PROCESSING OF 3D VIDEO OBJECTS AT DIFFERENT LEVELS OF QUALITY

Christian Weigel, Marco Rittermann, Günter Horna, Leif Lennart Kreibich

Institute of Media Technology - Technische Universität Ilmenau
Postfach 10 05 65, 98684 Ilmenau, Germany
E-mail: christian.weigel@tu-ilmenau.de

ABSTRACT

In this paper we present a system for the image-based processing of natural 3D video objects at different levels of quality. Firstly, the usage of 3D video objects and their generation using an image-based method are explained. Such methods still require many manual operations. In order to reduce this expenditure several automatic processes are employed. This improved production is presented and illustrated in this paper. The time for generating 3D video objects can be reduced up to 50%. Another prerequisite for a graded production is an objective quality assessment. For this, we present a methodology which can be adapted to any kind of 3D video objects.

1. INTRODUCTION

The evolutions in the fields of television and computer technology as well as the development of new standards have lead to a new type of media. In comparison to the classical 2D video representation content can now be created, transmitted, and displayed using an object based approach as described by the MPEG-4 standard [6]. In such an approach all types of media are coded and transmitted separately. The scene is composed at the user's side within a 2D or 3D context. This allows new kinds of applications since it enables the user to interact with the scene. For example, in a 3D scene the user can freely navigate around and interact with objects of interests. This freedom can be employed by a number of interesting applications like a virtual fair, a video conferencing system, or an interactive music show [7].

The interactive scene can be composed of synthetic objects like mesh representations as well as natural video objects. Currently these natural video objects are basically two dimensional and rectangular. In order to enhance scene realism in MPEG-4 methods are defined to code arbitrarily shaped video objects. Although these objects seem to appear more realistic they are still not sufficient for a complete immersion of the user which is required for the applications mentioned above. Thus, for highly immersive applications 3D video objects are required.

2. IMAGE-BASED 3D VIDEO OBJECTS

For the applications previously mentioned the quality of the 3D video objects is the key for the plausibility of the scene. There are different approaches to create such video objects that lead to different qualities of the object. To create a 3D video object additional information is required that needs to be obtained by new methods of capturing the scene. This yields a more complex process of acquisition described in Section 3.

In the work described in this paper an image-based method with implicit geometry usage is employed to create an arbitrary virtual view of the 3D video object. The object of interest (currently humans) is captured with a multi-view system in front of a blue screen. The convergent multi-view setup is a weak one where neither intrinsic nor extrinsic camera parameters are known a priori. Thus, a self calibration process is required. In order to obtain the fundamental matrix which describes the geometric relation between any of two images point correspondences are needed. The fundamental matrix can be estimated using the RANSAC algorithm [1] with at least eight of these correspondences. Once the fundamental matrices for each pair of cameras are calculated the generation of virtual views is reduced to a one dimensional search of corresponding points and subsequent morphing. In order to further facilitate this process the images with their respective scanlines are rectified by a pre-warp process. After this process all scanlines are aligned horizontally and the intermediate view can be calculated by a simple linear morphing algorithm. Once the view is calculated it needs to be de-rectified by a post-warp process in order to obtain the correct intermediate image. The whole process is based on a three step algorithm as proposed in [2].

3. PRODUCTION CHAIN

3.1. Workflow

In Figure 1 the workflow for the production of 3D video object as it was applied in [3] is shown. Different camera setups are considered for a multi perspective acquisition of a natural object. With respect to object based applications a convergent, horizontal configuration (cp. Figure 2) is used. At any

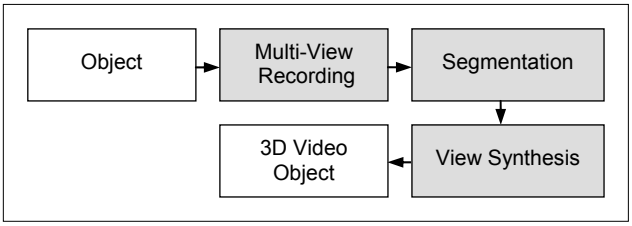

Figure 1 - Workflow of 3D video object generation.

rate it is helpful to use consistent recording equipment in order to simplify the postproduction.

After the capturing process only the object without background of the recorded scene is required. Concerning this matter a segmentation for every sequence has to be conducted, e.g. via chroma-keying.

