D30.2 Technical Report #2 on 3D Coding Techniques

Project Number: 511568
Project Acronym: 3DTV
Title: Integrated Three-Dimensional Television – Capture, Transmission and Display

Deliverable Nature: R
Number: D30.2
Contractual Date of Delivery: M29
Actual Date of Delivery: M30
Report Date: 21 February 2007
Task: WP9
Dissemination level: PU
Start Date of Project: 01 September 2004
Duration: 48 months
Organisation name of lead contractor for this deliverable: FhG-HHI

Name of responsible: Aljoscha Smolic (smolic@hhi.de)

Editor: Aljoscha Smolic (smolic@hhi.de)

21 February 2007

3D Coding Techniques
TC2 Technical Report 2

EDITOR

Aljoscha Smolic (**FhG-HHI**)

Contributing Partners to Technical Report:

Bilkent University (**Bilkent**)
Fraunhofer Gesellschaft zur Förderung der angewandten Forschung e.V. (**FhG-HHI**)
Tampere University of Technology (**TUT**)
Informatics and Telematics Institute, Centre for Research and Technology Hellas (**ITI-CERTH**)
Koç University (**KU**)
Middle East Technical University (**METU**)
University of Hannover (**UHANN**)
Technische Universitaet Berlin (**TUB**)

REVIEWERS

Georgios Triantafyllidis (**ITI-CERTH**)
Gozde Bozdagi Akar (**METU**)
Nikolce Stefanoski (**UHANN**)
Michael Droese (**TUB**)

Project Number: 511568
Project Acronym: 3DTV
Title: Integrated Three-Dimensional Television – Capture, Transmission and Display

TABLE OF CONTENTS

Executive Summary

Progress of research in WP9 during the last year is reported in this deliverable.

The NoE plays a worldwide leading role in multi-view video coding (MVC). Researchers from the NoE have been leading the corresponding standardization activities in MPEG and the JVT. Technology proposals from the NoE performed best within the related MPEG Call for Proposals, thus the fundament of this new standard has been developed within the NoE. Participation in MPEG and the JVT will be continued. Besides these more applied standardization activities, also basic research on MVC has been continued, resulting in various publications.

Research on 3D mesh compression has been continued successfully as well. One focus was on dynamic mesh compression. A progressive and a rate-distortion optimized coder have been developed in close collaboration. These codecs perform very well in international comparison as shown in the resulting publications. Recently, MPEG restarted activity on specification of a new standard for compression of dynamic 3D meshes, which was initiated by contributions from the NoE. It is planned to contribute to the development of this standard in related Core Experiments. A second line of research targets image-based compression of 3D meshes. This work has reached a good standard as proven by the related publications.

The NoE performed pioneering research on compression of holograms during the first 18 months. Unfortunately, this research was paused during the last year. However, Plzen University recently started studying these issues and will contribute results to the next Technical report.

The NoE performed pioneering research in the area of multiple description coding (MDC) for 3D. So far only a few publications about MDC for 3D meshes were known. In this area research in the NoE has reached a very good standard and significant contributions have been made. In addition MDC for stereo was first considered in the NoE. Future research will include studies of performance over real transmission channels, such as DVB-H transport, as well as more general error resilience for 3D meshes.

Watermarking for 3D is also a pioneering research area. Contributions have been made to watermarking of 3D mesh models that has been studied to some extend outside the NoE as well. In addition watermarking of image-based rendering data has been investigated within the NoE for the first time. Such algorithms will be very important for future protection of image-based 3DTV data. Future research will continue along these lines to ensure future protection of 3D data.

In general conclusion, the research on 3D compression within the NoE reached a high level of excellence, for both scientific fundamentals and more applied development of international standards. Partners formed a strong network and benefit from each others expertise and resources.

1. Introduction

This Technical Report 2 summarizes the research results in WP9 during the last year. Research in the NoE converged more and more into joint activities, where partners benefit from collaboration and synergies. The work has been clustered into the following high priority research areas:

- Multi-view video coding,
- 3D mesh compression,
- Hologram compression,
- Multiple description coding for 3D,
- Watermarking for 3D.

This list includes well established but nevertheless highly active fields like for instance 3D mesh compression as well as highly innovative topics like hologram compression. In the latter for instance the NoE was doing pioneering work. Worldwide only two groups outside the NoE are known to work on such issues (National University of Ireland in Maynooth and Kwangwoon University in Seoul Korea). Unfortunately, this path of research was paused during this reporting period (see below in 2.3). Multi-view video coding is related to compression of any kind of pixel data, such as video, stereo video, multi-view video, but also associated per-pixel depth data, etc. This wide field is partially well established but partially also very innovative, and in any case highly actual. For instance MPEG started to work on a new standard for multi-view video coding and partners from the NoE play a leading role in development of this standard. 3D mesh compression is related to this type of classical computer graphics representations of 3D data. Also here partners contribute to activities within MPEG. Usage of holographic data for communications is very new as explained before. Multiple description coding is studied for the first time with focus on 3D within the NoE. Also work on watermarking of various 3D data is, besides for 3D meshes, for the first time being done within the NoE.

For all these topics groups of institutions have been formed that carry out joint research in an integrated and collaborative manner. This includes exchange of know-how, software, data, researchers and students, common usage of equipment and infrastructure, with focus on common publications. Links have also been established to other WPs along the 3D-TV chain.

Section 2 of this document gives an analysis of the results so far. These are brief summaries each related to a specific part of the annex. The annex contains all the descriptions in detail. This may be published or unpublished papers, MPEG documents or other types of reports. The analysis in section 2 is intended to extract the essence of all of this work and to put all into a concise context. The assessment is done by research topic listed above by a “summary, conclusions and plans” section for each. An overall summary and conclusions can be found in section 3.

2. Analysis of Results

This section gives an analysis of the work and results of the NoE within the reporting period as a collection of abstracts. The more detailed descriptions of each topic can be found in the related annex. Research in WP9 has been clustered into the main subjects multi-view video coding (MVC), 3D mesh compression, hologram compression, multiple description coding (MDC) for 3D, and watermarking for 3D. Each of the related sub-sections in the following includes overall summary, conclusions, and plans for the related research topic as a whole.

2.1. Multi-view video coding (MVC)

Multi-view video coding (MVC) is related to compression of any kind of pixel data, such as video, stereo video, multi-view video, but also associated per-pixel depth data, etc. It is a key technology serving a wide range of applications, such as free viewpoint video (FVV) and 3DTV. More precisely, MVC is characterized by efficient methods for source and channel coding for multi-view video and related data, as disparity maps and camera parameters.

A main focus of work during this reporting period was the development of the new MPEG standard for MVC. Participants from the NoE submitted proposals in response to the Call for Proposals, performed Core Experiments, provided reference software, edited standards documents, and chaired working groups. Also the new MPEG standard for the video plus depth format MPEG-C Part-3 has been developed with strong participation from the NoE. Several conference publications listed below are as well related to this work. Further research was focused on scalability for MVC, stereo video coding including quality evaluation issues, and bit allocation for MVC. In the following abstracts regarding these topics are given, followed by an analysis of results, conclusions and future plans.

2.1.1. Extending Single-view Scalable Video Coding to Multi-view based on H.264/AVC

Authors: Michael Dröse, Carsten Clemens, and Thomas Sikora

Institutions: Technical University of Berlin

Publication: IEEE Int. Conf. on Image Processing (ICIP'06), Atlanta, GA, USA, 08.10.2006 - 11.10.2006

An extension of single-view scalable video coding to multiview is presented in this paper. Scalable video coding is recently developed in the Joint Video Team of ISO/IEC MPEG and ITU-T VCEG named Joint Scalable Video Model. The model includes temporal, spatial and quality scalability enhancing a H.264/AVC base layer. To remove redundancy between views a hierarchical decomposition in a similar way to the temporal direction is applied. The codec is based on this technology and supports open-loop as well as closed-loop controlled encoding.

The advantage of this approach lies in its compatibility to the state of the art single-view video codec H.264/AVC and its simple decomposition structure. Encoding a base view using H.264/AVC syntax, any standard single-view decoder is able to decode the data. The

hierarchical decomposition structure allows efficient access to all views and frames inside a view. This is especially important for video-based rendering and multi-view displays, which have different requirements.

The chosen decomposition structure also supports parallel processing. Gain in objective as well as subjective quality was achieved for some test sequences using a single layer. The results were compared to JSVM 5.1 (simulcast).

