Uncalibrated Multiview Synthesis based on Epipolar Geometry Approximation
Digest of Technical Papers

Young Ju Jeong[1,2], Hyoseok Hwang[1], Dongkyung Nam[1], and C.-C Jay Kuo[2]
[1]Samsung Advanced Institute of Technology, Korea
[2]Viterbi School of Engineering, University of Southern California, USA
young.jeong@samsung.com

***Abstract*—In this work, we propose an efficient multiview rendering algorithm that takes uncalibrated stereo as the input. First, the epipolar geometry of multiple viewpoints is analyzed for multiview display. Then, the camera pose for an arbitrarily selected viewpoint is estimated by algebraic approximation. Finally, by exploiting rectification homographies and disparities of rectified stereo, one can determine multiview images with their estimated camera poses. It is shown by experimental results that the proposed multiview synthesis algorithm can provide well calibrated results without warping distortion.**

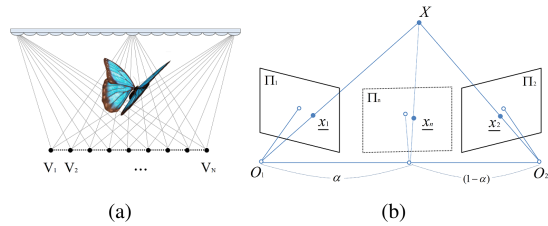

Fig. 1. (a) Lenticular lens multiview display and its viewpoints, (b) epipolar geometry of stereo and viewpoint V_n.

I. Introduction

The rapid development of 3D display technologies allows consumers to enjoy the 3D world visually through different display systems such as the stereoscopic, multiview, light field, and holography displays. As a pre-processing step, the 3D content should be manipulated properly to fit for each display type to offer the best user experience. The most common 3D source is the stereo input. It is difficult to capture ideal stereo images with the same rotation and translation parameters. For the stereoscopic display, non-ideal stereo images do not result in serious artifacts since people cannot perceive distortion within a certain range; namely, 0.7% ± 0.5 of the vertical disparity difference [1]. However, non-ideal stereo images deteriorate the perceived visual quality of other 3D displays seriously. This is because traditional stereo matching and view synthesis algorithms are vulnerable to non-ideal stereo images. To address this problem, a proper calibration process is essential to a 3D display system. One solution is calibration transformation that however yields unwanted image warping distortion [2]. In this paper, we propose a robust multi-view rendering algorithm that can provide well calibrated results without warping distortion.

II. Proposed Algorithm

In this section, we propose a multiview rendering algorithm based on uncalibrated input stereo images. First, we analyze multiview display viewpoints (MDVs) and their epipolar geometry model. Next, camera poses are estimated based on an algebraic approximation of the epipolar geometry model of MDVs. Finally, multiview images are effectively generated using rectification homographies and disparities of rectified stereo.

A. Multiview Display Viewpoints

According to the lenticular lens or parallax barrier design, MDVs are aligned in one line. Specifically, we show the lenticular lens multiview display and its linear composed viewpoints, V_1 to V_N, in Fig. 1(a).

B. Epipolar Geometry of MDVs

To generate images from uncalibrated stereo for the purpose of multiview display, the viewpoints should be located in the same line as MDVs. In Fig. 1(b), we depict the epipolar geometry of MDVs, where the left and right image planes are denoted by Π_1 and Π_2, respectively. The point,X, in the world coordinate becomes $\underline{x_1}$ in the left image and $\underline{x_2}$ in the right image. It is straightforward to derive the following

$$\underline{x_1} = K[R_1 \quad t_1]X, \tag{1}$$

$$\underline{x_2} = K[R_2 \quad t_2]X, \tag{2}$$

where $K[R_i \quad t_i]$, i=1 and 2, is a geometric transformation matrix and Ri and ti are the associated rotational matrix and translational vector, respectively

