MULTIMEDIA COMMUNICATIONS TECHNICAL COMMITTEE
IEEE COMMUNICATIONS SOCIETY
http://www.comsoc.org/~mmc

E-LETTER

Vol. 8, No. 3, May 2013

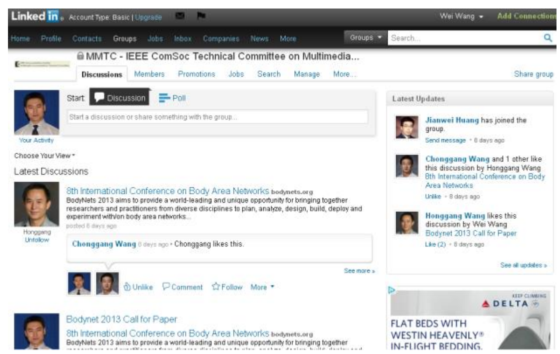
IEEE COMMUNICATIONS SOCIETY

CONTENTS

Message from MMTC Chair

Dear MMTC colleagues:

It is really a great pleasure for me to serve as the North America vice-chair for this vital ComSoc Committee during the period 2012-2014! As part of my duties, I have been devoted to promoting the MMTC's visibility and the benefits of all MMTC members.

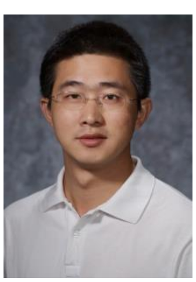


The first action is to create MMTC's social media footprints. I thank Dr. Honggang Wang from University of Massachusetts Dartmouth and Dr. Wei Wang from South Dakota State University whom I was working with closely in the past months to build MMTC group pages in Facebook and Linkedin. Both pages are developed to promote information dissemination and exchange among MMTC members. Facebook social media group provides a casual information dissemination/exchange environment, while the Linkedin group will provide such information in a more formal and professional settings. MMTC will plan to post official announcements and notices such as E-Letter and R-Letter updates. I enourage all of you to join and leverage both social groups to keep updated with MMTC and exchange information such as call for papers related to MMTC community, MMTC meeting reminders in major conferences, and hot topic discussion on multimedia telecommunication related technologies. You can join and connect to both groups using the following links. Please consult with Dr. Wei Wang (wei.wang@sdstate.edu), the MMTC administrators of these two groups, if you encounter any problems. We look forward to receiving your feedback and suggestions on both groups.

- MMTC Group on Facebook: http://www.facebook.com/pages/IEEE-ComSoc-MMC-Technical-Committee/292463107529841
- MMTC Group on LinkedIn: http://www.linkedin.com/groups?home&gid=4874863&trk=anet_ug_hm

The second action I was taking is to promote journal special issues for MMTC members. We received a good number of initial plans/proposals from MMTC Interest Groups (IG): 3DPRCIG (3D Rendering, Processing and Communications), ASPIG (Acoustic and Speech Processing for Communications), IMVIMEC (Image, Video & Mesh Coding), MCDIG (Multimedia Content Distribution: Infrastructure and Algorithms), MENIG (Multimedia Services and Applications over Emerging Networks), MSIG (Media Streaming), QoEIG (QoE for Multimedia Communications), and SecIG (Media Processing and Communications Security). In the following months, I will work with Dr. Yonggang Wen from Nanyang Technological University to coordinate full proposals from these IGs and assist contacting with corresponding journals on behalf of MMTC. As a contributing MMTC member, if you are intested in getting involved in journal special issues, I encourage you to contact IGs chairs and us.

I would like to thank all the IG chairs and co-chairs for the work they have already done and will be doing for the success of MMTC. I hope that any of you will find an IG of interest to get involved in our community!



Chonggang Wang

North America Vice-Chair
Multimedia Communications TC of IEEE ComSoc

EMERGING TOPICS: SPECIAL ISSUE ON 3D VIDEO: CONTENT CREATION, CODING, TRANSMISSION AND RENDERING

Guest Editors: Petros Daras, CERTH/ITI, Greece, Maria G. Martini, Kingston University, UK, Tasos Dagiuklas, Hellenic Open University, Greece

daras@iti.gr, M.Martini@kingston.ac.uk, dagiuklas@eap.gr

Recent technology developments on 3D Video (3DV) enable many revolutionary aspects that allow for more interaction with the viewer, such as Free Viewpoint Video (FVV) or Free Viewpoint Television (FVT). The capabilities, as well as the design, of a 3DV/FVV system are determined by the choice of the appropriate 3D representation of the visual content. In general, there are three representation options: i) image-based, ii) image plus depth-based and iii) geometry-based (i.e., 3D reconstructions). The choice of the 3D representation specifies the capturing requirements (e.g., number and position of cameras necessary) and on the other hand it determines the coding schemes, the transmission & rendering algorithms and capabilities such as the navigation range, robustness to occlusions, interactivity, degree of immersion, etc.

This special issue of E-Letter focuses on the recent progresses of the whole chain of 3D Video, i.e., from capturing to visualization until the QoE. It is the great honor of the editorial team to have six leading research groups, from both academia and industry laboratories, to report their solutions for meeting these challenges and share their latest results. Since this E-Letter is organized by three IEEE IGs, the first two papers are coming from the IG on 3D rendering, Processing and Communications, the other two from the IG on Image, Video and Mesh coding and the last two from the IG on QoE

The first article, *"An embedded 3D camera on FPGA"*, authored by Stefano Mattoccia *et al.*, from University of Bologna, Italy, focuses on 3D capturing and content creation. The paper presents the architecture of an embedded 3D camera based on stereo vision that entirely fits into a low cost FPGA without external memory. This 3D camera delivers dense depths maps and RGB images at high resolution and frame rate with minimal power requirements.

The second article, *"A Novel Planar Layered Representation for 3D Content and Its Applications"*, contributed by Burak Özkalayc and A. Aydın Alatan, METU, Turkey", introduces a novel 3D representation for multiview rendering applications. The proposed representation is a single reference based approach, such as layered depth video. The layers are not extracted based on visibility checks, but according to geometric models and the scene geometry is approximated by a planar space. The proposed 3D content representation is expected to be an input to object-based 3D compression algorithms.

The third article is contributed by Carl J. Debono, Sergio de Faria with the title *"3D Video Coding: Current State and Research Directions"*. This is a survey paper where the two basic 3D representations (multi-view and multi-view plus depth) are thoroughly analyzed and the corresponding coding schemes are explained. Furthermore, the main research challenges are identified as follows: a) the geometry information in the depth videos can be better exploited and therefore research in geometric transforms for motion and disparity compensation can help to reduce bandwidth requirements; b) encoding of depth data needs to better preserve the edges and thus unequal error protection can be applied; c) better up scaling techniques are required at the receiver, which could be realized by exploiting the texture information; d)the geometry in the depth information can be further used to obtain faster coding such as increase skip modes by mode prediction.

Giovanni Petrazzuoli *et al.* from Institut Mines-Télécom/Télécom ParisTech – CNRS/LTCI, presented a study on how Distributed Video Coding (DVC) -inspired techniques can help in the design of effective systems for 3D video services, in the fourth article entitled *"Enabling Immersive Visual Communications through Distributed Video Coding"*. The paper targets Immersive visual communications and analyzes the challenges of using DVC in the context of Interactive Multi-view Video Streaming (IMVS) and of compression of depth maps for multi-view video plus depth (MVD) representations.

Jing Li, Marcus Barkowsky and Patrick Le Callet from Ecole Polytechnique de l'Université de Nantes present recent advances in 3D QoE standardization activities from fora such as ITU, SMPTE and IEEE. In the area of 3D QoE assessment, the proposed methodologies, consider three basic perceptual dimensions: Picture Quality, Depth Quality, and Visual Comfort.

The last article of this special issue is entitled "Quality of Experience in 3D Video" by Ilias Politis from University of Patras and Varuna De Silva from University of Surrey. The paper presents the current state of research on measuring the 3D Quality of Experience (QoE), the factors that influence it and provides an overview of the future of QoE as the 3D technology matures and evolves further.

While this special issue is far from delivering a complete coverage on this exciting research area, we hope that the six invited letters give the audiences a taste of the main activities in this area, and provide them an opportunity to explore and collaborate in the related fields. Finally, we would like to thank all the authors for their great contribution and the E-Letter Board for making this special issue possible.

Petros Daras is an electrical and computer engineer (Diploma '99, MSc '02, PhD '05) graduated from the Aristotle University of Thessaloniki, Greece. He is a Researcher Grade B (Assoc. Prof.), in the Information Technologies Institute of the Centre for Research and Technology Hellas and head of the Visual Computing Lab (http://vcl.iti.gr). His main research interests include processing, retrieval, and recognition of 3D objects, 3D object reconstruction and coding, 3D image analysis, 3D medical image processing and segmentation, and bioinformatics. He acts as a reviewer for many major IEEE journals and serves as a TPC member in more than 50 international conferences. He has published more than 100 papers in international conferences and journals. Dr. Daras is a senior member of IEEE, chair of the IEEE Interest Group on Image, Video and Mesh coding and key member of the IEEE Interest Group on 3D rendering, Processing and Communications.

Maria Martini received the Laurea in electronic engineering (summa cum laude) from the University of Perugia (Italy) in 1998 and the Ph.D. in Electronics and Computer Science from the University of Bologna (Italy) in 2002.
She is a Reader (Assoc. Prof.) in the Faculty of Science, Engineering and Computing in Kingston University, London. An IEEE Senior Member, she has served as editor and reviewer for several IEEE international journals and as chair and member of the organizing and programme committee of several international conferences. She serves as Chair (2012-2014) of the IEEE Interest Group on 3D Rendering, Processing, and Communications. Her research interests include wireless multimedia networks, cross-layer design, 2D/3D error resilient video, 2D/3D video quality assessment, and medical applications. She is the inventor of several patents on wireless video.

Tasos Dagiuklas received the received the Engineering Degree from the University of Patras-Greece in 1989, the M.Sc. from the University of Manchester-UK in 1991 and the Ph.D. from the University of Essex-UK in 1995, all in Electrical Engineering. He is Assistant Professor at the Department of Computer Science, Hellenic Open University, Greece. Past positions include academic positions at TEI of Mesolonghi and University of Aegean Greece. Dr Dagiuklas is a Vice-Chair for IEEE MMTC QoE WG and Key Member of IEEE MMTC MSIG and 3DRPC WGs. He is also an active member of IEEE P1907.1 Standardiazation WG. He is a reviewer for journals such as IEEE Transactions on Multimedia, IEEE Communication Letters and IEEE Journal on Selected Areas in Communications. His research interests include 2D/3D video transmission/adaptation/rate control across heterogeneous wireless networks, P2P video streaming and service provisioning across Future Internet architectures. He is a Senior Member of IEEE and Technical Chamber of Greece.

An Embedded 3D Camera on FPGA

Stefano Mattoccia, Marco Casadio, Ilario Marchio,*
Michele Borgatti, Giovanni Neri, Vitaliano Gobbi
*University of Bologna, Bologna, Italy, *www.vision.deis.unibo.it/smatt*

1. Introduction

In recent years, motivated by the increasing consumer demand, many devices for 3D content creation have been proposed. Such devices often referred to as RGB-D cameras/sensors, enable delivering of dense depth maps and a color images of the observed scene in real-time. These devices rely on different technologies, however they can be roughly classified in two categories: active and passive sensors.

Active sensors infer depth projecting in the sensed environment specific patterns or emitting modulated light. An example of RGB-D camera based on pattern projection is the Microsoft Kinect while examples of RGB-D cameras based on emission of modulated light are sensors based on the *time-of-flight* (ToF) technology. A detailed analysis and review of these technologies can be found in [1]. Since its presentation in 2010 the Kinect sensor gained a lot of popularity thanks to the accuracy of the inferred depth maps and its low cost. For these reasons the Kinect is widely used for 3D content creation in many applications. Compared to current ToF cameras, the Kinect provides higher resolution depth maps and images both at VGA resolution.

The Kinect and ToF sensors have specific limitations, however both are unable to provide reliable depth measurements in regions where, respectively, the pattern or the emitted light vanishes (i.e. is *absorbed* by the sensed object). Moreover, devices of the same type may interfere when sensing the same area and, currently, these devices are not able to deal with sunlight.

On the other hand, passive RGB-D cameras rely on standard image sensor without any perturbation of the sensing environment. Stereo vision is a well known technique [2, 3] in computer vision to infer dense depth maps using two or more synchronized image sensors. This technology, by solving the correspondence problem (i.e. identifying the same point of the scene projected in the two or more image sensors of the stereo setup) and by knowing the *intrinsic* parameters of the sensors (e.g. focal lengths, principal points) and the *extrinsic* parameters of the stereo rig (i.e. the relative position of the image sensors in the 3D space), allows inferring depth. Compared to active cameras, this technology has mainly problems in poorly textured regions because the identification of corresponding points becomes ambiguous and hence unreliable. Nevertheless, being passive, these devices do not interfere when sensing the same area and the resolution of the inferred depth maps and images are constrained only by the image sensor technology deployed. Moreover, passive cameras are well suited to indoor as well as outdoor environments. However, solving the correspondence problem is a computationally demanding task and several computing architectures have been used to speed it up.

In this paper we briefly outline the architecture of an embedded 3D camera based on stereo vision that entirely fits into a low cost FPGA without external memory. This 3D camera delivers dense depths maps and RGB images at 752x480 at more than 30 frame per second (fps) with minimal power requirements.

2. Stereo vision

Stereo vision is a well known technique aimed at inferring 3D information from two or more images sensing the same scene from different viewpoints. This topic received a lot of attentions in the last decades and extensive reviews can be found in [2, 3].

In this paper we'll consider a *binocular* setup (i.e. made of two image sensors). Given two images, referred to as reference R and target T, being able to find corresponding points (i.e. projections of the same point of the scene in R and T), stereo vision allows obtaining depth by means of a simple triangulation [2]. However, a practical stereo vision system requires also the following preliminary steps: (offline) calibration and (runtime) stereo rectification of the stereo pair acquired by the image sensors.

Stereo calibration [7] is an offline procedure required once aimed at determining the intrinsic and extrinsic parameters of the stereo camera. These parameters can be determined by means of an offline calibration algorithm [7] that infers the intrinsic parameters, such as focal length and principal point, that characterize each camera. Moreover, calibration enables to infer the position of each camera with respect to a reference coordinate system (e.g. set in one of the two image sensors). The calibration algorithm infers intrinsic and extrinsic parameters by analyzing a known (typically planar) pattern (e.g. a checkerboard with known square size) in different positions. This procedure is required only once and can be easily done offline on a standard PC.

Once inferred the intrinsic and extrinsic parameters a function that removes lens distortions and transforms the raw image acquired by the image sensors in *standard form* images can be obtained. In a stereo vision system in standard form, corresponding points, if exist, lie on the same *scanline* in the two images. Once determined (offline) the rectifying function, at runtime, each new stereo pairs acquired by the two image sensors can be put in standard form to tackle the correspondence problem. Finding corresponding points in the two images is a challenging task and many algorithms have been proposed in the literature. According to the taxonomy proposed in [2, 3] most approaches perform four steps (*cost computation*, *cost aggregation*, *disparity optimization* and *refinement*) and algorithms can be roughly classified in *local* approaches and *global* approaches. The latter approaches are in most cases iterative and hence not suited to an FPGA system without external memory. Conversely, the former class mainly relies on cost aggregation deploying, on a point basis, a simple *Winner Takes All* (WTA) strategy [2,3]. The simplest approach aggregates matching costs on a fixed area centered in the points under examination but more sophisticated and effective methods that aggregates costs according to image content have been proposed in recent years [2,4,5]. In general, most local algorithms have a very simple computational structure and a small memory footprint typically well suited to our constrained computing architecture made of a low cost FPGA without any external device except for the communication controller (e.g. USB or Giga Ethernet). Once determined corresponding points (encoded in the *disparity map*) by means of an appropriate stereo matching algorithm, the position of each point in the 3D scene can be obtained by knowing the *baseline* b, the distance between the two image sensors, and the focal length *f* of the stereo camera (see [1,7] for details). Both parameters b and f can be inferred by the offline calibration procedure. In the next section we'll briefly outline the architecture of the 3D camera.

3. Architecture of the embedded 3D camera

The architecture of our embedded 3D camera is depicted in Figure 1. The two global shutter image sensors (manufactured by Aptina) deliver 752x480 images at a maximum frame rate of 60 fps. In our system the image sensors can be monochrome or color (with Bayer pattern) and they are perfecty synchronized in hardware. The resulting stereo stream and clock is transmitted through two LVDS (Low Voltage Differential Signaling) channels to the FPGA-based computing platform. This enables, when required, to put the computing platform (i.e. the FPGA) and the two image sensors at distances of some meters. Moreover, exploiting the same link between the two image sensors, the baseline can be easily set at a value appropriate for the specific application of the 3D camera.