2.1.2. Core Experiment on Simplified Prediction Structure

Authors: Michael Dröse, and Carsten Clemens

Institutions: Technical University of Berlin

Publication: ISO/IEC JTC1/SC29/WG11, Montreux, Switzerland, 03.04.2006 - 07.04.2006

This MPEG contribution is dealing with an encoder optimization. A simplification of the prediction structure as utilised in the reference software has been investigated. The proposed structure omits inter-view prediction at higher temporal levels, i.e. only pictures which cannot be temporarily predicted are inter-view predicted. The experiments have been carried out on the data set of the "Call for Proposals" at ISO/MPEG. The quality is kept constant, whereas the increase in bit rate is measured. As an outcome of this experiment it was shown, that most of the coding gain comes from those pictures, which do not include a temporal reference.

2.1.3. Core Experiment on a Dyadic Inter-view Prediction Structure

Authors: Michael Dröse

Institutions: Technical University of Berlin

Publication: ISO/IEC JTC1/SC29/WG11, Klagenfurt, Austria, 17.07.2006 – 21.07.2006

The experiment is dealing with a modification of the prediction structure towards dyadic hierarchical inter-view prediction. RD-performance was compared to the MVC reference model of ISO/MPEG. Negligible gain in RD-performance was achieved for certain sequences. Especially sequences with a small baseline and therefore a small disparity range could show some gain.

2.1.4. Temporal and Spatial Scaling For Stereoscopic Video Compression

Authors: A. Aksay, C. Bilen, E. Kurutepe, T. Ozcelebi, G. Bozdagi Akar, M. R. Civanlar, A. M. Tekalp

Institutions: Koç University (KU), Middle East Technical University (METU)

Publication: IEEE EUSIPCO 2006, Florence, Italy, Sept. 2006

In stereoscopic video, it is well-known that compression efficiency can be improved, without sacrificing PSNR, by predicting one view from the other. Moreover, additional gain can be

achieved by subsampling one of the views, since the Human Visual System can perceive high frequency information from the other view. In this work, we propose subsampling of one of the views by scaling its temporal rate and/or spatial size at regular intervals using a real-time stereoscopic H.264/AVC codec, and assess the subjective quality of the resulting videos using DSCQS test methodology. We show that stereoscopic videos can be coded at a rate about 1.2 times that of monoscopic videos with little visual quality degradation.

2.1.5. Towards Compound Stereo-Video Quality Metric: A Specific Encoder-Based Framework

Authors: A. Boev, A. Gotchev, K. Egiazarian, A. Aksay, G. Bozdagi Akar

Institutions: Tampere University of Technology (TUT), Middle East Technical University (METU)

Publication: IEEE SSIAI 2006, Denver, Colorado, USA, March 2006

We suggest a compound full-reference stereo-video quality metric composed of two components: a monoscopic quality component and stereoscopic quality component. While the former assesses the trivial monoscopic perceived distortions caused by blur, noise, contrast change etc., the latter assesses the perceived degradation of binocular depth cues only. We use the structural similarity index as a measure for perceptual similarity and design a multiscale algorithm for obtaining a perceptual disparity map and a stereo-similarity map to be used in the suggested metric. We verify the performance of the metric with subjective tests on distorted stereo images and coded stereo-video sequences with a final aim to build a perceptually-aware feedback for a H.264 based stereovideo encoder.

2.1.6. Content-Aware Bit Allocation in Scalable Multi-View Video Coding

Authors: Nükhet Özbek, A. Murat Tekalp

Institutions: Ege University, Koç University

Publication: MRCS workshop, Istanbul, Sept. 2006

We propose a new scalable multi-view video coding (SMVC) method with content-aware bit allocation among multiple views. The video is encoded off-line with a predetermined number of temporal and SNR scalability layers. Content-aware bit allocation among the views is performed during bitstream extraction by adaptive selection of the number of temporal and SNR scalability layers for each group of pictures (GOP) according to motion and spatial activity of that GOP. The effect of bit allocation among the multiple views on the overall video quality has been studied on a number of training sequences by means of both quantitative quality measures as well as qualitative visual tests. The number of temporal and SNR scalability layers selected as a function of motion and spatial activity measures for the actual test sequences are “learned” from these bit allocation vs. video quality studies on the training sequences. SMVC with content-aware bit allocation among views can be used for multi-view video transport over the Internet for interactive 3DTV. Experimental results are provided on stereo video sequences.

2.1.7. Scalable Multi-view Video Coding for Interactive 3DTV

Authors: Nükhet Özbek, A. Murat Tekalp

Institutions: Ege University, Koç University

Publication: Proc. ICME 2006, International Conference on Multimedia and Expo, Toronto, Ontario, Canada, July 2006

A standard for scalable video coding (SVC) is currently being worked on by the ISO MPEG Group. Work on standardization of multiple-view video coding (MVC) has also recently started under the ISO MPEG. Although there are many approaches published on SVC and MVC, there is no current work reported on scalable multi-view video coding (SMVC). This paper presents new coding structures for scalable stereo and multi-view video coding. The proposed structures are implemented as extensions to the JSVM software and resulting bitrates and PSNR are demonstrated. SMVC can be used for transport of multiview video over IP for interactive 3DTV by dynamic adaptive combination of temporal, spatial, and SNR scalability according to network conditions.

2.1.8. 3D Video and Free Viewpoint Video – Technologies, Applications and MPEG Standards

Authors: Aljoscha Smolic, Karsten Mueller, Philipp Merkle, Christoph Fehn, Peter Kauff, Peter Eisert, Thomas Wiegand

Institution: Fraunhofer HHI

Publication: Proc. ICME 2006, International Conference on Multimedia and Expo, Toronto, Ontario, Canada, July 2006

An overview of 3D and free viewpoint video is given in this paper with special focus on related standardization activities in MPEG. Free viewpoint video allows the user to freely navigate within real world visual scenes, as known from virtual worlds in computer graphics. Examples are shown, highlighting standards conform realization using MPEG-4. Then the principles of 3D video are introduced providing the user with a 3D depth impression of the observed scene. Example systems are described again focusing on their realization based on MPEG-4. Finally multi-view video coding is described as a key component for 3D and free viewpoint video systems. The conclusion is that the necessary technology including standard media formats for 3D and free viewpoint is available or will be available in the near future, and that there is a clear demand from industry and user side for such applications. 3DTV at home and free viewpoint video on DVD will be available soon, and will create huge new markets.

2.1.9. Efficient Compression of Multi-View Video Exploiting Inter-View Dependencies Based on H.264/MPEG4-AVC

Authors: P. Merkle, K. Müller, A. Smolic, T. Wiegand

Institution: Fraunhofer HHI

Publication: Proc. ICME 2006, International Conference on Multimedia and Expo, Toronto, Ontario, Canada, July 2006

Efficient Multi-view coding requires coding algorithms that exploit temporal, as well as inter-view dependencies between adjacent cameras. Based on a spatiotemporal analysis on the multi-view data set, we present a coding scheme utilizing an H.264/MPEG4-AVC codec. To handle the specific requirements of multi-view datasets, namely temporal and inter-view correlation, two main features of the coder are used: hierarchical B pictures for temporal dependencies and an adapted prediction scheme to exploit inter-view dependencies. Both features are set up in the H.264/MPEG4-AVC configuration file, such that coding and decoding is purely based on standardized software. Additionally, picture reordering before coding to optimize coding efficiency and inverse reordering after decoding to obtain individual views are applied. Finally, coding results are shown for the proposed multi-view coder and compared to simulcast anchor and simulcast hierarchical B picture coding.

2.1.10. Multi-view Video Coding Based on H.264/AVC Using Hierarchical B-Frames

Authors: K. Müller, P. Merkle, H. Schwarz, T. Hinz, A. Smolic, T. Wiegand

Institution: Fraunhofer HHI

Publication: Proc. PCS 2006, Picture Coding Symposium, Beijing, China, April 2006

New applications such as 3D video and free viewpoint video require efficient compression of multi-view video, where the same scene is captured by several synchronized cameras. In a statistical analysis it is shown that multi-view video contains a large degree of inter-view statistical dependencies in addition to the temporal statistical dependencies that can be exploited for video compression. Based on this analysis an efficient prediction structure is presented that applies hierarchical B pictures in temporal and inter-view prediction. The multi-view video data are reorganized into a single uncompressed video stream that is fed into a standard H.264/MPEG4-AVC encoder, thus the resulting bitstream is standard-conforming. The presented results including objective and subjective evaluation prove the efficiency of our approach. It has been selected by MPEG as reference for the development of a new standard for multi-view video coding that is expected to be available in early 2008.

