3D Reconstruction of Dynamic Scenes with Multiple Handheld Cameras

Supplementary Material

Hanqing Jiang[1,2], Haomin Liu[1], Ping Tan[2], Guofeng Zhang[1⋆], Hujun Bao[1]

[1]State Key Lab of CAD&CG, Zhejiang University
[2]National University of Singapore

1 Comparison with Temporal Smoothing on 'Boy' Example

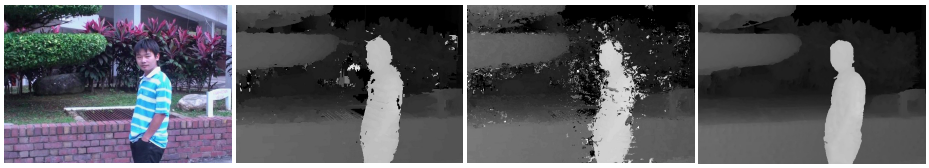

Fig. 1. Comparison results with temporal smoothing on 'Boy' example. (a) One source frame. (b) The initialized disparity map. (c) The disparity optimization result using the traditional temporal smoothing by 3D NURBS curve fitting, which contains much noise and outliers. (d) Our spatio-temporal disparity optimization result, which successfully correct the errors in (b).

Most previous methods such as [3, 1] simply enforced temporal coherence constraint individually on each single camera for depth refinement. Disparities of corresponding pixels across multiple frames were smoothed using linear interpolation or curve fitting. This optimization method is not robust as shown in Fig. 1(c), because a simple smoothing does not essentially help the inference of true disparity values and is very sensitive to outliers caused by either inaccurate depth estimate or false correspondences by optical flow errors. In comparison, our spatio-temporal optimization method aims to infer the true disparities by combining color and geometry coherence constraints among multiple cameras in different time instances. Comparing to traditional temporal smoothing methods, our method is much more robust, as demonstrated in Fig. 1(d).

⋆ Corresponding Author: Guofeng Zhang (zhangguofeng@cad.zju.edu.cn)

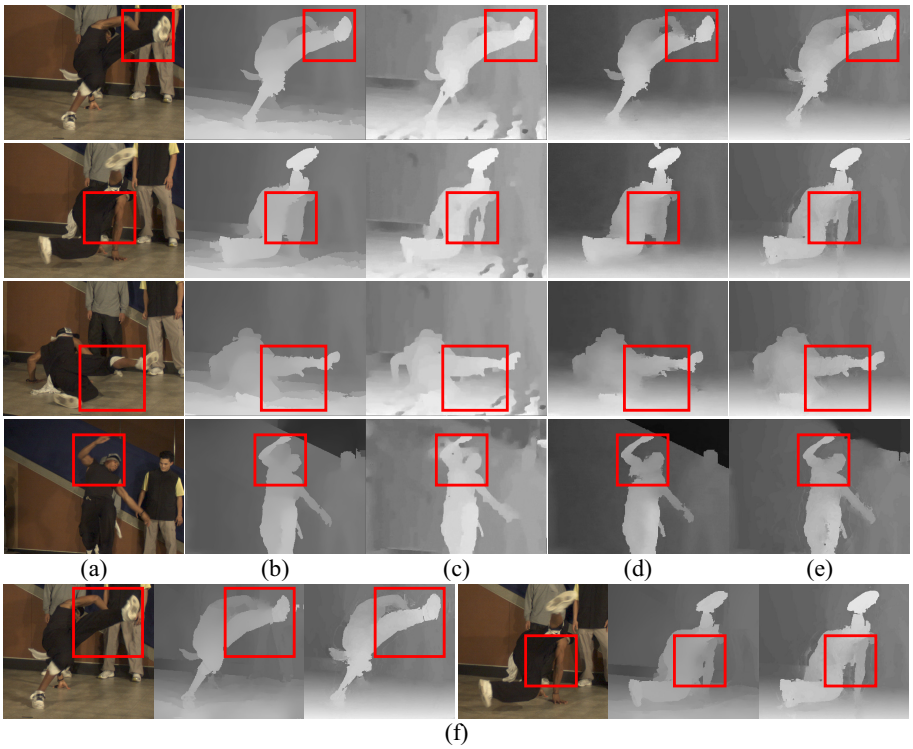

Fig. 2. Comparison of our results with [5], [1], [2] and [4] for the 'breakdancing' dataset. (a) Four selected source frames. (b) Microsoft depth results [5]. (c) The depth results of [1]. (d) The depth results of [2]. (e) Our results. (f) Two source frames (left), the results of [4] (middle) and our results (right).

2 Comparison Results of Microsoft Research 3D Video Dataset

Fig. 2 shows the comparison of our results with the results of [5], [1], [2] and [4] for the Microsoft Research breakdancing dataset. Different from our other examples, this example contains 8 videos captured by different synchronized cameras with relative smaller baselines. Even for this regular data, our method still produces more smooth and accurate depth results than [5, 1, 4], especially for the static background regions. The dynamic body of the dancer is also reconstructed with better depth quality for some representative frames. For most of the frames, our results are quite comparable with the most recent state-of-the-art work [2, 4]. However, our method can also work on very few ($2 \sim 3$) cameras with much wider baselines (the occlusions are quite complex and serious), which are very difficult for other methods like [4], as illustrated in Fig. 5(d) of our paper.

References

1. Larsen, E.S., Mordohai, P., Pollefeys, M., Fuchs, H.: Temporally consistent reconstruction from multiple video streams using enhanced belief propagation. In: ICCV. pp. 1–8 (2007)
2. Lei, C., Chen, X.D., Yang, Y.H.: A new multiview spacetime-consistent depth recovery framework for free viewpoint video rendering. In: ICCV. pp. 1570–1577 (2009)
3. Tao, H., Sawhney, H.S., Kumar, R.: Dynamic depth recovery from multiple synchronized video streams. In: CVPR. pp. 118–124 (2001)
4. Yang, M., Cao, X., Dai, Q.: Multiview video depth estimation with spatial-temporal consistency. In: BMVC (2010)
5. Zitnick, C.L., Kang, S.B., Uyttendaele, M., Winder, S., Szeliski, R.: High-quality video view interpolation using a layered representation. ACM Transactions on Graphics 23, 600–608 (August 2004)