

360° Panoramic Video from a 3-Camera Rig

S. Morteza Safdarnejad, Xiaoming Liu and Jingfeng Liu

Abstract In this work, a system for creation of 360° panoramic video from a low-cost 3-Camera rig is presented. To provide 360° coverage with 3 cameras, cameras equipped with fisheye lens are utilized. Fisheye lens introduces extra challenges in stitching videos from the cameras, due to considerable distortion near the borders of field of view. First, we calibrate each camera and extract mapping parameters to correct for the fisheye effect. Then, to speed up the process, these mappings are consolidated into the mappings calculated in stitching phase, to form the composite mappings. Finally, video sequences are read from cameras, warped using the composite mappings, and corrected for color uniformity to generate the 360° panoramic video.

Keywords Panoramic video · Video stitching

1 Introduction

360° panoramic videos are gaining more and more popularity due to the rich information they provide from the captured environment. Generally, for generation of the panoramic videos, scene is captured by an appropriate number of cameras and resultant videos are stitched together to form a uniform and seamless panoramic video. For full coverage of the 360°, generally a large number of cameras are used which results into costly camera rigs and costly data acquisition modules [8]. In this paper, we introduce a low-cost system for generation of 360° panoramic videos from a 3-camera rig, as shown in Fig. 1. In this scenario, to provide the 360° coverage, small cameras equipped with fisheye lens are packed into a tight and small rig.

S. Morteza Safdarnejad · X. Liu (✉)
Michigan State University, East Lansing, USA
e-mail: liuxm@cse.msu.edu

J. Liu
LinkSprite Technologies, Inc., Wuhan, China

Fig. 1 The proposed 3-camera rig for generation of 360° panoramic videos



Fisheye lens introduces considerable distortion, especially near the boundary of field of view of the camera. Although in the final panoramic video only about 140° of coverage of the fisheye camera is needed to form a seamless 360° panoramic video, in the phase of stitching mappings estimation, the distortion causes difficulties. So, effective fisheye distortion removal is required. Thus, we first calibrate each camera and extract mapping parameters to correct for the fisheye effect. After calibration, the rig is fixed in space and a calibration image is captured from each camera. These images are stitched together using a spherical compositing surface, to generate the stitching mappings. During the operational phase, videos are captured from each camera, and fisheye removal and stitching mappings are applied on each frame to map the video content to the global coordinate of the 360° videos. To speed-up the operational phase, for each camera, fisheye removal mapping and stitching mapping are consolidated into a single mapping in the calibration phase. Finally, warped videos are corrected for color uniformity and rendered into a seamless 360° panoramic video. The resultant video may be rendered in a flat surface for display on regular display devices or prepared appropriately to be displayed using 360° panoramic video services like YouTube 360°.

2 Proposed Framework

Since the proposed panoramic system is composed of only 3 cameras, fisheye lens cameras are used. Fisheye lenses provide very large wide-angle views, and thus, provide enough overlap between the cameras which is required for successful stitching of the frames from the cameras. There are algorithms for registration of

frames with minimal or no overlap between the frames given jointly moving cameras, such as the seminal work of Caspi and Irani [3]. However, our experiments reveal that these methods are not very accurate. More importantly, for the case of small or no overlap between frames, it is not possible to refine the results using bundle adjustment, as will be described later, to ensure that the 360° panoramic video forms a seamless realistic videos.

For realization of this work, capabilities of OpenCV [1] libraries were found to be appropriate. Main stages of the proposed framework are discussed in the following sections.

2.1 Fisheye Lens Calibration

Each fisheye camera is separately calibrated to remove fisheye lens distortion. At this stage, a checker board pattern is used. For calibration, since 180° coverage of the fisheye lens leads to huge distortions at the boundary, different Matlab and OpenCV libraries are tried to find the best choice for calibration, with maximum usable angular coverage after calibration. OpenCV built-in calibration libraries were finally found to have superior performance. These libraries use low-order polynomial models for radial distortion correction. Figure 2 shows how the calibration removes the fisheye lens distortion. As a side effect, this distortion removal decreases field of view of the camera from about 180 to about 140, however, still these coverage provides enough overlap for estimation of stitching parameters.