2.1.11. Applications, requirements and technical solution for the encoding of depth and parallax maps

Authors: Arnaud Bourge, Fons Bruls, Jan van der Meer, Yann Picard, Chris Varekamp, Aljoscha Smolic, Thomas Wiegand, Martin Borchert, Martin Beck

Institutions: Philips, Fraunhofer HHI, 3D-IP

Publication: ISO/IEC JTC1/SC29/WG11, Doc M12942, Bangkok, Thailand, January 2006

In this joint work with Philips and 3D-IP a proposal for a new MPEG standard for 3DTV was developed. The proposed format is based on the ATTEST approach to 3DTV using video and depth data for 3D rendering. There was great interest from industry inside MPEG for such a

simple, short-term and backwards compatible solution to enable 3DTV products and services. Specification of a corresponding standard has been initiated from this input contribution (*ISO/IEC 23002-3 Auxiliary Video Data Representation*, see below in 2.1.15). The document is not included in the annex.

2.1.12. Multiview Coding using AVC

Authors: Karsten Müller, Philipp Merkle, Aljoscha Smolic, Thomas Wiegand

Institutions: Fraunhofer HHI

Publication: ISO/IEC JTC1/SC29/WG11, Doc M12945, Bangkok, Thailand, January 2006

This is a proposal developed by Fraunhofer HHI as response to the MPEP Call for Proposals on Multiview Video Coding. The core of the proposal is an efficient inter-view/temporal prediction structure that applies hierarchical B pictures in both dimensions. The output of the encoder is a fully H.264/AVC compliant bitstream (see also 2.1.9 and 2.1.10). Out of the 8 submitted proposals it performed best in formal subjective tests. It was therefore selected by MPEG as basis for the further development of the MVC standard (*ISO/IEC 14496-10:2005/Amd.4 Multiview Video Coding,* see below in 2.1.14). The document is not included in the annex.

2.1.13. Comparative Study of MVC Prediction Structures

Authors: Philipp Merkle, Aljoscha Smolic, Karsten Müller, Thomas Wiegand

Institutions: Fraunhofer HHI

Publication: Joint Video Team (JVT) of ISO/IEC MPEG & ITU-T VCEG, Doc JVT-V132, Marakech, Morocco, January 2007

This contribution reports results of a comparative study of different MVC prediction structures, with regard to compression efficiency and complexity. The results show that simplified structures perform very well compared to full MVC while being far less complex. Most of the gain of MVC compared to simulcast comes from inter-view prediction of anchor frames (i.e. those that are coded in I mode in independent coding of the views). Depending on the application, efficient solutions in terms of compression and complexity can be found. This has implications for a lot of issues currently under study in MVC (processing, memory, delay, parallel processing, random access, etc.). For instance, it might not be necessary to define new syntax that simplifies aspects full MVC (as currently under study), since simplified structures can be used instead. The document is not included in the annex.

2.1.14. Development of MPEG Standard for Multiview Video Coding

Authors: MPEG, including Aljoscha Smolić and others

Institutions: MPEG, including Fraunhofer HHI and others

Publication:

Bangkok, January 2006
M12727 AHG on 3D Video Coding
N7779 Subjective test results for the CfP on Multi-view Video Coding
N7798 Description of Core Experiments Multiview Video Coding

Montreux, April 2006
M13035 AHG on Multiview Video Coding
N8017 Request for 14496-10:2006/Amd.4 Multiview Video Coding
N8018 Technology under Study for Buffer Management and High-level Syntax in Multiview Video Coding
N8019 Description of Core Experiments in Multiview Video Coding
N8064 Multiview Video Coding Requirements

Klagenfurt, July 2006
M13394 AHG on Multiview Video Coding
N8218 Multiview Video Coding Requirements
N8244 Joint Multiview Video Model (JMVM) 1
N8245 JMVM 1 Software
JVT-T100 Overview Presentation on MVC
JVT-T207 Common Test Conditions for Multiview Video Coding

Hangzhou, October 2006
N8458 Working Draft 1 of ISO/IEC 14496-10:2005/Amd.4 Multiview Video Coding
N8459 Joint Multiview Video Model (JMVM) 2
N8460 JMVM 2 Software
JVT-U016 AHG Report: JMVM text editing
JVT- U211 Common Test Conditions for Multiview Video Coding

Marakech, January 2007
N8753 Working Draft 2 of ISO/IEC 14496-10:2005/Amd.4 Multiview Video Coding
N8754 Video Joint Multiview Video Model (JMVM) 3
N8755 JMVM 3 Software
JVT-V016-Q AHG Report: JMVM & JD text editing

These are working documents of MPEG and the JVT for the development of the MVC standard. This includes specifications (Working Draft, JMVM), reference software, ad-hoc group reports, test conditions, requirements as well as other reports and documents. They have been prepared under the lead of and with strong influence from researchers of the NoE. The work has been moved to the JVT of MPEG and VCEG, and the final standard is scheduled to be available in 2008. The documents are not included in the annex.

2.1.15. Development of MPEG Standard for Auxiliary Video Data Representation

Authors: MPEG, including Aljoscha Smolić, Christoph Fehn, and others

Institutions: MPEG, including Fraunhofer HHI and others

Publication:

Bangkok
M12731 AHG on Depth Disparity Representation
N7898 WD of ISO/IEC 13818-1:200X/AMD 2 (SSVon2)
N7818 Working Draft 1.0 of ISO/IEC 23002-3 Auxiliary Video Data Representation

Montreux
N8037 Request for Subdivision: 23002-3 Auxiliary Video Data Representation
N8038 Text of ISO/IEC CD 23002-3 Auxiliary Video Data Representation
N8039 White Paper: Auxiliary Video Data Representation
N8094 Text of ISO/IEC 13818-1:200X/PDAM2 (Auxiliary Video Data)

Klagenfurt
N8259 Text of ISO/IEC FCD 23002-3 Auxiliary Video Data Representation
N8371 Text of ISO/IEC 13818-1:200X/FPDAM2 (Auxiliary Video Data)

Hangzhou
N8482 Study Text of ISO/IEC FCD 23002-3 Representation of Auxiliary Video and Supplemental Information

Marakech, January 2007
N8768 Text of ISO/IEC FDIS 23002-3 Representation of Auxiliary Video and Supplemental Information

These are working documents of MPEG for the development of the standard for video plus depth. Researchers from the NoE strongly influenced the development of this specification finally called MPEG-C Part-3 - Representation of Auxiliary Video and Supplemental Information. The standard has been finalized (FDIS in January 2007) and enables 3DTV. The documents are not included in the annex.

2.1.16. Analysis, conclusions, plans

The Call for Proposals (CfP) on MVC was prepared under strong influence from researches of the NoE on both, management and technical level as reported in the previous Technical Report. Responses to the CfP have been submitted to MPEG by Fraunhofer HHI and TU Berlin, as well as by 6 other institutions from industry and academia around the world (4 from Asia, 2 from the US). The proposal submitted by Fraunhofer HHI (using extended H.264/AVC with combined inter-view/temporal prediction based on hierarchical B pictures) performed best in formal subjective tests conducted by MPEG and was therefore chosen as the starting point for the development of the new standard. The 1st standard text and reference software was provided by Fraunhofer HHI.

Proposals from other institutions have been evaluated in core experiments and partially integrated. There is a strong participation of companies and academic institutions from Asia and the US within this process, only few Europeans are active. Technical work is focused on high-level syntax elements (e.g. for improved random access, buffer management, parallel processing), as well as on new coding tools (e.g. view interpolation prediction, illumination compensation, disparity vector prediction).

Since the standard will be an amendment to H.264/AVC (ISO/IEC 14496-10:2005/Amd.4 Multiview Video Coding) technical work has been moved to the JVT where researchers from the NoE play a leading role as well. Further work on the MVC standard will be one main focus of research in the NoE for the future as well. This may also include work on future extensions such as compression of multi-view video plus depth.

The specification of the video plus depth format MPEG-C Part-3 has been finalized under strong participation of researchers from the NoE. It now enables 3DTV. E.g. Philips plans to bring equipment based on this specification to the mass market in 2008. The specification may be extended to a layered depth image representation. In that case researchers from the NoE will again participate.

Outside MPEG/JVT basic research has been performed in the direction of scalable MVC, scaling for stereo video, quality metrics for stereo video, and bit allocation for MVC, resulting in several publications at prestigious conferences. This research will be continued and exploited in complete systems and applications.

2.2. 3D mesh compression

This research group is dealing with aspects of compression of static and dynamic 3D meshes. 3D meshes are a common form of representation for 3D objects. They are used in scene description languages like VRML97 or MPEG-4 BIFS and their rendering and texturing is supported by a lot of multimedia hardware. Recently, even mobile devices like smartphones are equipped with graphics chips which support rendering of 3D meshes. Such developments further the dissemination of static and dynamic 3D content. Developments like this stimulate not only a permanent improvement of existing applications, but they facilitate the development of new applications, like 3DTV. Compression of static and dynamic 3D content gets crucial importance in this context.

