3rd International Conference on 3D Systems and Applications
Seoul, Korea, 20~22 June 2011.

Trilateral-Filter-Based Depth Interpolation for Occlusion Handling in Stereo Vision

Yen-Chieh Lai, Chung-Te Li, Chien Wu, Chao-Chung Cheng, and Liang-Gee Chen
Graduate Institute of Electronics Engineering, National Taiwan University

Abstract

In this paper, we propose a trilateral-filter-based depth post processing to handle the occlusion points in order to enhance the quality of depth information and overcome the constraints in stereo vision. According to the experimental result, the depth quality is greatly enhanced especially in the occlusion regions and object boundaries. Proposed algorithm can provide high quality depth information from stereo image pair for the application, such as multi-view 3DTV.

Keywords--- Stereo vision, occlusion handling, depth interpolation.

1. Introduction

Stereo vision is a popular issue in computer vision. We can apply stereo matching technique to acquire the depth information from a stereo image pair for the application, such as multi-view 3DTV. In Fig.1, we show the flow of the depth generation application in multi-view 3DTV. In computer vision and image processing, the stereo matching is regarded as the problem of finding an optimal disparity vector assignment for every pixel. Finding an optimal assignment can be formulated in the Markov random field (MRF) framework as an energy minimization problem. The energy minimization problem can be defined as:

$$\{l_p\} = \arg\min \left\{ \sum_{p\in P} E_d(l_p) + \sum_{(p,q)\in G} E_s(l_p, l_q) \right\} \qquad (1)$$

where P is the set of all nodes and G is the specified neighborhood system. We assign the optimal label set $\{l_p\}$ to minimize the energy function. The energy function has two terms: a data term E_d and a pairwise smoothness term E_s. E_d is the cost to penalize the inconsistence between the disparity value and the observed data. E_s is the discontinuity cost between the neighborhood nodes. Several efficient global optimization techniques such as graph cuts [1] and belief propagation [2] [3](BP) can find the stronger optimal solution than local methods.

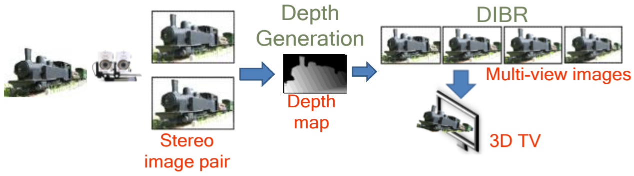


Fig. 1 The flow of generating multiple-view video from stereo video.

However, there are still some basic constraints and problems in stereo image pairs, such as stereo vision occlusion, two-camera setting parameter deviation, etc.

Therefore, if we only consider the matching correspondence between the disparity value and the observed data, some pixels are probably induced to the unreliable depth information due to the interference of occlusion and noise. If we aspire the higher quality of depth information from stereo images, we need overcome the basic constraints.

Therefore, we proposed a trilateral-filter-based depth interpolation algorithm to provide a depth information post-processing algorithm to handle the occlusion regions and revise the unreliable depth information for the better performance.

The rest of the paper is organized as follows. In Section 2, we present the proposed trilateral-filter-based depth interpolation process. In Section 3, we perform the proposed message passing operation to the stereo matching. Finally, we conclude this paper in Section 4.

2. Proposed algorithm

In this section, we introduce the proposed trilateral-filter-based depth interpolation algorithm. Our post-processing can be applied after any depth generation algorithm. For any initial depth information from stereo depth generation algorithm, we can apply our depth post-processing to provide the better depth performance. Our proposed algorithm can be composed of two steps: (1) Detecting the location where the unreliable points happen in the initial depth map and output one reliable map. (2) Trilateral-filter-based depth interpolation.

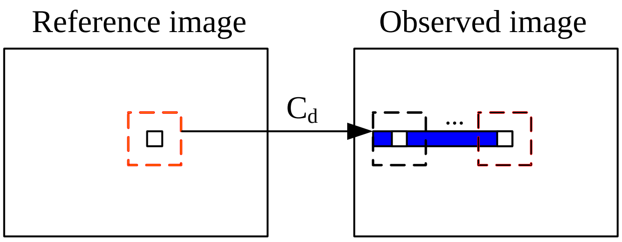
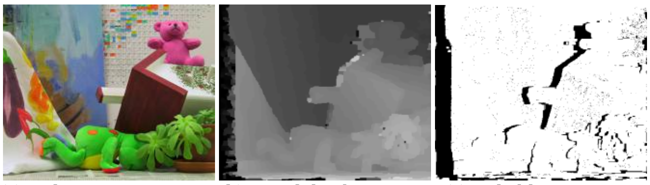


Fig. 2 The local matching correspondence process between reference image and observed image.



