A PROPOSAL FOR THE QUALITY ASSESSMENT OF 3D VIDEO OBJECTS

Marco Rittermann

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

ABSTRACT

In this paper I propose a methodology for the assessment of natural 3D video objects. Such objects become an important type of media objects in audiovisual 3D scenes which allow views from different viewpoints. For the assessment, a universally valid model for all kinds of 3D video objects, irrespective of their processing is introduced. This model shows that only the generated views with their masks can be compared with a representative selection of typical reference views or reference 3D video objects. The quality features of 3D video objects are pointed out. There are typical distortions, for example occurring while view synthesis. Some methods for the detection of such errors which consider the masking information are proposed. In future work I intend to collapse the founded quality parameters to a 3D video object quality metric (3DVQM) which has to be verified by subjective tests.

1. INTRODUCTION

In conventional audiovisual applications, visual content is represented by a single video stream. A novel approach is the content-based description of audiovisual content where all parts are coded separately as media objects, like in the MPEG-4 standard defined by the ISO [1]. The media objects can be composed in a 2D or a 3D scene. A typical example for a 2D composition is a video which consists of a sprite as background and some arbitrarily shaped video objects in the foreground. A 3D composition of several media objects can be prepared by an object-based production, e. g. a virtual studio production [2]. A coded 3D scene with media objects in it allows the user to navigate across the scene or to select from given viewpoints. But a major problem is the missing third dimension of natural video objects. When navigating across the scene the perspective onto the video object becomes incorrect [3]. Therefore, natural 3D video objects are required. There are several techniques for 3D video in development and they are investigated in MPEG [4], [5]. A uniform description of 3D video objects is necessary in order to compare them. For this, a common model of 3D video objects is introduced in Section 2 of this paper. The development of 3DVO techniques needs metrics for their evaluation. Such metrics have not been available because it is a new topic with unresolved questions: What can serve as ground truth? What quality features does a 3DVO show? A methodology for the comparison and quality assessment is proposed in Sections 3 and 4. Naturally, some existing assessment techniques for 2D video and 2D shapes can be applied for that issue.

2. THREE-DIMENSIONAL VIDEO OBJECTS

2.1. Characteristics

The major characteristic of 3D video objects is the availability of different perspectives at every given time. Ideally, the object can be viewed from every point around it. The MPEG describes 3D video objects as results of multiple view video. They comprise shape and appearance [5]. This shape can be represented in different ways (polygon meshes, implicit surfaces, depth images, or multiple layered depth images).

2.2. Capturing and Processing

The recording of 3D video objects requires much expenditure. In most cases a convergent set-up of a good few cameras is used. This can be supplemented or replaced by a new type of camera which records also depth values for every pixel. But also other new technologies for 3D video recording are in development, e. g. the use of LIDAR (Light detection and Ranging). However, the different 3D video recording techniques result in different outcomes (e. g. multiple view video or depth maps).

Naturally, the different recording techniques but also varying applications require different schemes of processing. One class of processing schemes is image-oriented and another class is shape-oriented [6] – [10].

A typical example for an image-oriented processing is the synthesis of intermediate views using morphing methods. In a first step the cameras' views incl. their shapes are rectified in order to morph one-dimensionally along the epipolar lines. This needs the fundamental matrix which is an outcome of the fixed calibration or of the self calibration. In another step the depth values are calculated by comparing corresponding points. The last step is the view synthesis. For a given viewpoint the resulting view is calculated by the rectified views, the correspondences, and the depth map.

2.3. Common Model

A scientific examination of 3D video objects requires a uniform description of them. In Figure 1, a model is introduced which is applicable to all types of 3D video objects [11]. This model starts at the recorded object, i. e. any natural three-dimensional object, e. g. an actor. The recording always yields a primary representation which consists of the cameras' output. In the next step, this unprocessed output has to be prepared for the synthesis of arbitrary views. The view synthesis uses the viewpoint's data and results in a 3D video object. In this proposal the term *3DVO* stands for the generated video sequence according to one specified examination.