During the second period research in static and dynamic 3D mesh compression was continued. The focus was now on one hand on the development of a complete coder for geometry of static 3D meshes based on techniques from image compression. On the other hand research was conducted to improve the coding efficiency of already existing coders for dynamic 3D meshes, which were developed during the first period of research. Furthermore, existing coders were extended by new features, like spatial scalability. All together the following prospective topics were covered during the second period:

- image based compression of static 3D meshes (I-frame compression)
- spatially scalable predictive compression of dynamic 3D meshes (P-frame compression)
- rate-distortion optimized predictive compression of dynamic 3D meshes (P-frame compression)

The results of this work are reflected by the following documents.

2.2.1. Connectivity-Guided Predictive Compression of Dynamic 3D Meshes

Authors: Nikolce Stefanoski, Jörn Ostermann

Institutions: Universität Hannover

Publication: IEEE Int. Conf. on Image Processing (ICIP'06), Atlanta, GA, USA, 08.10.2006 - 11.10.2006

We introduce an efficient algorithm for real-time compression of temporally consistent dynamic 3D meshes. The algorithm uses mesh connectivity to determine the order of compression of vertex locations within a frame. Compression is performed in a frame to frame fashion using only the last decoded frame and the partly decoded current frame for prediction. Following the predictive coding paradigm, local temporal and local spatial dependencies between vertex locations are exploited. In this framework we present a novel angle preserving predictor and evaluate its performance against other state of the art predictors. It is shown that the proposed algorithm improves up to 25% upon the current state of the art for compression of temporally consistent dynamic 3D meshes.

2.2.2. Scalable Linear Predictive Coding of Time-Consistent 3D Mesh Sequences

Authors: Nikolce Stefanoski, Patrick Klie, and Jörn Ostermann

Institutions: Universität Hannover

Publication: 3DTV-CON, The True Vision Capture, Transmission and Display of 3d Video, 7-9 May 2007

We present a linear predictive compression approach for time-consistent 3D mesh sequences supporting and exploiting scalability. The algorithm decomposes each frame of a mesh sequence in layers employing patch based mesh simplification techniques. This layered decomposition is consistent in time, producing a time-consistent multiresolution representation of a mesh sequence. Following the predictive coding paradigm, local temporal and spatial dependencies between layers and frames are exploited for compression. Prediction is performed vertex-wise from coarse to fine layers exploiting the motion of already encoded 1-ring neighbor vertices for prediction of the current vertex location. It is shown that a predictive exploitation of the proposed layered configuration of vertices can improve the compression performance upon other state-of-the-art approaches for compression of time-consistent 3D mesh sequences.

2.2.3. Nonrectangular wavelets for multiresolution mesh analysis and compression

Authors: Kıvanç Köse, A. Enis Çetin, Uğur Güdükbay, Levent Onural

Institutions: Department of Electrical and Electronical Engineering, Department of Computer Engineering, Bilkent University

Publication: SPIE Defense and Security Symposium, April 9-13 2006, Orlando, FA, USA

We propose a new Set Partitioning In Hierarchical Trees (SPIHT) based mesh compression framework. The 3D mesh is first transformed to 2D images on a regular grid structure. Then, this image-like representation is wavelet transformed and SPIHT is applied on the wavelet domain data. The method is progressive because the resolution of the reconstructed mesh can be changed by varying the length of the 1D data stream created by SPIHT algorithm. Nearly perfect reconstruction is possible if full length of 1D data is received.

2.2.4. Dikdörtgensel Olmayan Dalgacık Dönüşümüne Dayalı Çok Çözünürlüklü Üç Boyutlu Model Analizi ve Sıkıştırılması (Nonrectangular wavelets for multiresolution mesh analysis and compression)

Authors: Kıvanç Köse, A. Enis Çetin, Uğur Güdükbay, Levent Onural

Institutions: Department of Electrical and Electronical Engineering, Department of Computer Engineering, Bilkent University

Publication: SIU 2006, Sinyal İşleme ve İletişim Uygulamaları Kurultayı, 17-19 Nısan, Antalya, Turkey

We propose a new Set Partitioning In Hierarchical Trees (SPIHT) based mesh compression framework. The 3D mesh is first transformed to 2D images on a regular grid structure. Then, this image-like representation is wavelet transformed and SPIHT is applied on the wavelet domain data. The method is progressive because the resolution of the reconstructed mesh can be changed by varying the length of the one-dimensional data stream created by SPIHT algorithm. Nearly perfect reconstruction is possible if all of the data stream is received.

2.2.5. Connectivity-guided adaptive wavelet transform for mesh analysis and compression

Authors: Kıvanç Köse, A. Enis Çetin, Uğur Güdükbay, Levent Onural

Institutions: Department of Electrical and Electronical Engineering, Department of Computer Engineering, Bilkent University

Publication: submitted to Signal Processing: Image Communication

We propose a new connectivity-guided adaptive wavelet transform based mesh compression framework. The 3D mesh is first transformed to 2D images on a regular grid structure by performing orthogonal projections onto the image plane. Then, this image-like representation is wavelet transformed using a lifting structure employing an adaptive predictor that takes advantage of the connectivity information of mesh vertices and wavelet domain data is encoded using “Set Partitioning In Hierarchical Trees” (SPIHT) method. The proposed approach is progressive because the resolution of the reconstructed mesh can be changed by varying the length of the 1D data stream created by the SPIHT algorithm. Almost perfect reconstruction is possible if the full length of the 1D data stream is received.

2.2.6. Rate-Distortion Optimization in Dynamic Mesh Compression

Authors: K. Müller, A. Smolic, M. Kautzner, T. Wiegand

Institution: Fraunhofer HHI

Publication: IEEE Int. Conf. on Image Processing (ICIP'06), Atlanta, GA, USA, 08.10.2006 - 11.10.2006

Recent developments in the compression of dynamic meshes or mesh sequences have shown that the statistical dependencies within a mesh sequence can be exploited well by predictive coding approaches. Coders introduced so far use experimentally determined or heuristic thresholds for tuning the algorithms. In video coding rate-distortion (RD) optimization is often used to avoid fixing of thresholds and to select a coding mode. We applied these ideas and present here an RD-optimized mesh coder. It includes different prediction modes as well as an RD cost computation that controls the mode selection across all possible spatial partitions of a mesh to find the clustering structure together with the associated prediction modes. The structure of the RD-optimized D3DMC coder is presented, followed by comparative results with mesh sequences at different resolutions.

2.2.7. Rate-Distortion-Optimized Predictive Compression of Dynamic 3D Mesh Sequences

Authors: K. Müller, A. Smolic, M. Kautzner, P. Eisert, T. Wiegand

Institution: Fraunhofer HHI

Publication: Signal Processing: Image Communication, Volume 21, Issue 9, October 2006, Pages 812-828

Compression of computer graphics data such as static and dynamic 3D meshes has received significant attention in recent years, since new applications require transmission over channels and storage on media with limited capacity. This includes pure graphics applications (virtual reality, games) as well as 3DTV and free viewpoint video. Efficient compression algorithms have been developed first for static 3D meshes, and later for dynamic 3D meshes and animations. Standard formats are available for instance in MPEG-4 3D Mesh Compression for static meshes, and Interpolator Compression for the animation part. For some important types of 3D objects, e.g. human head or body models, facial and body animation parameters have been introduced. Recent results for compression of general dynamic meshes have shown that the statistical dependencies within a mesh sequence can be exploited well by predictive coding approaches. Coders introduced so far use experimentally determined or heuristic thresholds for tuning the algorithms. In video coding rate-distortion (RD) optimization is often used to avoid fixed thresholds and to select the optimum prediction mode. We applied these ideas and present here an RD-optimized dynamic 3D mesh coder. It includes different prediction modes as well as an RD cost computation that controls the mode selection across all possible spatial partitions of a mesh to find the clustering structure together with the associated prediction modes. The general coding structure is derived from statistical analysis

of mesh sequences and exploits temporal as well as spatial mesh dependencies. To evaluate the coding efficiency of the developed coder, comparative coding results for mesh sequences at different resolutions were carried out.

2.2.8. 3D Model Compression Using Image Compression Based Methods

Authors: K. Kose

Institution: Bilkent University

Publication: Master's Thesis at Bilkent University, January 2007

A Connectivity-Guided Adaptive Wavelet Transform based mesh compression algorithm is proposed. On the contrary to previous work, the proposed method uses 2D image processing tools for compressing the mesh models. The 3D models are first transformed to 2D images on a regular grid structure by performing orthogonal projections onto the image plane. This operation is computationally simpler than parameterization. Also it is not needed to find cuts on the 3D models to transform them to 2D images.

The neighborhood concept in projection images is different from 2D images because two connected vertex can be projected to isolated pixels. Connectivity data of the 3D model defines the interpixel correlations in the projection image. Two pixels have correlation if their respective vertices have a connection between them. So by using connectivity information, pixels can be predicted from each other.