C. Algebraic Approximation of MDVs

The camera pose of Π_n for viewpoint V_n is associated with translational vector $(1-\alpha)t_1 + \alpha t_2$ and rotational matrix R_n. We have the following proposition.

$$\underline{x_n} \approx (1-\alpha)\underline{x_1} + \alpha \underline{x_2}. \tag{3}$$

Proposition: Let $\underline{x_n}$ is an image point of viewpoint V_n of X. If $R_1$and R_2 are sufficiently similar (which will be quantified in the body of proof), it can be well approximated by

*proof.*By substituting Eqs. (1) and (2) to Eq. (3), we obtain:

$$\underline{x_2} = K[R_2 \quad t_2]X, \tag{4}$$

978-1-4799-7543-3/15/$31.00 ©2015 IEEE

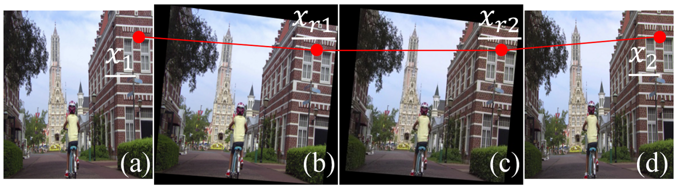

Fig. 2. (a) The left image, (b) the rectified left image, (c) the rectified right image, and (d) the right image.

$$\underline{x_n} = K[(1-\alpha)R_1 + \alpha R_2 \quad (1-\alpha)t_1 + \alpha t_2]X, \tag{5}$$

since K is a linear transform. Thus, this proposition implies: (i) $t_n \approx (1-\alpha)t_1 + \alpha t_2$ and (ii) $R_n \approx (1-\alpha)R_1 + \alpha R_2$. Condition (i) is clearly met as shown in Fig. 1(b). In the following, we would like to argue that condition (ii) is also valid. Let $\Delta R = R_2 - R_1$. We have $R_n = (R_2 R_1^{-1})^{\alpha} R_1$, which can be expressed as:

$$\begin{aligned}(R_2R_1^{-1})^{\alpha} &= (I + \Delta R R_1^{-1})^{\alpha} \\ &= I + \sum_{k=1}^{\infty} \frac{\alpha(\alpha-1)\cdots(\alpha-(k-1))}{k} (\Delta R R_1^{-1})^{\alpha}.\end{aligned} \tag{6}$$

If $max_i|\lambda_i(\Delta R)| < 1$, the summand decays at a rate faster than $O((max_i|\lambda_i(\Delta R)|)^k/k)$. Under such a circumstance, we can approximate the summation only with the first-order term as

$$\begin{aligned}(R_2R_1^{-1})^{\alpha}R_1 &\approx (I + \alpha \Delta R R_1^{-1})R_1 \\ &= (1-\alpha)R_1 + \alpha R_2.\end{aligned} \tag{7}$$

D. Implementation

We implement the multiview synthesis algorithm based on a pair of stereo images using rectification homographies and the disparity map from rectified stereo image. The disparity map of rectified stereo can be determined in an efficient manner. An example is given in Fig. 2, where $\underline{x_1}$, and $\underline{x_2}$ are uncalibrated matching points and $\underline{x_{r1}}$ and $\underline{x_{r1}}$ are their corresponding rectified image points. We have

$$\underline{x_{r1}} = H_1 \underline{x_1}, \quad \underline{x_{r2}} = H_2 \underline{x_2}, \tag{8}$$

where H_1 and H_2 are homography matrices which can be calculated using a rectification algorithm [3]. Because stereo images are rectified, stereo matching can be implemented in an efficient manner [4]. The disparity of $\underline{x_{r1}}$ and $\underline{x_{r2}}$ have only horizontal disparity term, which is denoted by $d(\underline{x_{r2}})$. Mathematically, we have

$$\underline{x_{r2}} = \underline{x_{r1}} + [d(\underline{x_{r2}}) \quad 0 \quad 0]^T, \tag{9}$$

