D32.2 Technical Report #2 on 3D Telecommunication Issues

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

Deliverable Nature: R
Number: D32.2
Contractual Date of Delivery: M29
Actual Date of Delivery: M30
Report Date: 20 February 2007
Task: WP10
Dissemination level: PU
Start Date of Project: 01 September 2004
Duration: 48 months
Organisation name of lead contractor for this deliverable: KU

Name of responsible: Murat Tekalp (mtekalp@ku.edu.tr)

Editor: Murat Tekalp (mtekalp@ku.edu.tr)

20 February 2007

3D Telecommunication Issues
TC3 Technical Report 2

EDITOR

Murat Tekalp (**KU**)

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**)
Technische Universitaet Berlin (**TUB**)

REVIEWERS

Erdal Arıkan (**Bilkent**)

Gözde Bozdağı Akar (**METU**)

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

TABLE OF CONTENTS

Executive Summary

The main difficulty in the deployment of multi-view video services, including 3D and free-viewpoint TV, appears to be the large bandwidth requirement associated with transport of multiple video streams. There are two potential avenues for transport of 3DTV: i) over existing DVB infrastructure, and ii) over the Internet Protocol (IP). For the former, an MPEG standard using the so-called video-plus-depth encoded representation inside the MPEG-2 transport stream is in the final stages [9-10]. A multitude of multi-view encoding and streaming strategies using RTP/UDP/IP or RTP/DCCP/IP exist for the latter, which constitute the main subject of this report. Video streaming architectures can be classified as i) server to single client unicast, ii) server multicasting to several clients, iii) peer-to-peer (P2P) unicast distribution, where each peer forwards packets to another peer, and iv) P2P multicasting, where each peer forwards packets to several other peers. Multicasting protocols can be supported at the network-layer or application layer. Main research issues are: i) determination of the best video encoding configuration for each streaming strategy – multi-view video encoding methods provide some compression efficiency gain at the expense of creating dependencies between views that hinder random access to views; ii) determination of the best rate adaptation method – adaptation refers to adaptation of the rate of each view as well as inter-view rate allocation depending on available network rate and video content, and adaptation of the number and quality of views transmitted depending on available network rate and user display technology and desired viewpoint; iii) packet-loss resilient video encoding and streaming strategies as well as better error concealment methods at the receiver; iv) best peer-to-peer multicasting design methods, including topology discovery, topology maintenance, forwarding techniques, exploitation of path diversity, methods for enticing peers to send data and to stay connected, and use of dedicated nodes as relays. Several testbeds are being built in the process of this research to demonstrate these concepts over local area and wide area networks. The first testbed, server unicasting to multiple clients with stereoscopic displays, has been demonstrated. Second testbed, a server unicasting to a client with a head-tracking 3D stereoscopic display, which selectively streams only the required views and some data for side views, is in the works. Finally, a peer-to-peer multicast distribution testbed is being planned.

1. Introduction

There are two alternative transport architectures for 3DTV signals: the DVB architecture and the Internet Protocol architecture. The "video-plus-depth" data representation can be employed to build 3DTV transport evolutionarily on the existing DVB infrastructure. This representation uses a regular video stream enriched with so-called depth maps providing a Z-value for each pixel. The final 3D images are re-constructed at the receiver side by using depth-image-based rendering (DIBR). MPEG has established a special standardization activity that focuses on 3DTV using video-plus-depth representation.

The Internet Protocol (IP) architecture is proving to be very flexible in accommodating a wide range of communication applications as can be seen from the ongoing replacement of classical telephone services by voice over IP applications. Transmission of video over the Internet is currently an active research and development area where significant results have already been achieved. There are already video-on-demand services, both for news and entertainment applications, offered over the Internet. Also, 2.5G and 3G mobile network operators started to use IP successfully to offer wireless video services. Transport of 3DTV signals over IP packet networks seems to be a natural extension of video over IP applications.

Video streaming architectures can be classified as i) server to single client, ii) server multicasting to several clients, iii) peer-to-peer (P2P) distribution, where each peer forwards packets to another peer, and iv) P2P multicasting, where each peer forwards packets to several other peers. Multi-view video streaming protocols can be RTP/UDP/IP, which is the current state of the art, and RTP/DCCP/IP, which is the next generation protocol. Multicasting protocols can be supported at the network-layer or application layer.