Since the correlation between the pixels is dependent on being connected, not being near to each other, the wavelet transforms used in image processing do not give good results on this representation. Connectivity-Guided Adaptive Wavelet Transform is defined to take advantage of interpixel correlations in the image-like representation. Using the proposed wavelet transform the pixels in the detail subbands are predicted from their connected neighbors in the low-pass subbands of the wavelet transform.

The resulting wavelet data is encoded using either "Set Partitioning In Hierarchical Trees" (SPIHT) or JPEG2000. SPIHT approach is progressive because different resolutions of the mesh can be reconstructed from different partitions of SPIHT bitstream. On the other hand, JPEG2000 approach is a single rate coder. The quantization of the wavelet coefficients determines the quality of the reconstructed model in JPEG2000 approach. The JPEG2000 approach can be converted to a progressive algorithm if the projection image is broken up into tiles and those tiles are coded separately.

Simulations using different basis functions show that lazy wavelet basis gives better results. The results are improved using the Connectivity-Guided Adaptive Wavelet Transform with lazy wavelet filterbanks. SPIHT based algorithm is observed to be superior to JPEG2000 based mesh coder and MPEG-3DGC in rate-distortion. The results can be improved by finding better projection methods so that less number of vertices are lost.

2.2.9. Analysis, conclusions, plans

Dynamic 3D meshes are represented as a series of static meshes called key frames. They are denoted as time-consistent if connectivity is not changing throughout the mesh sequence, whereas vertices are changing their location from frame to frame, describing the evolution of the mesh geometry in time. These type of dynamic meshes are used as input for the coders addressed in this documents.

In these documents techniques were investigated, developed, and improved which are necessary to efficiently encode a dynamic 3D mesh, i.e. a sequence of static 3D meshes. Dependencies within a single frame and between subsequent frames are exploited for compression. In this TR our focus was on the investigation and improvement of the compression performance compared to the results from first research period. Dependencies should be exploited within and between frames in the manner IPPP… . This means that the first frame is an I-frame, i.e. compression is performed using only itraframe dependencies, while subsequent frames are called P-frames and are compressed by one-directional prediction from previous already encoded frames.

A complete coder for compression of the vertex locations of I-frames was developed, which is described in the documents 2.2.3, 2.2.4, 2.2.5, and 2.2.8. Here static 3D meshes are first transformed into an image like representation and then compressed using wavelet-techniques, which were successfully applied in the area of image compression.

For compression of P-frames two coders, which were developed during the first research period, are further improved and extended. The first coder encodes vertex locations by applying clustering and sophisticated DPCM encoding of the clustered motion vector field of a dynamic mesh. Octree-clustering is applied in order to group motion vectors between two subsequent frames, assuming that motion vectors located in one cube or cluster of the octree describe similar motion. In the second period of research different coding modes are introduced for compression for single motion vectors and clusters. The procedure of octree-clustering was improved by employing rate-distortion criteria for determining the optimal mode and granularity of the octree, in order to compress all motion vectors in the most efficient way. A more in depth description of this approach can be found in the documents 2.2.6 and 2.2.7.

The second coder, based on connectivity-guided predictive compression (2.2.1), was extended during the second research period to a spatially scalable coder. Now it employs connectivity based mesh simplification in order to define a time-consistent spatial decomposition in layers of the whole dynamic 3D mesh. Already encoded vertex locations of previous frames and already encoded layers of the current frame are exploited for prediction of vertex locations. Here the assumption is exploited that vertices in the neighbourhood of a given vertex describe similar motion. Documents 2.2.1 and 2.2.2 discuss this coder more detail.

In the second period of research we showed that significant improvements in compression efficiency are still possible. A sophisticated rate-distortion analysis and even an extension to spatial scalability can lead to additional gains in bit-rate, due to a better exploitation of inter-vertex dependencies in time and space. In future we want besides I-frames and P-frames to support also B-frames, i.e. to employ bi-directional predictive coding, in order to improve coding efficiency even more. Coding could be performed in groups of meshes where

dependencies are exploited like follows: IBBBPBBB. An application of one transform to a whole group of meshes in spatial and temporal direction in order to decorrelate from spatio-temporal dependencies is an alternative approach, which should be examined and could lead to even higher gains than predictive approaches.

Currently we are using as input for our coders mostly synthetic mesh sequences generated with 3D animation tools, like Blender, 3ds Max, etc. In our future research we intend to intensify our collaboration with the subgroup "Time-varying surface representation" from WP8, which is working on the representation of real-world dynamic 3D meshes. First consultations in this direction already exist with our partners from Pilzen. Among other we intend to work together on "dynamic mesh simplification" using Plzen's know-how in this area in order to improve the feature of spatial scalability in our coder. Collaborations with TUT are already existing in the area of static 3D mesh compression, profiting from their rich expertise in image compression and wavelets.

2.3. Hologram compression

No progress has been made in this area, since the corresponding research at BIAS has been completed and terminated. The results are reported in the previous Technical Report. New ideas for compression of digital holograms have been created at the University of Plzen. The research has been started and results will be reported in the next Technical Report.

2.4. Multiple description coding (MDC) for 3D

In a typical scenario, compressed 3D data content such as multi-view video or 3D dynamic meshed has to be transmitted over error-prone channels. Therefore, error resilience should be utilized in order to ensure robust transmission. Multiple description coding (MDC) is a promising error resilience technique, applicable also to 3D data. It relies on splitting the single data source into two or more, essentially redundant, descriptions in such a way that it can be restored with acceptable quality by any of these subsets while by receiving more descriptions higher fidelity up to perfect reconstruction, is achieved.

During the second research period, the research group on MDC has focused on designing and studying MDC techniques for true 3D sources of information. MDC for stereoscopic video and MDC for meshes (3D geometry) have been addressed in more details. For the case of stereoscopic video, the recently developed h.264-based multiview codec has been modified to include error resilient features and two MDC schemes have been suggested for various amount of inter-view correlation. For the case of MDC for 3D geometry, two wavelet-based MDC schemes have been developed, one utilizing forward error correction codes and another utilizing splitting and duplicating wavelet coefficient trees.

2.4.1. Schemes For Multiple Description Coding Of Stereoscopic Video

Authors: Andrey Norkin, Anil Aksay, Cagdas Bilen, Gozde Bozdagi Akar, Atanas Gotchev, and Jaakko Astola

Institutions: Tampere University of Technology (TUT), Middle East Technical University (METU)

Publication: MRCS 2006, Istanbul, Turkey, Sept. 2006. Lecture Notes in Computer Science, vol. 4105, pp. 730-737, Springer-Verlag Heidelberg.

This paper presents and compares two multiple description schemes for coding of stereoscopic video, which are based on H.264. The SS-MDC scheme exploits spatial scaling of one view. In case of one channel failure, SS-MDC can reconstruct the stereoscopic video with one view low-pass filtered. SS-MDC can achieve low redundancy (less than 10%) for video sequences with lower inter-view correlation. MS-MDC method is based on multi-state coding and is beneficial for video sequences with higher inter-view correlation. The encoder can switch between these two methods depending on the characteristics of video.

2.4.2. Wavelet-Based Multiple Description Coding of 3-D Geometry

Authors: Andrey Norkin, Mehmet Oguz Bici, Gozde Bozdagi Akar, Atanas Gotchev, and Jaakko Astola

Institutions: Tampere University of Technology (TUT), Middle East Technical University (METU)

Publication: Visual Communications and Image Processing 2007, 28 January – 1 February 2007, San Jose, California, USA

In this work, we present a multiple description coding (MDC) scheme for reliable transmission of compressed three dimensional (3-D) meshes. It trades off the reconstruction quality for error resilience to provide the best expected reconstruction of 3-D mesh at the decoder side. The proposed scheme is based on multiresolution geometry compression achieved by using wavelet transform and modified SPIHT algorithm. The trees of wavelet coefficients are divided into different sub-sets. Each description contains the coarsest mesh and a number of wavelet coefficient tree subsets coded with different rates. The original mesh can be reconstructed with acceptable quality from any subset of the received descriptions. The proposed algorithm provides flexible number of descriptions and is optimized for varying packet loss rates (PLR) and channel bandwidth.

2.4.3. Multiple Description Coding Of 3D Geometry with Forward Error Correction Codes

Authors: M. Oguz Bici, Andrey Norkin, Gozde Bozdagi Akar, Atanas Gotchev, and Jaakko Astola

Institutions: Tampere University of Technology (TUT), Middle East Technical University (METU)

Publication: 3DTV-CON, The True Vision Capture, Transmission and Display of 3d Video, 7-9 May 2007, Kos Island, Greece

This work presents a multiple description coding (MDC) scheme for compressed three dimensional (3-D) meshes based on forward error correction (FEC). It allows flexible allocation of coding redundancy for reliable transmission over error-prone channels. The proposed scheme is based on progressive geometry compression, which is performed by using wavelet transform and modified SPIHT algorithm. The proposed algorithm is optimized for varying packet loss rates (PLR) and channel bandwidth. Modeling RD function considerably decreases computational complexity of bit allocation.