Additional to at least eight correspondences for each pair of cameras the process of the view-synthesis requires camera path information to calculate virtual views. The accurate generation of this data is utterly important to obtain high quality 3D video objects. Thus, a deeper contemplation of this aspect is pointed out in the next sections.

3.2. Challenges

The accuracy of the estimation of the fundamental matrix heavily influences the quality of 3D video objects. Therefore it is useful to propose possibilities which results in an arbitrary number of correspondences retrieved from the source-images. Moreover, the amount of time spent to determine these correspondences should be reduced to a minimum. Significant features (e. g. detectable in clothes) are suited to detect correspondences. Accordingly, the characteristics of the foreground object are important to retrieve favourable point coordinates.

3.3. Automation of Self Calibration

Different image processing software offers a basic approach for a manual extraction of correspondences. In addition, the algorithm of the view synthesis module requires a certain input format of the determined coordinates. So far the formatting had to be done manually, too. By using a completely manual method an acceptable quality of the 3D video object could only be achieved with an immense effort of time.

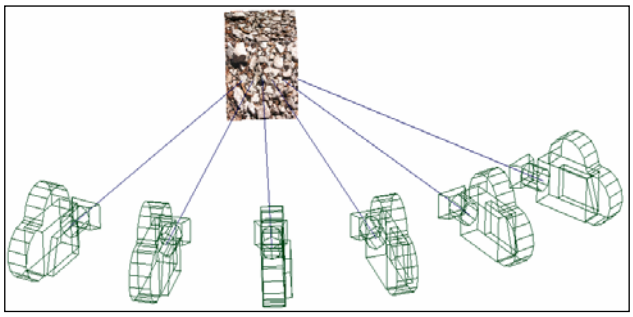

Figure 2: Multi-view setup with six cameras.

Within the scope of [3] an application was developed to enhance this process. The required coordinates can be retrieved in the needed format simply by mouse-clicking suitable features in correspondent images. This results in an enormous reduction of production time. In terms of a further time reduction and an improved estimation the potential of manual feature extraction is almost tapped. As a result, automatic solutions are expedient.

In terms of further processing the output of available software for feature detection and matching is of less consistency when dealing with natural foreground objects (like actors). In order to obtain satisfactory results predefined features of reference objects should be detected instead of natural image features. In addition, automatic procedures yield the retrieved coordinates with sub-pixel accuracy which again raises the 3D video object quality.

Due to the development of a further application it is possible to retrieve correspondences automatically by detecting four corners in every frame of a reference checkerboard sequence. Hence, a moving pattern must be recorded once by all cameras previously to the actual scene. In order to evaluate the automatic estimation visually some scanlines are displayed in the last frame of every sequence (see Figure 3). These scanlines represent the epipolar lines in the rectified image. For best performance the number of correspondences to be estimated can be adjusted.

3.4. Preparation of Rendering

For the calculation of virtual viewpoints the applied view synthesis algorithm needs information about the rendering path. Therefore, orientation and position of the virtual camera are described by four parameters for every frame of the resulting 3D video object. In order to minimize the effort for generating these large amounts of data a tool was developed in [3]. The values for each parameter and frame are set by interpolation between selectable initial and final values.

3.5. Enhanced View Synthesis

The process of view synthesis is basically accomplished by a simple linear morphing algorithm. Nonetheless, a high quality virtual image requires several steps of optimization during this process. One important step is the disparity error compensation. It removes vertical portions from the disparity values that were introduced by the rectification process. Furthermore it corrects the disparity values according to a relative depth. In subsequent steps the disparity map is calculated for the current virtual view and treated with different optimization tools as depicted in Figure 4. This is necessary to further enhance the quality of the 3D video object.

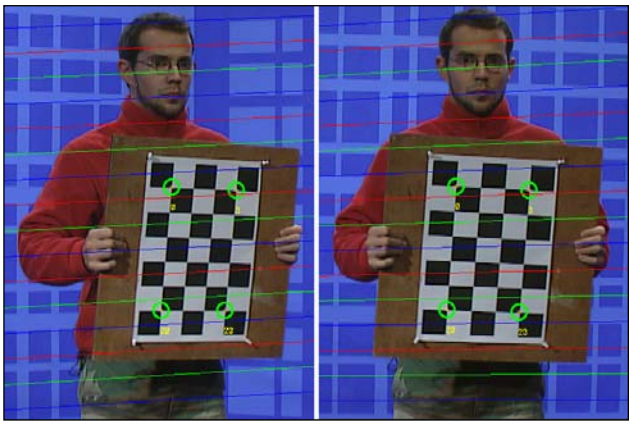

Figure 3 - Scanlines displayed in correspondent frames of a calibration sequence.

