51. IWK

Internationales Wissenschaftliches Kolloquium
International Scientific Colloquium

PROCEEDINGS 11-15 September 2006

FACULTY OF ELECTRICAL ENGINEERING AND INFORMATION SCIENCE

INFORMATION TECHNOLOGY AND ELECTRICAL ENGINEERING - DEVICES AND SYSTEMS, MATERIALS AND TECHNOLOGIES FOR THE FUTURE

TECHNISCHE UNIVERSITÄT
ILMENAU

Impressum

Herausgeber: Der Rektor der Technischen Universität Ilmenau
Univ.-Prof. Dr. rer. nat. habil. Peter Scharff

Redaktion: Referat Marketing und Studentische Angelegenheiten
Andrea Schneider

Fakultät für Elektrotechnik und Informationstechnik
Susanne Jakob
Dipl.-Ing. Helge Drumm

Redaktionsschluss: 07. Juli 2006

Technische Realisierung (CD-Rom-Ausgabe):
Institut für Medientechnik an der TU Ilmenau
Dipl.-Ing. Christian Weigel
Dipl.-Ing. Marco Albrecht
Dipl.-Ing. Helge Drumm

Technische Realisierung (Online-Ausgabe):
Universitätsbibliothek Ilmenau
ilmedia
Postfach 10 05 65
98684 Ilmenau

Verlag: isle
Verlag ISLE, Betriebsstätte des ISLE e.V.
Werner-von-Siemens-Str. 16
98693 Ilrnenau

ISBN (Druckausgabe): 3-938843-15-2
ISBN (CD-Rom-Ausgabe): 3-938843-16-0

Startseite / Index:
http://www.db-thueringen.de/servlets/DocumentServlet?id=12391

51st Internationales Wissenschaftliches Kolloquium
Technische Universität Ilmenau
September 11 – 15, 2006

Christian Weigel

Towards a 3D-TV System on the Basis of Image-Based Rendering Methods

1. Introduction

Conventional television is one of the most widespread electronic entertainment and information system of the world. Anyway, its design forces the user to participate in the program only passively. New multimedia standards, such as MPEG-4 [1] were developed in order to bring interaction to television and other multimedia presentations. The usage of virtual scene descriptions and coding of arbitrarily shaped two-dimensional video facilitate new ways of interaction and immersion for the user. 3D video objects are a new possibility to increase the realism of a scene further. The generation of such objects requires a number of new processes in the production chain. The production chain described in this paper relies on image based rendering methods (IBR) with implicit geometry [2].

2. System Design

Acquisition

The module responsible for the acquisition process uses six CCIR/PAL-resolution cameras. With respect to object based applications we decided to use a convergent, horizontal configuration. A workstation that stores the captured frames uncompressed accompanies each camera. The user can control the whole process centrally by a software tool that is connected via Ethernet. The whole system is shown in Figure 1.

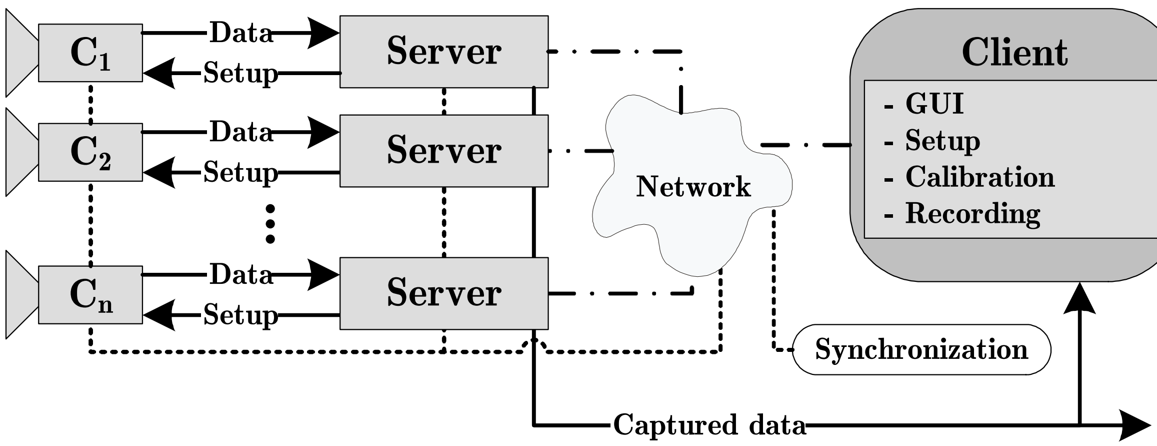

Figure 1 Acquisition System

In order to get synchronized frames for each instance of time for each of the cameras a timecode generator sends the current capture time to the servers via ethernet.

Preprocessing

After the capturing process only the object without background of the recorded scene is required. Concerning this matter segmentation for every sequence is conducted. Currently, this task is accomplished by out of the shelf composition software. Anyway, this approach needs manual operation and takes a lot of time. Therefore we started activities to implement a color-based segmentation that does both, run fully automatically and perform at interactive rates.

The next task of this module is to determine the geometric relations and the projection parameters of the cameras. These are required in order to model a mathematical construct that enables us to synthesize novel views of the object. Therefore, the preprocessing tool needs at least eight pixel correspondences for each pair of images. The accurate generation of this data is utterly important to obtain high quality 3D video objects. The tool facilitates interactive input of the correspondences as well as automatic retrieval using a calibration pattern recorded beforehand. Using these or even more sparse correspondences the fundamental matrix for each pair of cameras can be estimated by a robust RANSAC algorithm [5]. Additionally, when using the tensor-based view synthesis, the module estimates the seed tensor [3]. It is planned to outsource this task to the acquisition module since it is required only once for each camera setup.

