Video Stitching with Extended-MeshFlow.

Paper ID: 1199

Kai Chen, Jian Yao†, Binbin Xiang and Jingmin Tu

School of Remote Sensing and Information Engineering, Wuhan University, Wuhan, Hubei, China

†E-Mail: jian.yao@whu.edu.cn Web: http://cvrs.whu.edu.cn/

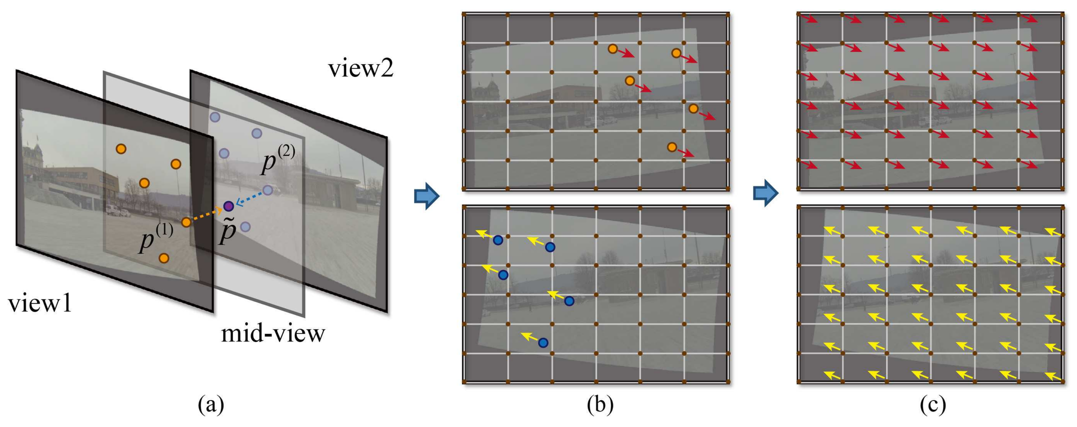
1. Introduction

A method is proposed to stitch multiple videos captured with a fixed camera rig. Compared with the case of freely moving cameras, videos with fixed rig are less complex, but they are more widely applied in practical applications, such as hyperlapse photography, virtual reality, and mixed reality. As the camera rig is usually imperfect, video stitching under such a condition is still a problem that has not been completely solved. In this paper, first, an extended meshflow motion model is designed to warp adjacent video views to the common central view to eliminate spatial misalignments. Secondly, collecting the motions located on the same mesh vertex forms the vertex profile, which are smoothed adaptively from temporal consistency. The proposed method achieves a balance between the spatial alignment and temporal consistency and finally produces a better video stitching result.

2. Extended MeshFlow

Feature points are detected and matched in the overlapping region, from which the extended meshflow model is estimated. Motion vectors located on feature positions are interpolated to mesh vertexes by **Multilevel B-Spline Approximation**.

$$\begin{cases} v_{p^{(1)}} = \tilde{p} - p^{(1)} = \dfrac{p^{(2)} - p^{(1)}}{2} \\ v_{p^{(2)}} = \tilde{p} - p^{(2)} = \dfrac{p^{(1)} - p^{(2)}}{2} \end{cases}$$



3. Video Stitching

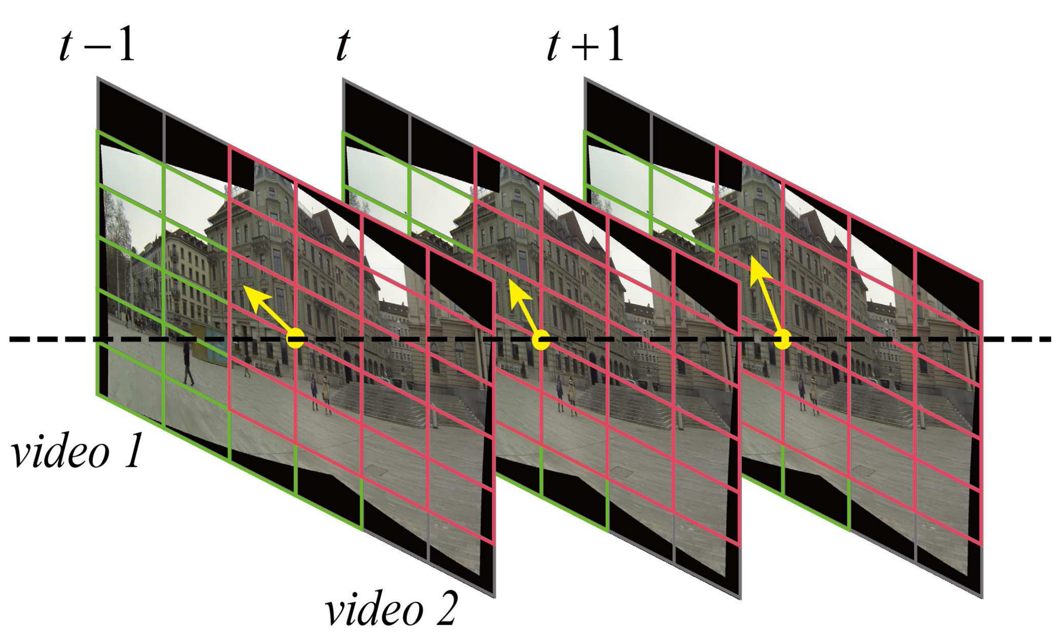
The extended meshflow of each video is temporally optimized by adaptively smoothing all vertex profiles that this video holds.

