IJIBC 19-2-6

Virtual Reality Image Shooting for Single Person Broadcasting with Multiple Smartphones

Sutanto Edward Budiman and Suk-Ho Lee[†]

†Dept. of Computer Engineering, Dongseo University, Korea
petrasuk@gmail.com

Abstract

Nowadays, one-person media broadcasting has become popular, and with the progress of this popularity, multimedia techniques which can support such broadcasting are also becoming more and more advanced. One of the most emerging multimedia technique used in this field is the virtual reality technology which sets the one-person media broadcasting environment as a virtual reality environment. However, as such an environment requires instruments of high cost, it is not easy for normal individuals to constitute such environments. Therefore, in this paper we propose how to construct virtual reality-like panoramas with a multiple of smartphones. For this purpose, we designed a special rig which can hold firmly 8 smartphone cameras which have overlapping view of the environment such that panorama stitching becomes possible. To reduce the computation cost, we precomputed the homography matrices, and used 1-D pointer structures to store the computed coordinate values.

Keywords: One person media, Virtual reality, Smartphones, Multimedia, Panorama

1. Introduction

Single-person media produces content that conforms to personalised tastes and interests, as well as a variety of materials that have not been seen on terrestrial broadcasting. In other words, providing various materials such as eating (eating broadcasting), cooking (cooking broadcasting), studying (studying broadcasting), game broadcasting, and talking (broadcasting reading stories and articles) that were not covered by conventional broadcasts is increasing interest in media content. Changes in the content environment according to single-person media are not limited to Korea. In China, "Wang Hong," which refers to people who have influence through online media, is rapidly emerging as an icon in China's consumer market. The added value of China's Wanghong industry in 2016 is about 58 billion yuan, far exceeding China's total box office of 44 billion yuan, showing the value and growth of Wanghong in the market.

Although single-person media is becoming increasingly popular as a new industry, competition to secure

high ratings among single creators calls for a new type of content to be produced. VR(Virtual Reality) contents based on real-life video that can show various parts of the object of interest from various points of view are one of the good types of content that meets the requirements. Most of the researches in virtual reality has been ongoing on the topic of how to control the virtual environment and control over time events in VRML based on the Web environment or multimedia synchronization in SMIL [1]-[4].

However, it is not easy for single-person media and small-scale broadcasting producers to overcome the barriers to producing VR contents that require high-cost VR filming equipment and high-level know-how to manipulate post-editing equipment. Therefore, we propose an enviromental setting for single-person media producers based on the use of several smartphones that can be used instead of VR cameras so that the content can be easily generated and edited by anyone. That is, we propose an omni-directional VR content shooting solution for MCN broadcasting.

The proposed VR content shooting solution uses smartphones instead of the VR cameras to shoot VR images. The solution includes general stitching and web upload solution to allow businesses to commercialization. Figure 1 shows the overall concept of the single-person media solution.
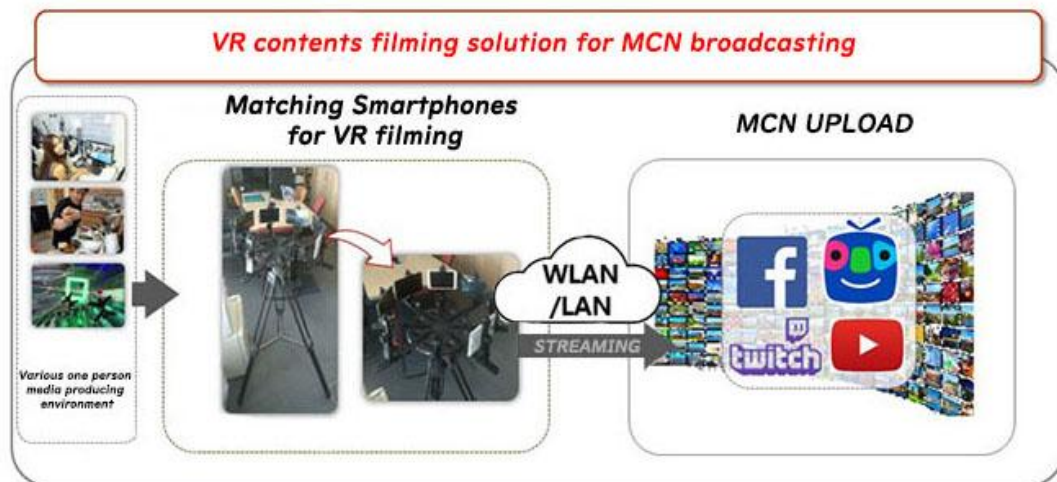