(a) Color image (b) Initial depth map (c) Reliable map

Fig. 3 The example of reliable map: In (c), the black points are the unreliable points.

2.1 Reliable map generation

First, for the initial depth information from stereo images, we utilize two-directional occlusion detection techniques[4], such as Left-Right-checking (LRC) to explore the occlusion points location. Left-Right-checking (LRC) can be regarded as the double-checking of the consistence in depth information between the stereo images. The detailed definition is in the following:

$$\text{Error} = x_R - \left(x'_L + d^L_{x'_L}\right) \qquad (2)$$

$$x'_L = x_R + d^R_{x_R} \qquad (3)$$

where x_R is the horizontal position in the reference image, $d^R_{x_R}$ is the initial disparity value in x_R.

If the error is over the threshold, the occlusion points happen, and we regards the occlusion points as the unreliable points in the reliable map.

Moreover, we check the local correspondence value between the disparity value and the observed data with the initial disparity *d*, if the value C_d is higher than the threshold, we also regard it as the unreliable point. In Fig.2, we show the the local correspondence computation. We estimate a reliable map *O(x,y)* for every pixel. For unreliable point, we assign 0 to *O(x,y)*, and the others are assigned to 1. We show the reliable map in Fig. 3 to indicate where the unreliable points happen.

2.2 Trilateral-filter-based depth interpolation

Finally, we apply the trilateral-filter to interpolate the revised depth value for unreliable points. We compute the revised depth value based on the spatial and luminance proximity. Besides, we exclude the unreliable point depth information by considering the reliable map.

where *R(x,y)* is the revised depth value for unreliable points, *S* is the neighborhood for the unreliable pixel *(x,y)*, *D(x,y)* is the initial depth value, *I(x,y)* is the luminance value, and $B(x,y,x_j,y_j)$ is the spatial and luminance proximity function.

$$R(x,y) \triangleq \frac{\sum_{(x_j,y_j)\in S} O(x,y)B(x,y,x_j,y_j)D(x,y)}{\sum_{(x_j,y_j)\in S} O(x,y)B(x,y,x_j,y_j)}. \qquad (4)$$

$$B(x,y,x_j,y_j) \triangleq \exp\left(-\frac{|(x_j-x)|^2}{2\sigma_x^2} - \frac{|(y_j-y)|^2}{2\sigma_y^2} - \frac{|I(x_j,y_j)-I(x,y)|^2}{2\sigma_i^2}\right). \qquad (5)$$

3. Experimental Results

In this section, we present our experimental result and compared with the unhandled image. We use datasets from the Middlebury vision website[5].

Our proposed algorithm not only revise the unreliable depth information based on the spatial and luminance proximity in the neighborhood regions but also exclude the interference by the unreliable points.