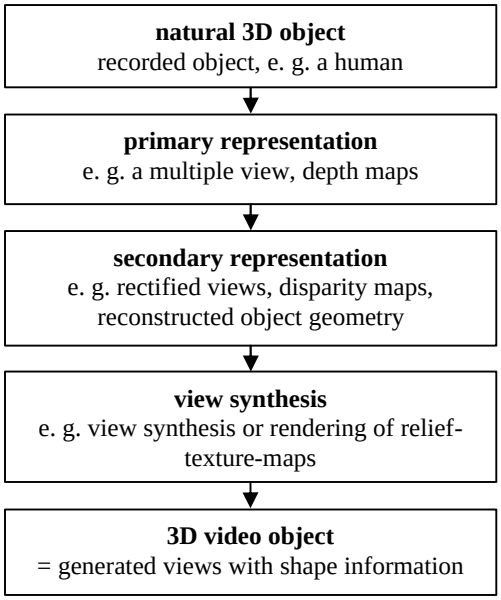

Figure 1 – Model of the acquisition of 3D video objects

This model is used to evaluate different 3D video objects representing one natural object. This is described in the next section.

3. EVALUATION OF 3D VIDEO OBJECTS

3.1. Quality Features

In order to evaluate a 3DVO the quality features of its generated views have to be determined.

- Such a generated video sequence shows the typical impairments of 2D video, too (e. g. blurring, noise).
- The object's shape of the generated views may be distorted which is very annoying. (In Figure 2, a typical distortion of the shape is to be seen down right.)
- There are typical distortions within the object, e. g. along epipolar lines. (In Figure 2, typical distortions at epipolar lines are to be seen in the area of the mouth.)
- The generated view may be taken from a wrong perspective. The deviation may be static, depending on time, or depending on the viewpoint.

Depending to the processing used, typical distortions will appear (e. g. warping methods produce distortions along epipolar lines). Another class of distortions is a result of limitations of the recording system (e. g. occlusions).

When watching a 3D video object it may be included badly into the 3D environment (e. g. because of a rim). Another typical error is a differing depth of focus. These are important quality features of the composition but not for the individual 3DVO. Therefore, these quality features are not considered here.

3.2. Ground Truth

A comparison at the level of the primary or secondary representation (see Fig. 1) is not possible if differently generated 3D video objects have to be compared. Furthermore, this evaluation would not include the synthesis. But even at the level of the generated views a 3DVO has many appearances at a given time because it can be viewed from various points. These facts show that approaches for the evaluation of conventional 2D video (Full-Reference FR, Reduced-Reference RR, No-Reference NR [12]) can not be applied in their original sense.

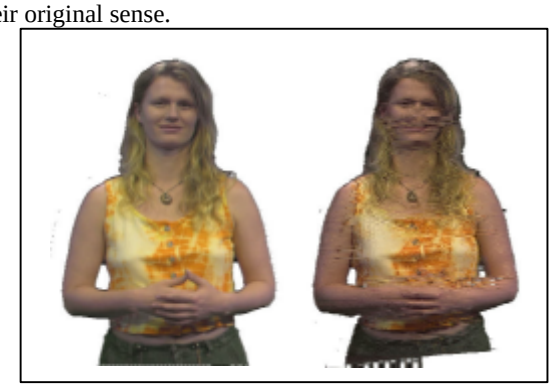

Figure 2 – Reference view and 3DVO view

A common solution is a comparison of generated views with reference views (incl. shape information) whereas both view sequences relate to the same examination of the object (see Fig. 2). Such a comparison results in evaluation of the recording and rendering/synthesis as a whole.

3.3. Previous Work

There are many metrics for the quality evaluation of conventional video [12], [13]. They are for instance regression-based, feature-extraction-based, or vision-calibrated [12]. The correlation with the subjective assessment is up to 0.94 [13]. Some principles of the feature extraction and the merging to a quality metric can be adopted to the quality assessment of 3D video objects.

By the introduction of arbitrarily shaped video the assessment of segmentation and the resulting shapes became necessary. In [14]–[16] some metrics are presented which allow an objective evaluation with or without a ground truth. A shape evaluation is important for 3D video objects, too,

because the generated view of a 3DVO contains shape information.