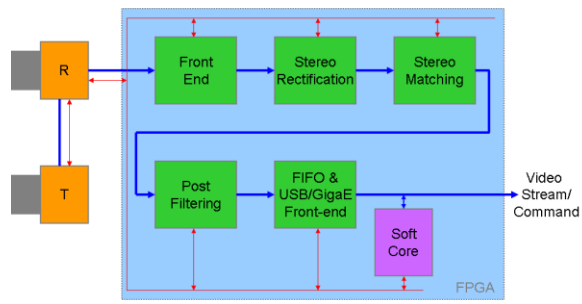


Figure 1 - Overview of the overall 3D camera

The stereo stream and the clock are then processed by a front end module implemented into the FPGA logic that translates the LVDS serial stream to a parallel stereo stream. This is than sent to the next modules (image rectification, stereo correspondence, etc.). It is worth observing that to meet our constraints (i.e. a single FPGA without additional external memory) we need to massively apply a *stream processing* methodology. This means that the pixel provided by the image sensors, at a maximum frequency of 27 MHz in the specific case of Aptina's sensors deployed, must be processed as soon as they are available in the front end module in order to minimize the memory footprint within the FPGA. Our target devices belong to the Xilinx Spartan 6 family, ranging from model 45 to 150 [6]. The Spartan 6 model 45 has 43 K logic cells and about 260 KB of fast internal memory, referred to as block RAM, organized in 116/232 independent blocks. Therefore, according to our constraints, efficient stream processing approach is mandatory.

Once the two raw images are processed by the front end the next step consists in removing lens distortions and putting the raw images in standard form. Rectification can be seen as a mapping, according to the parameters inferred during calibration between the raw stereo pairs and the rectified stereo pairs (i.e. stereo pairs with *epipolar* lines aligned with imaged scanlines and without lens distortions). To accomplish this task we warp the two raw images keeping as low as possible buffering requirements. As output, we have two image streams in lockstep (i.e. at each clock the rectification module sends out the same pixel for the two images).

The next step tackles the correspondence problem, also referred to as stereo matching, in order to identify the same projection of the scene in the two image planes. However, in practical application scenarios the two images are photometrically distorted and this would lead to unreliable correspondences. Therefore, to deal with this problem, we apply a robust image transformation to the stereo pair. We have tested different approaches and those based on the analysis of the relative ordering between pixels such as the Census transform are quite effective. Once we have applied a robust matching measure to the incoming images, the stereo matching module indentifies corresponding points. To this aim we have mapped to the FPGA a local algorithm that aggregates cost within a fixed image patch. Despite its simplicity this algorithm poses a significant challenge when implemented on an FPGA-based system such as those considered in our setup. Moreover, despite its well-known problems [2], this algorithm performs quite well in many practical applications. Nevertheless, we have also implemented more effective algorithms on the same target platform not discussed here for the lack of space. Independently of the stereo matching algorithm used, the correspondence problem is not always reliable due to occlusions or other artifacts in the images. Therefore, a module, referred to as *post-filtering*, that identifies unreliable correspondences follows the stereo matching module. Unreliable correspondences are identified by analyzing the input images detecting regions with lack of texture as well as the correspondences found by the previous module exploiting the so called *uniqueness constraint*. The post-filtering module also performs 1/16 subpixel depth estimation.

The resulting disparity map at subpixel resolution and the RGB images are sent to a module that contains a FIFO and an interface to the communication controller. Our system currently supports USB 2.0 and Giga Ethernet interfaces. A system with a USB 3.0 interface is currently under development.

The processing platform also includes a supervisor implemented into the FPGA logic as a soft core. This module programs the image sensors and initializes the processing modules within the FPGA. Moreover, the supervisor also enables fast bidirectional communications between the 3D camera and the host.
The overall system described so far has been mapped to Spartan 6 FPGA, model 45 an up, and delivers depth maps and RGB images at more than 30 frames per second with stereo pairs up to 752x480 pixels. Finally, we point out that the overall device, including the two image sensors shown in Figure 2, can be self powered by a standard USB port enabling for 3D content creation in low power scenarios such as those involving mobile devices.

Figure 2 – Current prototype of the 3D camera

4. Preliminary experimental results

In this section we report, in Figures 3 and 4, preliminary results concerned with real-time rectification and matching using as computing platform a Spartan 6 model 45 and as communication channel the USB 2.0 interface. Figure 3 shows the reference rectified image (sporting monochrome sensors) and Figure 4 shows the disparity map, encoded with a *colormap* that assigns warmer colors to points closer to the camera and colder colors to points far from the camera, computed by the FPGA in real-time at 30fps.

Figure 3 – Real-time operation: rectified reference image with monochrome image sensors

5. Conclusions

In this paper we have briefly described the architecture of a RGB-D camera based on stereo vision technology. The whole system delivers depth maps and RGB images in real-time and fits into a single low cost

FPGA. The passive technology deployed by this camera enables unconstrained setup for 3D content creation also in application scenarios characterized by low power requirements such those involving mobile devices. Moreover, being based on a passive technology, this camera is well suited in environments with multiple passive or active 3D sensors.

Figure 4 – Preliminary results computed in real-time, with respect to the reference image shown in Figure 3, by the 3D camera

References

[1] C. Dal Mutto, P. Zanuttigh, G. M. Cortelazzo, Time-of-flight Cameras and Microsoft Kinect, Springer, 2012

[2] R. Szeliski, Computer Vision: algorithms and applications, Springer, 2010

[3] D. Scharstein and R. Szeliski, A taxonomy and evaluation of dense two-frame stereo correspondence algorithms, Int. Journal of Computer Vision, 47(1/2/3):7-42, April-June 2002

[4] F. Tombari, S. Mattoccia, L. Di Stefano, E. Addimanda, Classification and evaluation of cost aggregation methods for stereo correspondence, Int. Conference on Computer Vision and Pattern Recognition (CVPR 2008)

[5] S. Mattoccia, S. Giardino, A. Gambini, Accurate and efficient cost aggregation strategy for stereo correspondence based on approximated joint bilateral filtering, Asian Conference on Computer Vision (ACCV2009), 2009

[6] Xilinx, Spartan 6 Family, DS160 available at www.xilinx.com

[7] R. Hartley and A. Zisserman, Multiple view geometry in computer vision, Cambridge University Press, 2004

[8] R. Zabih, J. Woodfill, Non-parametric local transforms for computing visual correspondence, ECCV 1994

Stefano Mattoccia is Assistant Professor at the School of Engineering and Architecture of the University of Bologna, DISI (Department of Computer Science and Engineering). He received an MSc Degree in Electronic Engineering and a PhD in Electronic Engineering and Computer Science from the University of Bologna. His research interests include stereo vision algorithms and applications, 3D sensor fusion and registration, image matching and mapping of computer vision algorithms on FPGAs, GPUs and embedded devices.

Marco Casadio received his Master's Degree in Information Engineering from the University of Bologna in 2012. He works as Temporary Research Fellow at DISI (Department of Computer Science and Engineering), University of Bologna. His research interests include Computer Vision, Software and hardware design and development.

Ilario Marchio received his Master's Degree in Information Engineering from the University of Bologna, Italy, in 2012. He is interested in the implementation of computer vision algorithms on FPGAs.

Michele Borgatti received the M.Sc. EE (with honors) and the Ph.D. EECS from the University of Bologna, Italy, in 1994 and 1998, respectively. He held senior management positions at STMicroelectronics and UPEK, Inc. and has been consulting multiple Corporations in wireless communications and automotive. His research interests include system and integrated circuit design for signal processing, multimedia portable applications, reconfigurable computing and hardware/software co-design.

Giovanni Neri received his MSc.Degree (Electronic Engineering) from the University of Bologna, Italy, in 1969 with Hons (100/100). From 1969 he held various academic positions and he is currently Full Professor in Electronic Computers at the Dipartimento di Ingegneria dell'Energia Elettrica e dell'Informazione "Guglielmo Marconi" of the University of Bologna (DEI), Italy. His research interests include Computer Vision, Computer Architecture and Electronic Signatures.

.

Vitaliano Gobbi got his Degree in Electronic Engineering from the University of Bologna in 2012 and since then he collaborates with the computer vision group at DISI. He is interested in any field concerning computer science and electronic design especially those focused on FPGA applications.

A Novel Planar Layered Representation for 3D Content and Its Applications

Burak Özkalaycı and A. Aydın Alatan
Department of Electrical and Electronics Eng., METU,Turkey
bozkalayci@gmail.com, alatan@eee.metu.edu.tr

1. Introduction

3D display technologies have quickly advanced during the last decade and reached to the consumers initially as 3DTVs with glasses in a two-view stereoscopic format. The next breakthrough for 3DTV applications is expected as glass-free multi-view 3D experience on autostereoscopic and/or holographic displays [1]. In the two-view case, 3D content is still provided in the conventional video format; however, the experiments conducted for the multiview amendment of MPEG-4 AVC (MVC) showed that the encoding rate almost increases linearly by the number of views [2]. Hence, for a feasible multiview 3D application, novel 3D representations, which are efficiently compressible and "3D experience friendly", are strictly required.

An extensive analysis of 3D representation techniques is available in [3]. Among them, *Depth Image Based Rendering* (DIBR) techniques seem to be one of the key technologies. As long as the scene geometry is available, DIBR methods might generate visually satisfactory novel views of the scene. Instead of containing all the views of the 3D content, the depth information of the scene is expected to reduce the redundant texture information of the scene for an efficient multiview representation.

The pioneering solution for DIBR based approach is the *N-view+N-depth* or *Multi View plus Depth* (MVD) representation. In MVD, a limited number of views with their depth maps are utilized and multiview of the scene is generated by DIBR techniques. MPEG community has announced a “Call for Proposal on 3D Video Coding” [4] in order to set the standardization of the 3D video transmission, and the main exploration experiments of the proposal is achieved on 2 and 3-video plus depth data sets. The exploration experiments show that the multiple referenced scene geometry information satisfies high quality novel view rendering, while keeping the number of views as low as 2 or 3.

An alternative representation for the multiview content is Layered Depth Video (LDV) [5, 6]. In LDV, the texture information of the multiview content is accumulated to a single reference view with its depth information. The layers are then extracted consecutively by the visibility checks. In comparison to MVD, LDV is quite promising for avoiding redundant texture and depth information by merging them to a single reference view. A hybrid approach of MVD and LDV is also proposed as *Depth Enhanced Stereo* (DES) in [7].

For assessing the compression performance of these representations, the main criterion should be the 3D experience quality of the viewer. Hence, the distortion metric of a 3D video should be redesigned by this perspective. Especially, the compression of the depth information must be handled in the guidance of the DIBR artifacts. In the literature, there is a growing concern for efficient depth compression in novel view rendering [8-10] and suitable depth distortion metrics for a better 3D experience [11,12]. The experiments support that preserving the depth discontinuity regions are crucial for high quality DIBR results.

This letter introduces a novel 3D representation for multiview rendering applications. The proposed representation is a single reference based approach, such as LDV. However, the layers are not extracted based on visibility checks, but according to geometric models. Each layer is modeled as a plane in 3D space by corresponding texture information. Hence, the scene geometry is approximated by a planar space. Next, the motivation of the proposed representation with a sketch of the approach will be introduced and some preliminary experimental results are also presented.

2. Planar Layered Representation

The initial motivation of the proposed representation is based on the problem of analyzing the distortion relations between the geometry and texture. The quantization parameter used in the conventional video compression techniques is explicitly correlated with the objective video quality metrics. Borrowing the conventional compression techniques to MVD compression problem leads to the question “what is the optimal quantization parameter pair for texture and depth compression?” Since DIBR operation brings various non-linear relations, it is hard to explicitly track the novel view rendering quality performance by the distortion measures on texture and depth. In [13], the rate-distortion optimal quantization parameter pairs of the texture and depth modalities are obtained by a brute force parameter search. In order to handle this problem, novel depth distortion metrics, which estimate the distortion on rendered novel views, are also proposed in the literature [11,12].
At this point, the proposed 3D content representation aims to decouple the geometric distortions from texture

distortions for novel view rendering applications. Since the ultimate aim of efficient transmission of a 3D content is creating a pleasant 3D experience for 3DTV or free-view TV applications, we propose handling the visual/textural distortions and geometric distortions separately. In this context, by geometric distortion, it is meant changing the depth and geometry of the objects in the scene, but not changing their textural properties. As an extreme geometric distortion example, the depth ordering based pseudo 3DTV representation [14] might be stated. Since using a limited number of disparity levels might still create a meaningful 3D perception, we firmly believe that there is still some room for distorting the scene geometry for an efficient 3D representation, while maintaining a pleasant 3D experience. On the other hand, the ultimate manipulation of scene geometry should be object based. After extracting the texture information of the objects in the scene, their relative depths can be modified in order to find an optimal representation for the constraints of a feasible application, such as bandwidth, 3D perception, etc. This kind of manipulation is similar to manipulating the 3D objects in a computer graphics environment.

Since precisely extracting objects from captured views is difficult, except for computer generated scenes, the proposed 3D representation segments scene regions which are represented as planar surfaces in 3D space. The planar scene regions are extracted by minimizing the geometric distortion (approximately modeling distortions in 3D perception), and using minimum number of planar layers with smooth boundaries (efficient compressibility).

The proposed planar representation fits 3D planes to the conjugate 3D point cloud representation of the N-view plus N-depth data. The plane fitting problem is solved in a MRF-like energy minimization framework as given below.

$$\arg\min_{f} \lambda_d \sum_{v=1}^{N} \sum_{p \epsilon I_v} D_v(f_p) + \lambda_s \sum_{v=1}^{N} \sum_{pq \epsilon \mathcal{N}_{3D}} V_{pq}(f_p, f_q) + \lambda_l \sum_{L \subseteq \mathcal{L}}^{N} \delta_L(f)$$

The data cost term, D_v, which is the geometric distortion of the solution, is regularized by two additional cost terms. The first one is the smoothness cost term V_{pq}, that helps to obtain object-like segments by forcing spatially neighboring pixels in 3D space to be on the same plane. The second regularization term is the labeling cost which forces the solution to be composed of minimum number of planar models, i.e, enforcing *minimum description length* (MDL) principle. The minimization problem of the cost energy can be solved approximately by Graph Cut (GC) algorithm [15]. An Expectation-Maximization based search technique, named PEARL [16], is also utilized in order to sample the continuous 3D plane parameter space efficiently.

The planar approximation of 3D point cloud of MVD data, segments the 3D scene into planar layers. The constraints utilized during the model fitting step intend to coincide with the layer boundaries as much as by the object boundaries. Considering the extracted planar layers as 3D objects, it is possible to assign texture information to them. This aim is the main motivation of the texture extraction of the proposed representation. The texture representation of the layers is based on single reference view LDV representation. The pixels on each view that are assigned to the same planar model are 3D warped to the reference view by a visibility check in order to accumulate the texture information of the planar layer. 3D warping of pixels is performed according to ground truth depth values of the MVD data. Hence, the texture extraction step is not distorted by the geometric distortions of the planar representation, which is crucial for the ultimate aim of uncoupling the distortions in geometry and texture. After obtaining the texture information of each layer, the proposed representation turns out to be a set of images with a known layer boundary and planar model which determines the depth value of each pixel. A typical representation is illustrated in Fig. 1 for *Teddy* dataset.

The novel view rendering for the proposed representation is performed by simply warping the planar textures to the desired image plane. The layer warping can be implemented efficiently and precisely by texture rendering techniques of the computer graphics by the analytical planar models of the proposed representation. Since the depth variations across a layer exploit the continuum of a planar model, the novel layered representation is very effective for artifact-free novel view rendering, which is promising to be the most important advantage of the proposed scheme.

In Figure 2 the planar layer based representation and the novel view rendering results of the *Dancer* dataset is given. The mean depth reconstruction error is 0.94

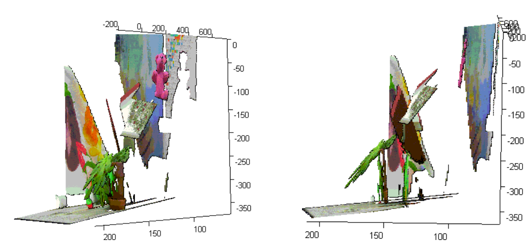


Figure 1. The textured planar layers in 3D space from two different point of views for *teddy*.

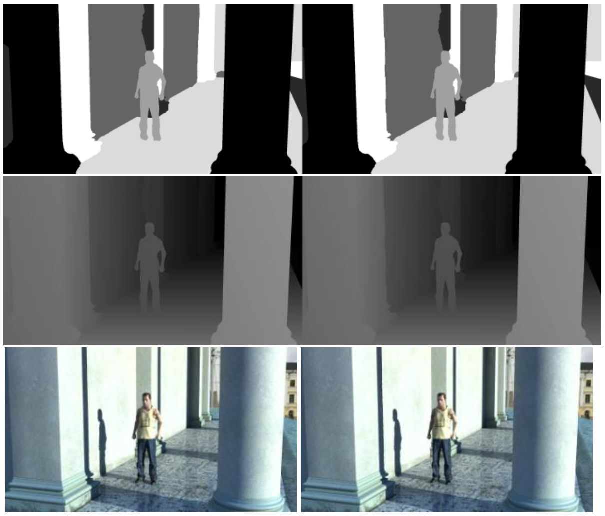


Fig. 2. Top row is the planar layer assignments, mid row is the planar depth reconstruction of the views, and bottom row is the two novel view rendering results.

and the number of planar layers utilized in the representations is only 6. The given novel views are located at 0.25 and 0.75 of the baseline between the left and right views.

