A Dense Stereo Matching Algorithm with Occlusion and Less or Similar Texture Handling

Hehua Ju and Chao Liang

College of Electronic Information and Control Engineering,
Beijing University of Technology, Beijing 100124, P.R. China
lc_s2009@emails.bjut.edu.cn

Abstract. Due to image noise, illumination and occlusion, to get an accurate and dense disparity with stereo matching is still a challenge. In this paper, a new dense stereo matching algorithm is proposed. The proposed algorithm first use cross-based regions to compute an initial disparity map which can deal with regions with less or similar texture. Secondly, the improved hierarchical belief propagation scheme is employed to optimize the initial disparity map. Then the left-right consistency check and mean-shift algorithm are used to handle occlusions. Finally, a local high-confidence strategy is used to refine the disparity map. Experiments with the Middlebury dataset validate the proposed algorithm.

1 Introduction

Considering the great advances in digital camera technology coupled with the convenience, compactness and relatively low costs, stereo vision is widely used. Stereo vision can reconstruct the 3D geometry of a scene from two or more images of the same scene taken from different views, and then perform localization, mapping and measurement. The core of the stereo vision is the stereo matching, which is to find a unique mapping between the points belonging to two images of the same scene. Generally the camera geometry is known and the images can be rectified, then the stereo matching reduces to a simplified problem, where points in one image can correspond only to points along the same scanline in the other image. However, due to illumination, image noise and occlusion, the stereo matching is still a great challenge.

There exists lots of works on the stereo matching. Scharstein and Szeliski [1] provide an exhaustive comparison of stereo matching algorithms. Generally, these algorithms can be roughly classified into two categories: the local algorithm and the global algorithm. The local algorithms usually utilize local measurements such as image intensity, and aggregate information from multiple around pixels. The simplest method is to minimize the matching error within rectangular windows of fixed size [2]. The accuracy of these methods depends on the size and shape of the window. Some modified algorithms are also proposed, which utilize multiple windows [3, 4] or adaptive windows [5] to minimize the error and give performance improvements. But occlusion and region with less or similar texture are still open problem for local algorithm.

Global algorithms for the stereo matching problem define a global cost function by making some assumptions on the disparity map. The global cost function generally

M. Kurosu (Ed.): Human-Computer Interaction, Part V, HCII 2013, LNCS 8008, pp. 171–177, 2013.

includes terms for local property matching, smoothness terms, and in some cases, penalties terms for occlusions. Then the most efficient minimization scheme is chosen according to the form of the cost function. Finally taking the results of the local algorithm as initial values, the global cost function can be solved. One typical global algorithm is based on dynamic programming techniques [6]. It can obtain good results at occlusion regions, while easily suffers from "streaking" artifacts due to its inconsistency between scanlines. The other popular global algorithms, such as the belief propagation (BP) [7] and Graph Cut [8], can improve the accuracy since they perform the minimization in two dimensions. However since the cost function is a highly nonlinear function, it is hard to obtain the global optimal solution. Additionally, these methods are computationally expensive.

In this paper, a new stereo matching algorithm is presented. The proposed algorithm first use cross-based regions to compute an initial disparity map which can deal with regions with less or similar texture. Secondly, the improved hierarchical belief propagation scheme is employed to optimize the initial result, where a stability factor is introduced into the hierarchical belief propagation scheme to balance the ratio of the data term and smoothness term in BP technique for every pixel. Then, the left-right consistency check and mean-shift are applied to handle occlusion pixels. Finally, a local high-confidence strategy is used to refine the obtained disparity map. Experiments with the Middlebury dataset validate the proposed algorithm.

The paper is organized as follows. Section 2 presents the computation of the initial disparity map with cross-based regions to handling with the less or similar texture region. Section 3 introduces the global optimization for the initial disparity map with the improved hierarchical BP. The proposed occlusion handling and disparity refinement are introduced in Section 4. Experimental results are reported in Section 5. Section 6 concludes this paper.

2 Initial Disparity Map with Cross-Based Regions

For the global stereo matching algorithm, the selection of the initial disparity map generally plays an important role. The accuracy of the local algorithms varies with the size and shape of the matching window. In region low texture region, the matching window should be large enough, while in region with similar texture the matching window should be small enough. So in this paper, we use an adaptive window determined by the cross-based scheme to compute the initial disparity map.