4. TECHNIQUES FOR THE 3DVQM

4.1. Procedure

Several steps are necessary for the quality assessment of 3D video objects. In the first step a calibration is useful to allow the extraction of quality features. Such a calibration is known from conventional video evaluation but that is only a spatial and temporal registration of a few pixels. A wrong perspective may result in a translation error of several pixels or in a wrong extension of the object. The calibration results (e. g. a detected focus error) are representing quality features and they are necessary for the calibration as well. The calibrated views can be used for the computation of statistical parameters and the detection of distortions. In the last step all parameters w_i will be collapsed to one 3D video object quality metric: 3DVQM. The single determined quality parameters have to be combined:

$$\text{3DVQM} = \sum_{i=0}^{n} b_i \cdot w_i$$

This weighting has to be verified by subjective tests (orientated to [17]). In Figure 3, the steps to the 3DVQM are shown.

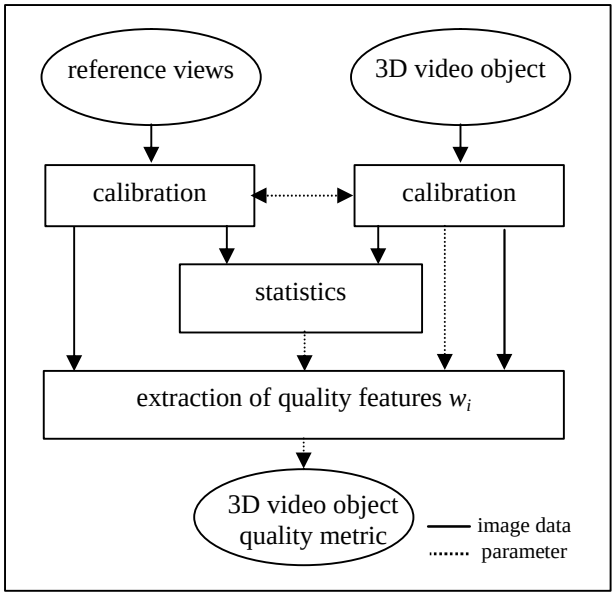

Figure 3 – Course of quality assessment

4.2. Correspondence Analysis

Correspondences analysis is applied for several purposes. It is used for the calibration as well as the extraction of distortions. Hierarchical block matching offers the possibility to extract different features at certain levels. For instance, a translation can be detected at level 0 and distortions can be detected at higher levels (see Figure 4)

Figure 4 – Detected distortions in block matching level 5

4.3. Shape Comparison

The shapes of the generated views can not be calibrated by pixel-based methods. The shape can be described by its contour function. The one-dimensional DFT of the contour function allows detecting a translation or a scaling.

$$T(q) = \sum_{p=0}^{P-1} t(p) \cdot e^{-\frac{2 \cdot p \cdot p \cdot q}{P} j}$$

The absolute values of DFT coefficients are invariant to translations ($q>0$) and scalings result in a scaling of the coefficients.

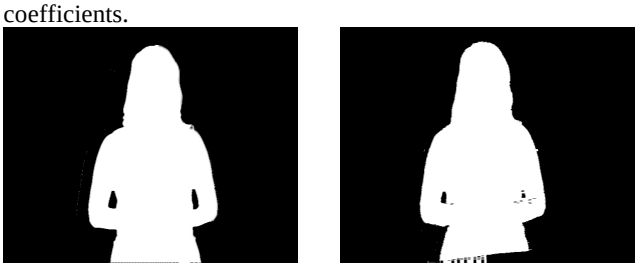

Figure 5 – Reference and disturbed shape

	r_o	r_1	r_2	r_3	r_4	r_5	r_6	r_7	r_8
R	442	140	18.2	10.8	5.4	7.2	3.8	4.3	2.2
D	408	143	18.6	7.6	5.8	8.4	3.4	2.6	4.1

Table 1 – Fourier coefficients of reference *R* and disturbed shape *D*

In Figure 5, a reference shape and a disturbed shape are shown. In Table, 1 the Fourier coefficients (0…8) of both contour functions are listed. The deviation of r_0 detects a translation of the whole shape. The deviations of the higher coefficients (r_5…r_8) detect impairments along the curve.

4.4. Statistic Parameters

A 3DVO shows typical quality features of a conventional video, too. Therefore, some parameters of 2D video can be used for the calibrated view. Especially, the differences of the spatial and temporal information (*SI*, *TI*; see [18], [19]) are a basis for quality parameters. For instance, distortions along epipolar lines produce deviations of SI. Table 2 shows a comparison of a reference 3DVO with 4 impaired 3DVO.