Next, based on the previously obtained calibration data, the module computes and optimizes a dense disparity map. These dense correspondences are required for both, the synthesis base on interpolation and the synthesis based on trifocal pixel transfer. The accuracy of the dense disparity map is determined by a number of factors of which the most problematic are occlusions and wrongly assigned correspondences. There are a number of algorithms trying to cope with these problems [6]. Our system currently employs a pixel to pixel stereo algorithm introduced in [7]. In order to use the algorithm the images are rectified beforehand by a planar rectification using homographies which are obtained from the fundamental matrices. A sample disparity map is shown in Figure 2.

Figure 2 Rectified Images and Disparity Map

Synthesis and Display

The system supports two kinds of synthesis of the virtual view. The first one is based on linear view morphing [4]. It can display virtual views which lie within the sector of a circle spanned by the camera setup. The second method employs trilinear tensors [3]. It allows much more flexibility because of its ability to even extrapolate from the original views. The novel view is always created based on two cameras. From these the disparities are estimated as described in the section before. Then, the texture information (pixel values) from one camera or two cameras can be used to transfer them to the novel view by solving linear equations which are the result of tensor equations. If two textures are used, both results are blended together linearly or by Hermite-Blending depending on the distance of the virtual camera from original camera. Some results are shown in Figure 3.

Figure 3 Novel Views Created by Tensor Based Pixel Transfer

After generating the virtual views the 3D video object can be put into a synthetic scene based on the MPEG-4 standard for example. Currently, the system only supports pre-calculated paths since it is not working at interactive rates. An example is shown in Figure 4.

Figure 4 A 3D Video Object in a Synthetic MPEG-4 Scene

3. Implementation

The implementation of the system is based on the so called ReVOGS (Realistic Video Object Generation System). ReVOGS provides modules for any type of required processing steps and can be easily extended for new demands. The system is based on the idea to connect small functional modules within a graph that can then be performed easily at runtime. The structure of the graph can simply be assembled by writing an XML file. Therefore, no programming knowledge is required in order to use the system. New functionality in terms of new modules can easily be implemented using an API provided by the ReVOGS system. Thus the system provides a very flexible basis for higher level application as outlined in Figure 5.

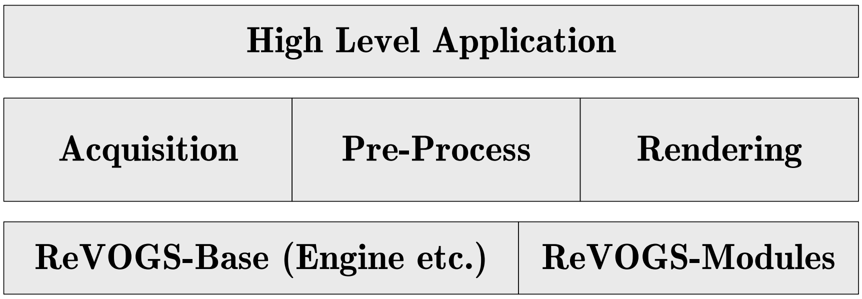

Figure 5 ReVOGS Software System

4. Conclusion and Future Work

In future work the system will be improved to work at interactive rates. Furthermore, the acquisition and pre-processing steps will be merged to one application. Another goal is the improvement of the disparity estimation and synthesis algorithms in order to obtain a higher quality of the 3D video objects.

5. Acknowledgment

This work is supported by EC within FP6 under Grant 511568 with the acronym 3DTV.

References:
[1] International Organization for Standardization: ISO/IEC 14496-1:2001.1+AMD1+COR1: Information technology - Coding of audio-visual objects - Part 1: Systems. International Organization for Standardization, Geneva, Switzerland, 2002.
[2] H.-Y. Shum, S. B. Kang and S.-C. Chan, Survey of image-based representations and compression techniques, IEEE Transactions on Circuits and Systems for Video Technology}, vol. 13, pp. 1020--1037, Nov. 2003.
[3] S. Avidan and A. Shashua.Novel View Synthesis by Cascading Trilinear Tensors. IEEE Transactions on Visualization and Computer Graphics (TVCG), 4(4), 1998.
[4] S. M. Seitz and C. R. Dyer, View Morphing, Proc. SIGGRAPH 96, 1996, 21-30.
[5] M. A. Fischler, R. C. Bolles, “Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography”, Comm. of the ACM, Vol. 24, pp 381-395, 1981.
[6] D. Scharstein, R. Szeliski. “A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms”, In Proceedings of the IEEE Workshop on Stereo and Multi-Baseline Vision, Kauai, HI, Dec. 2001.
[7] S. Birchfield and C. Tomasi, “Depth discontinuities by pixel-to pixel stereo,” in ICCV, 1998, pp. 1073–1080. [Online]. Available: citeseer.ist.psu.edu/article/birchfield96depth.html

Authors:
Dipl.-Ing. Christian Weigel
Technische Universität Ilmenau, Institut für Medientechnik , Postfach 10 05 65
98684 Ilmenau
Phone: +49 (0) 3677 692609,
Fax: +49 (0) 3677 691255
E-mail: christian.weigel@tu-ilmenau.de