It is reasonable to assume that pixels with similar colors are in the same plane and have the same disparity. According to the cross-based scheme [9], the matching window of a pixel $\mathrm{P} = (\mathrm{x_p}, \mathrm{y_p})$ can be determined by two orthogonal segmentations: the vertical segmentation $\mathrm{V(P)}$ and the horizontal segmentation $\mathrm{H(P)}$

$$\begin{cases} H(p) = \left\{(x,y) \mid x \in \left[x_p - x_p^l, x_p - x_p^r\right], y \in y_p\right\} \\ V(p) = \left\{(x,y) \mid x \in x_p, y \in \left[y_p - y_p^l, y_p - y_p^r\right]\right\} \end{cases} \tag{1}$$

The details of determine the matching window are as follows. First take P as the start point to compute the vertical length of the matching window. Compare the color difference between the point p_1 and the point $p_2 = (x_p - i, y_p)$ with the threshold (τ) for $i \in [1, r]$, where r is related to the maximum window radius.

$$S(p_1, p_2) = \begin{cases} 1, & \max_{c \in \{R,G.B\}} (|I_c(p_1) - I_c(p_2)| \leq \tau) \\ 0, & otherwise \end{cases} \tag{2}$$

when $S(p_1 - p_2) = 1$, the length of vertical length of the matching window V_p is added by 1. So the vertical length of the window can be computed by performing this operation in up and down direction. Secondly, determine the matching window by the horizontal segmentation. Take each point on the determined vertical segment as start to perform segmentation in left and right direction along the corresponding horizontal line using the strategy similar to equation (2). The obtained matching window $U(p)$ can be regarded as a set of horizontal segments $H(q)$.

$$U(p) = \bigcup_{q \in V(p)} H(q) \tag{3}$$

With the truncation model

$$e(x, y, d) = \min\left(\sum_{c \in \{R,G,B\}} \left| I_c^l(x, y) - I_c^r(x-d, y) \right|, T\right) \tag{4}$$

where T is a threshold for truncation, the cost function can be defined as

$$C(x, y, d) = \frac{1}{\left\| U_d(x, y) \right\|} \sum_{(x,y) \in U_d(x,y)} e(x, y, d) \tag{5}$$

where $\|U_d(x, y)\|$ is the number of pixels in both matching windows and

$$U_d(x, y) = \{(x, y) \mid (x, y) \in U_l(x, y), (x-d, y) \in U_r(x, y)\} \tag{6}$$

U_l and U_r is the matching window in left and right image respectively.

To speed up the matching, the orthogonal integral image technique is used to construct the cost function. It needs the following 4 steps to construct the cost function.

- Step 1. Compute the horizontal integral, $S^H(x, y)$, of the line corresponding to an interest point with cost function $e_d(x, y)$.

$$S^H(x, y) = \sum_{0 \leq m \leq x} e_d(m, y) = S^H(x-1, y) + e_d(x, y) \tag{7}$$

where $S^H(-1, y)$=0.

- Step 2. Compute the value of the cost function corresponding each line in the matching window.

$$E_d^H(x, y) = S^H(x + h_e, y) - S^H(x - h_s - 1, y) \tag{8}$$

where h_e and h_s denotes the end and start of each horizontal segment respectively.

- Step 3. Compute the vertical integral, $S^V(x,y)$

$$S^V(x,y) = \sum_{0\le n\le y} E_d^H(x,n) = S^V(x,y-1) + E_d^H(x,y) \tag{9}$$

- Step 4. Obtain the cost function

$$E_d(x,y) = S^V(x,y+v_e) - S^V(x,y-v_s-1) \tag{10}$$

Finally, with the winner-take-all strategy, a disparity map can be obtained.

3 Global Optimization with the Improved Hierarchical BP

To refine the results, we apply an improved hierarchical BP algorithm to optimize the initial disparity map. Generally, the cost function $\mathrm{E(p,d)}$ for the BP algorithm has the following form [7]

$$\mathrm{E(p,d)} = \mathrm{E}_d(\mathrm{p,d}) + \mathrm{E}_s(\mathrm{p,d}) \tag{11}$$

where $\mathrm{E}_d(\mathrm{p,d})$ is the data term and $\mathrm{E}_s(\mathrm{p,d})$ is the smoothness term. Noting that in some regions, the smoothness does not need. So an efficient weighting the rate of the smoothness is introduced, and the cost function of the BP is modified

$$\mathrm{E(p,d)} = \mathrm{C}_f(\mathrm{p}) \cdot \mathrm{E}_d(\mathrm{p,d}) + \frac{1}{\mathrm{C}_f(\mathrm{p})+0.5} \cdot \mathrm{E}_s(\mathrm{p,d}) \tag{12}$$

where $\mathrm{C}_f(\mathrm{p})$ is the stability factor is defined by the best disparity $\mathrm{C}_b(\mathrm{x,y})$ and the second best disparity $\mathrm{C}_s(\mathrm{x,y})$

$$\mathrm{C}_f(\mathrm{p}) = \frac{|\mathrm{C}_b(\mathrm{x,y}) - \mathrm{C}_s(\mathrm{x,y})|}{\mathrm{C}_b(\mathrm{x,y})} \tag{13}$$