For this, typical distortions during the morphing process were simulated by insufficient correspondences. The difference of all SI values was computed as:

$$\Delta SI = \frac{1}{256 \cdot n} \sum_{i=1}^{n} \sum_{k=1}^{256} \left| SI_{3DVOcorr,i,k} - SI_{ref,i,k} \right|_{trunc}$$

(256 blocks per video object plane)

no.	1	2	3	4
subj. assessment	-1	-0.5	-1.5	-3
ΔSI	0.297	0.246	0.441	4.37

Table 2 – Subjective assessment and changing of SI for four 3DVO with defective morphing

5. CONCLUSIONS

In this paper I have proposed a methodology for the quality assessment of 3D video objects. There are several methods for the acquisition of 3D video objects in development which are basically different. Therefore the object's representations are not comparable. I have presented a universally valid model for all types of 3D video objects. Furthermore, I have introduced a methodology to compare such video objects to reference views and assess them. In order to calibrate the views and compute quality parameters some methods have been proposed. Hierarchic block matching can be used to find translation errors at level 0 and distortions of the view synthesis at higher levels. The one-dimensional DFT of the shape's contour function is suitable to detect translations and scalings regardless of the pixel representation. Statistic parameters allow evaluating for instance the dynamic quality features. In the future work we intend to collapse the quality parameters to a 3D video object quality metric: 3DVQM.

6. REFERENCES

[1] 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

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

[3] O. Grau, M. Price, and G. A. Thomas, "Use of 3-D Techniques for Virtual Production (PROMETHEUS)", SPIE Conference, San Jose (USA), 2001

[4] A. Smolic and D. Mc Cutchen, "Effecient Representation and Coding of Omni-Directional Video Using MPEG-4", Proceeding WIAMIS 2003, 4th European Workshop on Image Analysis for Multimedia Interactive Services, London (UK), 2003

[5] International Organization for Standardization ISO/IEC JTC 1/SC 29/WG 11, "Applications and Requirements for 3DAV", Document N5877, Trondheim (Norway), 2003

[6] M. M. de Oliveira Neto, "Relief Texture Mapping", Ph.D. Thesis University of North Carolina at Chapel Hill (USA), 2000

[7] M. Rittermann and M. Schuldt, "3D Television Production Based on MPEG-4 Principles", The 11th International Conference in Central Europe on Computer Graphics, Visualization, and Computer Vision; Plzen (Czech Republic), 2003

[8] A. Smolic, C. Fehn, and K. Müller, "MPEG 3DAV – Video Based Rendering for Interactive TV Applications", 10. Dortmunder Fernsehseminar, Dortmund (Germany), 2003

[9] S. M. Seitz and C. R. Dyer, "View Morphing", Proceedings of SIGGRAPH 96, pp. 21-30, USA, 1996

[10] J.-R. Ohm and K. Müller, "Incomplete 3D - Multiview Representation of Video Objects", IEEE Transactions on Circuits and Systems for Video Technology, special issue on SNHC, pp. 389-400, March 1999

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

[12] Sarnoff Corporation, "Measuring Image Quality: Sarnoff's JNDmetrix Technology". Technology Overview, USA, 2002

[13] Video Quality Experts Group (VQEG), "Validation of Objective Models of Video Quality Assessment, Phase II", Draft Final Report, 2003

[14] P. Correia and F. Pereira, "Standalone Objective Evaluation of Segmentation Quality", Proc. Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS), Tampere (Finland), 2001

[15] P. Correia and F. Pereira, "Objective Evaluation of Video Segmentation Quality", IEEE Transactions on Image Processing, Vol. 12, No. 12, pp. 186-200, February 2003

[16] C. Erdem Eroglu and B. Sankur, "Performance Evaluation Metrics for Object-Based Video Segmentation", EUSIPCO 2000: 10th European Signal Processing Conference, pp. 917-920, Tampere (Finland), 2000

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

[18] American National Standards Institute (ANSI), "ANSI T1.801.03-1996: Digital Transport of One-Way Video Signals – Parameters for Objective Performance Assessment", 1996

[19] International Telecommunication Union (ITU), "ITU-T Recommendation J.144 - Objective perceptual video quality measurement techniques for digital cable television in the presence of a full reference", Recommendation, 2001