2.4.4. Analysis, conclusions, plans

Extensive study has been done for MDc of stereoscopic video and 3D meshes. For MDC of stereoscopic video, two new approaches have been introduced. These approaches produce balanced descriptions and provide stereoscopic reconstruction with acceptable quality in case of one channel failure for the price of moderate redundancy (in the range of 10-50%). Both approaches provide drift-free reconstruction in case of description loss. The performance of these approaches depends on characteristics of stereoscopic video sequence. The approach called SS-MDC performs better for sequences with lower inter-view correlation while MS-MDC approach performs better for sequences with higher inter-view correlation. The criterion for switching between the approaches is used by the encoder to choose the approach that provides better performance for this sequence.

A novel approach for MDC of 3-D meshes is introduced which is based on *progressive geometry compression* scheme. Wavelet transform is applied to the semi-regular remeshed model, and wavelet coefficient trees are coded with SPIHT algorithm. Our algorithm exploits redundancy in a form of redundant partially-coded trees of wavelet coefficients and a duplicated coarsest mesh. Two strategies of grouping the trees were studied, and spatially close grouping was chosen as the one providing better visual quality. To decrease the amount of computations at the preparatory stage, D-R function modeling can be exploited with the price of only slight decrease in network performance. The algorithm is capable of providing flexible number of descriptions and is optimized for varying packet loss rate. Graceful degradation of quality is achieved in presence of increasing packet loss rate. The reconstruction results show good visual perception quality.

As a further development of the previous approach, we have proposed an MDC-FEC algorithm for coding 3D-meshes. Stronger FEC is allocated to the beginning of SPIHT stream, while less or no FEC bits are allocated to the end of bit-stream. The algorithm generates multiple descriptions and is optimized for changing packet loss rate and channel bandwidth. Using the Weibull model instead of the real RD function during optimization considerably decreases the time needed for bit allocation while achieving similar average reconstruction quality.

Regarding future research, we plan to optimize the proposed MDC stereo video approaches and to study their performance over real transmission channel, such as DVB-H transport. As far as 3D geometry is concerned, our plans include optimization of the coarse mesh bit allocation and partitioning coarsest mesh vertices into different sets in order to further decrease the amount of redundancy. We shall consider more advanced channel model and packetization strategy which better emulate the behavior of a real network and further investigate tree grouping strategies.

2.5. Watermarking for 3D

The utilization of 3D information brings copyright problems for the content providers. In this situation, there is a strong necessity for developing techniques, which protect the copyright of the original 3D data and prevent unauthorized duplication or tampering. However, it should be noted that 3D scene representation is totally digital, hence all digital right management (DRM) solutions, which are applied to other digital modalities, can also be applied to 3D scene description. For instance, in a typical encrypted TV broadcast, a content provider might provide the media containing the 3D scene information to the customer, whereas the customer can use his/her valid license to play the media (Fig. 1).

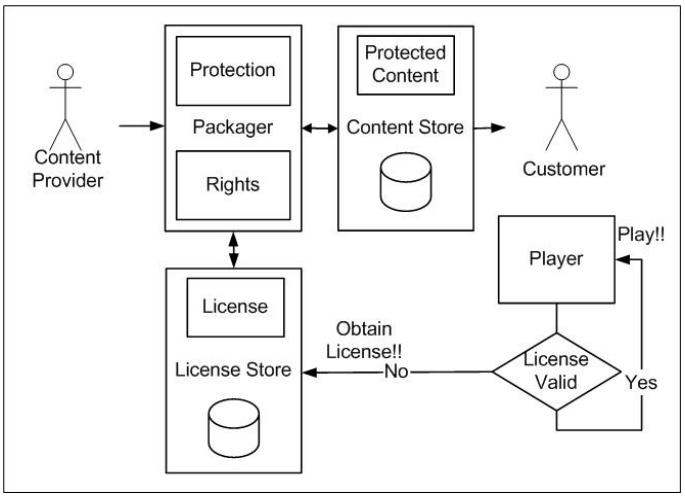

Fig. 1. *A typical DRM system*

An exception of the easy migration of general purpose DRM solutions of 3D data can be encountered in the field of data hiding (watermarking). Although, many approaches have been proposed for protecting the rights of content owners for their 2D multimedia data by hiding content owner information invisibly into the data, these methods can not be easily upgraded in case of 3D content. In other words, the watermarking technology for 2D media is on its way for reaching to a maturity, while 3D watermarking can still be accepted to be immature, although the great number of applications in which 3D content is utilized and consumed. Hence, the work of the 3D watermarking group of 3DTV NoE is completely devoted to the technologies for 3D watermarking.

In general, the term, *3D watermarking*, implies the watermarking of any representation of a 3D scene. The watermarking technology aims protection of a representation for a scene by embedding hidden data into the main components of the representation and extracting the watermark from the resulting components of a scene after any applications. Regarding the main components of the representations, a categorization can be achieved as given in Fig. 2. While the representations in the left side of this graph are in general dependent on the geometrical structure of the scene, the representation in the rightmost part is purely based on the images that are captured by the cameras appropriately located in the scene.

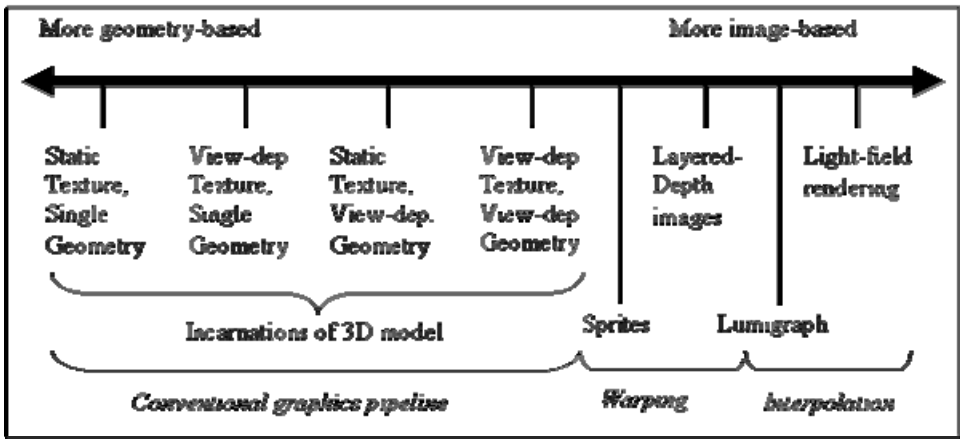

Fig. 2. *Categorization of Scene Representations*

In this context, the research group on 3D watermarking has identified the following prospective topics for research: First, a detailed state-of-the-art and future direction survey of 3d watermarking has been conducted. A scheme for embedding watermarking with a hologram has been also investigated. Another topic of research is the free-view point image/video. The target is that the watermark can be extracted successfully from an image generated from an arbitrary view at any position and rotation. Finally, a novel 3d watermarking scheme able to resist geometric attacks has been also proposed based on 3D Generalized Radon Transform: Radial Integration Transform (RIT) and the Spherical Integration Transform (SIT).

2.5.1. 3D watermarking: Techniques and Directions

Authors: Alper Koz1, George A. Triantafyllidis2 and A. Aydin Alatan1

Institutions: 1 Middle East Technical University (Turkey), 2Informatics and Telematics Institute, Thessaloniki, Greece

Publication: submitted as a book chapter to 3D TV Edited book in preparation

In this chapter, a complete literature survey on 3D watermarking techniques is presented. The chapter categorizes the 3D watermarking techniques according to the main components of the major representation techniques, as geometry-, texture- and image-based watermarking. The requirements and problems in each branch of 3D watermarking are briefly reviewed in the manuscript.

A detailed literature survey on the state-of-the-art 3D watermarking methods forms the core of the chapter. It is realized that the existing 3D scene watermarking methods mainly focus on the watermarking of 3D geometry data, which are mostly represented with mesh structures. These methods are classified, as spatial domain methods, where the watermark is embedded into the geometric values of the geometric primitives, such as the coordinates of the points, length of a line, area of a polygon, volume of a polyhedron, and transform domain methods, where the watermark is embedded into the resulted coefficients, after a 3D geometry-based transformation is applied to the 3D geometry data. Pros and cons of each category are summarized. Briefly, transform domain methods are more appropriate for finding the

significant parts of the 3-D object, hence more robust against the compression and noise attacks. On the other hand, spatial domain approaches are easier to implement and robust against the geometrical attacks like cropping.