In this paper, the data term takes the cost function defined in section 2. The smoothness term has the following form

$$\mathrm{E}_s^i(\mathrm{p,d}) = \sum_{\mathrm{q}\in\mathrm{N(p)}} \mathrm{M}_{q\to p}^i(\mathrm{d}) \tag{14}$$

where i is the iteration number, $\mathrm{M}_{q\to p}^i(\mathrm{d})$ denotes the message between the pixel p and its neighbor $\mathrm{q} \in \mathrm{N(p)}$, and the jump cost, $\mathrm{E}_j(\mathrm{d}_p,\mathrm{d})$, expresses the discontinuity

$$M_{q\to p}^i(d) = \arg\min_{d_q}(E_d(p,d_q) + \sum_{s\in N(q), s\notin p} M_{q\to p}^{i-1}(d_q) + E_j(d_q,d)) \tag{15}$$

$$E_j(d_q,d) = \min(\theta_{bp}, \rho\,|\,d_q - d\,|) \tag{16}$$

where θ_{bp} is the truncation value to limit the increase of the cost.

4 Disparity Map Refinement

In real applications, occlusion inevitably occurs. The occlusion results in mismatches in the obtained disparity map [1]. To handle this problem, the left-right consistency check is performed to detect the occluded points.

$$\mathrm{D}_l(\mathrm{x},\mathrm{y}) = \mathrm{D}_r(\mathrm{x} - \mathrm{D}_l(\mathrm{x},\mathrm{y}),\mathrm{y}) \tag{17}$$

where D_l, D_r is the left and right disparity respectively. Denote the pixel which does no satisfy equation (17) as the occluded points. Once the occluded points are detected, use the mean-shift algorithm [10] to segment the left image into different regions. Then according to segment, assign each occluded points the disparity which corresponds to its segment region.

To further improve the performance, the local high-confidence voting technique [11] can be used to refine the obtained disparity map. At first, smooth the image with the 3×3 median filter. Then for the left image, compute cross-based region image and calculate the histogram of disparity values in each region. Take the largest value of the histogram as the disparity of this region. Finally, filter the final disparity map with a median filter to remove the noise.

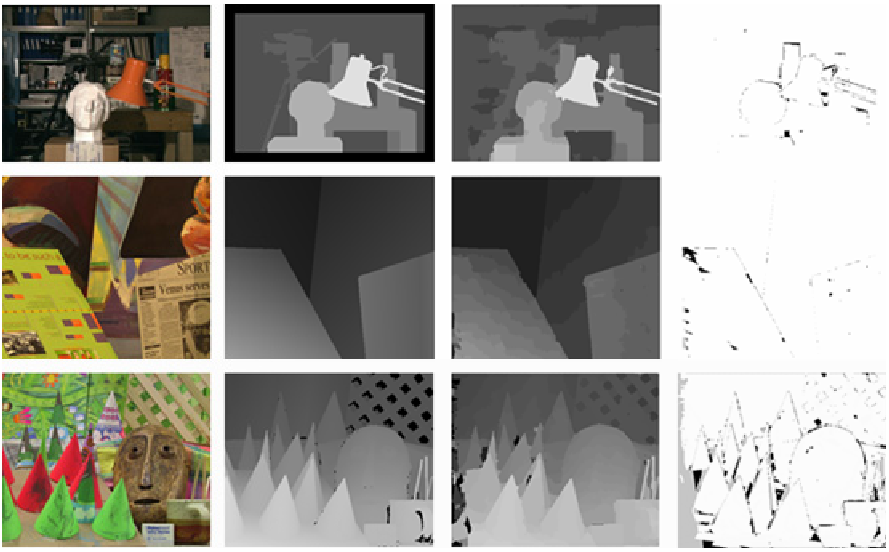
5 Experiments

The proposed algorithm is validated with the Middlebury dataset [1]. Parameters in the proposed algorithm are set as follows: the maximum window radius $r = 17$, the threshold $\tau = 20$, the truncation value $T = \max\_d + 5$, $\theta_{bp} = 2\max\_d/16$, and $\rho = 1.0$, where $\max\_d$ is the largest disparity of the stereo images pair.