Figure 1. Concept of the Single-person VR media solution

2. Construction of rig for VR image shooting

Since the proposed solution uses smartphones of varying sizes and camera module locations rather than dedicated VR cameras, it requires the design of variable Rigs that are appropriate for VR imaging. Furthermore, since the solution is aimed for single-person media and small-scale broadcast producers, the cost of this Rig should not be too high.

To this end, we designed an original-type, variable-type smartphone bearing Rig so that it utilize existing smartphone girders, with cross-section bars, and triangles. This kind of Rig is also required because if the Rig is not used the images taken by the multiple phones are not aligned well, which results in a stitching of bad alignment as can be seen in Fig. 2. Figure 3a shows the smartphone bearing system which can obtain 360-degree live VR images using multiple smart phones. In order to make a live VR, a 360-degree VR league for eight smart phones is built as shown in Figure 3. Figure 3b is an example of a VR shooting station developed by the proposed task that indexes the images taken by many different types of smartphone cameras arranged at an angle of 45 degrees, to take pictures of the surroundings. Depending on how the camera arrangement is organized for VR image shooting, the cameras can be indexed by different numbers.
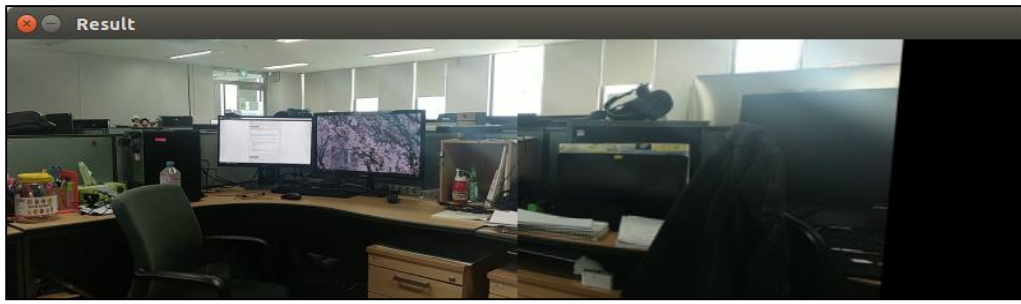
Figure 2. Showing the stitching result of bad alignment of the multiple cameras
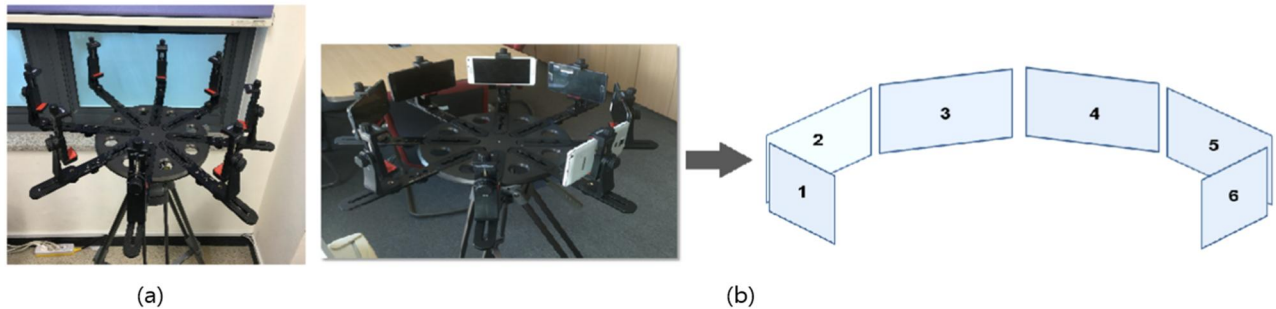

