Research Activities in 3D Content Generation

Matej Nezveda, Margrit Gelautz
Institute of Software Technology and Interactive Systems
Vienna University of Technology
Vienna, Austria
{matej.nezveda, margrit.gelautz}@tuwien.ac.at

Werner Zellinger, Bernhard Moser
Knowledge-Based Vision Systems
Software Competence Center Hagenberg
Hagenberg, Austria
{werner.zellinger, bernhard.moser}@scch.at

Florian Seitner
emotion3D GmbH
Vienna, Austria
seitner@emotion3d.tv

Christian Kapeller
cmotion GmbH
Vienna, Austria
christian.kapeller@cmotion.eu

***Abstract*— This report describes work on 3D content generation performed under the Austrian 3D-ConTourNet participation project Hyperion3D. We review key processing steps such as disparity computation, inter-camera mapping and virtual view synthesis, and give information on a publicly available database for evaluating 3D visual discomfort.**

Keywords—trifocal camera; 3D film; stereo matching; view synthesis; 3D database; visual discomfort; 3D-ConTourNet

I. INTRODUCTION

We describe work on 3D content generation performed under the Austrian COST participation project Hyperion3D. The project deals with the development of a system for 3D content production that consists of a high-end main camera which is accompanied by a compact assistant stereo rig.

In Section II we describe our camera set-up and give an overview of the workflow and algorithms we designed to convert the trifocal input stream into disparity/depth maps and virtual views derived from them. We discuss some peculiarities of the inter-camera mapping process in the context of our L-shaped camera alignment. The selection of virtual viewpoints allows adjusting the disparity range to be presented to the viewer on a particular display. Suitable strategies have been developed to suppress visual discomfort which may be experienced by the viewer depending on the scene's depth (and motion) characteristics and the properties of the chosen stereoscopic display.

In Section III we summarize work on a related depth recommendation system whose evaluation in a user study resulted in a publicly available database containing stereo videos along with subjective quality scores.

II. TRIFOCAL SYSTEM FOR 3D CONTENT GENERATION

We have developed a trifocal system for high-quality 3D content generation (see Fig. 1). In this section, we give a brief overview of the system's architecture and its main components. We refer the interested reader for more details on this trifocal system to our corresponding publication [1].

A. *Trifocal Camera System*

Our camera system involves a high-end main camera C_M (i.e., Arri ALEXA [2]) with a compact assistant stereo rig (i.e., pair of sinaCAM HDC1-100 camera heads [3]) consisting of two assistant cameras C_{A1} and C_{A2} attached to it. An important characteristic of our system is its L-shaped configuration, as visible in Fig. 1. This configuration makes the whole system flexible in terms of camera positioning and handling.

B. *Rectification and Color Matching*

We consider view M of camera C_M as our primary source of information. Thus, during rectification and color matching our objective is to preserve this camera view. To this end, we first align view A_1 of camera C_{A1} with view M and in a second step view A_2 of camera C_{A2} with the rectified view A_1 (i.e., only A_1 and A_2 are transformed during rectification). In particular, we perform an uncalibrated rectification using [4] based on SURF [5]. For color matching, we fit the color histograms of both A_1 and A_2 to the color histogram of M using global color histogram alignment as proposed in [6].

C. *Disparity Computation*

We perform a stereo-based disparity computation between views M and A_1 and between views A_1 and A_2. To this end, we have combined an adapted cost volume filtering framework previously developed by us in [7] and a hierarchical matching framework within a unified framework.

For stereo matching, we perform the following four basic steps: First, we store the dissimilarities based on the census transform [8] between both views for each pixel at each allowed disparity level in a cost volume. Second, we perform an edge-aware filtering of the cost volume using [9]. Third, we generate an initial disparity map by selecting for each pixel the disparity with the lowest cost. Fourth, we refine the initial disparity map using the approach of [7] and apply an additional sub-pixel enhancement according to [10].