3. Discussion

The proposed 3D content representation is expected to be an input to an object-based 3D compression algorithm. For the evaluation of the proposed representation in 3D applications the conventional metrics for novel view rendering, such as PSNR comparison is not a fair method since geometric distortions are allowed for the proposed representation. The changes in geometry inherently drops the PSNR scores, whereas the 3D experience and visual quality can be kept at a higher level. As an ultimate solution for the proposed method, the scene objects might be coherently segmented into the planar layers and the geometry is wisely distorted so that the 3D experience is kept satisfactory in an efficient representation. Hence, we argue that this representation also brings the question of “what is the optimal geometric distortion of a 3D content for an efficient representation, while keeping 3D experience satisfactory?” In order to advance in this perspective, the effects of geometric distortions during 3D perception should be studied subjectively.

References

[1] H. Jegou, M. Douze, and C. Schmid, “Improving bag-of-features for large scale image search,” in *IJCV*, vol. 87, no. 3, pp. 316-336, 2010.

[2] P. Merkle, et al. “Efficient prediction structures for multiview video coding,” IEEE Transactions on CSVT, vol. 17, pp. 1461-1473, 2007.

[3] A. A. Alatan, et al. "Scene representation technologies for 3DTV—a survey" IEEE Transactions on CSVT, vol. 17, pp. 1587-1605, 2007

[4] ISO/IEC JTC1/SC29/WG11, "Call for Proposals on 3D Video Coding Technology," MPEG output document N12036, March 2011.

[5] J. Shade, S. Gortler, L. He, and R. Szeliski, “Layered depth images,” SIGGRAPH, pp. 231-242, 1998.

[6] K. Muller, et al. “Reliability-based Generation and View Synthesis in Layered Depth Video”, IEEE 10th Workshop on Multimedia Signal Processing, pp. 34-39, 2008

[7] A. Smolic, et al. “An overview of available and emerging 3D video formats and depth enhanced stereo as efficient generic solution,” PCS, pp. 1–4, 2009.

[8] M. Maitre, “Shape-adaptive wavelet encoding of depth maps,” PCS, pp. 4-7, 2009.

[9] A. Sánchez, G. Shen, and A. Ortega, “Edge-preserving depth-map coding using graph-based wavelets,” 43th Asilomar Conf., pp. 578-582, 2009.

[10] P. Merkle, Y. et al. “The effects of multiview depth video compression on multiview rendering,” Signal Processing: Image Communication, vol. 24, pp. 73-88, 2009.

[11] W. Kim, A. Ortega, P. Lai, and D. Tian, “Depth map distortion analysis for view rendering and depth coding,” ICIP, pp. 721-724, 2010.

[12] G. Tech, et al. “3D Video Coding using the Synthesized View Distortion Change,” PCS, pp. 25–28, 2012.

[13] K. Muller, A. Smolic, K. Dix, and P. Merkle, “Coding and intermediate view synthesis of multiview video plus depth,” ICIP, pp. 741-744, 2009.

[14] E. Turetken and A. A. Alatan, “Temporally Consistent Layer Depth Ordering via Pixel Voting for Pseudo 3D Representation,” 3DTV Conference, pp. 1-4, 2009.

[15] Y. Boykov, et al. “Fast approximate energy minimization via graph cuts,” IEEE PAMI, 2001.

[16] H. Isack and Y. Boykov, “Energy-based geometric multi-model fitting,” IJCV, vol. 97, no. 2, pp. 123-147, 2012.

Burak Özkalaycı received his B.S. and M.S degrees from Middle East Technical University, Turkey, all in electrical engineering, in 2003, 2006 respectively. He is currently a PhD student in the same department.

A. Aydın Alatan received the B.S. degree from Middle East Technical University, Turkey in 1990, the M.S. and D.I.C. degrees from Imperial College of Science, Medicine, and Technology, U.K., in 1992, and the Ph.D. degree from Bilkent University, Turkey, in 1997, all in electrical engineering. Since August 2000, he is a faculty member at Electrical and Electronics Engineering Department in Middle East Technical University.

3D Video Coding: Current State and Research Directions

Carl J. Debono[1], Sergio de Faria[2]
[1]*Department of Communications and Computer Engineering, University of Malta, Malta*
[2]*Instituto de Telecomunicações, ESTG / Instituto Politécnico de Leiria, Portugal*
c.debono@ieee.org, sergio.faria@co.it.pt

1. Introduction

Latest developments in network infrastructures together with advancements in video capturing equipment and better signal processing and coding algorithms have increased demands for more multimedia services. Amongst these is the development of immersive multimedia experiences. To present a 3D representation of a scene, two (stereoscopic) or more (multi-view) views of the same scene must be captured and transmitted over a channel, which is usually bandwidth-limited. Using a Multi-View Video (MVV) representation allows a number of applications, including Free-View Video (FVV) and 3D TeleVision (3D TV) 0. Furthermore, commercially available stereoscopic and auto-stereoscopic 3D displays are accessible and content availability is increasing. This content is being consumed through 3D Blu-ray disks, 3D broadcasts and the Internet [1].

The capturing and transmission of multi-view videos demands huge bandwidths, something which is not readily available. Hence, efficient 3D video coding is imperative for this technology to succeed. Current 3D broadcasts rely mainly on transmission of stereoscopic video encoded using H.264/AVC and transmitted using MPEG-2 transport streams. However, in order to obtain better 3D TV quality, exploit auto-stereoscopic displays, and allow FVV applications more views need to be available at the receiver. Furthermore, to allow real-time content transmission and use it in portable devices, the complexity of the encoding and decoding schemes must be limited.

In this paper, we will look at the current advancements in 3D video coding and highlight the research directions in this field.

2. 3D Video Representations

This section gives an overview on current 3D video representations, namely frame compatible and simulcast, stereo, and those expected to be used in the near future: Multi-View (MV) and Multi-View plus Depth (MVD).

Currently, 3D video communication is based on stereo and its transmission is mostly performed through frame compatible formats or simulcast schemes. In frame compatible formats, both views of a stereo pair are accommodated into a single frame, generating a single video stream, by means of sub-sampling each frame in space or in time. In simulcast video transmission each view is independently encoded into a stream, at the full resolution, having the final bitrate proportional to the number of encoded views.

In order to have a more immersive 3D experience some devices are allowed to display multiple views (multi-view). But, encoding multi-view video requires a larger bandwidth to transmit it, which grows with the number of views. Less bandwidth would be required is some views were artificially interpolated in the receiver. Actually, it is also possible to encode only 2 or 3 views and create the remaining intermediate views, if additional scene geometry information is used, like depth maps. This depth information increases the capability of rendering and solving occlusions between views [2]. This format is labeled Multi-View plus Depth, and is expected to be used by the future 3D devices, namely auto-stereoscopic displays.

3. Multi-view Video Coding

Both frame compatible format and simulcast used for stereo video communication involve a single view video encoder, like H.264/AVC. However, the result is poor, as in the former the resolution is lost and in the latter the compression is inefficient.

The key point is to exploit the inter-view correlation of the multi-view video sequences. Multi-view Video Coding (MVC) is a method that exploits spatial, temporal and inter-view redundancies present in video captured at the same instant from multiple cameras. The most important step into multi-view compression has been done with standardization of the *Multiview High Profile* and the *Stereo High Profile*, in the amendment (Annex H) of the H.264/AVC, to support Multi-view Video Coding, which was been the adopted video compression on the 3D Blu-ray discs. Using inter-view prediction tools, this scheme achieves 25-30% gains in bitrate in coding the auxiliary view [3], in comparison to H.264/AVC. A typical frame structure and inter-frame/view coding dependency is illustrated in Figure 1.

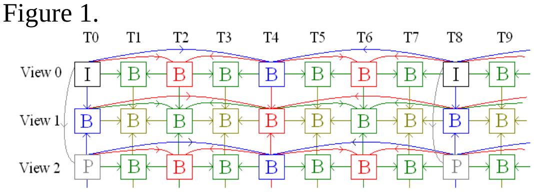


Figure 1 Hierarchical Bi-Prediction MVC structure (GOP=8)

Motion estimation has been extended to estimate a macroblock in a frame from another viewpoint. Thus, frames from other views can be included in the reference list, allowing for more accurate predictions in motion and disparity.

In order to promote bit rate savings, asymmetrical view coding can also be performed, where one view is encoded with less quality than the other, for example by means of coarse quantization or by reducing the spatial resolution [5]. Results of subjective picture quality evaluation suggested that, as compared to 2D, a 20~25% bitrate increase for the auxiliary view provides good subjective quality for 3D HD video applications.

The recently approved standard for high efficiency video coding (HEVC) [6] has emerged with a new set of coding tools, which have increased the compression efficiency. Evaluation tests for quality comparison against H.264/AVC demonstrated that HEVC is able to increase the compression by about 50% at the same quality [7]. HEVC simulcast yet outperforms H.264 MVC, due to its powerful tools. However, the computational complexity is increased in comparison to H.264/AVC in between 2 to 10 times more in the encoder, but is approximately the same in the decoder. The increase in complexity for multi-view encoding will be much higher.

The extension (Annex F) of HEVC for multi-view video coding (MV-HEVC) will enable HEVC-based high quality video coding, hence providing an expected reduction of 50% in the bit rate required by previous products and services for 3D television and 3D Blu-ray discs (H.264 MVC). As H.264 MVC, MV-HEVC takes advantage of the inter-view predictions tools to exploit redundancy between views, also surpassing the HEVC simulcast [8].

4. Multi-view Video-plus-Depth Coding

Depth-based representation is emerging as an important class of 3D format, as it allows the generation of virtual views through depth-based image rendering (DBIR) techniques. Thereby, it enables display-independent solutions, as the displays may generate the desirable number of views. Through view synthesis prediction schemes, the encoder achieves more efficient compression by exploiting the depth information, typically compressed at 20% of the texture video. Although the depth data is not directly output to a display and viewed, maintaining the fidelity of the depth information is very important. It has a drastic influence on the view synthesis quality, due to the geometric information provided by the depth.

The depth data can contain coding errors which result in wrong pixel shifts in synthesized views, leading to annoying artifacts, especially along object boundaries.

Joint video and depth coding is the current research path to enable a high compression scheme for multi-view video-plus-depth format. Beyond the use of inter-view prediction techniques, that can be applied to both video and depth independently, there is some block-level information, such as motion vectors, that may be similar for both data types and thus can be shared. In the context of multi-view coding, adjacent views may be warped in relation to the reference view in order to reduce the residual error between views. As video coding algorithms are typically designed to preserve low frequencies and image blurring, it is required to develop techniques to deal with the characteristics of depth signals, keeping the edge information fidelity.

Some preliminary results for the 3D-HEVC have shown this scheme achieves a Bjøntegaard Delta (BD) bitrate gain of 22,7% over MV-HEVC and 47,6% over HEVC simulcast, considering the synthesized videos and an overall bit rate of video and depth [9]. 3D-HEVC is an extension of HEVC with block-level tools, where the base view is fully compatible with HEVC and dependent views use additional tools that exploit the correlation of motion and mode parameters between base and dependent views.

In the design of joint video and depth coding the quality optimization for the synthesized views is an important issue. Instead of evaluating the decoded objective quality, against an uncoded reference, the 3D video-plus-depth format additionally demands a good quality of the intermediate synthesized views. As usually the original intermediate view does not exist, a comprehensive subjective evaluation is carried out. This is important as new types of errors may occur, like pixel shifts, regions with wrong depths or outworn object boundaries at depth edges [10]. The uncoded references require high quality depth data and a robust view synthesis method, such that a reference image looks indistinguishable from the original. Experimental results show that intermediate synthesized views present lower image quality, when the allocated bit rate for depth maps is unbalanced for the view coding [10].

Besides these results, all of the dependencies between video and depth information are still under evaluation in terms of their compression, rendering capabilities, as well as compatibility and complexity of future coding algorithms. It is also possible that single depth and asymmetrical frame sizes for view and depth will be supported in future standards.

5. Research trends in 3D Video Coding

Current research efforts are in producing better quality 3D video while minimizing bandwidth requirements and time for encoding and decoding. Pattern matching together with predictive coding algorithms are being applied to 3D images and video to exploit more redundancies in the data. The geometry information in the depth videos can be better exploited and therefore research in geometric transforms for motion and disparity compensation can help reduce bandwidth requirements further. Encoding of depth data needs to better preserve the edges and thus unequal error protection can be applied. This can lead towards a scalable coding approach. Some techniques downscale the depth information to preserve bandwidth, however this means that better upscaling techniques are required at the receiver [1], which could be done by exploiting the texture information. The geometry in the depth information can be further used to obtain faster coding such as increase skip modes by mode prediction 0.

References

[1] Y. S. Ho, and K. J. Oh, “Overview of multimedia video coding,” in *Proc. 14th Int. Workshop on Systems, Signals, & Image Processing & EURASIP Conference Focused on Speech & Image Processing, Multimedia Communications & Services*, pp. 5-12, Maribor, Slovenia, Jun. 2007.

[2] A. Vetro, A. Tourapis, K. Müller, and T. Chen “3D-TV content storage and transmission,” *IEEE Transactions on Broadcasting, Special Issue on 3D-TV Horizon: Contents, Syst. and Visual Perception*, vol. 57, no. 2, pp. 384-394, Jun. 2011.

[3] K. Müller, P. Merkle, and T. Wiegand, “3D video representation using depth maps,” in *Proc. of the IEEE, Special Issue on "3D Media and Displays"*, vol. 99, issue 4, pp. 643-656, Apr. 2011.

[4] T. Chen, Y. Kashiwagi, C.S. Lim, T. Nishi, “Coding performance of Stereo High Profile for movie sequences,” in *ITU-T & ISO/IEC JTC1/SC29/WG11 Doc. JVT-AE022*, London, UK, 2009.

[5] T. Chen, and Y. Kashiwagi, “Subjective picture quality evaluation of MVC Stereo High Profile for full-resolution stereoscopic high-definition 3D video applications,” in *Proc. IASTED Conference on Signal and & Image Processing*, Maui, Hawai, USA, Aug. 2010.

[6] ITU Video Coding Experts Group and MPEG ISO/IEC JTC1/SC29/WG11, "High Efficiency Video Coding" ITU-T Rec. H.265 and ISO/IEC 23008-2 (HEVC), 2013.

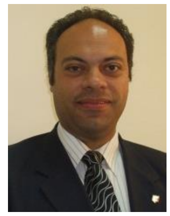
[7] G. J. Sullivan, J. Ohm, W. Han and T. Wiegand, “Overview of the High Efficiency Video Coding (HEVC) standard,” *IEEE Trans. Circuits and Systems for Video Technology*, Dec. 2012.

[8] M. Domański, T. Grajek, D. Karwowski, K. Klimaszewski, J. Konieczny, M. Kurc, A. Łuczak, R. Ratajczak, J. Siast, O. Stankiewicz, J. Stankowski, and K. Wegner, “Coding of multiple video+depth using HEVC technology and reduced representations of side views and depth maps,” in *Proc. Picture Coding Symposium*, Krakow, Poland, May 2012.

[9] A. Vetro, and D. Tian, “Analysis of 3D and multiview extensions of the emerging HEVC standard,” in *Proc. of SPIE Conf. on Applications of Digital Image Processing XXXV*, Paper 8499-33, San Diego, USA, Aug. 2012.

[10] S. Yea, and A. Vetro, “View synthesis prediction for multiview video coding,” *Signal Processing: Image Communication*, vol. 24, no. 1+2, pp. 89-100, 2009.

[11] B. W. Micallef, C. J. Debono, and R. A. Farrugia, “Exploiting depth information for efficient multi-view video coding,” in *Proc. IEEE International Conference on Multimedia* and *Expo*, Barcelona, Spain, Jul. 2011.



Carl James Debono (S’97, M’01, SM’07) received his B.Eng. (Hons.) degree in Electrical Engineering from the University of Malta, Malta, in 1997 and the Ph.D. degree in Electronics and Computer Engineering from the University of Pavia, Italy, in 2000.

Between 1997 and 2001 he was employed as a Research Engineer in the area of Integrated Circuit Design with the Department of Microelectronics at the University of Malta. In 2000 he was also engaged as a Research Associate with Texas A&M University, Texas, USA. In 2001 he was appointed Lecturer with the Department of Communications and Computer Engineering at the University of Malta and is now an Associate Professor. He is currently the Deputy Dean of the Faculty of ICT at the University of Malta. He is a senior member of the IEEE and served as chair of the IEEE Malta Section between 2007 and 2010. He has served on various technical program committees of international conferences and as a reviewer in journals and conferences. He is also a member of the management committee of the COST Action IC1105 - 3D Content Creation, Coding and Transmission over Future Media Networks (3D-ConTourNet) where he chairs the 3D Media Coding Working Group. His research interests are in wireless systems design and applications, multi-view video coding, resilient multimedia transmission, and modeling of communication systems.