The proposed solutions in the literature to the main attacks in geometry watermarking are examined separately, in order to give a more concise evaluation on 3D geometry watermarking. The robustness of the methods against main attacks is also compared. Although it is difficult to state the superiority of a method against another, due to the lack of a unique 3D scene representation, coding scheme and a benchmarking on 3D geometry watermarking, some methods are found out to be more promising, based on their robustness against possible common attacks on 3D geometry.

One important consequence of the chapter is to realize that the existing 3D watermarking methods in the literature mostly work on the 3D geometry-based representation of 3D scenes. As pointed before, in such a representation, the scene geometry is modeled by using meshes, point clouds or voxels and then, texture and reflectance maps are used as the additional descriptions on top of the model. Currently, such methods are mostly proposed to protect the geometry description of a 3D object that is used in computer graphics applications. However, considering that the geometry-based representations might also be a suitable framework for the future technologies, such as 3D TV, some suitable extension of these methods can also be a good candidate for the solution of the copyright problem and other related problems such as authentication, content labeling, and content protection, in the coming applications.

On the other hand, some alternative representation techniques for 3D scenes such as image based modeling and rendering (IBR) have been rapidly developed in the recent years. By capturing a set of images from different viewpoints of a scene, these techniques are designed to reproduce the scene correctly at an arbitrary view point. Compared to the geometry-based models, this approach is more advantageous, since images are easier to obtain, simpler to handle and more realistic to render. Noting that the user can also can record a personal video for his/her arbitrarily selected view and misuse this content, this technology makes the copyright problem for the image based represented scenes more apparent. There are also a number of pioneering works in this new area of 3D watermarking.

2.5.2. Quantization Index Modulation Based Watermarking Using Digital Holography

Authors: O. Erman Okman, G. Bozdagi Akar

Institutions: Middle East Technical University (METU)

Publication: Journal of Optical Society of America A. 24, 243-252, Jan. 2007.

In this paper, we propose a blind watermarking method where the watermark is a hologram itself. In the proposed approach, the quantized phase of the hologram is embedded into the wavelet-transformed host image using quantization index modulation (QIM). In the detection stage, wavelet transform of the watermarked image followed by a minimum distance decoder is used. The proposed method is blind since it requires only information about the quantizers and the parameters of the hologram recording process. The robustness of the proposed technique is tested against several attacks such as filtering, compression, occlusion and

cropping. Moreover, we discuss the impact of quantization of the hologram on the reconstruction quality.

2.5.3. Quantization Index Modulation Based Watermarking Using Digital Holography

Authors: O. Erman Okman, G. Bozdagi Akar

Institutions: Middle East Technical University (METU)

Publication: Optics and Photonics, 13-17 August 2006, in San Diego, California USA.

In this paper, we propose a blind watermarking method where watermark is chosen as the hologram of the signal to be embedded. In the proposed approach the quantized phase of the hologram is embedded into an image using quantization index modulation (QIM). The robustness of the proposed technique is tested against several attacks such as filtering, compression, occlusion and cropping. Also the effects of quantization of the hologram on the reconstruction quality are demonstrated.

2.5.4. 3D watermarking based on 3-D Generalized Radon Transform: Radial Integration Transform (RIT) and the Spherical Integration Transform (SIT)

Authors: P. Daras and G.A Triantafyllidis

Institutions: ITI-CERTH

Publication: working draft, to be submitted

This work describes a novel method for 3-D model watermarking based on the 3-D Generalized Radon Transform. Two forms of the GRT are implemented: (a) the Radial Integration Transform (RIT), which integrates the 3-D model's information on lines passing through its center of mass and contains all the radial information of the model, and (b) the Spherical Integration Transform (SIT), which integrates the 3-D model's information on the surfaces of concentric spheres and contains all the spherical information of the model. Combining RITs and SITs properties, a completely invariant transform, in terms of scaling and rotation, is created. The proposed 3D model watermarking is robust against geometric attacks such as rotation, translation and uniform scaling.

2.5.5. Free View Watermarking for Free View Television

Authors: Alper Koz, Cevahir Çığla, A. Aydin Alatan

Institutions: Middle East Technical University

Publication: IEEE International Conference on Image Processing, Atlanta, GA, USA, October 8-11, 2006.

The recent advances in Image Based Rendering (IBR) has pioneered a new technology, *free-view television,* in which TV-viewers select freely the viewing position and angle by the application of IBR on the transmitted multi-view video. Noting that the TV-viewer might also record a *personal* video for this arbitrarily selected view and misuse this content, it is apparent that copyright and copy protection problems also exist and should be solved for free-view TV. In this paper, we focus on this problem by proposing a watermarking method for free-view video. The watermark is embedded into every frame of multiple views by exploiting the spatial masking properties of the Human Visual System (HVS). Assuming that the position and rotation for the imagery view is known, the proposed method extracts the watermark successfully from an arbitrarily generated image. In order to extend the method for the case of an unknown imagery camera position and rotation, the modifications on the watermark pattern due to image based rendering operations are also analyzed. Based on this analysis, a camera position and homography estimation method is proposed considering the operations in image based rendering. The results show that the watermark detection is achieved successfully for the cases in which the imagery camera is arbitrarily located on the camera plane.

2.5.6. Watermarking for Free-View Television via Homography based Virtual Camera Location Estimation

Authors: Alper Koz, Cevahir Çığla, A. Aydin Alatan

Institutions: Middle East Technical University

Publication: submitted as a correspondence to IEEE Transactions on Information Forensics and Security

In the upcoming consumer electronic product, *free-view TV,* viewers might record *personal* videos for their arbitrary view (angle) selections. Hence, it is apparent that some solutions to the copyright and copy protection problems should be proposed against misusing such content. The method in this manuscript embeds the same watermark into the transmitted multiple-views of the free-view TV after exploiting the spatial masking properties of the Human Visual System. The extraction of the watermark from any virtual view is achieved after correlating this frame by a "rendered watermark", which is obtained through specific image rendering algorithms, after determining the location of the virtual camera through a homography between focal and image planes. The encouraging simulation results promise not only a novel method, but also a new direction for the watermarking research.

2.5.7. Free View Watermarking for Free View Television

Authors: Alper Koz, Cevahir Çığla, A. Aydin Alatan

Institutions: Middle East Technical University

Publication: to be submitted to IEEE Transactions on Image Processing

In the upcoming consumer electronic product, *free-view TV,* viewers might record *personal* videos for their arbitrary view (angle) selections. Hence, it is apparent that some solutions to the copyright and copy protection problems should be proposed against misusing such

content. The method in this manuscript embeds the same watermark into the transmitted multiple-views of the free-view TV after exploiting the spatial masking properties of the Human Visual System. The extraction of the watermark from any virtual view is achieved after correlating this frame by a "rendered watermark", which is obtained through specific image rendering algorithms, after determining the location of the virtual camera through a homography between focal and image planes. The encouraging simulation results promise not only a novel method, but also a new direction for the watermarking research.

2.5.8. Analysis, conclusions, plans

Since the geometry is the main content of a 3D scene in most of the 3D applications, the watermarking research mostly focuses on the copy and copyright protection of the geometry information of a scene. The general scheme for the geometry watermarking is given in Fig. 3. The watermark is imperceptibly embedded to the object and the watermarked object is delivered to the channel instead of the original one. The channel can include any distortions on the geometry that can happen due to any malicious or non malicious use of the object. Then, the watermark is tried to be obtained from the tested object in the extraction part.

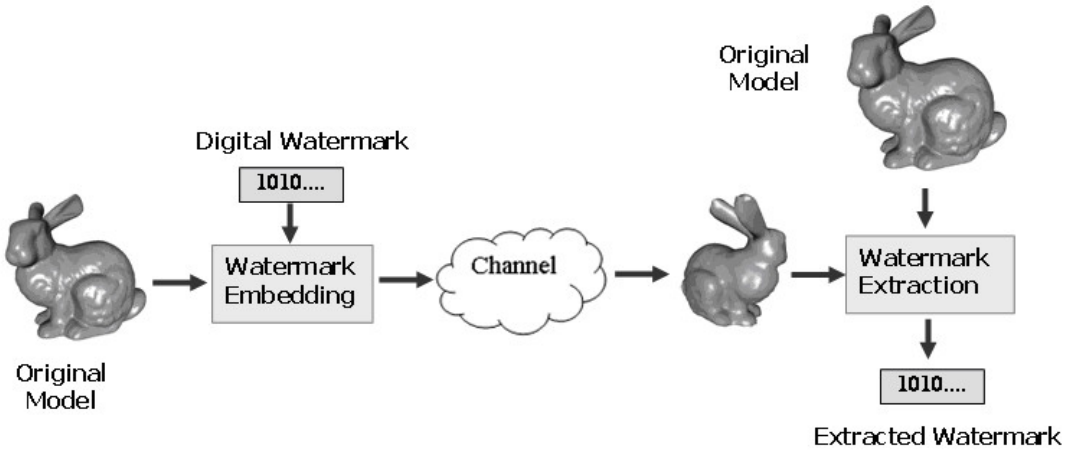

Fig. 3. *A general scheme for 3D geometry watermarking*

In this context, we proposed a method for 3D model watermarking based on the 3D Generalized Radon Transform. Two forms of the GRT are implemented: (a) the Radial Integration Transform (RIT), which integrates the 3D model's information on lines passing through its centre of mass and contains all the radial information of the model, and (b) the Spherical Integration Transform (SIT), which integrates the 3D model's information on the surfaces of concentric spheres and contains all the spherical information of the model. Future work will focus on selecting the right combination of transforms in order to succeed robust 3d watermarking.