For hierarchical matching, we build a Gaussian pyramid with three layers. Each layer has half the resolution of its preceding layer. Then, we compute an initial disparity map for the coarsest layer using the described stereo matching

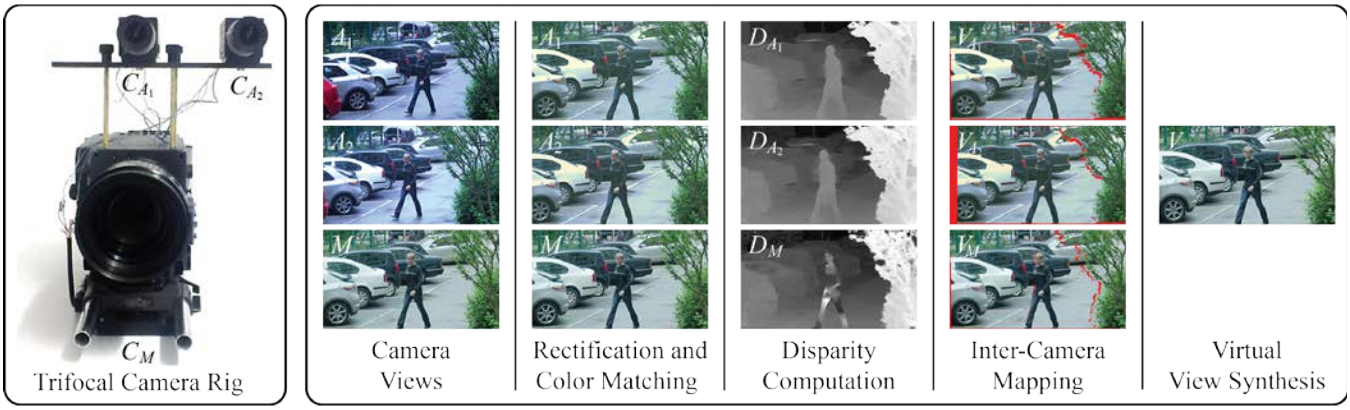

Fig. 1. Trifocal system pipeline. A camera system consisting of one main camera C_M and two assistant cameras C_{A1} and C_{A2} captures three time synchronous views M, A_1 and A_2. These views are first rectified and color matched and then their disparity maps D_M, D_{A1} and D_{A2} are computed. Finally, the rectified and color matched views are mapped via the computed disparity maps to virtual camera views V_M, V_{A1} and V_{A2} which are located at the same position and fused to a final virtual view V.

algorithm. In the next two finer layers, we bound the number of disparity levels according to the results of the specific previous layer. Thus, we obtain a disparity map for the final layer which is based on the priors of the coarser layers.

D. Inter-Camera Mapping and Virtual View Synthesis

We use the stereo-derived disparity maps between the views M, A_1 and A_2 to synthesize a virtual view V using depth-image-based rendering similar to [11]. However, since no disparity map exists between M and V, we first generate such a disparity map. To this end, we define a virtual camera view A_3 that is located on the same vertical position as M and on the same horizontal position as A_2. Then, we vertically map the disparity map $D^h_{A1\rightarrow A2}$ to camera view M and obtain $D^h_{M\rightarrow A3}$. Finally, by scaling the disparity map $D^h_{M\rightarrow A3}$ we can shift the position of virtual view V. Fig. 2 outlines this setup.

During our experiments we found that the quality of the disparity maps is significantly lower between views M and A_1, which are captured with different cameras, than between views A_1 and A_2, which are captured with identical cameras. Thus, instead of using direct mapping based on disparity maps between views A_1 and M, we introduce indirect mapping where horizontal disparity maps between A_1 and A_2 can be converted to vertical disparity maps between A_1 and M.

For virtual view synthesis, we map all three camera views M, A_1 and A_2 to the same virtual viewpoint and obtain three virtual views V_M, V_{A1} and V_{A2}. Note that we can here use either direct or indirect mapping. Then, we fuse all three virtual views to a final virtual view V. Finally, we fill missing pixels after fusion using “horizontal copy background” as described in [12].

III. VISUAL DISCOMFORT DATABASE

The selection of a suitable depth budget, that is maximum and minimum disparity values that are present in the stereo pair and their adjustment relative to the screen plane, has an important effect on the quality of the perceived 3D impression. We developed a depth grading recommendation tool that seeks to minimize visual discomfort by applying an optimization scheme which is based on a human visual attention model in conjunction with the scene’s estimated depth and motion characteristics. The proposed model incorporates the assumption that the object of interest should be located in depth near the screen plane and undergo only limited depth jumps. More details on the formulation in terms of a linear optimization problem are given in [13].

In order to evaluate the performance of the depth recommendation system, we conducted a user study in which the levels of image quality, depth quantity, and visual comfort were rated by 14 subjects on a set of available and self-recorded stereoscopic test videos. The stereo videos (see Fig. 3) along with subjective quality scores are publicly available at the SCCH 3D Visual Discomfort Database (www.scch.at/en/id-3d-visual-discomfort-database).