Figure 3. Constructed smartphone bearing Rig for VR shooting

3. VR Image reconstruction process

For the first frame of each smartphone the calculation of the homography matrices will be performed. The calculation for the first frame takes a lot of time due to the calculations of these matrices, but after the initial calculation the homography matrices will be cached and read in so that the panoramic composition of the images can be composed without the repeated calculation of the homography matrices. Instead of calculating the matrix required for the homography transformation, the transferred coordinate values are composed into look-up tables to take out the coordinate values of each frame. Furthermore, the coordinate values are arranged into 1-D arrays rather than a 2-D array to avoid the double memory access as required by double pointers, and therefore, using a single pointer array for a single access to the memory as shown in Fig. 4.
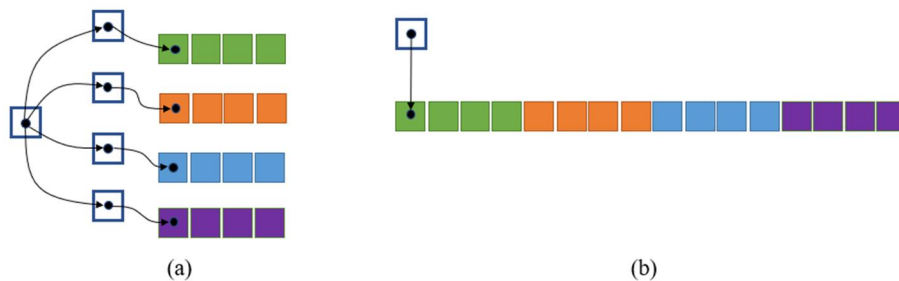

Figure 4. Change of the double pointer memory access (a) to the single memory access (b).

The time required for the double pointer and single pointer access is compared in Table 1.

Table 1. Comparison of access time for single pointer and double pointer access

	Single pointer access to 1-D array	Double pointer access to 2-D Array
Times (second)	0.132063627243042	0.29457926750183105

When stitching is continuously performed it can happen that there appears a gap in between the images due to the misalignment. The upper row in Fig. 5 shows how a gap appears between the second and the third frame. In this case, we computed the corners of the images and matched the keypoints by the FLANN method. By using the feature and matching keypoints, we can estimate the transformation matrix which corrects the stitching results.
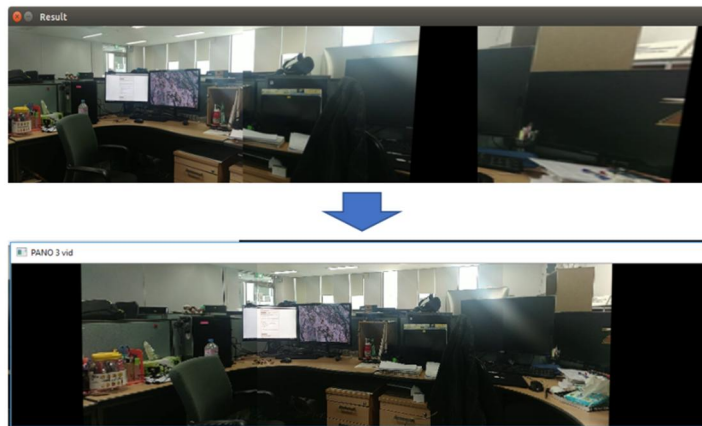