3.6. Results

Within [3] efforts were made to improve not only the production process but also the quality of 3D video object. Because of the wide automation in the field of extracting correspondences and generating a navigation path the developed applications lead to a reduced production time (up to 50%). In most cases the fundamental matrix is now precisely estimated so satisfactory results of 3D video objects can be achieved with high probability. In this context it is possible to evaluate the accuracy of estimation by analyzing scanlines. This direct feedback allows fine tuning of the estimation process.

Improvements in the algorithm of the view synthesis may also result in more advanced quality of the 3D video object. A possibility to control these adjustments without interfering extrinsic influences is to compare the calculated viewpoints with a simulated reference.

In order to achieve a real-time production of 3D video objects the time costs of every generation step need to be examined. The reduction of these costs goes along with a change of the quality. Thus, further research on the influence of every production step on the achieved quality needs to be done.

4. Different levels of quality

4.1. Requirements to 3D Video Objects

Requirements to 3D video objects have to be stated regarding the intended application:

- Area of possible viewpoints (angle, distance)
- Inclusion of the object into the scene (illumination, depth of focus etc.)
- Quality of the object

The area of possible viewpoints depends on the camera setup used for the acquisition. Ideally, multi-view setups record

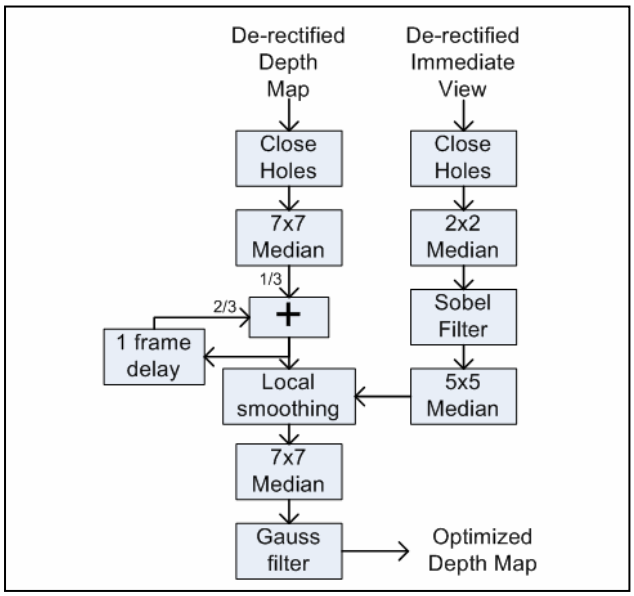

Figure 4 - Depth map optimization

from every direction. However, this is not necessary for every application.

A poor inclusion occurs if the object does not seem to belong to its environment, for instance a bright object in a dark environment or vice versa.

Naturally, a 3D video object can show distortions. Due to the low maturity of the generation processes of 3D video objects often severe distortions occur. These distortions and quality features have to be determined in order to assess the quality of these objects [4], [5].

These requirements show the variety of features of 3D video objects. Looking at the expenditure of the production of such objects it is useful to produce accordingly to the requested properties.

4.2. Limitation of Possible Viewpoints

A typical application according to synthetic natural hybrid coding (SNHC) is the integration of natural video objects into a 3D environment [6], [7]. Depending on the environment not every viewpoint is meaningful. In Figure 2 a typical camera setup restricted to 75 degrees around the object is shown. Of course, this limitation will be the same for the usage of the 3D video object.

Another limitation is given by the nearest allowed distance to the object. If it would be viewed closer the pixel may become apparent. Cameras having a resolution according to ITU-R BT.601 are suitable to record typical 3D video objects in not more than a medium shot or a medium close-up [3].