Sérgio M. M. de Faria (M’93–SM’08) was born in Portugal, in 1965. He received the engineering degree and the M.Sc. in electrical engineering from the Universidade de Coimbra, Portugal, in 1988 and 1992 respectively, and the Ph.D. degree in electronics and telecommunications from the University of Essex, Essex, U.K., in 1996.

Since 1990, he has been with the Department of Electrical Engineering, Escola Superior de Tecnologia e Gestão, Instituto Politécnico de Leiria, Portugal. He is a researcher and Head of group of Multimedia Signal Processing - Leiria in Instituto de Telecomunicações, Portugal. He is senior member of the IEEE. He has edited 2 books and is the author of 5 book chapters and 11 journal papers. He holds 2 patents and he has more than 76 papers published in international conferences. He has been responsible for various funded projects, national and international, He is also a member of the management committee of the COST Action IC1105 - 3D Content Creation, Coding and Transmission over Future Media Networks (3D-ConTourNet). His current research interests include digital signal and image processing, namely, image and video compression, motion representation, and 3D video coding.

Enabling Immersive Visual Communications through Distributed Video Coding

Giovanni Petrazzuoli, Marco Cagnazzo, Béatrice Pesquet-Popescu, Fréderic Dufaux
Institut Mines-Télécom/Télécom ParisTech – CNRS/LTCI
{petrazzu, cagnazzo, pesquet, dufaux}@telecom-paristech.fr

1. Introduction

The realism of video communication tools has been steadily increasing in the last years, thanks to the advances in several fields: efficient video compression techniques allow conveying high-resolution video over network connections; high-dynamic-range imaging promises to further improve the perceived quality of videos; higher resolution and frame-rate systems are under way. In spite of all these improvements, a further enhancement of the visual communication experience is expected in the next years. One of the most anticipated new features is the addition of depth representation to visual conferencing systems. The huge expectation related to 3D video [1] is testified by the standardization effort produced by the research community: the MPEG group is finalizing the normalization of the multi-view video plus depth (MVD) representation [1]. This format consists in sending a few views, each with an associated "depth map", i.e. a single-component image representing the distance of each pixel to the camera. This can be obtained by using a certain number of so-called range cameras. The MVD format enables the generation of high quality virtual views (i.e. views that did not exist in the original video) by Depth Image Based Rendering (DIBR) [3], and thus would make it possible to have new and exciting interactive services [4] such as Free Viewpoint Television (FVT) [5] where users have the impression of being present at a real event.

Obviously, MVD has a huge redundancy: not only in time, as ordinary mono-view video, but also among views (*inter-view* correlation) and between views and depth maps (*inter-component* correlation). All these kinds of redundancies have to be exploited in order to reduce the storage space on the server and the bandwidth used for transmission [6]. These requirements are equivalent in the context of non-interactive scenarios (like TV broadcasting): the entire video stored on the server will be sent to the user. An alternative, interesting paradigm of multiple-views video is the Interactive Multiview Video Streaming (IMVS) one [7][8][9]. IMVS enables the client to select interactively the views that he/she wants to display. Given this constraint, the server will send only the data needed to display the views according to the switch pattern decided by the user. However, the video is first encoded and stored in a server and afterwards it is sent to the clients. The user will choose a pattern of viewpoints which is not known at the encoding time, and that will possibly change from a user to another. Nevertheless, we would like to minimize both the storage space demanded by the compressed video and the bandwidth needed to interactively send the requested view to the user. These requirements are conflicting in the case of IMVS, which makes the problem challenging. The fact that the user trajectory of viewpoints is unknown at the encoding time makes it difficult to employ differential coding to reduce the transmission rate.

Distributed video coding (DVC) could be an effective solution for the IMVS problem, since inter-image correlation can be exploited even if the encoder ignores the actual reference image. However DVC of MVD video involves some new challenges. In the following we introduce our recent work in using DVC in the context of IMVS and of compression of depth maps for MVD.

2. Interactive Multiview Video Streaming

In order to optimize the bandwidth use for IMVS, the server should store a huge number of versions of the original video, given by all the possible switches among views (redundant P-frames). On the other hand, in order to reduce the storage space, one could just encode all the video pictures as *Intra* frames (I-frames). However this would result in an inefficient use of the channel bandwidth. A possible solution [7], proposed in the case of multiview video without depth, is to use Distributed Video Coding (DVC). In DVC [10][11], the correlation among images is not exploited by temporal prediction at the encoder. The available frames are used to produce an estimation of the current one (for example, by motion-compensated temporal interpolation). This estimation, also called *Side Information* (SI) is considered as a noisy version of the original image. Therefore, it is improved by sending some parity bits to the decoder. Of course, the parity bits do not depend on the SI: whatever the estimation of the current image, the correction bits are always the same. This means that the server just needs to store and send a set of parity bits for a given picture. In [7], the authors find an optimal combination of I-frames, Redundant P-frames and M-frames for the view switching, but they are interested in the case of multi-view video coding without the depth information.

However, we remark that to the best of our knowledge, only a few papers deal with IMVS in Multiview Video plus Depth (MVD) with the constraint of ensuring that the video playback is not interrupted during switching.

In our work [8] we propose several methods for allowing arbitrary switches between two views in an MVD stream. We use the available depth-maps to produce DIBR estimation of a given image in the target view; then this estimation may or may not be corrected by parity bits. According to whether the DIBR is performed immediately (i.e. in the time instant of the view switch) or on the previous image (in advance), we have four switch strategies, in addition to the case of no-DIBR and the one where the first image of the GOP is estimated with DIBR. The best strategy depends on the position of the switching instant with respect to the GOP structure of the target view. For example, if the switch instant corresponds to an Intra image, there is no need for DIBR, while if it is in the middle of the GOP the advance DIBR + parity bits will give the best performances.

In summary, we have found the best strategy according to the temporal position of the switch. According to our study, the best single strategy is advance DIBR + parity bits: it allows an average bit-rate saving of 3.4% (Bjontegaard delta rate, [12]) with respect to the simple no-DIBR technique. Further rate savings can be obtained if the strategy is selected adaptively with respect to the temporal instant [8].

3. Wyner-Ziv Depth Map Coding

When compressing multiple video plus depth streams, one should always try to exploit not only spatial, temporal, and inter-view correlation, but also the correlation existing between textures and depth maps.

This observation is still valid when one considers *distributed* coding of MVD. In our work [9] we consider the distributed compression of depth maps. We develop several techniques for compressing depths and exploiting inter-component correlation. More precisely, we proposed new techniques for producing the side information of the current depth image. The reference scheme is the Discover SI generation method [11], used directly on the depth images. Discover performs a temporal interpolation assuming straight trajectories for object motion. A first improvement to this scheme is to consider high-order interpolation motion estimation (HOMI, [13]) for object trajectories. HOMI needs an initialization for object trajectory, and usually the linear motion of Discover is a good initialization. However, the depth images are not well suited for the motion estimation process, since they are very flat. Therefore better results can be obtained by using the motion trajectory computed on the corresponding texture images: we exploit the inter-component correlation as the movement of an object is pretty much the same in the two domains. However, since sometimes depth motion can differ from texture motion, we propose a further improvement: the trajectories are initialized with Discover on the texture, but then they are refined using HOMI on the depth data. This scheme is shown in Figure 1..

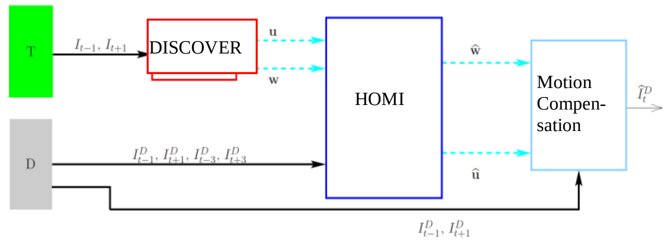


Figure 1. The proposed scheme depth SI generation

The proposed technique allows gaining up to more the 11% (Bjontegaard delta rate) with respect to the simple Discover SI generation of depth maps.

4. Conclusion

Immersive visual communications are one of most awaited technical innovations of the next few years. Several challenging problems must be solved before efficient solutions can be proposed. This study shows how DVC-inspired techniques can help in the design of effective systems for 3D video services.

References

[1] F. Dufaux, B. Pesquet-Popescu, M Cagnazzo (eds.): Emerging Technologies for 3D Video. Wiley, 2013

[2] Call for Proposals on 3D Video Coding Technology, ISO/IEC JTC1/SC29/WG11 Doc. N12036, Geneva, CH, March 2011.

[3] C. Fehn, "A 3D-TV Approach Using Depth-Image-Based Rendering (DIBR)," in *Proc. of 3rd IASTED Conference on Visualization, Imaging, and Image Processing*, Benalmadena, Spain, Sept. 2003, pp. 482–487

[4] C. Fehn, P. Kauff, O. Schreer, and R. Schäfer, "Interactive virtual view video for immersive TV applications," in *Proc. of Intern. Broadcast Conf*, 2001, pp. 14–18.

[5] M. Tanimoto, M.P. Tehrani, T. Fujii, and T. Yendo, "Free-viewpoint TV," *Signal Processing Magazine, IEEE*, vol. 28, no. 1, pp. 67 –76, 2011.

[6] M. Flierl and B. Girod, "Multiview video compression," Signal Processing Magazine, IEEE, vol. 24, no. 6, pp. 66–76, Nov. 2007.

[7] G. Cheung, A. Ortega, and N. M. Cheung, "Interactive streaming of stored multiview video using redundant frame structures," Image Proc., IEEE Trans., 2010.

[8] G. Petrazzuoli, M. Cagnazzo, F. Dufaux, B. Pesquet-Popescu. "Using distributed source coding and depth

image based rendering to improve interactive multiview video access". In *IEEE International Conference on Image Processing*, vol. 1, pp. 605-608, September 2011. Bruxelles, Belgium.

[9] G. Petrazzuoli, M. Cagnazzo, F. Dufaux, B. Pesquet-Popescu. "Wyner-Ziv coding for depth maps in multiview video-plus-depth". In *IEEE International Conference on Image Processing*, vol. 1, pp. 1857-1860, September 2011. Bruxelles, Belgium.

[10] B. Girod, A. Aaron, S. Rane, and D. Rebollo-Monedero, "Distributed video coding," Proc. IEEE, vol. 93, no. 1, pp. 71–83, Jan. 2005.

[11] X. Artigas, J. Ascenso, M. Dalai, S. Klomp, D. Kubasov, and M. Ouaret, "The DISCOVER codec: Architecture, techniques and evaluation, picture coding symposium," in Coding of Audio-Visual Objects, Part 10: Advanced Video Coding, 1st Edition, 2007

[12] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," in VCEG Meeting, Austin, USA, Apr. 2001.

[13] G. Petrazzuoli, M. Cagnazzo, and B. Pesquet-Popescu, "High order motion interpolation for side information improvement in DVC," in *IEEE Int. Conf. on Acoustics Speech and Signal Processing*, 2010, pp. 2342 –2345.

Giovanni Petrazzuoli obtained the PhD degree from Telecom-ParisTech in January 2013, with a thesis on distributed video coding for multi-view and multi-view plus depth video. His research interests also cover interactive streaming and shape adaptive coding.

Marco Cagnazzo obtained the Ph.D. degree from Federico II University and the University of Nice-Sophia Antipolis, Nice, France in 2005. Since 2008 he has been Associate Professor at Telecom ParisTech within the Multimedia team. His research interests are scalable, robust, and distributed video coding, 3D and multi-view video coding, multiple description coding, and video delivery over MANETs. He is the author of more than 70 journal articles, conference papers and book chapters, and a co-editor of the book "Emerging Technologies for 3D Video: Creation, Coding, Transmission, and Rendering", Wiley Eds., (to appear in 2013).

Dr. Cagnazzo is an Area Editor for Elsevier Signal Processing: Image Communication and Elsevier Signal Processing. He is an IEEE Senior Member, a Signal Processing Society member and an EURASIP member.

Beatrice Pesquet-Popescu received the Ph.D. thesis from the École Normale Supérieure de Cachan in 1998. Since Oct. 2000 she is with Télécom ParisTech, first as an Associate Professor, and since 2007 as a Full Professor, Head of the Multimedia Group. She is also the Scientific Director of the UBIMEDIA common research laboratory between Alcatel-Lucent Bell Labs and Institut Mines Télécom. She serves as an Editorial Team member for IEEE Signal Processing Magazine, and as an Associate Editor for IEEE Trans.CSVT, IEEE Trans Multimedia, IEEE Trans. Image Processing and for Elsevier Signal Processing: Image Communication. She holds 23 patents in video coding and has authored more than 270 book chapters, journal and conference papers in the field. She is a co-editor of the book to appear "Emerging Technologies for 3D Video: Creation, Coding, Transmission, and Rendering", Wiley Eds., 2013. Beatrice Pesquet-Popescu is an IEEE Fellow.

Frédéric Dufaux is a CNRS Research Director at Telecom ParisTech. He is also Editor-in-Chief of Signal Processing: Image Communication. He received his M.Sc. in physics and Ph.D. in electrical engineering from EPFL in 1990 and 1994 respectively. Frédéric has over 20 years of experience in research, previously holding positions at EPFL, Emitall Surveillance, Genimedia, Compaq, Digital Equipment, MIT, and Bell Labs. He has been involved in the standardization

of digital video and imaging technologies, participating both in the MPEG and JPEG committees. He is currently co-chairman of JPEG 2000 over wireless (JPWL) and co-chairman of JPSearch. He is the recipient of two ISO awards for these contributions. Frédéric is an elected member of the IEEE Image, Video, and Multidimensional Signal Processing (IVMSP) and Multimedia Signal Processing (MMSP) Technical Committees.

His research interests include image and video coding, distributed video coding, 3D video, high dynamic range imaging, visual quality assessment, video surveillance, privacy protection, image and video analysis, multimedia content search and retrieval, and video transmission over wireless network. He is the author or co-author of more than 100 research publications and holds 17 patents issued or pending. He is a co-editor of the book to appear "Emerging Technologies for 3D Video: Creation, Coding, Transmission, and Rendering", Wiley Eds., 2013.

Recent Advances in Standardization on 3D Quality of Experience

Jing Li, Marcus Barkowsky, Patrick Le Callet
LUNAM Université, Université de Nantes, IRCCyN UMR CNRS 6597
Polytech Nantes, rue Christian Pauc BP 50609 44306 Nantes Cedex 3, France
{jing.li2,marcus.barkowsky,patrick.lecallet}@univ-nantes.fr

1. Introduction

For the last decades, video quality assessment has mostly tackled 2D video sequences. Technological advances were mostly tackling coding and transmission schemes while the display technology, especially in lab viewing environments, could be considered as transparent. Subjective assessment methodologies needed to be selected mostly with respect to the severity of the degradations. Typical examples are Absolute Category Rating with Hidden Reference (ACR-HR) from ITU-T P.910 for strong degradations experienced in networked multimedia scenarios, and Paired Comparison (PC) or Double Stimulus Continuous Quality Scale (DSCQS) from ITU-R BT.500 for near lossless scenarios such as satellite transmissions.

The introduction of affordable stereoscopic 3D viewing not only in digital cinemas but also in the home environment recently led to new technological developments. For example, in terms of standardization for coding and transmission requirements, the Digital Video Broadcasting Project (DVB) and the European Broadcasting Union (EBU) provide interim recommendations regarding producing, exchanging, archiving, distributing, coding, and transmitting 3D programs using 2D compatible or newly developed 3D infrastructure and transmission technologies.

In analyzing this production and broadcasting chain, one challenge is the measurement of the perceived video quality. for this emerging content. Three main issues were identified.

Firstly, the display device was no longer transparent enough for the observers. While restrictions of 2D devices were widely accepted, i.e. a flat screen with a colored bezel of a certain extent, standing on some surface at a certain viewing distance, 3D viewing introduced glasses based technologies with reduced brightness and reduced temporal or spatial resolution, for active shutter glasses or passive polarized technology respectively. Different 3D display technologies influence the viewing experience differently and none of them can be considered as mature as 2D displays.

Secondly, due to the “added value of depth”, the traditional one dimensional concept of “quality” for 2D video sequences is not sufficient to measure all sensations involved in a communication service for stereoscopic 3D. Quality of Experience (QoE) was proposed as a global term and it has been defined in different ways in literatures nowadays. A well approved definition has been provided in the Qualinet White paper summarizing various recent publications from more than 20 research groups [1]. QoE is defined as “the degree of delight or annoyance of the user of an application or service. It results from the fulfillment of his or her expectations with respect to the utility and / or enjoyment of the application or service in the light of the user’s personality and current state”. QoE in 3D is often split into three basic perceptual dimensions: Picture Quality, Depth Quality, and Visual Comfort. Higher level concepts, e.g. naturalness and presence, are then derived from these dimensions which ultimately lead to a global QoE value.