Figure 5. Correction of the stitching result with mis-alignment

All Figure 6 shows the diagram of the stitching process of the proposed system.
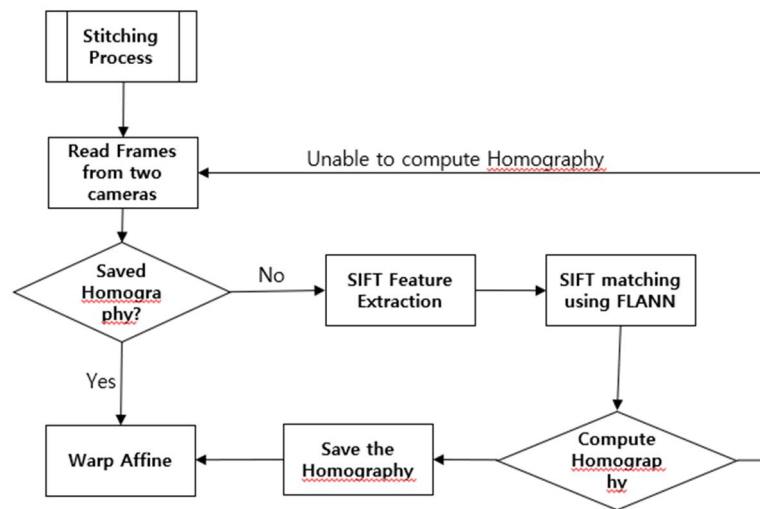

Figure 6. Overall diagram of the stitching process of the proposed system

Figure 7 shows the overall schematic diagram of the operations performed in the proposed system. For the first frames, all the calculated homography matrices are initially saved. After that all the frames obtained directly from the camera are subsequently stitched together with the cashed look-up table. Figure 7 shows the sequential stitching process between the several images obtained by the different cameras. The panoramic views is projected to a cylinder where the transformation uses the Brown's distortion model [5][6]. Every image taken from the cameras are converted to a cylindrical plane and are put on a big canvas which is updated over and over. For the initial cylindrical projection calculation it takes about 5.77 seconds for one image, but after that the calculation using the look-up table takes only 0.132 seconds per frame using 1-D array approach.
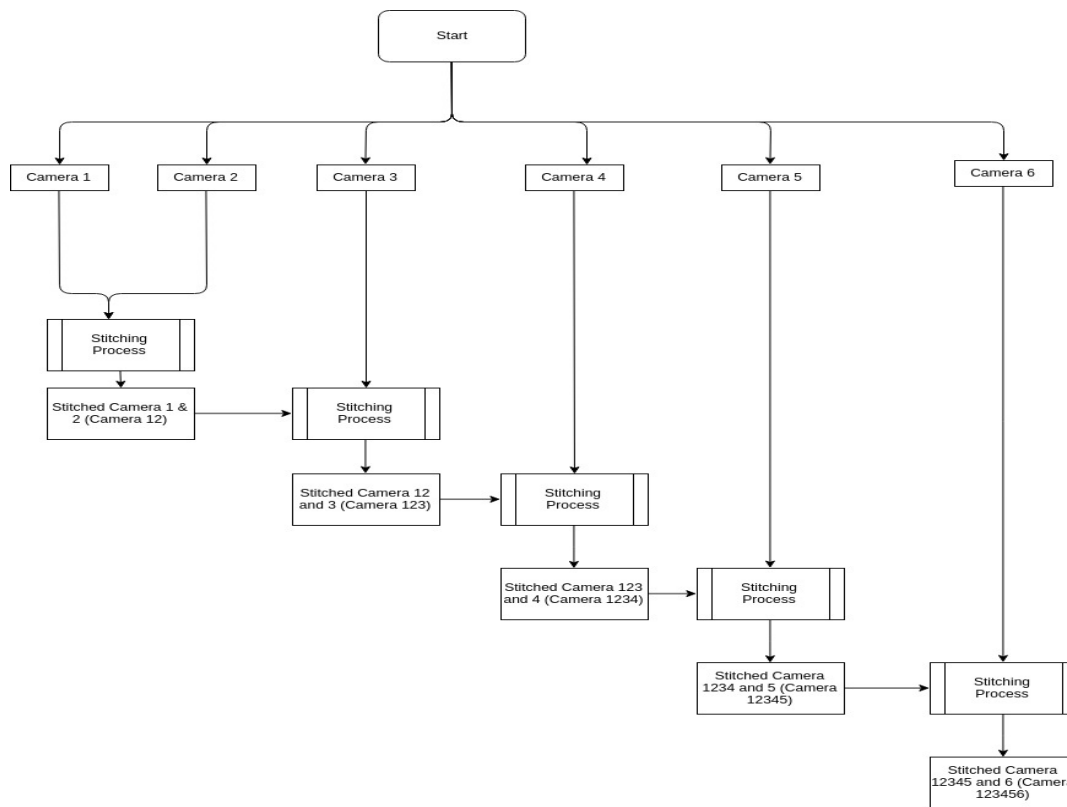