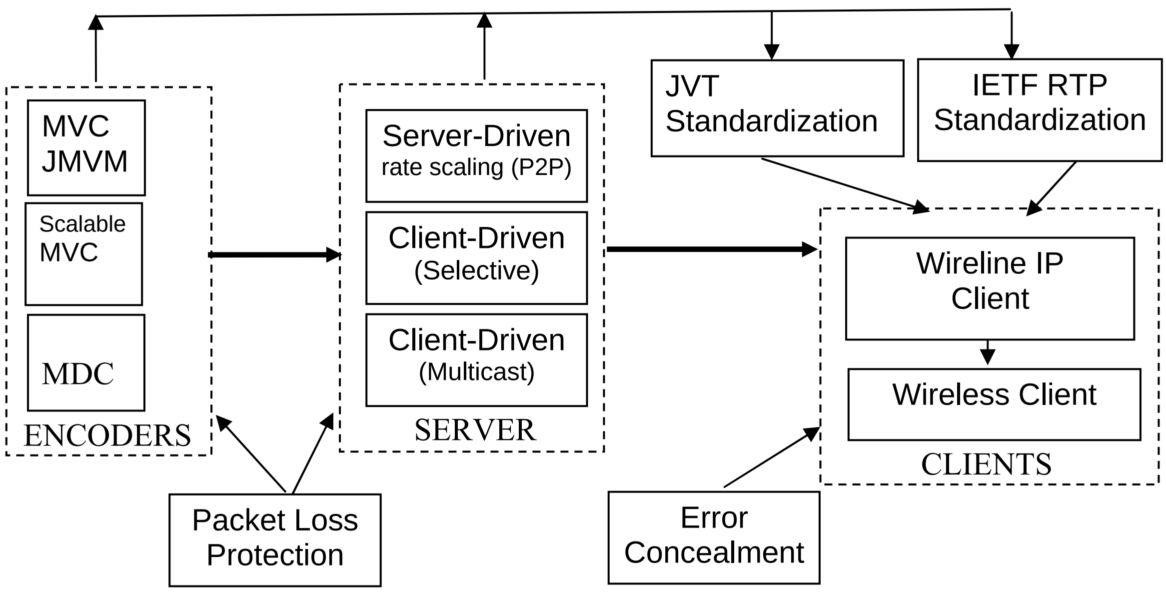

Figure 1: Block diagram of the framework and system for 3DTV transport over IP

We propose a complete framework for end-to-end transport of 3DTV over IP, which is shown in Figure 1. The proposed framework includes several options for video encoder/decoder, rate scaling and allocation strategies, error control and concealment schemes, streaming schemes and client support, as well as support for various 3D displays. Joint work carried on various components of this framework is discussed in Section 2. Section 3 presents the roadmap for future work.

2. Overview of the Results

We start with an overview of our work in 3D/multi-view video encoders in Section 2.1. Section 2.2 discusses transmission of encoded "video-plus-depth" data over the MPEG-2 transport stream. Alternatively, this representation and/or other multi-view encoded representations can be streamed over IP. The major research issues are multi-view video rate allocation and adaptation and packet loss protection and error concealment on the server side for various streaming strategies. We overview our work on rate adaptation in Section 2.3 (Server-driven rate adaptation) and in Section 2.4 (Client-driven selective streaming). Packet loss protection and error concealment are discussed in Section 2.5. Finally, we discuss implementation of a server and multiple clients for a demonstration testbed in Section 2.6.

2.1. 3D Video Encoding

Multi-view 3D video can be encoded implicitly, in the so-called video-plus-depth representation [1], or explicitly. Implicit encoding approach is discussed in 2.1.1 below. There are several strategies for explicit coding of multi-view video: 1) simulcast coding, 2) scalable simulcast coding, 3) multi-view coding, and 4) scalable multi-view coding. In the first two options, each view can be independently coded using the public-domain H.264 [2, 3] and SVC reference codecs [4, 5], respectively. We have developed two independent implementations for option 3, the JVT reference codec [6] and the 3DTV METU codec [7]. We also have multiple alternative implementations for option 4, the SMVC codec developed by Ege and Koc Universities [8], and the MVC base plus simulcast enhancement layers implementation of Koc University [9]. Our demo server implementation (see Section 2.6) can support these different codecs. It is also possible to encode and send depth maps for view interpolation at the client-side, although our demo server and clients do not currently support this option.

2.1.1. An Advanced 3DTV Concept Providing Interoperability and Scalability For a Wide Range of Multi-Baseline Geometries.

Authors: C. Fehn, N. Atzpadin, M. Müller, O. Schreer, A. Smolic, R. Tanger, and P. Kauff.

Institution: Fraunhofer HHI

Publication: Proc. of Int. Conf. on Image Proc., pp. 2961-2964, Atlanta, GA, USA, Oct. 2006

The paper discusses an advanced approach for 3DTV services that is based on the concept of an *N* x video-plus depth data representation. It particularly considers aspects of interoperability, scalability, and adaptability for the case that different multi-baseline geometries are used for multi-view capturing and 3D reproduction. In addition, it presents a method for the creation of depth maps and an algorithm for depth-image-based rendering related to the system approach.