We can estimate $\underline{x_{r1}}$ based on rectified images in Fig. 2(b) and Fig. 2(c) with a stereo matching algorithm (e.g., [5]). Finally, since

$$\underline{x_2} = H_2^{-1}[\underline{x_{r1}} + [d(\underline{x_{r2}}) \quad 0 \quad 0]^T], \tag{10}$$

one can calculate and $\underline{x_n}$ have only horizontal disparity term, which is denoted by in viewpoint V_n directly as

$$\underline{x_n} = (1-\alpha)\underline{x_1} + \alpha H_2^{-1}[\underline{x_{r1}} + [d(\underline{x_{r2}}) \quad 0 \quad 0]^T], \tag{11}$$

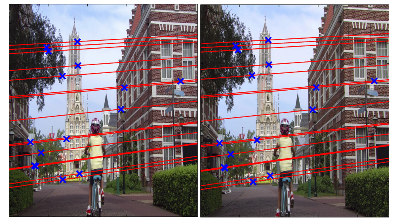

Fig. 3. Feature points and epipolar lines of left and right images.

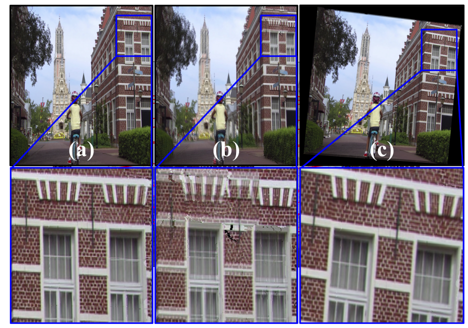

Fig. 4. Interpolated view using (a) the proposed algorithm, (b) uncalibrated stereo image, (c) calibrated stereo image.

III. Experimetal Results

We uses non-ideal stereo images as the input. Figure 3 shows input stereo images, their matching points and epipolar lines. The vertical disparity of this image is 3.085 pixels and it is less than the human detection limits 0.7 % $\pm$ 0.5, which is 7.56 $\pm$0.5 pixels [1]. However, since their epipolar lines are not parallel, we observe warping distortion after rectification.

Figure 4 shows image for V_n with $\alpha = 0.5$. The result of the proposed algorithm is well calibrated and it does not have the warping distortion as shown in Fig. 4(a). In contrast, blurring occurs in Fig. 4(b) due to the uncalibrated stereo input while warping distortion is observed in Fig. 4(c) due to rectification.

IV. Conclusion

An efficient multivew rendering algorithm with uncalibrated stereo images as the input was proposed in this paper. We derived a model for multiview viewpoints from the non-ideal stereo input and demonstrated the effectiveness of the proposed algorithm by comparing three interpolated views using three different algorithms.

References

[1] Y. Nojiri, H. Yamanoue, A. Hanazato, M. Emoto, and F. Okano, "Visual comfort/discomfort and visual fatigue caused by stereoscopic hdtv viewing," in *Electronic Imaging 2004*. International Society for Optics and Photonics, 2004, pp. 303–313.

[2] J. Mallon and P. F. Whelan, "Projective rectification from the fundamental matrix," *Image and Vision Computing*, vol. 23, no. 7, pp. 643–650, 2005.

[3] A. Fusiello and L. Irsara, "Quasi-euclidean uncalibrated epipolar rectification," in *Pattern Recognition, 19th International Conference on*. IEEE, 2008, pp. 1–4.

[4] D. Scharstein and R. Szeliski, "A taxonomy and evaluation of dense two-frame stereo correspondence algorithms," *International journal of computer vision*, vol. 47, no. 1-3, pp. 7–42, 2002.

[5] Y. J. Jeong, J. Kim, H. Y. Lee, and D.-S. Park, "Confidence stereo matching using complementary tree structures and global depth-color fitting," in *Consumer Electronics, IEEE International Conference on*. IEEE, 2013, pp. 468–469.