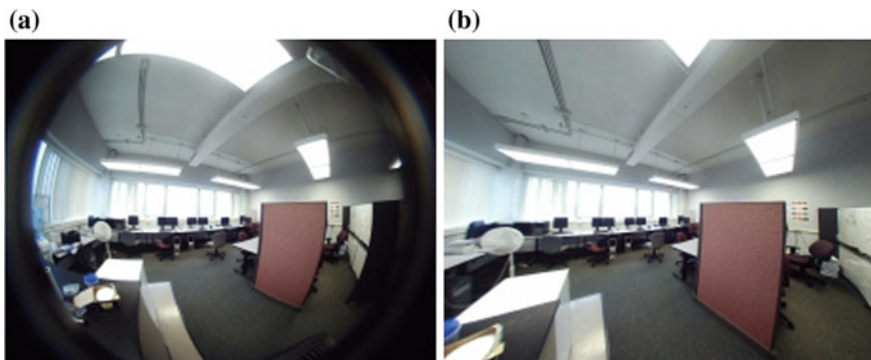


Fig. 2 **a** Input frame affected by fisheye lens distortion, **b** frame after removing fisheye lens distortion

2.2 *Estimation of Stitching Parameters*

To estimate the mapping from each video to the global 360° panorama coordinate, a calibrated image from each camera is used. This still image should be captured in a textured environment so that enough keypoint correspondence may be made between calibration images of all the 3 cameras. This is the most crucial step, since video stitching problem would be a generalization of this image stitching step. Also, since a single set of images are used to find the stitching parameters, the final video stitching results will lead to the best results when the environment depth is similar to the depth in the calibration images. So, if the target application involves capturing videos from objects far from the camera, the same kind of condition should be emulated for the parameter estimation stage.

For estimation of stitching mappings, first, SURF [2] keypoints are detected and matched [4] for adjacent cameras, via the calibration images. Given the matches, relative rotation and translation of cameras are estimated. To ensure that the mappings result into a seamless 360° panorama, bundle adjustment [6] is used to refine the mapping parameters.

Given the estimated parameters, in the operational phase, frames are continuously read from the 3 cameras and mapped to the global 360° panorama coordinate to form the panoramic video. To maintain synchronization of the videos from each camera compositing the final panoramic video, it is important to read 3 frames from all the 3 cameras simultaneously and store them, before starting processing of frames. Otherwise, if frames are read and processed sequentially, the delay in reading the frames caused by processing time will lead to synchronization artifacts in the panoramic video.

2.3 *Video Composition*

After finding the stitching parameters, a compositing surface should be selected to produce the final stitched image and video. This surface might be flat, cylindrical, spherical, etc. [5]. For the case of 360° panorama, compositing surface cannot be flat, since it leads to huge distortions on the boundaries and more importantly, the panorama cannot be wrapped to form a 360° coverage. In practice, flat panoramas start to look severely distorted if the field of view exceeds 90°. Thus, spherical or cylindrical surface is used, with better results achieved with spherical surface. Also, it is possible to composite the results on a dome to reproduce the stitched images as if they are created with a single fisheye lens camera, facing upward, perpendicular to the optical axes of the 3 cameras.

2.4 *Blending and Color Uniformity*

The frames from the 3-camera rig have some overlap, so, it is important to deal with the overlap area when composing the final stitched video. Combination of blending algorithms and exposure compensation algorithms [7] transform the stitched result into a color-uniform 360° video. For blending, multi-band blending is used [5].

2.5 *Preparation for YouTube 360°*

To enable playback of videos using YouTube 360° service, some meta data needs to be added to the resultant panoramic video to show the type of compositing surface used. For this purpose, “360° Video Metadata Tool” is used: <https://github.com/google/spatial-media/releases/download/v2.0/360°.Video.Metadata.Tool.win.zip>.

3 Results

In this section, we present stitching results for a sample set of frames from the 3 cameras. These frames are shown in Fig. 3. Figure 4 presents the stitched frames composited on spherical and cylindrical compositing surface. To illustrate how well the stitched frames match at the far ends to construct a full 360° panorama, we also present the stitching results using a dome as the compositing surface. As shown, the low-cost 3-camera rig equipped with fisheye lens provides an acceptable 360° panoramic system.