2.1.2. 3D Video and Free Viewpoint Video – Technologies, Applications and MPEG Standards

Authors: A. Smolic, K. Mueller, P. Merkle, C. Fehn, P. Kauff, P. Eisert, and T. Wiegand

Institution: Fraunhofer HHI

Publication: Proc. ICME 2006, Int. Conf. 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.

2.1.3. 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, Oct. 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).

2.1.4. 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 Tech. (TUT), Middle East Tech. 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 stereo video encoder.

2.2. Video-Plus-Depth in MPEG-2 Transport Stream

The basic video-plus-depth format described in Section 2.1.1 is currently being standardized within the MPEG (Motion Pictures Experts Group) consortium. This work was initiated by a joint effort of Philips and Fraunhofer HHI (partner of 3DTV Network of Excellence). The underlying idea was to only standardize the format itself – by means of metadata that conveys the meaning of the gray level values in the depth imagery – and some additional meta data required to signal the existence of an encoded depth stream to the receiver. The actual compression of the per-pixel depth information, on the other side, will not be defined explicitly such that every conventional MPEG video codec (e.g., MPEG-2, MPEG-2 Visual, or H.264/AVC) can be used.

The new standard will be published in two parts. The specification of the depth format itself will be called ISO/IEC 23002-3 (MPEG-C)[1], a method for transmitting video-plus-depth within a conventional MPEG-2 Transport Stream will become an amendment (Amd. 2) to ISO/IEC 13818-1 (MPEG-2 Systems). Both standards are finalized at the MPEG meeting in Marrakech, Morocco (January 2007). The details can be found in the standardization documents [10, 11].

[1] Fraunhofer HHI is working together with Philips as editors of this standardization document.

2.3. Server-Driven Rate Adaptation in Multi-View Video Streaming

In streaming video over the Internet, video rate must be adapted to the available throughput in order to avoid congestion. Furthermore, it is desirable that the video rate must be friendly with other TCP traffic. Rate adaptation of stereo and multi-view video differs from that of monocular video, since rate allocation between views offers new flexibilities [12, 13]. Several open-loop rate adaptation strategies for stereo and multi-view video at the server side are studied for UDP and DCCP protocols in the following.

2.3.1. Temporal and Spatial Scaling for Stereoscopic Video Compression

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

2.3.2. Rate Allocation between Views in Scalable Stereo Video Coding using an Objective Stereo Video Quality Measure

Authors: Nükhet Özbek, A. Murat Tekalp, and Turhan Tunali

Institutions: Ege University, Koç University

Publication: submitted to IEEE Trans. Circ. and Systems for Video Technology

It is well-known that in stereoscopic 3D video systems humans perceive good quality 3D video as long as one of the eyes sees a high quality view. Hence, in stereo video encoding/streaming, best rate allocation between views can be addressed by reduction of the spatial resolution, frame rate, and/or quantization parameter of the second view with respect to the first view. In this paper, we address selection of the rate allocation strategy between views for our recently developed scalable multi-view video codec (SMVC) to obtain the best rate-distortion performance. Since 3D video quality perception does not correlate well with the overall PSNR of the two views, we propose a new quantitative measure for stereo video quality as weighted combination of two PSNR values and a jerkiness measure. The weights are determined by means of correlating subjective quality test results and the objective measure scores on a set of test videos. DSCQS test methodology is used for subjective evaluation of stereo videos. Experimental results are presented to demonstrate how the objective and subjective 3D video quality varies for different choices of rate allocation between the views.

2.3.3. Optimal Packet Scheduling and Rate Control for Video Streaming

Authors: Eren Gurses, Gozde Bozdagi Akar, Nail Akar

Institutions: METU

Publication: SPIE Visual Communications and Image Processing (VCIP) 2007

In this paper, we propose a new low-complexity retransmission based optimal video streaming and rate adaptation algorithm. The proposed OSRC (Optimal packet Scheduling and Rate Control) algorithm provides average reward optimal solution to the joint scheduling and rate control problem. The efficacy of the OSRC algorithm is demonstrated against optimal FEC based schemes and results are verified over TFRC (TCP Friendly Rate Control) transport with ns-2 simulations.

2.4. Client-Driven Rate-Adaptation in Multi-View Video Streaming

There are also streaming strategies where rate adaptation must be done at the server using feedback from the client [14, 15, 16]. A streaming strategy for a single-user head-tracking display client is discussed in Section 2.4.1; an application layer multicast scenario is studied in Section 2.4.2.

