Consistent Depth Maps Recovery of Video via Object Segmentation

Chia Ju Ho Hong Shiang Lin Ming Ouhyoung

National Taiwan University

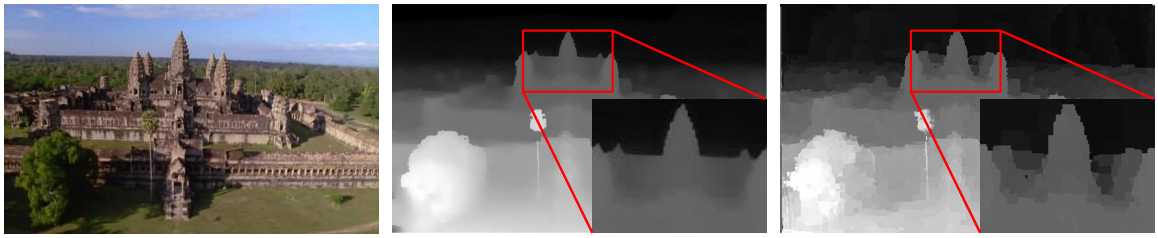

Figure 1: *(a) One frame of the input video. (b) The result of [Zhang et al. 2008]. (c) Our result.*

1 Introduction

Video depth reconstruction is a long-term problem in computer vision. A fair amount of previous work adopts Markov random field model in depth estimation formulation. Various local prior terms are included into global smoothness of those optimization frameworks. However, the estimated depth discontinuities are usually not consistent with object boundaries and intra-object smoothing is not well achieved. Such problems are addressed in color segment-based methods but cannot be solved well when color segments exist across object boundaries or are not temporarily consistent.

Motivated by object stereo [Bleyer et al. 2011], we propose an effective object-aware video depth estimation framework. Consistent object labels and depth values in time sequence gives better occlusion reasoning, thus reducing errors than color segment-based methods. We compare our result with [Zhang et al. 2008] and it shows better depth reconstruction on object boundaries (see Figure 1).

2 Our system

Our framework consists of two stages. In first stage, our system selects key frames of the input video, and then jointly estimates depths and object labels for each key frame. In second stage, those depth maps and object maps are warped to intermediate frames. And then the depth maps of intermediate frames are refined according to the optimization framework in [Zhang et al. 2008].

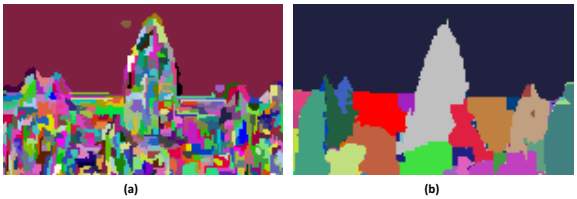

Figure 2: *(a) Color segmentation. (b) Object segmentation.*

Temporal Object-Depth Optimization The object-stereo system [Bleyer et al. 2011] initializes the object maps by disparity plane assignment, and detects occlusion region by comparing object labels. Since their system unifies object labels by warping the object map of the reference view to the other view, the object labeling of the other view heavily relies on the initial disparity map of the reference view. Our system projects the disparity planes of key frames to 3D space and merges transformed planes to unify object labels. The object labeling on each key frame still equals the disparity plane assignment on itself in this updating strategy. Therefore, the occlusion reasoning is more stable since it is supported by disparity initialization of multiple images.

Intermediate-frame Depths The object-aware disparity maps of key frames provide a good guidance of depth estimation on intermediate frames. Depth initialization of each frame is achieved by fusing warped disparity map of all key frames. The bi-directional depth fusion improves the depth estimation and reduces holes in occluded regions than that of mono-directional propagation. We adopt the energy minimization framework in [Zhang et al. 2008] to generate complete disparity maps and enhance their temporal consistency.

3 Result and Conclusion

Figure 1 shows one of our experimental cases. Note that the result of [Zhang et al. 2008] is directly from author's presentation. We implement their system for disparity map refinement on intermediate frames, and our result is presented in Figure 1 (c). Since the object segmentation better separates foreground and background than pure color segmentation around object boundaries (see Figure 2), the result shows that our method has better performance on the indicated regions. In conclusion, we have developed an object-aware video depth estimation system and the future work is to incorporate video object cut techniques [Li et al. 2005] to effectively computing layer assignment for each frame in a video for film post-production.

References

BLEYER, M., ROTHER, C., KOHLI, P., SCHARSTEIN, D., AND SINHA, S. 2011. Object stereo:joint stereo matching and object segmentation. 3081–3088.

LI, Y., SUN, J., AND SHUM, H.-Y. 2005. Video object cut and paste. *ACM Transactions on Graphics (TOG) 24*, 3, 595–600.

ZHANG, G., JIA, J., WONG, T.-T., AND BAO, H. 2008. Recovering consistent video depth maps via bundle optimization. 1–8.

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the Owner/Author.
SIGGRAPH 2013, July 21 – 25, 2013, Anaheim, California.
2013 Copyright held by the Owner/Author.
ACM 978-1-4503-2261-4/13/07