Fig. 3 Sample input frames

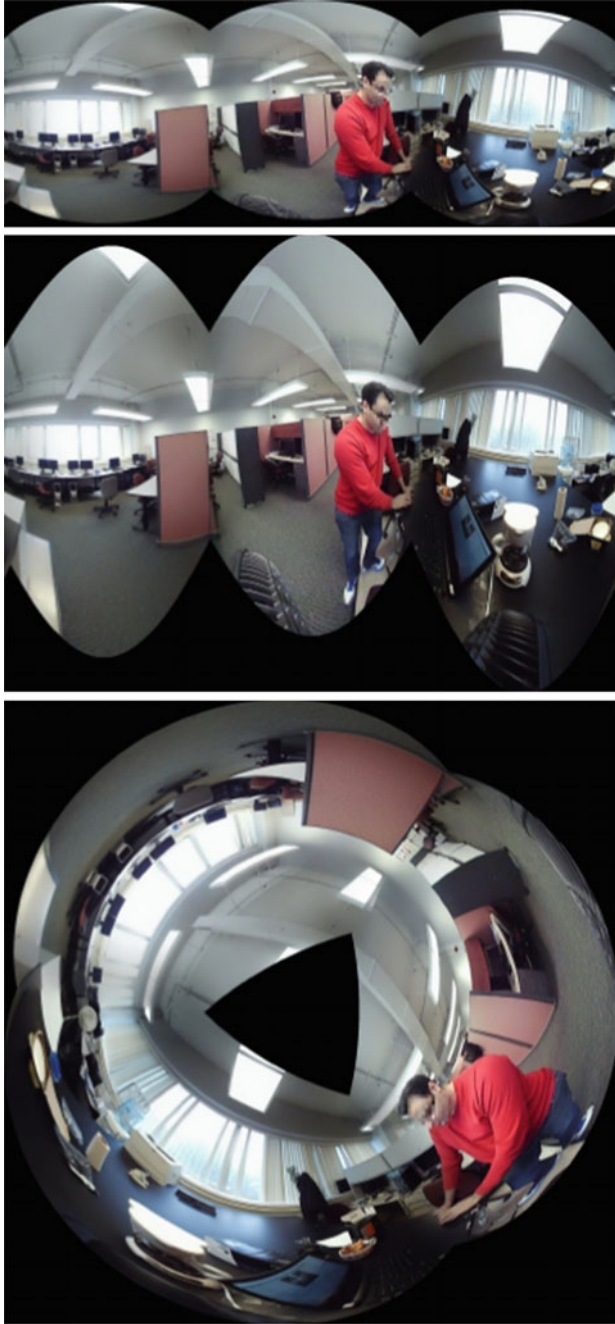


Fig. 4 Stitching results using different compositing surfaces, *top to bottom*: spherical, cylindrical, and dome compositing surfaces

4 Conclusion and Discussion

In this paper, a low-cost 3-camera rig equipped with fisheye lenses is proposed which enables generation of 360° panoramic videos. This low-cost system is realized via only 3 USB cameras, which omits the need for data acquisition system. Fishseye lens provides enough coverage and overlap between the cameras, so that the stitching parameters can be estimated reliably. However, it adds the fisheye lens distortion correction which is computationally expensive. To reduce this computational burden, it is possible to consolidate the distortion correction warping and the stitching mapping into a single mapping function for each camera.

Despite the reasonable performance, such a system which relies on offline and pre-calibration of the system, will be affected by parallax issue. Parallax issue will be more severe for the conditions far from the calibration condition. For the case of panoramic videos, due to computational costs and high efficiency required, parallax tolerant methods such as [9] are not feasible. Thus, for best performance, calibration should be performed in an environment with depth variations similar to the desired operational environment.

Acknowledgements This research was supported in part by Dongguan's Recruitment of Innovation and Entrepreneurship talent program.

References

1. G. Agam. Introduction to programming with openCv. Online Document, 27, 2006.
2. H. Bay, T. Tuytelaars, and L. Van Gool. SURF: Speeded up robust features. In Proc. European Conf. Computer Vision (ECCV), pages 404–417. Springer, 2006.
3. Y. Caspi and M. Irani. Aligning non-overlapping sequences. *International Journal of Computer Vision*, 48(1):39–51, 2002.
4. D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. Computer Vision*, 60(2):91–110, 2004.
5. R. Szeliski. Image alignment and stitching: A tutorial. *Foundations and Trends® in Computer Graphics and Vision*, 2(1):1–104, 2006.
6. B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon. Bundle adjustment modern synthesis. In *Vision algorithms: theory and practice*, pages 298–372. Springer, 1999.
7. W. Xu and J. Mulligan. Performance evaluation of color correction approaches for automatic multi-view image and video stitching. In *Computer Vision and Pattern Recognition (CVPR), 2010 IEEE Conference on*, pages 263–270. IEEE, 2010.
8. Y. Xu, Q. Zhou, L. Gong, M. Zhu, X. Ding, and R. K. Teng. High-speed simultaneous image distortion correction transformations for a multicamera cylindrical panorama real-time video system using FPGA. *Circuits and Systems for Video Technology, IEEE Transactions on*, 24(6):1061–1069, 2014.
9. F. Zhang and F. Liu. Parallax-tolerant image stitching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3262–3269, 2014.