Figure 7. Overall schematic diagram of the operations

We developed a program using the above stitching algorithm, which stitches and stores images from a self-produced VR crew. The program consists of a "Live video" reception, "Image stitching" processing unit and a "Video stitching" processing unit. The UI of the development program is shown in Fig. 8 and Fig. 9 for inside and outside taken frames. The first row in Fig. 8 and Fig. 9 show the 8 images obtained by 8 cameras with different angles. The third row shows the stitching result using the 8 images. The Image stitching processing unit performs a single stitching operation while the video stitching unit performs sequential stitching operations on the video frames. The above program can be connected to an editing system implemented on a tablet PC which has a console for controlling the playback of the VR images, and further to a PC for media broadcasting which can broadcast the 360-degree live VR image contents.
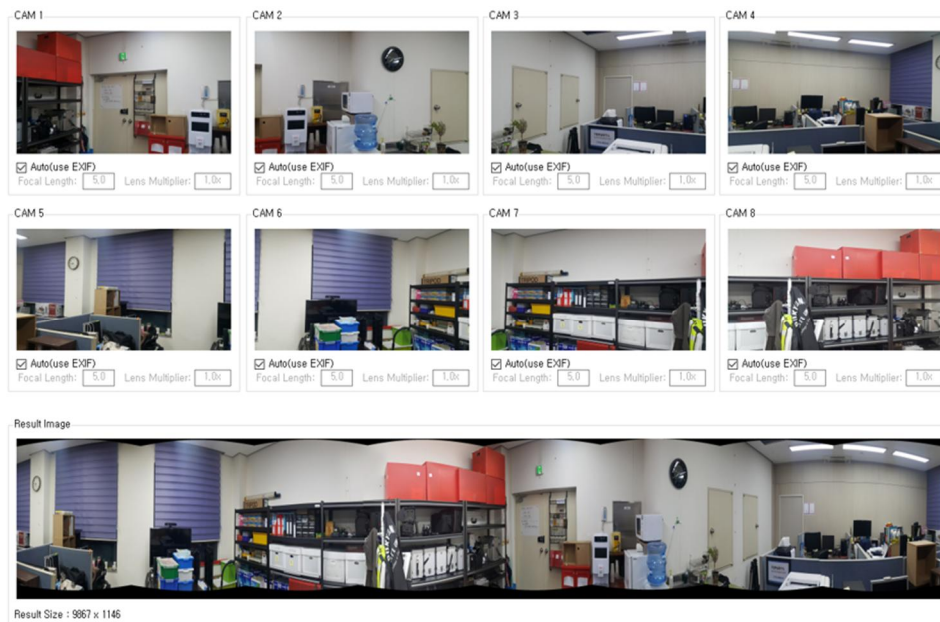
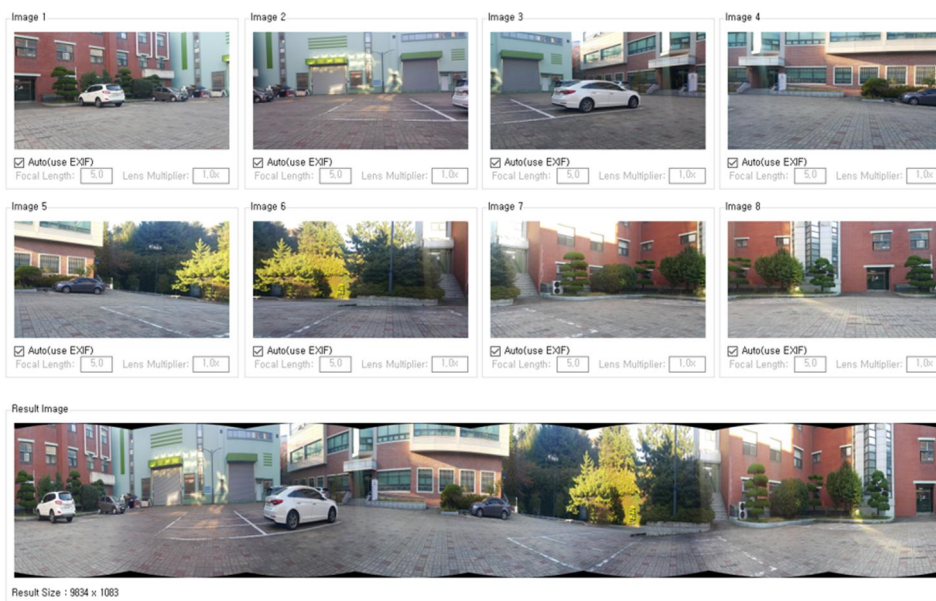**Figure 8. Developed UI program operating in indoor environment**