2.4.1. Client-Driven Selective Streaming of Multi-View Video for Interactive 3DTV

Authors: E. Kurutepe, M. R. Civanlar, and A. M. Tekalp

Institution: Koç University

Publication: submitted to IEEE Trans. on CSVT

We present a novel client-driven multi-view video streaming system that allows a user watch 3-D video interactively with significantly reduced bandwidth requirements by transmitting a small number of views selected according to his/her head position. The proposed scheme can be used to efficiently stream a dense set of multi-view sequences (light-fields) or wider baseline multi-view sequences together with depth information. The user's head position is tracked and predicted into the future to select the views that best match the user's current viewing angle dynamically. Prediction of future head positions is needed so that views matching the predicted head positions can be requested from the server ahead of time in order to account for delays due to network transport and stream switching. The system allocates more bandwidth to the selected views in order to render the current viewing angle. Highly compressed, lower quality versions of some other views are also requested in order to provide protection against having to display the wrong view when the current user viewpoint differs from the predicted viewpoint. An objective measure based on the abruptness of the head movements and delays in the system is introduced to determine the number of additional lower quality views to be requested. The proposed system makes use of multi-view coding (MVC) and scalable video coding (SVC) concepts together to obtain improved compression efficiency while providing flexibility in bandwidth allocation to the selected views. Rate-distortion performance of the proposed system is demonstrated under different experimental conditions.

2.4.2. Interactive Transport of Multi-view Videos for 3DTV Applications

Authors: Engin Kurutepe, M. Reha Civanlar, A. Murat Tekalp

Institution: Koç University

Publication: Packet Video 2006

In this paper we propose a novel method for transporting multi-view videos that aims to keep the bandwidth requirements on both end-users and servers as low as possible. The method is based on application layer multicast, where each end point receives only a selected number of views required for rendering video from its current viewpoint at any given time. The set of selected videos changes in real time as the user's viewpoint changes because of head or eye movements. Techniques for reducing the black-outs during fast viewpoint changes have been investigated. The performance of the approach has been studied through network experiments

2.5. Packet Loss Protection and Error Concealment

Congestion is the main cause of packet losses over the wired Internet. There is an ongoing collaboration between TUB and METU to combine the stereo video coder of METU and a modified decoder with an interface for error concealment based on a Gilbert-Elliot channel model. Carsten Clemens (TUB) visited METU, where he incorporated the MPEG-3DAV multiview codec and NS2 for channel simulation.

In contrast to the wired backbone, the capacity of the wireless channel is fundamentally limited by the available bandwidth of the radio spectrum and various types of noise and interference. Therefore, the wireless channel can be regarded as the "weakest link" of future multimedia networks and, hence, requires special attention, especially if mobility gives rise to fading and error bursts. This work between ITI-Certh and Bilkent concentrates on channel-error related issues in 3D video transmission. In particular, joint source and channel techniques have been developed for the efficient transmission of video streams over packet erasure channels, both in wired and wireless networks. Advanced channel coding techniques (such as Reed-Solomon, Turbo and LDPC codes) in conjunction with schemes which add unequal amount of redundancy to the data according to their importance are used to protect effectively visual information from packet losses.

Recently, a video coding scheme based on macroblock classification and unequal error protection of H.264/AVC streams has been developed by ITI-CERTH [17, 18]. Prior to transmission, macroblocks are classified into three slice groups by examining their contribution to video quality. Since the transmission scenarios are over packet networks, facing moderate to high packet loss rates, RS codes are used for channel protection. RS protection is selected for each slice group using a channel rate allocation algorithm based on dynamic programming techniques. The present method is the first utilizing the explicit mode of the H.264/AVC flexible macroblock ordering (FMO) in conjunction with channel coding techniques. The performance gain is attributed to the more efficient data organization of our scheme, which allows better error concealment without sacrificing coding performance, and to the finer protection of slice groups arising from our unequal error protection strategy.

ITI-CERTH and Bilkent University are collaborating on more advanced channel coding techniques for the effective handling of packet losses during the transmission of visual content over unreliable channels. The recent study of ITI in [17, 18] is extended to include multi-view video. The use of LT codes is investigated, which exhibit low encoding and decoding complexity and can adapt to the erasure rate of the packet networks due to their rateless property.

METU and Bilkent University are collaborating on different channel coding techniques and layered multiview coding for error resilient multiview video transmission. Unequal error protection (UEP) is applied to three different layers for stereo data. Optimal operating regions for UEP are observed via simulations.

2.5.1. Robust Transmission of H.264/AVC Streams Using Adaptive Group Slicing and Unequal Error Protection

Authors: N. Thomos, S. Argyropoulos, N. V. Boulgouris, and M. G. Strintzis

Institution: ITI-Certh

Publication: EURASIP Journal on Applied Signal Processing - Special issue on Advanced Video Technologies and Applications for H.264/AVC and Beyond, vol. 2006.