Thirdly, the subjective assessment methodology itself needs to be reconsidered. Observers feel comfortable with providing an absolute 2D video quality score because they know well what to expect from 2D videos. As 3D technology is not used on a daily basis, such an internal reference scale may not be available to observers in 3D stereoscopic experiments. Reliability and reproducibility become questionable which in turn calls for improvements in currently standardized testing methodologies. Therefore, either multi-dimensional scale value need to be used or direct comparison with the reference or between degraded sequences is required. Whether these methodologies provide similar results and may therefore be considered reliable and reproducible is still an open question.
Current standardization activities related to these three issues are presented in the following sections.

2. Display Characterization: SID’s ICDM standard

ITU-R BT.500 has been the reference standard for display specifications for the past decades. Its specifications mostly concerned brightness and contrast measurements which could be performed with calibrated luminance meters. Current stereoscopic 3D displays require further characterization on the newly added perception issues, concerning for example:

- The measurement of crosstalk, i.e. the amount of light that is perceived by one eye while it is destined for the other eye.
- Viewing angle measurements, which are

particularly important with autostereoscopic and polarized screens. The physical distance between the screen's pixel illumination and the view dividing material, for example, a parallax barrier, a lenticular array sheet, or a line polarizing foil, introduces a view position dependency which may even lead to the inversion of the stereoscopic views at extreme positions.

- The contrast ratio and the luminance uniformity, which are often influenced nowadays by display specific post-processing methodologies. For example, overdrive technology allows achieving a very fast response time in terms of black-to-white or grey-to-grey. For stereoscopic displays this may introduce view asymmetries.

The Society of Information Display (SID) recently published a standard created by its International Committee for Display Metrology (ICDM). The Information Display Measurement Standard (IDMS) contains methods for reproducible measurement of display variables for various display types as well as requirements for the measurement devices, such as angular resolution for luminance meters, a parameter particularly important for measuring autostereoscopic displays [2].
The standard focuses on measuring technological variables which impact Quality of Experience measurements in subjective assessments without quantifying their influence on perception

3. Multi-scale measurement of 3DTV broadcasting: ITU-R BT.2021 Recommendation
Subjective methodologies for the assessment of stereoscopic 3DTV systems including general test methods, grading scales and viewing conditions have been recently combined in ITU-R BT.2021[3]. Four subjective methods from ITU-R BT.500 are included in this recommendation, which are Single-Stimulus (SS), DSCQS, PC and Single Stimulus Continuous Quality Evaluation (SSCQE). These methods are suggested to measure the three primary dimensions of QoE independently: Picture Quality, Depth Quality and Visual Comfort. They are not suggested to be used for the assessment of naturalness, sense of presence, or overall QoE. As depth quality and visual comfort are new added terms for the traditional 2D methods, these methods should be used with a slight modification. For example, for assessing the visual comfort, the continuous scale is labeled with the attributes "Very comfortable", "Comfortable", "Mildly uncomfortable", "Uncomfortable", and "Extremely uncomfortable".

The viewing conditions are consistent with the 2D viewing conditions recommended by ITU-R BT.2022[4]. The selection of the test video sequences should be performed carefully as visual discomfort or visual fatigue might generate safety and health issues for the viewers. Experiments that aim particularly at measuring visual comfort are exempted from this rule. The maximum value of the disparity, the discrepancies between left and right images including geometric distortions, brightness discrepancy and crosstalk, and the change on distribution of the disparity, etc. are factors that need investigation during the video selection.

Furthermore, the vision screening process, instructions prior to the experiment, the training session, and the duration of the formal test are different from 2D studies in order to deal with visual discomfort issues.

4. Efforts towards international standardization in 3DTV
While some results of the world-wide research has been recently integrated into standards as shown above, further research is ongoing to refine the measurements and to prove their reliability and reproducibility. The 3DTV group of the Video Quality Experts Group is working on the three distinct projects. The first project aims at establishing a Ground Truth database for 3D QoE measurement methodologies by evaluating a reasonably large dataset with Paired Comparison methodology as it has been widely accepted that people may express their preference for one of a pair of sequences. The second project studies the influence factors of the viewing environment on reproducible QoE measurements with emphasis on video quality. The third project is the validation of objective models to predict the perceived quality of stereoscopic video sequences when the degradations are mostly related to video quality and not to visual comfort or depth issues. These efforts are stimulating discussions in the ITU-T Study Group 9 and the ITU-R Study Group 6.

Furthermore, studies are being progressed on scalable view-range representation for free viewpoint television (FTV). The Society of Motion Picture & Television Engineers (SMPTE) focuses on the standardization related to stereoscopic 3DTV in production environments, e.g., the exchange of 3D content amongst mastering facilities, and between a mastering facility and the ingest facility of a distribution system. This includes the specification of requirements for a data representation of disparity maps, methods to transport a pair of stereoscopic frames within single video frames, etc.[5].

IEEE P3333, which is a so-called individual-based project approved by the IEEE Standards Association, works on the quality assessment of 3D Displays, 3D Contents and 3D Devices based on Human Factors, such as photosensitive seizures, motion sickness and visual fatigue and identify and quantify the causes of those factors.

5. Conclusion

The reliable and reproducible assessment of 3D stereoscopic contents is progressing and intermediate standards have been established allowing the industry to perform the required quality assurance tasks. It should be noted that many guidelines provide rather conservative specifications that may reduce the public interest in 3D transmissions. Most of the cited standards and organizations are open to revisions when new evidence becomes available from scientific studies and cross-validations.

References

[1] P. Le Callet, S. Möller, A. Perkis, eds, "Qualinet White Paper on Definitions of Quality of Experience (2012)", European Network on Quality of Experience in Multimedia Systems and Services (COST Action IC 1003), Lausanne, Switzerland, Version 1.1, June 3, 2012

[2] Society for Information Display - International Committee for Display Metrology, "Information Display Measurements Standard (IDMS)", 2012

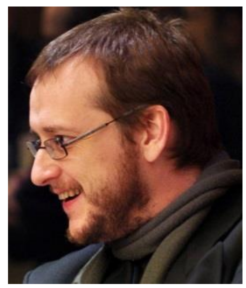
[3] International Telecommunication Union - Radiocommunication Sector, "Recommendation ITU-R BT.2021: Subjective methods for the assessment of stereoscopic 3DTV systems", 2012

[4] International Telecommunication Union - Radiocommunication Sector, "Recommendation ITU-R BT.2022: General viewing conditions for subjective assessment of quality of SDTV and HDTV television pictures on flat panel displays", 2012

[5] International Telecommunication Union - Radiocommunication Sector, "Report ITU-R BT.2160-2: Features of three-dimensional television video systems for broadcasting", 2012

Jing Li received her B.S. degree in Electronic Engineering from Xi'an University of Technology, Shaanxi, China and M.S. degree in Electronic Engineering from Xidian University, Shaanxi, China, in 2006 and 2010, respectively. The research interest during her master study was image and video quality assessment. Since 2010, she has been a Ph.D. student in the Image and Video Communications Group at IRCCyN at the University of Nantes. Her Ph.D. thesis focuses on subjective assessment methodologies on QoE, visual discomfort and quality of experience in 3DTV.

Marcus Barkowsky received the Dr.-Ing. degree from the University of Erlangen-Nuremberg in 2009. He joined the Image and Video Communications Group at IRCCyN at the University of Nantes in 2008, and was promoted to associate professor in 2010. His activities range from modeling effects of the human visual system, in particular the influence of coding, transmission, and display artifacts in 2D and 3D to measuring and quantifying visual discomfort and visual fatigue on 3D displays using psychometric and medical measurements. He currently co-chairs the VQEG "3DTV" and "Joint Effort Group Hybrid" activities.



Patrick Le Callet received M.Sc. degree PhD degree in image processing from Ecole Polytechnique de l'Université de Nantes. He was also student at the Ecole Normale Superieure de Cachan where he got the "Aggrégation" (credentialing exam) in electronics of the French National Education. Since 2003, he is teaching at Ecole Polytechnique de l'Université de Nantes in the Electrical Engineering and the Computer Science department where he is now Full Professor. Since 2006, he is the head of the Image and Video Communication lab at CNRS IRCCyN, a group of more than 35 researchers. He is mostly engaged in research dealing with the application of human vision modeling in image and video processing. His current centers of interest are 3D image and video quality assessment, watermarking techniques and visual attention modeling and applications. He is co-author of more than 190 publications and communications and co-inventor of 13 international patents. He has coordinated and is currently managing for IRCCyN several National or European collaborative research programs. He is serving in VQEG (Video Quality Expert Group) where he is co-chairing the "HDR Group" and "3DTV" activities. He is currently serving as associate editor for IEEE transactions on Circuit System and Video Technology, SPIE Journal of Electronic Imaging and SPRINGER EURASIP Journal on Image and Video Processing.

Quality of Experience in 3D video

Ilias Politis[1] and V. De Silva[2]

1. *Wireless Telecommunications Lab., University of Patras, 26500, Greece*
2. *I-Lab Multimedia Research Center, University of Surrey, Guildford, UK*

Abstract—Recently the entertainment industry, the multimedia content providers and the research community have been experiencing the decisive introduction and evolution of three-dimensional (3D) imaging. This far more complex and sophisticated than the traditional 2D, technology raises the major problem of how the user perceives quality of 3D video. This paper summarizes the current state of research on measuring the 3D Quality of Experience (QoE), the factors that influence it and provides an overview of the future of QoE as the 3D technology matures and evolves further.

I. INTRODUCTION

The effectiveness of distributed multimedia applications as well as mobile computing services, which seem to become dominant in the modern telecommunication era, is primary based on the networking protocols and communication systems that deliver content to the end-user. Research and development in those protocols and delivery system, is currently driven from a technical perspective being given to the user's benefit. However it is a fact that the aforementioned effectiveness of any multimedia presentation is ultimately measured by the user's multimedia experience in terms of information simulation and satisfaction.

Quality of Experience can be defined as the overall acceptability of an application or service strictly from the user's point of view. It is a subjective measure of end-to-end service performance from the user perspective and it is an indication of how well the network meets the user's needs [1]. Encompassing many different aspects QoE rivets on the true feelings of end users when they watch streaming video and podcasts, listen to digitized music and browse the Internet through a plethora of methods and devices [2].

In this paper the influencing factors of 3D QoE are examined in Section II, and an overview of the most up-to-date standardization activities for 3D video quality assessment methods and metrics is presented in Section III. The future challenges the 3D QoE will face are described in Section IV and Section V summarizes a number of open research issues and concludes the paper.

II. FACTORS AFFECTING 3D QoE

Quality of Experience involves a number of different factors that influence the viewing experience. These factors are generally grouped into three categories: Subjective factors, System factors and Context factors [3].

In the first category, subjective factors that influence the perceptual ability of the viewer include the physical, emotional and mental state of the viewer, as well as the human ability to comprehend and judge the stimuli. Therefore, the viewer's age, gender, visual capability, ability to focus and retain attention, are parameters that will influence the quality of visual perception. Additionally, the educational background of the viewer, the personality, experience, emotions and even socio-economic status, bias their opinion of quality.

In addition, many System factors related to the quality of the application or service and extending over the entire delivery chain, are affecting the perceived quality. In particular, parameters related to the content, temporal-spatial properties, color depth and texture, could influence the perceived quality. Moreover, this category also involves parameters related to encoding schemes, resolution, frame rate, audio-visual synchronization. In particular for 3D content, new aspects that influence QoE include depth perception, naturalness, sense of presence and visual discomfort. Moreover, network related parameters, including loss, delay, jitter, bandwidth, throughput are factors that could distort the visual quality. Finally, the end-user devices are influencing the QoE according to their specifications, characteristics and capabilities. Parameters like, the screen resolution, the display size, memory, battery life, user-friendliness of the interface and its mobility or portability, are also some of the devise related factors that will determine the level of quality that the viewer is experiencing.

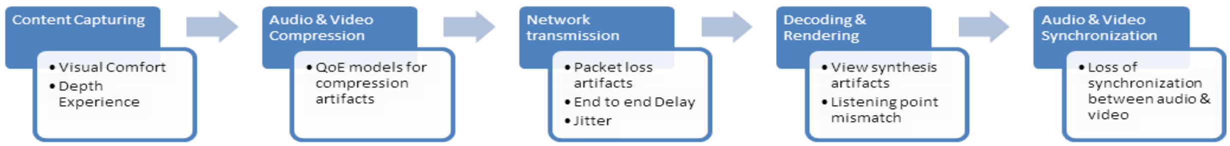


Fig. 1 Example of audiovisual content delivery chain and QoE measurements required at each process.

The third category comprises of parameters that are related to the physical environment (indoors, outdoors, mobile or stationary), the time of day and the duration of use, the subscription cost and the charging of the application or service and also factors that are related to the type of interaction with other users or devices (i.e. social media) or networks (i.e. peer-to-peer networking). Such Context related factors could condition the perceived quality.

A. QoE Measurement Framework

Some of the factors that influence the 3D QoE across the entire delivery chain are illustrated in Fig. 1. In particular, each block, which represents one step in the video delivery chain, the major factors related with the particular step have been identified. Therefore during content capturing it is important to consider factors such as visual comfort of the captured content and the sensation of depth. For example, depth/disparity variations in the scene should not exceed certain thresholds, and scene changes should be planned such that they do not cause visual discomfort. Compression of video with the aid of state-of-the-art codecs yield artifacts such as blurring and blocking. The effect of such artifacts should be carefully modeled to preserve subjective quality. With the approach of Visual Attention Modeling video encoding can take into account what part of the visual scene probably draws the viewers' attention. This information helps to improve subjective quality. Issues such as error concealment due to packet losses occurring due to congestion or fading also need to be considered during QoE modeling. To cater user requirements, such as arbitrary view point switching, the intermediate views need to be synthesized (i.e. rendered) from the available views. Depending on the quality of disparity map and the whole filling algorithm utilized, the rendered views will have different artifacts that affect the user perception. During audio rendering it is also important to measure listening point and viewpoint mismatch and its effect on the overall 3D perception.

III. ASSESSMENT OF 3D VIDEO QoE

Standardization activities for 3D QoE assessment are focusing in all stages of the 3D video delivery chain. Since the International Telecommunications Union (ITU) and International Standardization Organization (ISO) are responsible for standardizing most of the systems from coding to video display, they have also undertaken the standardization of quality assessment methods.

Particularly, ITU standardization work on 3D video QoE is handled by several groups. WP6C, a subgroup of Study Group 6 in the ITU-R, is responsible for developing recommended methods of quality assessment for broadcast services ITU-T Study Group 9 addresses questions related to IPTV and cable TV and Study Group 12 is responsible for recommendations on performance, quality of service (QoS) and quality of experience (QoE) for the full spectrum of terminals, networks and services. The Video Quality Experts Group (VQEG) is a joint effort of experts from various backgrounds and affiliations that combine their expertise on video quality assessment of communications and broadcasting video services. Curently, ITU-R WP6 and ITU-T SG9 are working on the following recommendations:

- ITU-T P.3D-sam – a work on 3D assessment methods over the current 3D environment [4]
- ITU-T J.3D-fatigue – a work on visual fatigue for 3D video and associated safety assessment [5]
- ITU-T J.3D-disp-req – a work on requirement for displays used during 3D video QoE measurement testing [6]

ITU-T SG12 is currently working on recommendation P.STAT, which addresses the methods, metrics and procedures for statistical evaluation, qualification and comparison of objective quality prediction models.

Additionally, VQEG is currently focusing in the following areas, namely: studying the influence of the sense of depth and comfort/viewing fatigue on 3D image quality, estimating the cross-talk effect on displays currently used for subjective tests and investigating assessment methodologies to determine the suitability of three-view video codecs.

Recently, ISO published the recommendation ISO/TR 9241-331 [7], which focus on the optical characteristics of autostereoscopic displays and summarize the assessment methods for determining viewing range.

It is apparent that the reliable QoE assessment of 3D video requires a new array of methodologies and metrics. Hence, the task is currently the focus of numerous research studies [8],[9]. The conducted experiments through new light on how the Human Vision System (HVS) perceives 3D content and underline the various challenges 3D quality assessment methods face, including visual fatigue, binocular rivalry and suppression and so on.

IV. QoE CHALLENGES

Higher resolution (i.e ultra-HD), advanced display screens (e.g. retina) exceed the perception acutance of the observers and can support higher frame rates and as a result fluent motion reconstruction. Such evolutions naturally raise the question of how the content quality can be measured and which is its added value to the Quality of Experience. Moreover, new assessment methods and metrics need to be designed in order to measure the higher dynamic ranges and the extended color gamut offer by todays and future consumer devices.

