Depth Map Estimation from a Single Video Sequence

Tien-Ying Kuo, Cheng-Hong Hsieh, and Yi-Chung Lo
Department of Electrical Engineering, National Taipei University of Technology, Taiwan, R.O.C.
tykuo@ee.ntut.edu.tw , chhsieh@image.ee.ntut.edu.tw , yclo@image.ee.ntut.edu.tw

Summary —**The goal of this paper is to develop a robust depth estimation method from a single-view video sequence. We utilize an estimated initial depth to establish a reference depth for further obtaining the reliable depth information, and then it is refined with a temporal-spatial filter. At first, we use adaptive support-weight block matching to extract disparity information from consecutive video frames. The disparity is compensated with the camera motion and then transformed to the initially estimated depth. Based on the initial depth, two kinds of depth maps, the propagation depth and the optical flow depth can be established. Finally, these three depth maps are fused together by using voting merger, and then applied with the superpixel segmentation and a temporal-spatial smoothing filter to improve the noisy depth estimation in the textureless region. The experiments show that the proposed method could achieve visually pleasing and temporally consistent depth estimation results without additional pre-processing and time-consuming iterations as required in other works.**

Index Terms —**Single-view video, 2D-to-3D, depth estimation, adaptive support-weight, depth propagation, optical flow, image segmentation, DIBR**

I. INTRODUCTION

In recent years, technology advances makes the human begin to pursue a more realistic visual experience. Our goal is to pursue a more natural, vivid video quality on stereo display. However, with the problem of insufficient of 3D stereo database, if we can develop a technology of 2D to 3D video conversion with high quality and efficiency, there will be such a great prospect in 3D category due to the huge database stored in 2D format.

The key of 2D to 3D video conversion is the accuracy of each video frame's depth. How to solve the problem is what we want to investigate in this study. In depth estimation, owing to the single view video contents lacking the decisive depth features, the way to find out the depth of the object will mostly turn to the motion vector, i.e. the larger the motion vector, the closer the distance of the object to the camera. But we must consider the exception of the irregular movement of camera, which may cause discontinuity in depth estimation.

This paper combines the spatial, temporal, and intensity information to estimate the depth. To consider that the movement of the camera may cause depth estimation error, our method added the global motion vector for correction, which enhances the consistency of video depth result. The detail of our algorithm will be introduced in next section.

This work was supported by the National Science Council under Grants: NSC 101-2221-E-027-078-MY2.

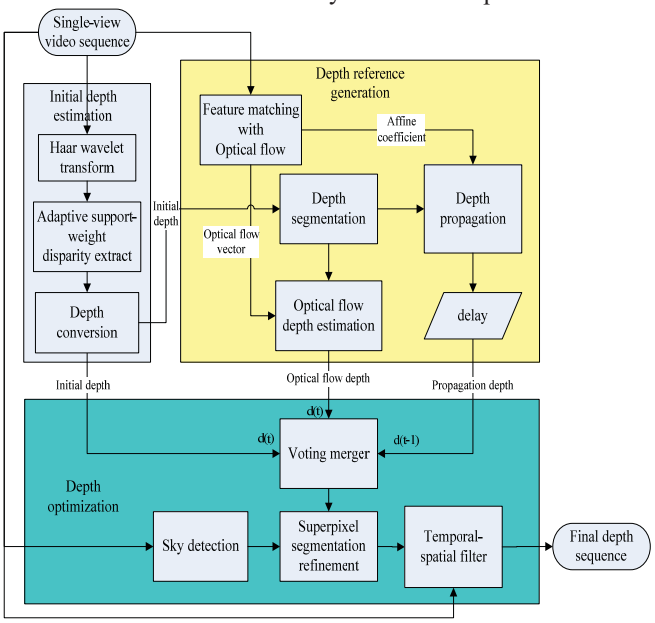


Fig. 1. Depth estimation block diagram

II. PROPOSED ALGORITHM

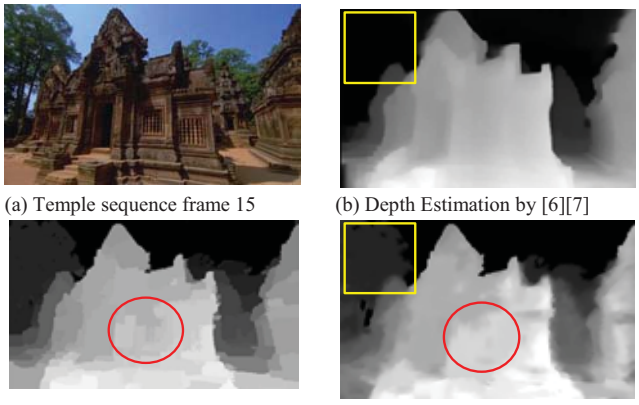
Figure 1 shows the flow chart of our video depth estimation algorithm. Our method can be divided into three steps, initial depth estimation, depth extraction for reference, and depth optimization. For initial depth estimation, Haar wavelet transform is used in the image pre-processing and the initial depth is computed by adaptive support weight method [1], which makes use of color similarity and geometric proximity. In the step of depth extraction for reference depth, propagation depth [2] can be calculated from the initial depth then. By the help of the optical flow's feature matching, we can derive the conversion relationship from the temporally adjacent image. With this conversion relationship, we can propagate depth information from initial depth to adjacent image's reference depth by block affine mapping. The strength of optical flow can also be served as another depth reference, too. Then, we take the optical flow depth [3], propagation depth and initial depth into voting merger. Finally, in order to obtain the more accurate depth, we apply the sky detection [4] and superpixel segmentation [5] to optimize the depth result.