We present a novel scheme for the transmission of H.264/AVC video streams over lossy packet networks. The proposed scheme exploits the error-resilient features of H.264/AVC codec and employs Reed-Solomon codes to protect effectively the streams. A novel technique for adaptive classification of macroblocks into three slice groups is also proposed. The optimal classification of macroblocks and the optimal channel rate allocation are achieved by iterating two interdependent steps. Dynamic programming techniques are used for the channel rate allocation process in order to reduce complexity. Simulations clearly demonstrate the superiority of the proposed method over other recent algorithms for transmission of H.264/AVC streams.

2.5.2. Robust Transmission of Multi-view Video Streams Using Flexible Macroblock Ordering and Systematic LT codes

Authors: S. Argyropoulos, A. S. Tan, N. Thomos, E. Arikan, and M. G. Strintzis

Institutions: Bilkent and ITI-Certh

Publication: submitted to 3DTV-CON 2007

The transmission of multi-view video coded streams over packet erasure networks is examined. The proposed scheme employs a fully compatible H.264/AVC multi-view video codec. Macroblock classi- fication into unequally important slice groups is achieved using the Flexible Macroblock Ordering (FMO) tool of H.264/AVC. Systematic LT codes are used for error protection due to their low complexity and advanced performance. The optimal slice grouping and channel rate allocation is determined by an iterative optimization algorithm based on dynamic programming. The experimental evaluation clearly demonstrates the validity of the proposed method.

2.5.3. Error Resilient Layered Stereoscopic Video Streaming

Authors: A. S. Tan, A. Aksay, C. Bilen, G.Bozdagi Akar, E. Arikan

Institutions: METU and Bilkent

Publication: submitted to 3DTV-CON 2007

In this paper, error resilient stereoscopic video streaming problem is addressed. Two different Forward Error Correction (FEC) codes namely Systematic LT and RS codes are utilized to protect the stereoscopic video data against transmission errors. Initially, the stereoscopic video is categorized in 3 layers. Then, a packetization scheme is used to increase the efficiency of error protection. A comparative analysis of RS and LT codes are provided via simulations to observe the optimum packetization and UEP strategies.

2.5.4. Schemes for Multiple Description Coding of Stereoscopic Video

Authors: A. Norkin, A. Aksay, C. Bilen, G. B. Akar, A. Gotchev, and J. Astola

Institutions: Tampere University of Tech. (TUT), Middle East Tech. 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.

2.5.5. A Full Frame Loss Concealment Method for Stereo Video

Authors: Cagdas Bilen, Anil Aksay, Gozde Bozdagi Akar

Institution: METU

Publication: submitted to ICIP 2007

In applications regarding video transfer or streaming over internet, the packet losses should be taken into consideration. Especially when dealing with low bitrate videos, packet losses may lead to the loss of an entire frame of the video. Several studies are found in the literature on frame loss concealment algorithms for monoscopic video but these methods are not directly applicable to the stereoscopic video. In this paper we propose a full frame loss concealment algorithm for stereoscopic sequences. The proposed method uses redundancy between the two views and previously decoded frames to estimate the lost frame. The results show that, the proposed algorithm outperforms the monoscopic methods when they are applied to the same view as they are simulcast coded

2.6. Test-Bed Implementation

METU and Koc University have developed an end-to-end prototype system for point-to-point streaming of multi-view video over UDP. The prototype server employs the METU codec and RTP packetization with application layer framing [19]. Multiple clients have been developed by modifying the VLAN client for different 3D displays, including i) the projection system at Koc University, ii) 17" lenticular display at Koc University and Bilkent, iii) 23" display hosted at TUT and METU, and iv) the Sharp 3D laptop. The prototype system currently operates over a LAN with no packet losses. This test-bed has successfully been demonstrated at the Annual Review Meeting in Koc University, Istanbul and at the IST event in Helsinki in November 2006.

2.6.1. End-to-End Stereoscopic Video Streaming System

Authors: S. Pehlivan, Anil Aksay, Cagdas Bilen, Gozde Bozdagi Akar, and M. Reha Civanlar

Institution: Koç University and METU

Publication: *IEEE ICME 2006, Toronto, Canada, July 2006.*

Today, stereoscopic and multi-view video are among the popular research areas in the multimedia world. In this study, we have designed a platform consisting of stereo-view capturing, real time transmission and display. At the display stage, end users view video in 3D by using polarized glasses. Multi-view video is compressed in an efficient way by using multi-view video coding (MVC) techniques and streamed using real time protocols on the sender side. The entire system is built by modifying available open source systems whenever possible. Receiver can view the content of the video built from multiple channels as mono or stereo depending on its display and bandwidth capabilities.