Multimodal Quality of Experience assessment requires a new set of metric and methods to be defined in order to measure not only visual quality but audio and interaction, as well. Additionally, a more extended set of human senses needs to be monitored including haptics and temperature, as there is yet no model that can provide an insight of how additional senses, or external factors like hunger or viewing posture, affect the quality of visual experience.

There is also the need for an objective measurement of the observer's state of mind that may provide valuable insight on the observer's immersion during 3D video content watching. Towards this end, Quality of Experience will need at some point to involve continuous health monitoring and measurement of essential perception factors including eye dryness, or binocular rivalry, in order to adapt the display parameters to the observer's requirements by lowering the depth amount or the resolution.

Finally, user-optimized 3D QoE may require the adapting of the temporal fluidity of the multimedia content (e.g. adapt the duration of action sequences), in order to increase user's immersiveness, based on monitoring observer's reactions and forecasting expectations and anticipations.

V. OPEN RESEARCH AREAS & CONCLUSION

Quality of Experience is not just video quality in the contrary, is an effort to measure the actual degree of immersiveness by using methods that step beyond everything QoS has ever managed to achieve. The challenges involved with the measurement of QoE can be found along the entire 3D video delivery chain. There is a variety of perceptual attributes that affect 3D video QoE including the overall image quality, naturalness, presence, perceived depth, comfort and immersiveness, which in turn are the result of technical, social and psychological factors. To gain an overall picture of QoE it is imperative to perform a set of extensive subjective experiments assessing different aspects of the system. However, these subjective measurements need to redesigned and adopt multi-scale methods that will account for multi-modal evaluation, spatial-temporal synchronization, interactive services, etc. Moreover, the environment that the subjective experiments are being performed needs to be redefined, in order to provide a more "at home" experience. The disadvantages of subjective experiments that may be expensive and time consuming could be compensated by objective measurement methods of 3D QoE that correlate well with subjective results. Thus it is necessary that these objective methods will be based on ground truth data and psycho-visual experiments that among others, will throw light on how humans judge depth cues, fuse multi-scale indicators and perceive 3D artifacts.

REFERENCES

[1] Nokia-Siemens Networks, "QoS-enabled IP networks for assured multiplay QoE," *Technical White Paper*, 2011.

[2] K.-T. Chen, C.-J. Chang, C.-C. Wu, Y.-C. Chang, and C.-L. Lei, "Quadrant of Euphoria: A crowdsourcing Platform for QoE Assessment," *in IEEE Network*, pp.28-35, Mar. 2010.

[3] Le Callet, P., S. Möller, and A. Perkis., "Qualinet white paper on definitions of quality of experience," 2012."

[4] Chen, Wei, Fournier Jérôme, and Jean-Charles Gicquel. "New 3D video subjective test method to consider in draft Recommendation ITU-T P. 3D-sam," 2012.

[5] **ITU-R WP6C, "**Programme production and quality assessment," ITU-T SG 9, 2011.

[6] ITU-T SG 9 (Study Period 2013) - Broadband cable and TV, "Draft New Recommendation J.3D-disp-req", Display requirements for 3D video quality assessment.

[7] ISO/TR 9241-331:2012, "Ergonomics of human-system interaction -- Part 331: Optical characteristics of autostereoscopic displays" Stage: 60.60, 2012.

[8] Quan Huynh-Thu; Le Callet, P.; Barkowsky, M., "Video quality assessment: From 2D to 3D — Challenges and future trends," *Image Processing (ICIP), 2010 17th*

IEEE International Conference on , vol., no., pp.4025,4028, 26-29 Sept. 2010.

[9] Urvoy, M.; Barkowsky, M.; Cousseau, R.; Koudota, Y.; Ricorde, V.; Le Callet, P.; Gutierrez, J.; Garcia, N., "NAMA3DS1-COSPAD1: Subjective video quality assessment database on coding conditions introducing freely available high quality 3D stereoscopic sequences," *Quality of Multimedia Experience (QoMEX), 2012 Fourth International Workshop on* , vol., no., pp.109,114, 5-7 July 2012

Ilias Politis (M'05) received his BSc in Electronic Engineering from the Queen Marry College London in 2000, his MSc in Mobile and Personal Communications from King's College London in 2001 and his PhD in Multimedia Networking from University of Patras Greece in 2009.

Currently he is a Postdoctoral Researcher at the Wireless Telecommunications Lab of the Electrical and Computer Engineering Department at the University of Patras, Greece. He is also an Adjunct Lecturer at the Dept. of Telecommunication Systems and Networks, Technical Educational Institute of Mesolonghi, Greece. He has been involved in FP7-ICT-ROMEO and FP7-ICT-FUTON projects and several national funded research projects. His research interests include immersive multimedia, quality of experience modeling, 3D video and multimedia networking.

Dr. Politis is a member of FITCE and of the Technical Chamber of Greece.

Varuna De Silva received the B.Sc. Engineering degree (First class) in Electronic and Telecommunications Engineering from the University of Moratuwa, Sri Lanka in 2007. After breifly working in the telecommunication industry for around one year, in 2008, he was awarded the Overseas Research Scholarships Award (ORSAS) by the Higher Education Funding Council of England to pursue his PhD at the I-Lab Multimedia Communications Research Group, at the University of Surrey in the United Kingdom, which he completed in September 2011.

Currently he is a work package leader of a Large Scale Integrated Project "ROMEO" funded by the European Commission. His research interests include 3D and multiview video processing, compression, transmission and 3D QoE related issues and have more than 25 publications on these areas.

Dr. De Silva is the recipient of the 2009 IEEE "Chester W. Sall Award" for the 2nd best transaction paper in IEEE Transactions on Consumer Electronics. He was recently selected as the recipient of the Vice Chancellors Award for the "Postgraduate Research Student of the year 2011" of the Faculty of Engineering and Physical Sciences at the University of Surrey.

INDUSTRIAL COLUMN: SPECIAL ISSUE ON SOFTWARE DEFINED NETWORKS AND MULTIMEDIA APPLICATIONS

Enabling Network Programmability for Supporting Multimedia Applications

Guest Editor: Konstantinos Samdanis, NEC Europe Ltd, Germany

konstantinos.samdanis@neclab.eu

Software Defined Networks (SDN) is an emerging technology, which has the potentially to introduce tremendous changes on nowadays networks in terms of architecture design, deployment of new services and network management. SDN introduces decoupling of the control plain and data plane allowing a logical centralized management entity to control efficiently the underlying abstracted network. SDN permits network programmability, network virtualization, scalable and ease mobility of services enhancing flexibility as well as simplified network management via abstraction and automation. SDN opens a new business horizon for carriers, allowing applications to make the best use of, or otherwise to "guide", the network infrastructure in order to achieve true customer satisfaction.

This special issue of E-Letter provides the results of some of the recent developments on SDN focusing on multimedia applications. It is the great honor to have five outstanding articles from leading industry and academia research that report the experience gained by addressing the current challenges in employing the SDN paradigm for enhancing the performance of multimedia applications.

The first article titled, *"Applying ALTO to Multimedia Traffic Optimization",* by Carlos Haya, Victor Lopez, Diego Lopez and Juan Pedro Fernandez-Palacios from Telefónica I+D, Spain, elaborates mechanisms to enable collaboration between the application layer and network layer, which is particularly useful for cloud and CDN applications especially when combining IT resource from multiple locations. The proposed mechanism is based on the ALTO protocol, which provides applications with abstract map information of the underlying network infrastructure. A prototype implementation based-on ALTO and PCE validates the proposed approach and demonstrates the performance benefits for a CDN environment.

The following article, *"Future Edge ICT Networks"* contributed by Antonio Manzalini from Telecom Italia, Italy, provides an overview of a decentralized network architecture and service provision based-on SDN and Network function Virtualization (NfV) paradigms. It envisions network functions being provided at the edge of the network as software in a flexible and scalable manner following the users' demands and Operators' policies. The author argues that positioning resources, closer to the users brings several advantages including performance improvements, especially for multimedia applications, while through a use case study analysis he derives a number of high-level requirements and an architecture concept that could potentially flourish new ICT ecosystems and business models.

The third article *"Software Defined Networking for Heterogeneous Networks",* by Marc Mendonca, Bruno Astuto A. Nunesy, Katia Obraczka and Thierry Turletti from University of California, Santa Cruz, US and INRIA France, explores the deployment and challenges of the current SDN paradigm under a heterogeneous wireless network environment considering wired, cellular or Wi-Fi infrastructures combined with infrastructure-less ad-hoc networks. The authors provide a technical insight considering different user connectivity scenarios and service optimizations for multimedia applications, before analyzing specific requirements and limitations.

Andreas Kunz, Surendran Palanimuthu, JaeSeung Song and Hans-Jörg Kolbe from NEC Europe Ltd, Germany and Nash Technologies, Germany contributed the forth article titled *"MC-SDN: Introducing OpenFlow in the 3GPP Evolved Packet Core"*, which introduces the use of SDN for enhancing the connection efficiency of specific services including multimedia applications in mobile networks. The authors analyzed how the use of OpenFlow can enable a coordination of the underlying transport network with the 3GPP Evolved Packet Core (EPC) protocols demonstrating via a prototype named Mobile Core SDN (MC-SDN) its use considering a local breakout scenario towards a media server inside the mobile core network.

The last article titled *"Kandoo: A Framework for Efficient and Scalable Offloading of Control Applications"*, by Soheil Hassas Yegeneh and Yashar Ganjali from University of Toronto, Canada provides a novel control plain framework called Kandoo to deal with frequent network events, which can overload the control plane raising scalability concerns for the SDNs paradigm. Kandoo is a two layer hierarchical control plain consisting of multiple local controllers and a logical centralized root controller; it is OpenFlow compatible and has the capability to offload frequent applications that do not require a network-wide state into local controllers.

While this special issue cannot cover the entire current state of the art in the field of SDN and multimedia applications, we hope that the five invited articles provided a representative taste to the audience of some of the main activities in the field increasing their awareness. We would like to thank all the authors for their great contribution and the E-Letter Board for making this special issue possible.

Konstantinos Samdanis is a Senior Researcher and backhaul standardization specialist at NEC Europe Ltd, Germany. He is leading a research activity on LTE network management, and is the editor of the Energy Efficient Mobile Backhaul work item at the Broadband Forum (BBF). His BBF activities also include SDN and Cloud for broadband multi-service networks, while his research interests focus on network virtualization and SDN for mobile cellular networks. In the past, Konstantinos delivered several tutorials on green communications on IEEE conferences and has contributed at 3GPP System Architecture Group 5 (SA5) in the area Energy Saving Management. He obtained his PhD in Mobile Communications and an MSc in Electronic Research from King’s College London. Priory he studied Electrical and Electronic Engineering at Brunel University, with major in Telecommunication Systems.

Applying ALTO to Multimedia Traffic Optimization

Carlos Haya, Victor Lopez, Diego Lopez and Juan Pedro Fernandez-Palacios
Telefónica I+D, c/ Don Ramón de la Cruz, 82-84, Madrid 28006, Spain.
{b.haya, vlopez, diego, jpfpg}@tid.es

1. Introduction

Emerging cloud-based applications require a closer collaboration between application layer and network layer [1]. These applications include real-time data backup, virtual machine migration, server clustering, or load reorganization. In these applications, it is required to allocate IT resources in multiple locations and to interconnect remote data centers (DC). Similarly, when accessing (mostly multimedia) content through Content Delivery Networks (CDN) users can be assigned to different injection points using network costs. As proposed in [4], a protocol is required to exchange information between the application and the network layer. These requirements for cross-stratum optimization are currently addressed by means of two different approaches:

- **Coordinated**: IT and Network controllers exchange abstract information to improve resource utilization.
- **Centralized**: Control Plane disseminates IT information so the Network Controller becomes an Enhanced Network Controller [6].

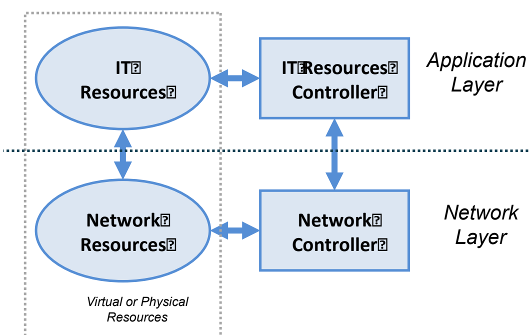


Figure 1 Cross-Stratum Optimization

2. ALTO Protocol

The ALTO protocol is defined as a service that provides network information to the application layer based on abstract maps of the underlying network. This information provides a simplified view, but it is useful to route the traffic of the application layers. ALTO Services enable Service Providers to share information about network locations and costs between them. The selection criteria to choose between two locations may depend of information such as maximum bandwidth, minimum cross-domain traffic, lower cost to the user, etc.

As the ALTO protocol exchanges information between network and application layer, ALTO defines Provider-defined Identifiers (PIDs) as the entrance points to the network. Each node in the application layer is known as an End Point (EP). Each EP is assigned to a PID, because PIDs are the entry point of the application in the network. Typically, multiple EPs are grouped in a single PID, as shown in Figure 2..

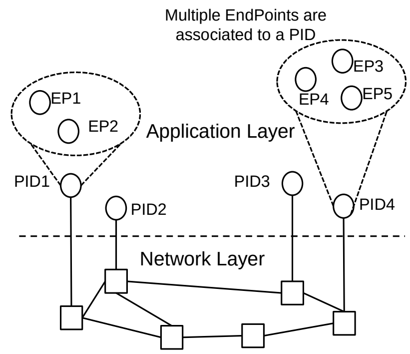


Figure 2 Application and network layer relationships in ALTO nomenclature

ALTO is defined as a client-server protocol [2]. ALTO clients use ALTO Service Discovery to find the closest ALTO Server. Once the ALTO client knows the location of its ALTO Server, it can send requests to the server. The ALTO server aggregates information from multiple systems to provide abstract, unified and useful network information. This information can come from routing protocols, external interfaces, dynamic network information, or provisioning policy. This abstracted information is the one sent to the ALTO client.

3. Application-Based Network Operations Architecture

Recently, IETF has proposed an architecture called "Application-Based Network Operations" (ABNO) (Figure 3) [3]. An ABNO controller is defined as an entity in charge of orchestrating the network in response to the requests from NMS, OSS, or applications or network changes. Thanks to this interface with the applications, CDNs can request bandwidth from the underlying network to configure the amount of capacity required at each moment. Thanks to the utilization of an active PCE in the ABNO architecture, the control plane in network elements can be triggered by the active PCE to modify the bandwidth of the connections, according to CDN requests.

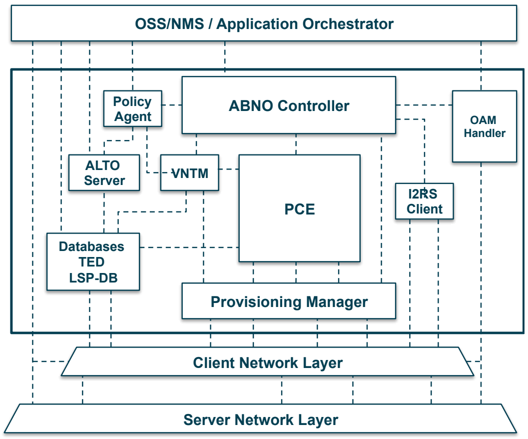


Figure 3 ABNO Architecture

4. Information Retrieval from the Network

The ALTO protocol does not define how the Network Map and Cost Map are updated. Different options are mentioned in [2].
At the time of this writing there is no description about which is the interface between databases and ALTO Server module in the ABNO architecture [3]. TED information cannot be used to give PID to PID connectivity to the ALTO server, because the TED has information about links of the network and their state. The ALTO server can retrieve information from the TED, but it would have to compute paths in this topology and translate such paths into ALTO costs. There are two solutions that can reduce the complexity of the ALTO server. The first option is to use LSP-DB to have this end-to-end information in the network. LSP-DB is populated with already established LSPs or possible connections that can be created in the network. This LSP-DB can give information about PID-to-PID connections, so ALTO does not have to perform path computation. However, LSP-DB does not contain information about LSP costs. This could be added when a new connection is created or when a possible path is included in the LSP-DB. The second option is to use PCE [3]. As a PCE is a computation entity in charge of computing paths in MPLS and GMPLS networks, the ALTO server could talk with the PCE to get dynamic network information. Thanks to the computation capability of the PCE, the ALTO server does not have to perform any routing process. Furthermore, path costs can be requested on demand from the PCE. According to RFC5541 [5], PCC can query the PCE for different objective functions. There are six objective functions defined in RFC5541, but other objective functions can be used in the PCE. Thanks to this functionality, the ALTO server can query possible paths between two PIDs receiving different paths depending on the objective function. The utilization of this option provides the ALTO server with different cost maps for different users. Moreover, since the PCE has real time information of the network, cost maps can be updated based on network conditions. Like in all previous options, the ALTO server has to translate this network information into ALTO costs.
Such cost map construction process can be done periodically or triggered by an event. Once the new cost map is updated, the ALTO Server can update this information so it is available for ALTO clients.