III. Experiment Result

The method proposed in this study was developed with programming languages C/C++ and OpenCV2.3.1 library; the test is on a PC platform with hardware setup of Intel(R) Core(TM)2 Duo CPU E8300 @2.83GHz, 4.0GB RAM. The test video sequences are selected from the relative literature.

Figure 2 shows the test result of Temple sequence of our methods in comparison with other works [6][7] and [8]. Although method [6] [7]'s result has good depth hierarchies, the depth change cannot be effectively estimated in smooth region, as shown in Figure 2 (b), where the boundary of the sky and roof does not meet the object's contour, and the coverage of object's depth is obviously larger than the real contour of the object. In addition, the most serious problem is that the method [6][7] did not estimate the depth of the background woods (yellow box); On the other hand, [8]'s result has the good background depth, but the depth map has blocky phenomenon (red circle), which may lead to poor depth hierarchies and temporal discontinuity as shown in Fig. 2(c). Instead, our proposed method in depth hierarchy is better than [8] and estimates more accurate depth background than [6] [7].

Figures 3(c) and (d) show the depth results of Road sequence. Comparing the result of our algorithm with paper [8], our depth result in sky region is much closer to real



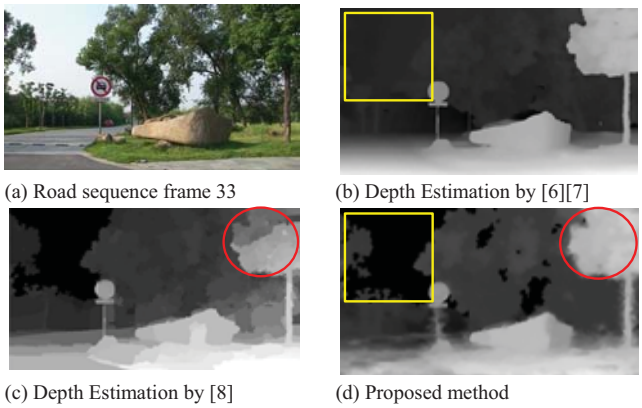
(a) Temple sequence frame 15 (b) Depth Estimation by [6][7]

(c) Depth Estimation by [8] (d) Proposed method

Fig. 2 Comparisons of Temple sequence

image. Observing the red circle, we can see that the tree's depth is miscalculated by [8], and the sky region does not fully be detected, too. [6] [7]'s result is the same as the test in Temple sequence that the depth of sky region disappeared again.

For the test in Road sequence of frame resolution 960x540, [6][7]'s method required 27 minutes to compute just one frame and [8] needed 5 minutes per frame, but our method only consumed 50 seconds per frame. Consequently, we have demonstrated that our algorithm can successfully and robustly handle different video sequences with a high efficiency.



(a) Road sequence frame 33 (b) Depth Estimation by [6][7]

(c) Depth Estimation by [8] (d) Proposed method

Fig. 3 Comparisons of Road sequence

IV. Conclusion

Our method uses adaptive support weight method to generate the initial depth information, and applies the global motion to calibrate the movement of the camera, which enhance the depth continuity between adjacent frames. Through segmentation-based depth propagation method, we can propagate depth information to neighboring frames efficiently and exclude the error caused by traditional depth propagation. After the establishment of reference depth to voting mechanism, we can extract the reliable depth information without using time-consuming method like global depth estimation and iterative depth correction, and still can get the visual comfort and temporally continuous depth estimation results.

REFERENCES

[1] Kuk-Jin Yoon, In-So Kweon, "Adaptive Support-Weight Approach for Correspondence Search," *IEEE Transactions on* Pattern Analysis and Machine Intelligence, vol. 28, no. 4, April 2006.

[2] W.-N. Lie, C.-Y. Chen, and W.-C. Chen, "2D to 3D video conversion with key-frame depth propagation and trilateral filtering," *Electron. Letters*, vol. 47, no. 5, pp. 319–321, Mar. 2011.

[3] J. Y. Bouguet, "Pyramidal Implementation of the Lucas Kanade Feature Tracker," *Intel Corporation, Microprocessor Research Labs*, 2000.

[4] S. Battiatoa, S. Curtib, M. L. Casciac, M. Tortorac, and E. Scordatoc, "Depth-map Generation by Image Classification," *Proc. SPIE on Three-Dimensional Image Capture and Applications*, vol. 5302, 95, April 2004

[5] P. Felzenszwalb and D. Huttenlocher, "Efficient Graph-Based Image Segmentation," *Int'l J. Comput. Vision*, vol. 59, no. 2, pp. 167-181, 2004.

[6] G. Zhang, J. Jia, T. Wong, H. Bao, "Consistent Depth Maps Recovery from a Video Sequence" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 974-988, June 2009.

[7] G. Zhang, J. Jia, T. Wong, H. Bao, "Recovering consistent video depth maps via bundle optimization," *IEEE Conference on Computer Vision and Pattern Recognition*, CVPR 2008.

[8] Sheng-Po Tseng; Shang-Hong Lai, "Accurate Depth Map Estimation from Video via MRF Optimization," *2011 IEEE* Visual Communications and Image Processing (VCIP), Nov. 2011