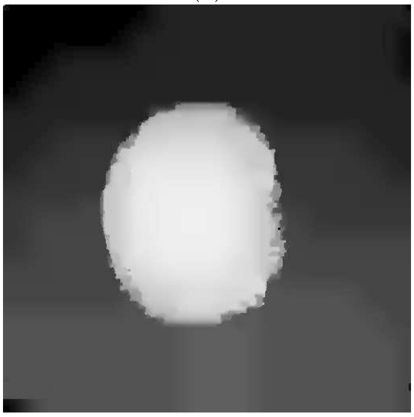
Baseball: In the first example, a baseball is on a quite flat surface as shown in Figure 3(a). Comparing to the belief propagation (BP) method, the proposed algorithm produces a much flat background that is more close to the ground truth. The estimated disparity surface of the baseball is also smooth but enough details are preserved by the proposed algorithm. Whereas, BP method produces a more accuracy disparity surface of the ball but the background is noisy. As the whole images contain large flat areas, the proposed algorithm can obtain the disparity map for these areas using interpolation only.

Tsukuba: This example has been widely used to evaluate stereo matching algorithm as it contains objects in difference depth and objects in different shapes. The proposed algorithm produces a disparity map of good quality shown in Figure **??**. This example proves that the proposed algorithm is able to recover details from coarse initial estimation.

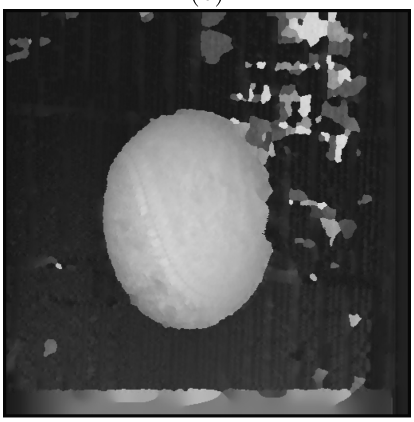
Road: This is a real work example and the original images are shown in Figure 1. The images contain trees, road, and buildings, etc. The illumination



(a)



(b)



(c)

Figure 3: The estimation of the disparity map: (a) Left image of the baseball. (b) The disparity map obtained by the proposed method. (c) The disparity map obtained by the BP method.

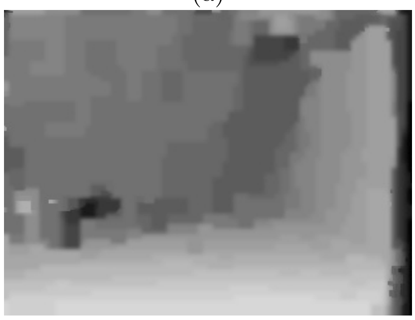
difference between the two images is also significant. Both disparity maps are obtained by the proposed algorithm with same conditions except the cost function. The disparity map as shown in Figure 5(a) is obtained using SAD-based cost function. Clearly the SAD-based cost function catches the repeated pattern of the road instead of the true dense disparity. While, ZSAD-based cost function is able to detect the true dense disparity, therefore it is robust to camera gain or bias.

(a)

(b)

Figure 4: The estimation of the disparity map: (a) Left image of the Tsukuba. (b) The disparity map obtained by the proposed algorithm.

(a)



(b)

Figure 5: The estimation of the disparity map: (a) SAD-based cost function. (b) ZSAD-based cost function.

5 Conclusion

This paper has presented a new coarse to fine approach for fast stereo matching. To get a good initial estimation of dense disparity map for finer layers, the dynamic programming (DP) is applied on the top layer. In each finer layer, a new dense disparity map is inherited from the coarser layer by interpolation. The new dense disparity map will be updated only in selected areas according to the local matching cost and the depth difference among neighbouring areas. The approach is evaluated using rectified stereo images and good results are achieved in terms of quality and running speed. Experimental results show that the ZSAD-based cost function is robust to illumination and camera gain or bias. The results also prove that the updating scheme is able to restore the discontinuity from a coarse disparity map.

References

[1] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *Internation Journal of Computer Vision*, vol. 47, no. 1/2/3, pp. 7–42, April-June 2002.

[2] L. D. Stefano, M. Marchionni, S. Mattoccia, and G. Neri, "A fast area-based stereo matching algorithm," 2002.

[3] C. Sun, "Fast stereo matching using rectangular subregioning and 3d maximum-surface techniques," *International Journal of Computer Vision*, vol. 47, no. 1/2/3, pp. 99–117, May 2002.

[4] J. Banks, M. Bennamoun, and P. Corke, "Non-Parametric Techniques for Fast and Robust Stereo Matching," in *Proceedings of IEEE TENCON - Speech and Image Technologies for Computing and Telecommunications*, 1997, pp. 365–368.

[5] H. H. B. R. C. Bolles and M. J. Hannah, "The JISCT Stereo Evaluation," in *Proceedings of DARPA Image Understanding Workshop*, 1993, pp. 263–274.

[6] R. Zabih and J. Woodfill, "Non-Parametric Local Transforms for Computing Visual Correspondence," in *Proceedings of ECCV*, 1994, pp. 151–158.

[7] C. L. Zitnick and T. Kanade, "A cooperative algorithm for stereo matching and occlusion detection," *IEEE Trans. on pattern analysis and machine intelligence*, vol. 22, no. 7, pp. 675–684, July 2000.

[8] O. V. Yuri Boykov and R. Zabih, "Fast approximate energy minimization via graph cuts," *IEEE Trans. on pattern analysis and machine intelligence*, vol. 23, no. 11, pp. 1222–1239, November 2001.

[9] D. Terzopoulos, "Regularization of inverse visual problems involving discontinuities," *IEEE Trans. on pattern analysis and machine intelligence*, vol. 8, no. 4, pp. 413–424, April 1986.

[10] P. N. Belhumeur, "A bayesian approach to binocular stereopsis," *International Journal of Computer Vision*, vol. 19, no. 3, pp. 237–260, 1996.

[11] S. Birchfield and C. Tomasi, "A pixel dissimilarity measure that is insensitive to image sampling," *IEEE Trans. on pattern analysis and machine intelligence*, vol. 20, no. 4, pp. 401–406, April 1998.

[12] S. B. R. I. J. Cox, S. L. Hingorani and B. M. Maggs, "A maximum likelihood stereo algorithm," *CVIU*, vol. 63, no. 3, pp. 542–567, 1996.

[13] I. S. K. Kuk-Jin Yoon, "Adaptive support-weight approach for correspondence search," *IEEE Trans. on pattern analysis and machine intelligence*, vol. 28, no. 4, pp. 650–656, April 2006.

[14] T. Kanade and M. Okutomi, "A stereo matching algorithm with an adaptive window: Theory and experiment," *IEEE Trans. on pattern analysis and machine intelligence*, vol. 16, no. 9, pp. 4920–932, September 1994.