Another work of the watermarking group of 3DTV NoE is dealing with the protection of a 3D scene that is represented and rendered by using the 2D images, which are taken from different view points in the scene. With the recent advances, the copyright and copy protection problems for the image-based represented scenes become more important. For instance, in the free-view TV application of IBR, a TV viewer might record a personal video for his/her

arbitrarily selected view and misuse this content. Therefore, a content provider should prove his/her ownership on the recorded media in such a case. Concerning with the robustness requirement, the watermark should not only be resistant to common video processing and multi-view video processing operations, it should also be extracted from a rendered video frame for an arbitrary view (see Fig. 4). In order to extract the watermark from such a rendered view, the watermark detection scheme should involve an estimation procedure for the imagery camera position and orientation, where the rendered view is generated. In addition, the watermark should also survive from image-based rendering operations, such as frame interpolation between neighbor cameras and pixel interpolation inside each camera frame. The future work will concentrate on the extension of the proposed method for the static scenes consisting of multiple depths and dynamic scenes.

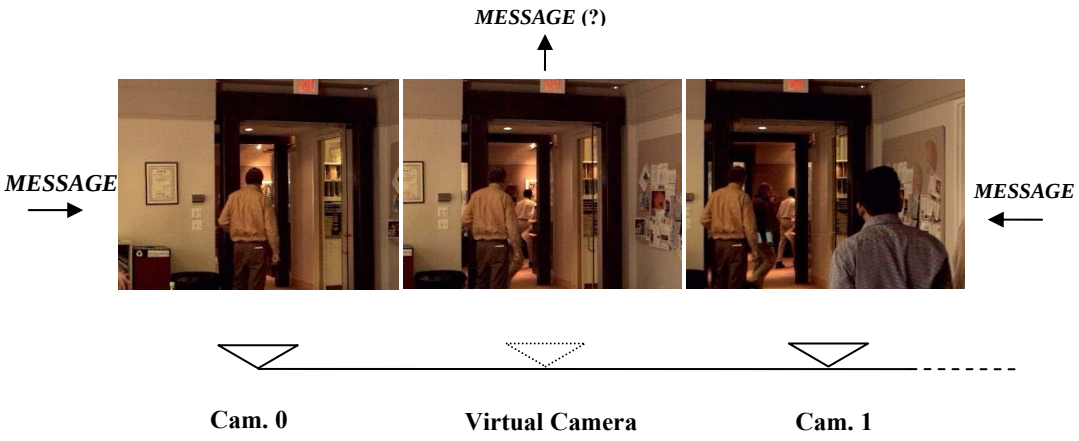

Fig. 4. *The watermarking problem for free-view television*

Finally, a blind watermarking method is proposed where the watermark is a hologram itself. In the proposed approach, the quantized phase of the hologram is embedded into the wavelet-transformed host image using quantization index modulation (QIM). In the detection stage, wavelet transform of the watermarked image followed by a minimum distance decoder is used. In future, this work will be extended in 3D data watermarking.

3. General Summary and Conclusions

Technical work on 3D compression is progressing very successfully. The partners' individual research converges more and more into real joint and integrated activities, taking benefit of each others expertise and resources. The work covers all areas identified to be important for 3D compression.

In the highly actual and important area of multi-view video coding the NoE plays a worldwide leading role, both in basic research as well as in standards activities. The latter includes management as well as technical work. It can be foreseen that this work will have a significant impact on future economy and society in a global sense. This work will of course be continued with maximum effort, including both the scientific fundamentals and the exploration in MPEG.

Although quite established compression of 3D meshes still leaves a lot of room for improvement. Achievements are reported for dynamic mesh compression. The activities on dynamic 3D mesh compression have been contributed to MPEG in order and stimulated standardization of such algorithms. Basic research as well as exploitation in MPEG will be continued in a joint manner. Mesh compression based on images is a second line of research in this area. Good results have been achieved and this activity will be continued as well.

Pioneering work has been done in the NoE in the area hologram compression. Holographic data will very likely play an important role in the future and researchers in the NoE were among the first worldwide to study compression aspects. Unfortunately, this work has been paused but will be continued in the future. Results will be reported in the next Technical Report.

Multiple description coding (MDC) for 3D is also a very new research area. Here the NoE is doing pioneering work that will be extended. Outside the NoE only MDC for 3D meshes has been studied. MDC for stereo and MVC have first been investigated within the NoE.

Watermarking of 3D data has been studied to some extend for the case of 3D meshes so far. Watermarking for other types of 3D data such as multi-view video or holographic data is being studied for the first time within the NoE. Also here the future work will benefit from close collaboration between the partners.

In general conclusion, the research on 3D compression within the NoE reached a high level of excellence, for both scientific fundamentals and more applied development of international standards. Partners formed a strong network and benefit from each others expertise and resources. The following table gives an overview of established collaborations within the different research areas.

AREA	**Collaborating Partners**
MVC	TUB, KU, METU, HHI
Stereo-Video Quality Metric	TUT, METU
3D Mesh compression	Hannover, Bilkent, HHI
Multiple Description Coding Of Stereoscopic Video / 3D geometry	TUT, METU
3D/free view point watermarking	CERTH, METU

4. Annex

4.1. Multi-view video coding (MVC)

4.1.1. Extending Single-view Scalable Video Coding to Multi-view based on H.264/AVC

4.1.2. Core Experiment on Simplified Prediction Structure

4.1.3. Core Experiment on a Dyadic Inter-view Prediction Structure

4.1.4. Temporal and Spatial Scaling For Stereoscopic Video Compression

4.1.5. Towards Compound Stereo-Video Quality Metric: A Specific Encoder-Based Framework

4.1.6. Content-Aware Bit Allocation in Scalable Multi-View Video Coding

4.1.7. Scalable Multi-view Video Coding for Interactive 3DTV

4.1.8. 3D Video and Free Viewpoint Video – Technologies, Applications and MPEG Standards

4.1.9. Efficient Compression of Multi-View Video Exploiting Inter-View Dependencies Based on H.264/MPEG4-AVC

4.1.10. Multi-view Video Coding Based on H.264/AVC Using Hierarchical B-Frames

4.2. 3D mesh compression

4.2.1. Connectivity-Guided Predictive Compression of Dynamic 3D Meshes

4.2.2. Scalable Linear Predictive Coding of Time-Consistent 3D Mesh Sequences

4.2.3. Nonrectangular wavelets for multiresolution mesh analysis and compression

4.2.4. Dikdörtgensel Olmayan Dalgacık Dönüşümüne Dayalı Çok Çözünürlüklü Üç Boyutlu Model Analizi ve Sıkıştırılması

(Nonrectangular wavelets for multiresolution mesh analysis and compression)

4.2.5. Connectivity-guided adaptive wavelet transform for mesh analysis and compression

4.2.6. Rate-Distortion Optimization in Dynamic Mesh Compression

4.2.7. Rate-Distortion-Optimized Predictive Compression of Dynamic 3D Mesh Sequences

4.2.8. 3D Model Compression Using Image Compression Based Methods

4.3. Hologram compression

None

4.4. Multiple description coding (MDC) for 3D

4.4.1. Schemes For Multiple Description Coding Of Stereoscopic Video

4.4.2. Wavelet-Based Multiple Description Coding of 3-D Geometry

4.4.3. Multiple Description Coding Of 3D Geometry with Forward Error Correction Codes

4.5. Watermarking for 3D

4.5.1. 3D watermarking: Techniques and Directions

4.5.2. Quantization Index Modulation Based Watermarking Using Digital Holography

4.5.3. Quantization Index Modulation Based Watermarking Using Holography

4.5.4. 3D watermarking based on 3-D Generalized Radon Transform: Radial Integration Transform (RIT) and the Spherical Integration Transform (SIT)

4.5.5. Free View Watermarking for Free View Television

4.5.6. Watermarking for Free-View Television via Homography based Virtual Camera Location Estimation

4.5.7. Free View Watermarking for Free View Television