4.3. Assessment of 3D Video Objects

In order to estimate the aimed quality most relationships between the processing and any quality metric have to be

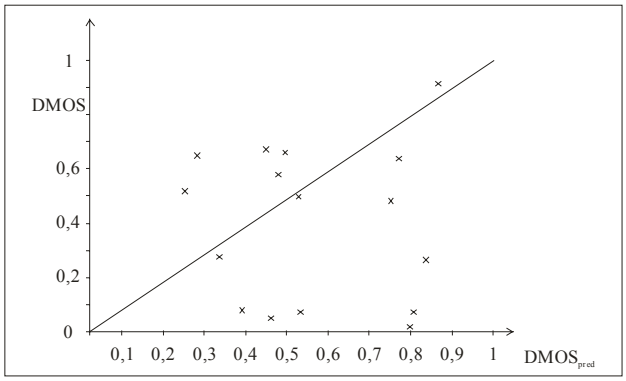

Figure 5 - Scatter diagram: DMOS and its prediction.

known. Such a quality metric requires an objective assessment of 3D video objects. In [4] and [8] a methodology to assess 3D video objects irrespective of their generation was shown. Using another (ideal) object as ground truth a 3D video object quality metric *3DVQM* based on regression to quality features has been modelled. The correlation with the subjective assessment will not be high when comparing different types of objects with different quality features (e. g. Spearman correlation: 0.1). Because of the regression-based method this 3DVQM can be adopted to a certain kind of 3D video objects and a certain quality feature. This can be reached by extended subjective tests (based on methods according to [9]). That is useful to find out relationships between the acquisition/processing and the attainable level of quality.

4.4. Example of Quality Assessment

In the following example the gradation of quality of 3D video objects was determined. These objects were generated by the image-based method explained in Section 2. The gradation was caused by different coefficients of the fundamental matrix. For this, different methods for the determination of correspondences were employed.

The employed 3DVQM was modelled with only one extended subjective test set. This was enough to attain a sufficient correlation between the subjective assessment (differential mean opinion score *DMOS*) and its prediction by the 3DVQM. In Figure 5 a scatter diagram of this assessment is shown. The Spearman correlation of this example was 0.45. In order to improve this correlation more extended subjective tests or improved feature extraction methods are necessary.

5. CONCLUSIONS

In this paper we presented a system for the image-based processing of natural 3D video objects at different levels of quality. Addressing the whole chain of the production process several methods were introduced in order to optimize and automate the required effort of time as well as the quality of the 3D video objects at different costs. An exemplary comparison between two levels of quality is depicted in Figure 6. In future development steps these methods will be enhance further to achieve a real time production of 3D video objects at high levels of quality.

6. REFERENCES

[1] 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.

[2] S. M. Seitz, "Image-Based Transformation of Viewpoint and Scene Appearance", PhD Thesis, pp. 24-26, USA, 1997

[3] G. Horna, L. L. Kreibich “3D-Videoobjektgenerierung mittels Multiview-Aufnahmen”, Student research project, Technische Universität Ilmenau, Ilmenau (Germany), 2004

[4] M. Rittermann, “A Proposal for the Quality Assessment of 3D Video Objects”, Proceedings of the 5th Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS), Lisbon (Portugal), 2004

[5] A. Smolic, K. Mueller, P. Merkle, T. Rein, M. Kautzner, P. Eisert, and T. Wiegand, “Free Viewpoint Video Extraction, Representation, Coding, and Rendering", IEEE International Conference on Image Processing, ICIP 2004, Singapore 2004

[6] International Organization for Standardization ISO/IEC JTC 1/SC 29/WG 11, “ISO 14496 Information Technology – Generic Coding of Audio-Visual Objects (MPEG-4)”, 1998pp

[7] H. Drumm, U. Kühhirt, M. Rittermann, U. Reiter, “Application Systems for MPEG-4”, IEEE International Symposium on Consumer Electronics ISCE’02, Erfurt (Germany), 2002

[8] M. Rittermann, "Quality Assessment of 3D Video Objects”, IEEE International Symposium on Consumer Electronics ISCE'03, Sydney (Australia), 2003

[9] International Telecommunication Union (ITU), "ITU-T Recommendation P.910 - Subjective Video Quality Assessment Methods for Multimedia Applications". Recommendation, 1999

Figure 6 – Different levels of quality with: manual search of correspondences (left) and automatic search (right).