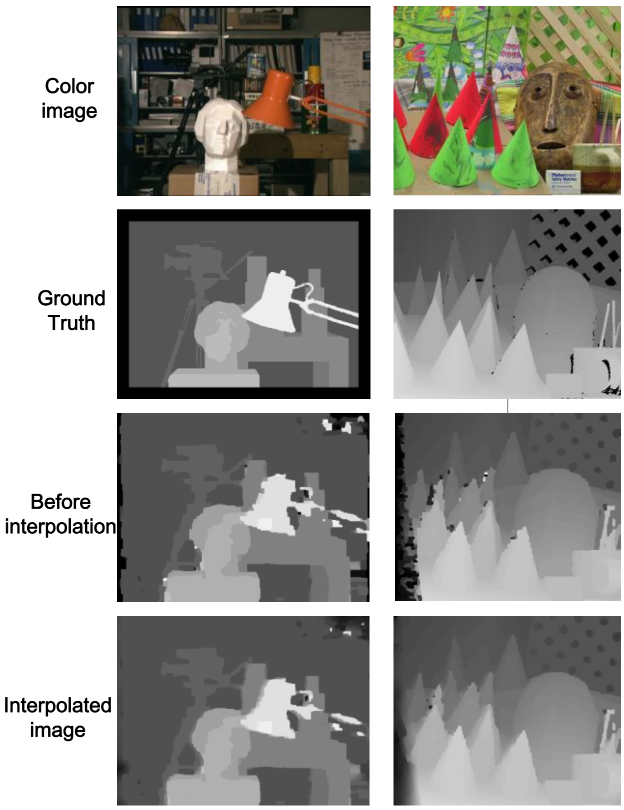


Fig. 4 The experimental result of Tsukuba and Cones

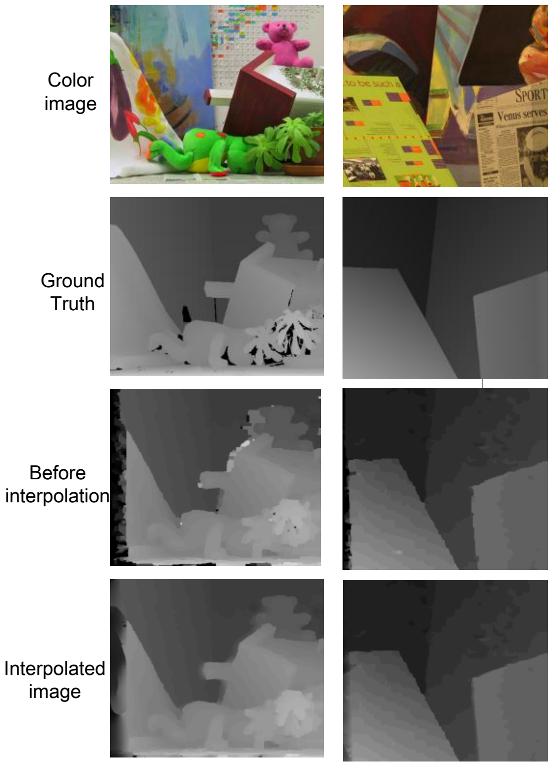


Fig. 5 The experimental result of Teddy and Venus

Compared to ground truth (PSNR)	Before interpolation	Interpolated
Avg.	27.95	30.08

Table. 1 The PSNR value compared to the ground truth

We evaluate the performance by the PSNR between the experimental result and ground truth depth map. We set the correspond error threshold as 20 in the Census matching transform[6] with matching block size 5x5. And we perform the trilateral-filter-based depth interpolation within the 9x4 region. In Fig. 4, 5, we show that the experimental results and the unhandled results.

We show the PSNR value compared to the ground truth in Table. 1. Our algorithm can provide 2 dB performance enhancement compared to the unhandled image.

4. Conclusions

We propose a trilateral-filter-based depth post processing to handle the occlusion points in order to enhance the quality of depth information and overcome the constraints in stereo vision. According to the experimental result, the depth quality is greatly enhanced especially in the occlusion regions and object boundaries. Proposed algorithm can provide high quality depth information from stereo image pair for the application, such as multi-view 3DTV.

5. References

[1] Y. Boykov, O. Veksler, and R. Zabih, “Fast approximate energy minimization via graph cuts,” in *IEEE transactions on Pattern Analysis and Machine Intelligence* , vol. 23, no. 11, pp. 1222-1239, 2001.

[2] W. T. Freeman, E. C. Pasztor, and O. T. Carmichael, “Learning the low level vision,” in *International Journal of Computer Vision* , vol. 70, no. 1, pp. 41-54, 2000.

[3] P. F. Felzenszwalb and D. R. Huttenlocher, “Efficient belief propagation for early vision,” *International Journal of Computer Vision*, vol. 70, no. 1, October 2006.W. T. Freeman, E. C. Pasztor, and O. T. Carmichael, “Learning the low level vision,” in *International Journal of Computer Vision* , vol. 70, no. 1, pp. 41-54, 2000.

[4] Egnal, G.; Wildes, R.P.; "Detecting binocular half-occlusions: empirical comparisons of five approaches," in *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol.24, no.8, pp. 1127-1133,2002.

[5] The Middlebury computer vision pages. http://vision.middlebury.edu/

[6] R. Zabih and J. Woodfill, “Non-parametric local transforms for computing visual correspondence,” in *Proc. ECCV*, pp. 150-158, 1994.