2.6.2. End-to-End Stereoscopic Video Streaming with Content-Adaptive Rate and Format Control

Authors: A. Aksay, S. Pehlivan, E. Kurutepe, C. Bilen, T. Ozcelebi, G. B.Akar, M. R. Civanlar, and A. M. Tekalp

Institution: METU and Koç University

Publication: to appear in Signal Processing: Image Communication, February 2007.

We address efficient compression and real-time streaming of stereoscopic video over the current Internet. We first propose content-adaptive stereo video coding (CA-SC), where additional coding gain, over that can be achieved by exploiting only inter-view correlations, is targeted by downsampling one of the views spatially or temporally depending on the content, based on the well-known theory that the human visual system can perceive high frequencies in 3D from the higher quality view. We also developed stereoscopic 3D video streaming server and clients by modifying available open source platforms, where each client can view the video in mono or stereo mode depending on its display capabilities. The performance of the end-to-end stereoscopic streaming system is demonstrated using subjective quality tests.

3. Conclusions and Future Directions

3.1. Conclusions

Research conducted within this period on the transportation of 3D content in the form of multi-view video leads to the following conclusions:

i) Various options for encoding multi-view video have been discussed. It is still an active research problem to determine the best coding configuration for the various streaming strategies that are discussed in Sections 2.3 and 2.4.

ii) Rate adaptation of multi-view video allows new flexibilities; i.e., possibility of inter-view rate adaptation, which exploits the human perception of stereoscopic and multi-view video in addition to temporal rate adaptation. Best multi-view video rate adaptation strategy will be further investigated in year 3.

iii) Multicasting is a promising direction for effective distribution of 3D content. Streaming strategies for the application layer multicast scenario should be studied further. These include server to client multicast and peer-to-peer multicast scenarios.

iv) Packet loss-resilience techniques at the server/sending-peer side, and loss concealment techniques at the client/receiving peer are important for robust streaming of 3D content over wide area networks.

v) A test-bed for streaming stereo video from a server to multiple clients has been build and demonstrated.

3.2. Future Directions

The following are research directions and planned collaborations for the next period:

i) Streaming over DCCP: The Datagram Congestion Control Protocol (DCCP) is a transport protocol that implements bi-directional unicast connections of congestion controlled unreliable datagrams [20]. DCCP provides reliable handshakes for connection setup/teardown and reliable negotiation of options. Besides handshakes and feature negotiation, DCCP also accommodates two different congestion control schemes, one of which is to be selected at connection startup time. These are TCP-like Congestion Control and TCP-Friendly Rate Control (TFRC) [21]. TFRC, which is identified by CCID3, is a form of equation-based flow control that minimizes abrupt changes in the sending rate while maintaining longer-term fairness with TCP. It is hence better suited for streaming media applications. During its operation, CCID3 calculates an allowed sending rate, called TFRC rate, by using the TCP throughput equation, which is provided to the sender application upon request. We propose to use CCID3 as the DCCP congestion control mechanism for transport of multi-view video [22]. The video rate shall be adapted to the TFRC rate that is calculated by CCID3 to obtain best results. Koc University (KU) is leading this work.

ii) Rate adaptation strategies: There are two possible rate adaptation strategies to match the TFRC rate: i) instantaneous rate adaptation/control in real-time non-scalable multi-view video encoding, and ii) real-time extraction of scalability layers from off-line encoded scalable multi-view bitstreams. KU will lead this work.

iii) Loss resilience and concealment: Information-theoretic approaches including channel coding strategies as well as multiple description video encoding shall be considered for better loss resilience. Better loss concealment methods will also be investigated. This is joint work between Bilkent, METU and ITI-CERTH.

iv) Peer-to-peer multicasting strategies: Given the success of Joost TV anywhere distribution model (http://www.joost.com/), we will continue investigating peer-to-peer multicasting strategies for transport of multi-view video. This is joint work between Technical University Berlin (TUB) and KU.

v) Testbed implementation: METU and KU are currently extending the prototype demo server-client system to support i) JVT MVC compatible coding/decoding, ii) error correction and concealment against packet losses, and iii) streaming over DCCP.

vi) Wireless client: TUT and KU will work towards developing a wireless client. TUT and Turku University have a comprehensive DVB-H simulator, so they can work with error traces corresponding to different channel scenarios and then study joint source-channel coding issues. The results will then be integrated to a streaming client application.

vii) Selective streaming to a head-tracking display client: We are also working to develop a prototype client-driven selective multicast streaming of multi-view video. This is joint work between TUB and KU.