5. Prototype implementation and validation

The goal of our implementation is to demonstrate both the use of ALTO services to reduce the traffic congestion and the possibility of using the PCE as the source of information for the ALTO Cost Maps.
We have deployed a CDN scenario with three End Points, each of them assigned to three different PIDs. One End Point is for an end user, while there are two video servers in each of the other End Points. The network topology for the experiment is the one shown at Figure 4. The entry point for the user is PID 3 and the video capable endpoints are allocated in PID 1 and PID 2. This information is loaded in the TED DB used by the PCE, setting the unreserved bandwidth to 900 and 700 respectively.

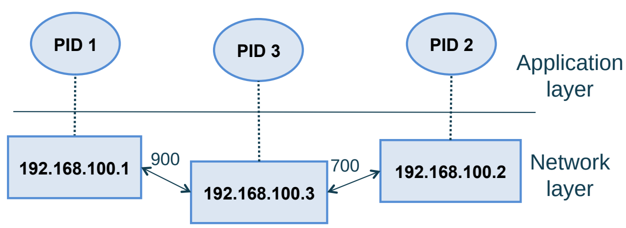


Figure 4 Prototype network topology

In the experiment, the end-user requests a video to the Content Provider, which detects that both servers can stream it to the user. As both endpoints can provide the video, Content Provider selects the endpoint based on the network information provided by the ALTO Server. In this demonstration, the ALTO Server is generating the Cost Map based on the unreserved bandwidth metric given by the PCE, as shown in Figure 5. ALTO Server sends periodically requests to the PCE between all PIDs in the network.

	PID 1	PID 2	PID 3
PID 1	1	1600	900
PID 2	1600	1	700
PID 3	900	700	1

Figure 5 ALTO Cost Map generated

One of the demo scenarios for this prototype contemplates the steps are presented below

0. The ALTO Server periodically queries the PCE to update it cost map.
1. The end-user requests a video from the Content Provider.
2. The Content Provider looks for the endpoints with enough IT resources, which are Video Server 1 (PID 1) and Video Server 2 (PID 2).
3. The Content Provider requests the cost map from the ALTO Server, and then selects the PID 1.
4. The content is streamed to the user from the Video Server 1.
5. Now we introduce congestion in the PCE TED DB sending an OSPF update message. Consequently, when the ALTO Server queries again the values of the cost map will change.
6. The user requests a video to the Content Provider. Although both Servers have enough IT resources, now the PID with the lowest value will be PID 2.
7. The video is streamed to the user from the selected server.

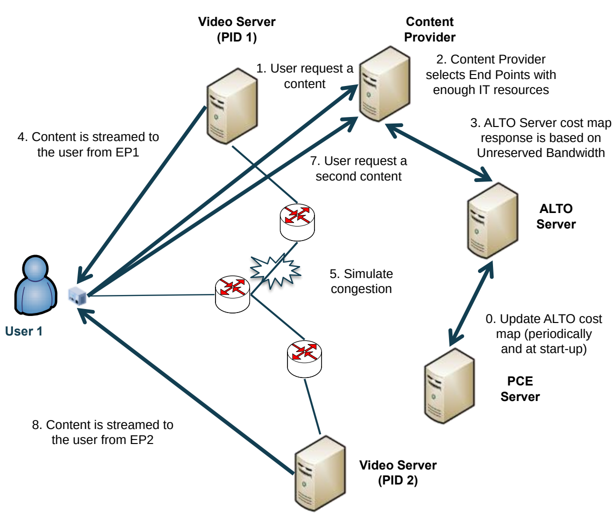


Figure 6 Prototype demo scenario

6. Conclusion

Cooperation between application and network layers is mandatory to have efficient resource utilization in cloud and CDN environments. The ALTO protocol is a perfect candidate to provide abstracted network information to the application layer. This work validates the utilization of ALTO and PCE for a CDN environment.

References

[1] L. M. Contreras, V. López, O. González de Dios, A. Tovar, F. Muñoz, A. Azañón, J.P. Fernández-Palacios, J. Folgueira: *Towards Cloud-Ready Transport Networks*, in IEEE Communications Magazine, September 2012, Vol. 50, Issue. 9, Pages. 48 - 55.

[2] R. Alimi, R. Penno, and Y. Yang. ALTO Protocol. IETF Internet-draft (work-in-progress), draft-ietf-alto-protocol-13, September 2013.

[3] D. King, A. Farrel, “A PCE-based Architecture for Application-based Network Operations”, IETF draft http://tools.ietf.org/html/draft-farrkingel-pce-abno-architecture-03

[4] Dhody, D.; Young Lee; Hui Yang, "Cross Stratum Optimization (CSO) Enabled PCE Architecture," IEEE International Symposium on Parallel and Distributed Processing with Applications (ISPA), pp.500-505, 2012.

[5] JL. Le Roux, JP. Vasseur and Y. Lee, “Encoding of Objective Functions in the Path Computation Element Communication Protocol (PCEP)”, RFC5541, June 2009.

[6] G. Landi et al, "A Network Control Plane architecture for on-demand co-provisioning of optical network and IT services", Future Network Mobile Summit 2012.

Carlos Haya is about to finish his B.E. in Computing Science and Mathematics from Universidad Autónoma de Madrid. He is doing an internship in Telefonica I+D since 2012 working on the design of new protocols used in metro core transport networks.

Víctor López received his M.Sc. and Ph.D. at the universities of Alcalá de Henares (2005) and Autónoma de Madrid (2009), respectively. In 2011, he joined Telefonica I+D as Technology specialist. His research interests include the integration of Internet services over optical networks.

Diego R. López is responsible for Technology Exploration activities within the GCTO Unit, Telefónica I+D. He earned his MSc and PhD at the universities of Granada and Seville, respectively. After a period in the academic sector, he joined Telefónica I+D in 2011, where currently works in Software Defined Networking, network abstractions, and identity and security issues.

Juan Pedro Fernández-Palacios Giménez received the MS in Telecommunications Engineering from Polytechnic University of Valencia in 2000. In Sept. of 2000 he joined TIDSA where he is currently leading the Metro and Core Transport Networks department. Currently he is coordinating the FP7 project IDEALIST as well as the standardization activities within CAON cluster, where he is acting as co-chair.

Future Edge ICT Networks

Antonio Manzalini
Telecom Italia – Strategy-Future Centre-Innovative Architectures
antonio.manzalini@telecomitalia.it

1. Introduction

This paper argues that the continuous advances in standard hardware technologies and open source software are creating the conditions for exploiting highly flexible and cost effective network architectures, both in the edge (i.e., in the aggregation and access) and in the core parts of the networks.

Initially, Internet architecture was based on the well-known end-to-end principle: packets were forwarded through the network, form source to destination, mainly unmodified. In current Internet, and communication networks, this model is not valid anymore: before reaching the destination traffic crosses several middle-boxes [1] such as Wide Area Network (WAN) optimizers, Network Address Translation (NAT), performance-enhancing-proxies, intrusion detection and prevention systems, any sort of firewalls, other application-specific gateways, etc.

There are quite a lot of middle-boxes deployed in current networks: not only these nodes are contributing to the so-called “network ossification”, but also they represent a significant part of the network capital and operational expenses (e.g., due to the management effort that they require). Basically, a middle-box is a stateful system supporting a narrow specialized network functions (e.g., layer 4 to 7) and it is based on purpose-built hardware (typically closed and expensive). Removing, or even reducing, the number of middle-boxes deployed in current networks would determine several advantages, such as cost savings and increased flexibility.

Technologies advances, even expressed through emerging paradigms such as Software Defined Network (SDN) [2] and Network function Virtualization (NfV) [3], are likely to offer the opportunity to develop, fully in software, middle-boxes’ network functions, and to allocate them dynamically according to Users’ needs and Operators’ policies. Some of these network functions are already developed in open source software and, in principle, could be executed on Virtual Machines (VM) running standard hardware servers (as the ones available today in some Data Centers for provisioning Cloud Computing services). For example there are some open source implementations of firewalls [4], load balancers [5], proxies and caches [6], monitoring and measurement [7], intrusion detection [8, 9], and ubiquitous NAT. Today there are still some concerns about the performance of network functions purely developed in software and running on standard hardware. Nevertheless said performance can be improved by utilizing hardware accelerators or by running multiple instances of the software (utilizing multiple processing elements) This is for further investigations.

Nevertheless, this paper goes beyond this vision and it argues that it will be even more effective and advantageous to execute these network functions mostly, if not completely, at the edge of the network, where an incredible amount of processing and storage resources are progressively accumulating. Actually the edge of the network is where “intelligence” has already started migrating, and it is where innovation is more urgently needed to overcome the “ossification” by improving networks and services infrastructures flexibility.

As a matter of fact, current “ossification” is creating a lot of limitations for the development and deployment of new network functionality, services, protocols, security designs, management policies and approaches, and other element that are essential to cope with the increasingly challenges of future ICT networks and services.

It is argues that the exploitation of SDN, NfV principles at the edge of current networks will transform this area into a fertile ground for the flourishing of new ICT ecosystems and business models.

Figure 1 is providing a scenario where the core part of the network is becoming (almost) stateless, whilst the execution of the stateful network functions is moved to the edge networks and to the Data Centers. Network functions are executed by processing resources (e.g., standard hardware) crossed by the traffic.

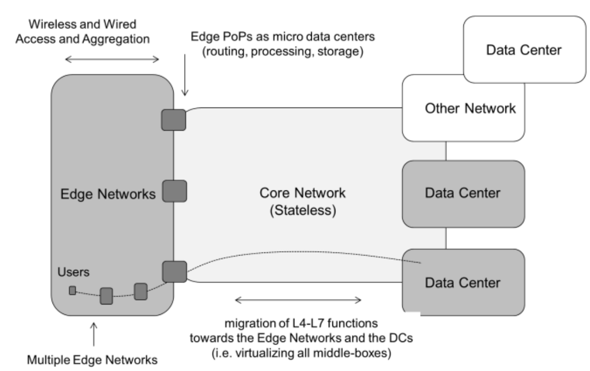


Figure 1: Vision of a stateless core network interconnecting stateful edges full of processing and storage resources

The outline of this paper is the following. In Section II provides some examples of use cases. Sections III focuses on the architectural aspects. Section IV elaborates about the techno economic impact and the business aspects. Finally, Section V draws some conclusions and discusses future work.

2. Examples of use -cases

This section describes three simple use-cases: all of them have in common the vision of the development and exploitation of a distributed network processing architecture, deployable at the edge of current networks. In particular, it is argued that the sheer number of nodes, devices and processing, storage systems being deployed at the edge, up to Users' premises, are offering an enormous processing and storage power. Using these resources, closer to the Users, to execute network functions and services will bring several advantages, not only in terms of flexibility but also in performance improvements: for example, the equation "Throughput = Window Size / RTT" is showing that the execution of services closer to the Users will allow improving the QoS (i.e., reducing RTT).

The requirements of these three use cases (and other ones under investigations) are offering interesting technical challenges which have to be solved to transform this vision into reality. Some of these requirements and challenges are listed in the following sections.

Personal Data and Services following Users

This use case concerns to the possibility that Users' personal data (e.g., stored in edge micro Data Centers) and services (e.g., executed by local edge resources) are following the Users when they are moving from one network attachment point to another one, even when they are moving across different edge networks (and as such the Core network has to be crossed).

In other words, it should be feasible to move data and VM executing services seamlessly with no impact on QoE perceived by the Users. It should be possible, also, for security or other policies to follow logically specific network applications (e.g., running on VMs).

It should be possible even to federate data and services associated to Users in order to build distributed virtual data centers at the edge (this is an ideal service, for example, for Universities, Enterprises, etc.), provided at costs which are a small fraction of traditional cloud computing services.

Harnessing idle resources at the Edge

Several start-ups ([10] represents a recent example) are starting offering storage and disaster resilience services using decentralized and distributed virtual resources, shared by Users. The concept of harnessing idle resources could be extended also to the processing idle power distributed at the edge, up to the Customers' premised. This would require the capability of properly orchestrating said local idle storage and processing resources when executing and provisioning network functions and other services. Examples of provisioned services will be CDN-like services, content sharing, aggregation, transformation, data collection, etc.

End-to-End Services across Edge Networks

A Service Provider may want to provide end-to-end services to Users who are attached to edge networks belonging or operated by different Network or Infrastructure Providers. This means that it should be possible even hooking and orchestrating heterogeneous network resources and functions for the provisioning of end-to-end services.

3. Overall Requirements and Architecture

Analysis of above use cases has brought to the main assumption that network functions and services should mainly be executed in ensembles of VMs, allocated and moved across a scalable distributed platform. This platform could be deployed at the edge (e.g., by designing a standard node architecture that could scale from Customer Premise Equipment up to Edge PoPs nodes). It might be advantageous, depending on the context and the executed functions, to complement the use of edge power also with virtual resources deployed in the Data Centers: in that cases, as mentioned, it should be necessary to put in place proper orchestration capabilities.

The edge distributed platform should be characterized by high flexibility, performance and self- adaptation at run-time (e.g. dynamic flocking of resources according to needs), harnessing and combining all unused resources (e.g. computing and storage power at end Users' home and in the edge micro data centers). It should be possible programming, allocating and moving a variety of virtual architectures (spanning across diverse edge networks or even across today DCs) on-demand, based on Users' requests, governance and biz requirements overcoming the current ossified networks structures. End Users should have access to a certain number of abstraction for programming, setting-up and tearing down, migrate and optimize their network functions and services (e.g. local traffic engineering, failure handling policy, local topology optimization, etc.) according their need and service level agreement.

It will be necessary also to have efficient systems and methods capable of creating optimal overlay paths which are changing continuously with network conditions: in turn this will mean solving constrained based optimization problems with local-to-global tradeoff.

Figure 2 is showing a simple example of the architectural concept, which could be prototyped and demonstrated by customizing open source software and standard hardware (e.g., x86).

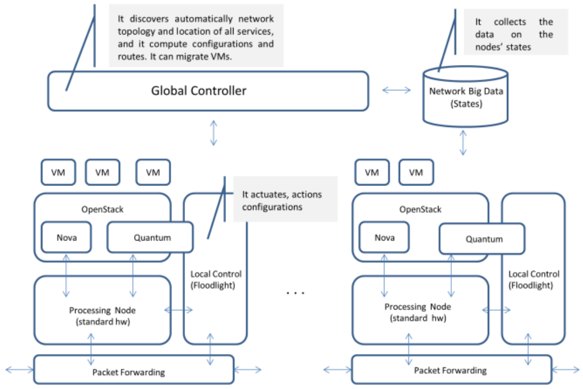


Figure 2: Example of prototyping architecture (using open source software and standard hardware, e.g. x86)

In particular, the figure 2 is showing how developing the prototypes of the Edge PoPs, which are seen as sort of micro Data Centers at the border between the core and the edge networks (e.g., a future evolution of today edge routers). The Global Controller is orchestrating the local controllers by using historic and real-time network data: these data have to be elaborated to create sort of maps needed to orchestrate virtualized functions and services in a reliable and efficient way. In other words the Global Controller should be capable of handling the so-called network states, which are stored in the Network Big Data (NBD). This NBD includes historic and real-time data collected over time from the network nodes and refreshed continuously.

This architectural concept should scale down up to the CPE, which again could be seen as a smaller Data Center. Hierarchies of global-local controllers should be able to ensure the global vs local orchestration of virtualized functions and services.

4. Socio-economic and Business Impacts

This paper argues that in the near future the edge will look like a distributed network processing architecture, deeply integrating processing, storage and networking resources. This will impact dramatically future networks evolution (even for the core segment), not only allowing cost-savings and improvements in the QoS, but even creating new business opportunities. Actually, the vision of this paper is following the thinking that technology and business developments will be more and more strictly intertwined in the future. A certain technology will be adopted not only if it is advantageous (reducing costs) and trusted but also if it will be able to enable desired business ecosystems (with the related foreseen business models); on the other hand, newly designed potential ecosystems will look for enabling solutions and technologies capable to bring them into reality. It is argued that incentives, cooperation and competition at the edge will boost the long-term value of networks: like in ecosystems, where evolution select the winning species, winning services will succeed, grow, and promote further investments, while losing ideas will fade away.

In order to capture all these implications, it will be necessary analyzing, modeling and simulating diverse cooperation-competition business strategies for Telco-ICT ecosystems where there are several interacting "species": e.g., not only N.Os and OTTs, but also other Players, e.g., Municipalities, Consumers electronics Providers, SME, etc. This is for further study.

5. Conclusions

This paper has proposed a vision arguing that the recent advances in standard hardware technologies and open source software are creating the conditions for exploiting highly innovative network architectures, both in the edge and in the core parts of the networks. The sheer number of nodes, devices and systems being deployed at the edge, up to Users' premises, are offering tomorrow an enormous processing and storage power. Using these resources, closer to the Users, to execute network functions and services will bring several advantages, both in term of improved performance and cost savings (e.g., determined by the removal of middle-boxes). Moreover, it is argued that the exploitation of these principles at the edge of current networks will transform this area into a fertile ground for the flouring of new ICT ecosystems and business models.