$$\mathcal{O}(\widetilde{V}(t)) = \sum_{r=t0}^{t-1} \|v^{\phi}(r) - v^{\phi-1}(r)\|^2$$

$$+ \|v^{\phi}(t) - v(t)\|^2 + \sum_{r=t_0}^{t-1} \omega_{t,r} \|v^{\phi}(t) - v^{\phi}(r)\|^2.$$

$$\omega_{t,r}^{(t)} = exp(-\|r-t\|^2/\sigma^2)$$

$$\omega_{t,r}^{(v)} = 1 - \frac{\|v(t) - v^{\phi-1}(r)\|}{\|v(t)\| + \|v^{\phi-1}(r)\| + \varepsilon},$$



4. Method validation

Comparisons between videos before and after proposed optimization scheme validates the proposed algorithm. The RMSE of one minus normalized cross correlation (NCC) is computed to evaluate the spatial alignment quality, and the temporal consistency is evaluatedin the frequency domain.

$$RMSE(I_i, I_j) = \sqrt{\frac{1}{N} \sum_{\pi} (1.0 - NCC(\mathbf{x}_i, \mathbf{x}_j))}$$

No.	Alignment Error		Consistency Score		No.	Alignment Error		Consistency Score	
	before	after	before	after		before	after	before	after
01	11.21	**9.90**	21.09	**25.25**	02	10.24	**9.06**	20.12	**23.09**
03	10.32	**9.49**	27.80	**28.53**	04	14.30	**10.41**	28.20	**28.96**
05	14.39	**10.69**	28.67	**28.96**	06	14.65	**12.34**	**22.60**	22.27

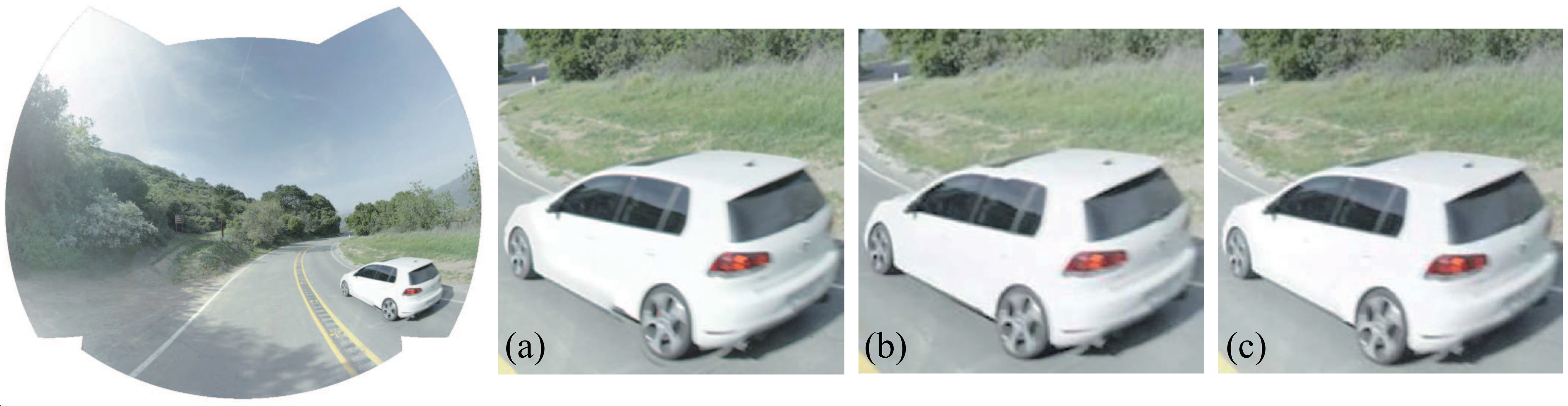
5. Compare with commerical softwares

Comparisions with three popular commerical softwares: VideoStitch Studio [1], Autopano 2.5 and Autopano 3.0 [2]. Autopano 3.0 is the improved version of Autopano 2.5, which utilizes D.WARP technique to eliminate spatial misalignment and improve temporal consistency.

VideoStitch Studio | Autopano 2.0 | Extended MeshFlow | t-1 | t | t+1

6. Compare with Perazzi's method [3]

The proposed method can produce very similar results even in some challenging scenes, but is more efficient than Perazzi's method. Without any additional optimization, the proposed method takes ~ 1 second to stitch a frame pair with up to $2K$ resolutions, while Perazzi's method resorts to dense optical flow and usually requires tens of seconds.



7. References

[1] Videostitch studio. `https://www.orah.co/software/videostitch-studio/`.

[2] Kolor autopano video. `http://www.kolor.com/autopano-video/`.

[3] F.Perazzi and A.Sorkine-Hornung. Panoramic video from unstructured camera arrays. In *Computer Graphics Forum*, volume 34, pages 57–68, 2015.