4. References

[1] C. Fehn, P. Kauff, M. Op de Beeck, F. Ernst, W. Ijsselsteijn, M. Pollefeys, L. Vangool, E. Ofek, and I. Sexton, "An evolutionary and optimised approach on 3D-TV," Proc. of IBC 2002, Int. Broadcast Convention, Amsterdam, Netherlands, Sept. 2002.

[2] ITU-T Rec. H.264, Advanced video coding for generic audiovisual services, March 2005. http://www.itu.int/rec/T-REC-H.264

[3] JM Reference Software, version JM 11.0, August 2006. http://iphome.hhi.de/suehring/tml/

[4] Joint Video Team of ITU-T VCEG and ISO/IEC MPEG, "Scalable Video Coding – Working Draft 1," Joint Video Team, *Doc. JVT-N020*, Hong-Kong, China, Jan. 2005.

[5] Joint Video Team of ITU-T VCEG and ISO/IEC MPEG, "Joint Scalable Video Model JSVM-4," Joint Video Team, *Doc. JVT-Q202*, Oct. 2005.

[6] Joint Video Team of ITU-T VCEG and ISO/IEC MPEG, "Joint Multiview Video Model (JMVM) 1.0," Joint Video Team, *Doc. JVT-T208*, July 2006.

[7] S. Pehlivan, Anil Aksay, Cagdas Bilen, Gozde Bozdagi Akar, and M. Reha Civanlar, "End-to-end stereoscopic video streaming system," Proc. IEEE Int. Conf. Multimedia and Expo (ICME), Toronto, Canada, July 2006.

[8] N. Ozbek and A. M. Tekalp, Scalable multi-view video coding for interactive 3DTV, Proc. IEEE Int. Conf. Multimedia and Expo (ICME), Toronto, Canada, July 2006.

[9] E. Kurutepe, MS Thesis, Koc University, Sept. 2006.

[10] ISO/IEC JTC 1/SC 29/WG 11. FPDAM of ISO/IEC 13818-1:200X/AMD 2 (Carriage of Auxiliary Video Data). Eds.: J. van der Meer and A. Bourge. WG 11 Document N8094, Klagenfurt, Austria, July 2006.

[11] ISO/IEC JTC 1/SC 29/WG 11. Study of ISO/IEC FCD 23002-3: Representation of Auxiliary Video and Supplemental Information. Eds.: A. Bourge and C. Fehn. WG 11 Document N8482, Hangzhou, China, October 2006.

[12] L. B. Stelmach, W. J. Tam, D. Meegan, and A. Vincent, "Stereo image quality: effects of mixed spatio-temporal resolution," IEEE Trans. Circ. Systems Video Techn., vol. 10, no. 2, pp. 188-193 (2000).

[13] N. Ozbek and A. M. Tekalp, Content-aware bit allocation in scalable multi-view video coding, Int. Workshop on MRCS (LNCS 4105), Istanbul, Sept. 2006.

[14] S. McCanne, V. Jacobson, and M. Vetterli, "Receiver-driven layered multicast," in SIGCOMM, New York, NY, ACM Press, pp. 117-130, 1996.

[15] E. Setton, J. Noh, and B. Girod, "Rate-distortion optimized video peer-to-peer multicast streaming," Proc. of ACM workshop on Advances in Peer-to-Peer Multimedia Streaming, New York, NY, ACM Press, pp. 39-48, 2005.

[16] E. Kurutepe, M. R. Civanlar, and A. M. Tekalp, A receiver-driven multicasting framework for 3DTV transmission, Proc. EUSIPCO, Antalya, Turkey, Sept. 2005.

[17] N. Thomos, S. Argyropoulos, N. V. Boulgouris, and M. G. Strintzis, "Error resilient transmission of H.264/AVC streams using flexible macroblock ordering", 2nd European Workshop on the Integration of Knowledge, Semantic, and Digital Media Techniques, EWIMT05, London, UK, November 2005

[18] N. Thomos, S. Argyropoulos, N. V. Boulgouris, and M. G. Strintzis, "Robust Transmission of H.264/AVC Streams Using Adaptive Group Slicing and Unequal Error Protection", EURASIP Jour. on Applied Signal Processing, Spec. issue on Advanced Video Technologies and Applications for H.264/AVC and Beyond, vol 2006, Feb. 2006.

[19] S. Pehlivan, MS Thesis, Koc University, August 2006.

[20] E. Kohler, M. Handley, and S. Floyd, "Datagram Congestion Control Protocol (DCCP)," IETF, March 2006, RFC 4340.

[21] S. Floyd, E. Kohler, and J. Padhye, "Profile for DCCP Congestion Control ID 3: TCP-Friendly Rate Control (TFRC)," IETF, March 2006, RFC 4342.