To make comparisons, the VariableCross [9] and the RealTimeBP [7] algorithms are also performed. Results are reported in Table 1. The measurement is computed for three subsets of an image. They are "nonocc": the subset of non-occluded pixels, "all": pixels that are either non-occluded or half-occluded, and "disc": pixels near the occluded areas [1]. It is obviously that the proposed algorithm outperforms the VariableCross and the RealTimeBP. Figure 1 shows results with the proposed algorithm. The proposed algorithm is also applied to an outdoor scene. Figure 2 shows experimental results of one pair of images. We can see for the unstructured outdoor, the proposed algorithm can also obtain good results.

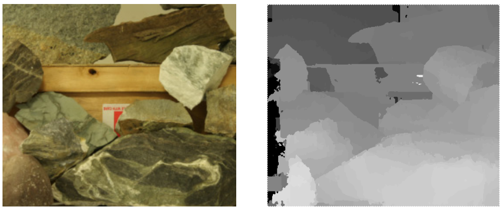
Table 1. Experimental results

Algorithm	Tsukuba			Venus			cones		
	nocc	all	disc	nocc	all	disc	nocc	all	disc
The proposed method	1.24	1.60	6.58	0.35	0.53	3.09	4.64	11.6	10.9
VariableCross	1.99	2.65	6.77	0.62	0.96	3.20	6.28	12.7	12.9
RealTimeBP	1.49	3.40	7.87	0.77	1.90	9.00	4.61	11.6	12.4



(a) left image (b) ground truth (c) obtained disparity (d) disparity errors

Fig. 1. Experimental results for Middlebury data set



(a) left image (b) obtained disparity

Fig. 2. Experimental results for a pair of images of outdoor scene

6 Conclusions

In this paper, a new dense stereo matching algorithm is proposed. The proposed algorithm first use cross-based regions to compute an initial disparity map which can deal with regions with less or similar texture. Secondly, the improved hierarchical belief propagation scheme is employed to optimize the initial disparity map. Then the left-right consistency check and mean-shift algorithm are used to handle occlusions. Finally, a local high-confidence strategy is used to refine the disparity map. Experiments with the Middlebury dataset validate the proposed algorithm.

References

1. Scharstein, D., Szeliski, R.: A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. Int'l Journal of Computer Vision 47(1), 7–42 (2002)
2. Okutomi, M., Kanade, T.: A multiple baseline stereo. IEEE Trans. PAMI 15(4), 353–363 (1993)
3. Geiger, D., Ladendorf, B., Yuille, A.: Occlusions and binocular stereo. In: Sandini, G. (ed.) ECCV 1992. LNCS, vol. 588, pp. 425–433. Springer, Heidelberg (1992)
4. Fusiello, A., Roberto, V., Trucco, E.: Efficient stereo with multiple windowing. In: Proceedings of IEEE CVPR, pp. 858–863 (1997)
5. Kanade, T., Okutomi, M.: A stereo matching algorithm with an adaptive window: theory and experiment. IEEE Trans. PAMI 16(9), 920–932 (1994)
6. Salmen, J., Schlipsing, M., Edelbrunner, J., Hegemann, S., Lüke, S.: Real-time stereo vision: Making more out of dynamic programming. In: Jiang, X., Petkov, N. (eds.) CAIP 2009. LNCS, vol. 5702, pp. 1096–1103. Springer, Heidelberg (2009)
7. Yang, Q.X., Wang, L., et al.: Real-time global stereo matching using hierarchical belief propagation. In: Proceedings of BMVC (2006)
8. Boykov, Y., Veksler, O., Zabih, R.: Fast approximate energy minimization via graph cuts. IEEE Trans. PAMI 23(11), 1222–1239 (2001)
9. Zhang, K., Lu, J., Lafruit, G.: Cross-based local stereo matching using orthogonal integral images. IEEE Trans. CSVT 19(7), 1073–1079 (2009)
10. Comaniciu, D., Meer, P.: Mean-shift: A robust approach toward feature space analysis. IEEE Trans. PAMI 24(5), 603–619 (2002)
11. Lu, J., Lafruit, G., Catthoor, F.: Anisotropic local high-confidence voting for accurate stereo correspondence. In: Proceedings of of the SPIE, vol. 6812, pp. 68120J–68120J-12 (January 2008)