IV. CONCLUSION

We have reviewed a workflow and algorithms for generating 3D content from a trifocal camera set-up, as part of our research activities connected to 3D-ConTourNet. Subjective quality evaluation in the context of visual (dis)comfort led to the implementation of a publicly available database with user-provided ratings. Future work might address the design of user studies to assess the quality of depth maps and derived novel views for highly dynamic scenes for which ground truth data are usually not available.

ACKNOWLEDGMENT

We wish to thank the Austrian Research Promotion Agency (FFG) for financial support under the project Hyperion3D (no. 840222) and the COST Action 3D-ConTourNet (IC1105) for funding the networking and dissemination activities.

REFERENCES

[1] F. Seitner, M. Nezveda, M. Gelautz, G. Braun, C. Kapeller, W. Zellinger, and B. Moser, “Trifocal system for high-quality inter-camera mapping and virtual view synthesis,” International Conference on 3D Imaging, pp. 1-8, 2015.

[2] Arri, ALEXA camera, http://www.arri.com/camera/alexa/cameras/camera_details/alexa, Accessed: 18-Mar-2016.

[3] sinaCam, HDC1-100 camera, http://www.sinacam.eu/specifications.html, Accessed: 18-Mar-2016.

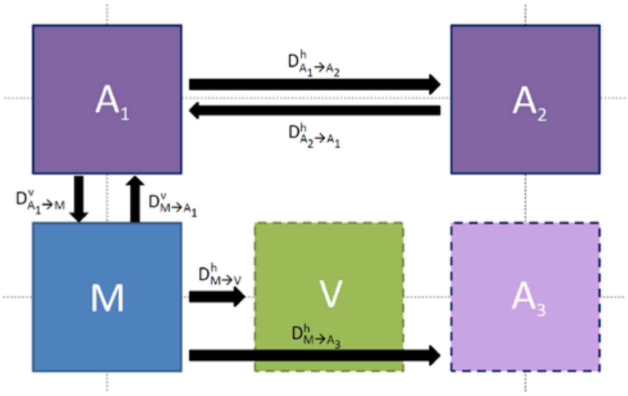

Fig. 2. Schematic illustration of our virtual view synthesis framework.

[4] H.-H. P. Wu and Y.-H. Yu, "Projective rectification with reduced geometric distortion for stereo vision and stereoscopic video," Journal of Intelligent and Robotic Systems, vol. 42, no.1, pp. 71-94, 2005.

[5] H. Bay, T. Tuytelaars, and L. Van Gool, "SURF: Speeded up robust features," European Conference on Computer Vision, pp. 404-417, 2006.

[6] J. P. Rolland, V. Vo, B. Bloss, and C. K. Abbey, "Fast algorithms for histogram matching: Application to texture synthesis," Journal of Electronic Imaging, vol. 9, no. 1, pp. 39-45, 2000.

[7] A. Hosni, C. Rhemann, M. Bleyer, C. Rother, and M. Gelautz, "Fast cost-volume filtering for visual correspondence and beyond," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 35, no. 2, pp. 504-511, 2013.

[8] R. Zabih and J. Woodfill, "Non-parametric local transforms for computing visual correspondence," European Conference on Computer Vision, pp. 151-158, 1994.

[9] C. Cigla and A. A. Alatan, "An efficient recursice edge-aware filter," Signal Processing: Image Communication, vol. 29, no. 9, pp. 998-1014, 2014.

[10] Q. Yang, R. Yang, J. Davis, and D. Nister, "Spatial-depth super resolution for range images," IEEE Conference on Computer Vision and Pattern Recognition, pp. 1-8, 2007.

Fig. 3. Test videos from the SCCH 3D Visual Discomfort Database. The database provides left and right stereo views along with user ratings regarding their perceived stereoscopic quality.

[11] K. Müller, A. Smolic, K. Dix, P. Merkle, P. Kauff, and T. Wiegand, "View synthesis for advanced 3D video systems," EURASIP Journal on Image and Video Processing, vol. 2008, no. 1, pp. 1-11, 2008.

[12] M. Eisenbarth, F. Seitern, and M. Gelautz, "Quality analysis of virtual views on stereoscopic video content," International Conference on Systems, Signals and Image Processing, pp. 333-336, 2012.

[13] W. Zellinger, B. Moser, A. Chouikhi, M. Nezveda, M. Gelautz, "Linear optimization approach for depth range adaption of stereoscopic videos," IS&T Electronic Imaging Stereoscopic Displays and Applications XXVII, 6 pp., 2016.