[22] N. Ozbek, B. Gorkemli, A. M. Tekalp, and T. Tunali, "Adaptive streaming of scalable stereoscopic video over DCCP," submitted to ICIP 2007.

5. Annex: Papers

1. C. Fehn, N. Atzpadin, M. Müller, O. Schreer, A. Smolic, R. Tanger, and P. Kauff, “An advanced 3DTV concept providing interoperability and scalability for a wide range of multi-baseline geometries,” Proc. of IEEE Int. Conf. on Image Processing (ICIP), pp. 2961-2964, Atlanta, GA, USA, October 2006.
2. A. Smolic, K. Mueller, P. Merkle, C. Fehn, P. Kauff, P. Eisert, and T. Wiegand, “3D video and free viewpoint video – Technologies, applications and MPEG standards,” Proc. IEEE Int. Conf. Multimedia & Expo (ICME), Toronto, Ontario, Canada, July 2006.
3. M. Drose, C. Clemens, and T. Sikora, “Extending single-view scalable video coding to multi-view based on H.264/AVC,” IEEE Int. Conf. on Image Processing (ICIP), Atlanta, GA, USA, Oct. 2006.
4. A. Boev, A. Gotchev, K. Egiazarian, A. Aksay, G. Bozdagi Akar, “Towards Compound Stereo-Video Quality Metric: A Specific Encoder-Based Framework,” IEEE SSIAI, Denver, Colorado, USA, March 2006.
5. A. Aksay, C. Bilen, E. Kurutepe, T. Ozcelebi, G. Bozdagi-Akar, M. R. Civanlar, and A. M. Tekalp, “Temporal and spatial scaling for stereoscopic video compression,” Proc. EUSIPCO, Florence, Italy, Sept. 2006.
6. N. Ozbek, A. M. Tekalp, and T. Tunali, “Adaptive Inter-View Rate Allocation for Scalable Stereo Video Coding using an Objective Stereo Video Quality Measure” submitted to IEEE Trans. on Circ. and Syst. for Video Tech., Dec. 2006.
7. E. Gurses, G. Bozdagi Akar, and N. Akar, “Optimal Packet Scheduling and Rate Control for Video Streaming,” SPIE Visual Comm. and Image Processing (VCIP), Jan. 2007.
8. E. Kurutepe, M. R. Civanlar, and A. M. Tekalp, “Client-Driven Selective Streaming of Multi-View Video for Interactive 3DTV,” submitted to IEEE Trans. CSVT, Dec. 2006.
9. E. Kurutepe, M. R. Civanlar, and A. M. Tekalp, “Interactive Transport of Multi-view Videos for 3DTV Applications, Proc. Packet Video 2006, China.
10. N. Thomos, S. Argyropoulos, N. V. Boulgouris, and M. G. Strintzis, “Robust Transmission of H.264/AVC Streams Using Adaptive Group Slicing and Unequal Error Protection,” EURASIP Jour. on Applied Signal Processing, Feb. 2006.
11. S. Argyropoulos, A. S. Tan, N. Thomos, E. Arikan, and M. G. Strintzis, “Robust transmission of multi-view video streams using flexible macroblock ordering and systematic LT codes,” submitted to 3DTV-CON, Kos Island, Greece, May 2007.
12. A. S. Tan, A. Aksay, C. Bilen, G.B.-Akar, and E. Arikan, “Error resilient layered stereoscopic video streaming,” submitted to 3DTV-CON, Kos Island, Gre., May 2007.
13. A. Norkin, A. Aksay, C. Bilen, G. B. Akar, A. Gotchev, and J. Astola, “Schemes for multiple description coding of stereoscopic video,” Proc. MRCS 2006, Istanbul, Turkey, Sept. 2006; Lec. Notes in Comp. Science, vol. 4105, pp. 730-737, Springer.

14. C. Bilen, A. Aksay, and G. Bozdagi-Akar, “A full frame loss concealment method for stereo video,” submitted to ICIP 2007.

15. S. Pehlivan, Anil Aksay, Cagdas Bilen, Gozde Bozdagi Akar, and M. Reha Civanlar, “End-to-End Stereoscopic Video Streaming System,” IEEE ICME 2006, Toronto, Canada, July 2006.

16. A. Aksay, S. Pehlivan, E. Kurutepe, C. Bilen, T. Ozcelebi, G. B.Akar, M. R. Civanlar, and A. M. Tekalp, “End-to-end stereoscopic video streaming with content-adaptive rate and format control,” Signal Processing: Image Communication, February 2007.

17. N. Ozbek, B. Gorkemli, A. M. Tekalp, and T. Tunali, “Adaptive streaming of scalable stereoscopic video over DCCP,” submitted to ICIP 2007.