Figure 9. Developed UI program operating in outdoor environment

4. Conclusion

In this paper, we proposed a single person VR media solution using only smartphones for low cost production of such contents. To fix the position of the smartphones we designed a rig which can held 8 smartphones at different even angles so that the images obtained by each smartphone camera have overlapping regions to be stitched together. For fast computation, we computed the homography matrices only for the first frames, and saved the transformed coordinates into a map. As the cameras are fixed with the proposed rig, the transformation need not be calculated again, which resulted in a fast panorama construction. Experimental results show that a clear and distortion-free panorama can be obtained even with this low cost system. We expect that general users use these type of panorama making system, so that VR panorama

videos for social network broadcasting become widely used.

References

[1] S.Mann, R. W. Picard, "Virtual bellows: constructing high-quality images from video," in Proceedings of the IEEE First International Conference on Image Processing, pp. 363-367, Nov. 13-16, 1994.
DOI: 10.1109/ICIP.1994.413336

[2] G. Ward, "Hiding seams in high dynamic range panoramas," in Proceedings of the 3rd Symposium on Applied Perception in Graphics and Visualization. ACM International Conference, pp. 150-150, Jul. 28-29, 2006
DOI: 10.1145/1140491.1140527

[3] S. Mann., "Compositing Multiple Pictures of the Same Scene," in Proceedings of the 46th Annual Imaging Science & Technology Conference, pp. 50-52, May 9-14, 1993

[4] S. Mann, C. Manders, and J. Fung, "The Lightspace Change Constraint Equation (LCCE) with practical application to estimation of the projectivity+gain transformation between multiple pictures of the same subject matter," in Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 481-484, Apr. 6-10, 2003 DOI: 10.1109/ICASSP.2003.1199516

[5] D. C. Brown, "Decentering Distortion of Lenses," Photometric Engineering, Vol.32, No. 3, pp. 444–462, 1966.

[6] Niklas Rydholm.,   Panoramic Video Stitching,   Master. Thesis. Linköpings universitet, Linköping, Sweden., June 2015.