The paper has described three use cases whose study has allowed the derivation of a number of high level requirements. An example of architectural concept has been proposed and will be prototyped by using open source software and standard hardware. Future steps includes also modeling and simulating diverse cooperation-competition business strategies for Telco-ICT ecosystems at the edge.

References

[1] IETF RFC3234 http://tools.ietf.org/html/rfc3234;
[2] White paper on "Software-Defined Networking: The New Norm for Networks" https://www.opennetworking.org/;
[3] White paper on "Network Functions Virtualisation"; http://portal.etsi.org/NFV/NFV_White_Paper.pdf NfV;
[4] SmoothWall Express: Express Open Source Firewall Project. http://www.smoothwall.org/;
[5] HAProxy - The Reliable, High Performance TCP/HTTP Load Balancer;
[6] Squid : Optimising Web Delivery. http://www.squid-cache.org/;
[7] Ntop: Open Source Network Trac Monitoring. http://www.ntop.org/;
[8] Bro Intrusion Detection System http://bro-ids.org/;
[9] SNORT http://www.snort.org/;

[10] http://www.symform.com/.

Antonio Manzalini received a M. Sc. Degree in Electronic Engineering from Politecnico of Turin. In 1990 he joined CSELT, which then became Telecom Italia Lab. He started his activities on research and development of technologies and architectures for future optical networks. In this RTD avenue, he was actively involved in leading positions in several Project funded by Eurescom and European Commission. He chaired two ITU-T Questions on transport networks. He published in about 80 papers and several book's chapters. He was member of technical and program committee of several IEEE Conferences. He is currently leading the activity "Smart Networks at the Edge" funded by the European Institute of Technology (EIT). He is part of the Strategy Department of Telecom Italia (Future Centre) addressing mainly RTD activities concerning future Telco-ICT networks and services technologies (e.g., Software Defined Networks, Network Function Virtualization, Autonomic/Cognitive, self-* networks)..

Software Defined Networking for Heterogeneous Networks

Marc Mendonça[1], Bruno Astuto A. Nunes[2], Katia Obraczka[1] and Thierry Turletti[2]
[1]University of California, Santa Cruz, USA, Email: {msm, katia}@soe.ucsc.edu
[2]INRIA, France, Email: {bruno.astuto-arouche-nunes, thierry.turletti}@inria.fr

1. Introduction

Software–Defined Networking (SDN) has been proposed as a way to programmatically control networks, making it easier to deploy new applications and services, as well as tune network policy and performance. The key idea behind SDN is to decouple the data- from the control plane by: (1) removing control decisions from the forwarding hardware, (2) allowing the forwarding hardware to be "programmable" via an open interface, and (3) having a separate entity called "controller" define by software the behavior of the network formed by the forwarding infrastructure, thereby creating a "software-defined network". OpenFlow [3] has been proposed as the de-facto standard protocol used for communication between the controller and "programmable" switches. The latter forward data according to a "flow table" containing an entry for each flow along with an action (or "rule") to be invoked when forwarding packets belonging to that flow. A switch's flow table is built based on the rules sent to the switch by the controller specifying how to forward data for a given flow. SDN techniques to–date largely target infrastructure–based networks, for example, those found in data centers.

Motivated by a vision of a fully connected world, we explore how SDN can be utilized to support both infrastructure- based and infrastructure–less networks. We also discuss the research challenges involved in augmenting the current SDN model to operate in heterogeneous networked environments. While previous work has examined the use of SDN in wireless environments, their scope has primarily focused on infrastructure-based wireless deployments (e.g., WiMAX, Wi- Fi access points). For example, the idea of a flexible wireless infrastructure supported by OpenFlow was introduced in [7], [6]. The use of OpenFlow in wireless mesh environments has been explored in [1], [2].

However, to our knowledge, no efforts have investigated the challenges and benefits offered by extending the SDN paradigm to heterogeneous networked environments. This paper aims at bridging this gap by exploring the use of software–defined networking in such heterogeneous environments. In Section 2, we examine example scenarios that would benefit from enabling SDN in heterogeneous networks. Section 3 discusses how the current SDN paradigm could be extended to operate in heterogeneous networked environments and the research challenges that will result.

2. User-Assisted Connectivity

We consider heterogeneous networked scenarios that include mobile end-user devices with limited or intermittent connectivity to the network infrastructure, i.e., wired-, cellular- or WiFi infrastructure, but are able to form ad-hoc networks with other nearby units. Additionally, some of the mobile units have multiple network interfaces (e.g., wired/802.11 or 802.11/cellular). In such environment, users connecting through their mobile devices may want to communicate and/or retrieve or store content in the "cloud". For this use case, we examine two scenarios, one in which SDN is not enabled (called the "traditional" scenario) and the other in which SDN is enabled. We identify and discuss the benefits of the SDN-enabled scenario to both the users and network providers. In our discussion, we assume that the mobile units have agreed to some form of external control insofar as routing decisions are concerned. This of course, raises several issues, which we discuss in Section III.

Let us consider that in the scenario depicted in Figure 1, a user "Alice" wishes to connect to the Internet to access the Web. Unfortunately, she is unable to connect to the infrastructure and joins an ad hoc network instead. Suppose that another user, "Bob", is connected to both the ad hoc network and an infrastructure-based wireless access network. In our SDN-based architecture, a device such as Bob's is considered a "Gateway" (GW) device.

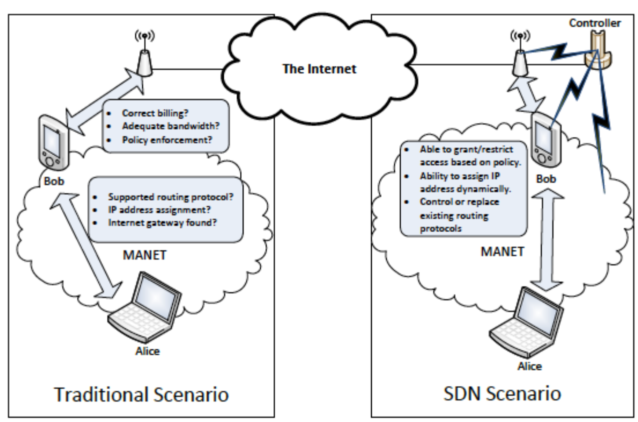


Figure 1. Heterogeneous SDN use case scenario

Traditional vs. SDN Enabled Scenarios

Traditional Scenario: Even if we assume the ad hoc network learns how to route to Bob as a gateway, and Bob allows his device to be used as a NAT router by strangers, the mobile data service provider is not aware of the existence of Alice. Bob's connection is not assigned additional bandwidth, possibly harming performance; the Internet Service Provider is not able to differentiate Alice from Bob and cannot apply any QoS rules, access restriction or any sort of polices over Alice without also applying it over Bob. Furthermore, Bob will be held responsible for Alice's traffic by the service provider for any possible data overages or illegal activity.

SDN Enabled Scenario: The service provider is made aware when Alice joins the ad hoc network. Therefore, it may decide to offer service to Alice via Bob and provision Bob's connection accordingly. The service provider may decide to sell Alice a temporary connection plan on the spot, or Alice may have an existing contract on another device. Available resources, past user behavior, or any number of factors can be used on deciding whether to offer service to Alice. The service provider is thus able to maintain control of its network policy while being granted an opportunity for additional business. Alice is able to seamlessly connect to the Internet using a service plan. For his part, Bob may be offered incentives by the service provider, while avoiding performance loss or being held liable for Alice's traffic.

Multiple Gateways

An extension to the base case previously discussed is a scenario with multiple gateways. For example, shortly after Alice joins, a user "Charlie" with access to wired infrastructure also connects to the ad hoc network. In the traditional scenario, traffic is routed solely based on how the MANET protocol handles multiple gateways. In the SDN scenario, the network capacity can be managed based on the policies of the service providers and the characteristics of the available resources. For example, Alice's traffic may continue to flow through the slower mobile data network instead of the wired network, because she only has a service plan with the mobile data provider. Alternatively, the mobile data provider may have an agreement with the wired network such that even Bob's traffic will flow through Charlie to either increase Bob's performance or reduce the load on the mobile data network.

Service Optimizations

In another possible situation, a group of users in the ad hoc network may be viewing the same content simultaneously (e.g., live streaming of a sport event). Using the base case from above, Bob is the link to the Internet from which the content originates. In the traditional scenario, any optimizations such as caches or CDN are performed either in the provider network or in the cloud; the result may be that Bob's link to the provider gets saturated with duplicate content. SDN enables routing policies to evolve and promotes the creation of new services; for example, it may be possible to reduce the strain on the limited infrastructure connectivity by caching and retrieving common content locally, or by creating multicast streams on-the-fly for live content.

3. Requirements And Challenges

Enabling SDN in heterogeneous networked environments raises several requirements and research challenges. We dis- cuss some of them below.

a) End-user device limitations: Unlike infrastructure-based networks, in infrastructure-less networked environments, such as multi-hop wireless networks, or MANETs, there is no real distinction between network elements (i.e., switches, routers) and end-user devices. The latter perform core network functions like routing and forwarding, as well as source and sink application traffic. Therefore, end devices should be able to communicate with controllers and understand how to handle traffic forwarding rules. But, because in these types of networks, devices are often limited in terms of power, processing, communication, and storage capabilities, protocol overhead should be minimized.

b) Gateway device incentives: From the use case scenario discussed in Section 2, it is clear that incentives are necessary to ensure collaboration between nodes in order for gateway devices to agree to forward traffic from other nodes. These new incentive schemes should be able to use the revenue collected through the new offered service and the bandwidth shared by the GW device to reward to contributing GW devices.

c) Resource discovery: Infrastructure-less networks tend to be heterogeneous in terms of the devices they interconnect and the links use to interconnect them. Therefore, a variety of factors should be considered when choosing an end device as gateway ranging from battery lifetime, network connectivity, and trust, to name a few. Clearly, a controller that learns this information would be better equipped to make decisions.

d) Control plane: Several of the independently-operated devices participating in an infrastructure-less network may not be SDN-capable and thus unable to communicate with a SDN controller. However, such devices could receive control information through some other protocol, for example, routing. This calls for a “hybrid” control plane that combines different ways to convey control information to non SDN-capable devices. In the use case example discussed in Section 2, Alice should still be able to connect to the Internet even if Bob is the only SDN-enabled device in the infrastructure- less network. This could be done through a standard MANET routing protocol such as OLSR.

e) Security: Though SDN can be used to improve network control and traffic policy enforcement, keeping the network secure and guaranteeing confidentiality, integrity, and availability is quite challenging, especially in the types of heterogeneous networks we are considering. In particular, in an infrastructure-less network with independently owned end devices also acting as forwarding nodes, it may be difficult to establish trust and ensure a secure channel end-to-end. Since a wide variety of threats ranging from jamming at the physical layer to worms at the application layer must be considered, solutions will likely need to take a multi-layered approach.

Although security in MANETs has been explored in the MANET community [5], security challenges are exacerbated by the heterogeneous SDN architecture, which needs to employ a distributed control plane using independently run controllers.

While a switch in an infrastructure-based network may easily be configured to securely connect to a pre-determined con- troller, devices and controllers in infrastructure-less networks must discover each other without prior knowledge of the network topology. Furthermore, it is not enough that control messages successfully and securely reach their destination; both endpoints must be able to trust each other, i.e., before accepting control, forwarding nodes need to be able to trust that the discovered controller is not malicious. Likewise, the controller must be able to trust that forwarding nodes that have accepted control are following instructions. For this trust to exist, mechanisms must be in place to ensure the legitimacy of nodes and controllers, the authenticity of the control traffic, and to verify that devices act as expected in response to instructions.

f) Distributed Control Plane: Heterogeneous networks may span multiple domains of control. As illustrated by our use case, an ad hoc network may have gateways connecting to two different infrastructure networks. While previous work [4] considered using a transparent proxy to allow multiple controllers, devices in an infrastructure-less network must be able to discover and connect to multiple controllers on their own as they may not be able to rely on an outside proxy.

g) Flexible rules and actions: Current specifications that target infrastructure-based networks often limit the types of rules that can be performed on flows, often due to performance or hardware constraints. Although the latest OpenFlow 1.3 specification already supports user-specified flow match fields, compliant switches do not have to support this feature and are only required to handle a small, pre-defined set of rules. Because of the inherent heterogeneity and limitations of wireless infrastructure-less networks, supporting flexible rules (e.g. flow matching on custom headers) is critical to enable SDN in these kinds of networks.

4. Conclusions

In this paper, motivated by the vision that future internets will comprise infrastructure–based and infrastructure– less networks, we explore the use of the Software–Defined Networking (SDN) paradigm in these so–called “heterogeneous” networked environments. To make the case for SDN in heterogeneous networks, we examine an application scenario in which SDN is a key enabling technology. We also identify the additional requirements imposed by the SDN paradigm and discuss the research challenges they raised.

References

[1] A. Coyle and H. Nguyen. A frequency control algorithm for a mobile adhoc network. In Military Communications and Information Systems Conference (MilCIS), Canberra, Australia, November 2010.

[2] P. Dely, A. Kassler, and N. Bayer. Openflow for wireless mesh networks. In Proceedings of 20th International Conference on Computer Commu- nications and Networks (ICCCN), pages 1–6. IEEE, 2011.

[3] N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, S. Shenker, and J. Turner. Openflow: enabling innovation in campus networks. ACM SIGCOMM Computer Communication Review, 38(2):69–74, 2008.

[4] R. Sherwood, M. Chan, A. Covington, G. Gibb, M. Flajslik, N. Handigol , T. Huang, P. Kazemian, M. Kobayashi, J. Naous, et al. Carving research slices out of your production networks with Openflow. ACM SIGCOMM Computer Communication Review, 40(1):129–130, 2010.

[5] H. Yang, H. Luo, F. Ye, S. Lu, and L. Zhang. Security in mobile ad hoc networks: challenges and solutions. Wireless Communications, IEEE, 11(1):38–47, 2004.

[6] K. Yap, M. Kobayashi, R. Sherwood, T. Huang, M. Chan, N. Handigol, and N. McKeown. Openroads: Empowering research in mobile net- works. ACM SIGCOMM Computer Communication Review, 40(1):125– 126, 2010.

[7] K. Yap, R. Sherwood, M. Kobayashi, T. Huang, M. Chan, N. Handigol, N. McKeown, and G. Parulkar. Blueprint for introducing innovation into wireless mobile networks. In Proceedings of the second ACM SIGCOMM workshop on Virtualized infrastructure systems and architectures, pages 25–32. ACM, 2010.

Marc Mendonca was a visiting student researcher at the INRIA Plane`te project in Sophia Antipolis. He recently received his M.S. in Computer Engi- neering from the University of California, Santa Cruz and previously completed his B.S. in Computer Engineering at the University of California, Santa Barbara.

Bruno Astuto A. Nunes is a Pos-doc at INRIA Sophia Antipolis, France. He received his B.Sc. in Electronic Engineering at the Federal University of Rio de Janeiro (UFRJ), Brazil, where he also com-pleted his M.Sc. degree in Computer Science. He received his PhD degree in Computer Engineering from UC Santa Cruz, USA.

Katia Obraczka is Professor of Computer Engi- neering at UC Santa Cruz. Before joining UCSC, she held a research scientist position at USC's Infor- mation Sciences Institute and a joint appointment at USC's Computer Science Department. Her research interests span the areas of computer networks, dis- tributed systems, and Internet information systems. She is the director of the Internetwork Research Group (i-NRG) at UCSC and has been a PI and a co-PI in a number of projects sponsored by govern- ment agencies (NSF, DARPA, NASA, ARO, DoE, AFOSR) as well as industry. Prof. Obraczka has edited one book, wrote a number of book chapters, and published over 200 technical papers in journals and conferences. She is co-recipient of the Best Paper Award at IFIP Networking 2007 for the paper ”On-Demand Routing in Disrupted Environments”. She is also co-recipient of one of three Best Paper Awards at IEEE MASS 2007 for the paper ”DYNAMMA: A DYNAmic Multi-channel Medium Access Framework for Wireless Ad Hoc Networks”. She has served on the Editorial Board of the Elsevier Ad Hoc Networks Journal. She has been involved with the organization of several technical conferences in her area including IEEE Global Internet, IEEE SECON, IEEE MASS, ACM Mobicom, and ACM Mobihoc, IEEE NetSciCom. She received the USC ISI's Meritorious Service Award in 1999. She is a senior member of the IEEE.

Thierry Turletti received the M.S. (1990) and the Ph.D. (1995) degrees in computer science from the University of Nice - Sophia Antipolis, France. He has done his PhD studies in the RODEO group at INRIA where he designed the Inria Videoconferenc- ing System (IVS). During the year 1995-96, he was a postdoctoral fellow in the Telemedia, Networks and Systems group at LCS, MIT and worked in the area of Software Defined Radio (SDR). He is cur- rently a senior research scientist at the Diana team at INRIA Sophia Antipolis. His current research interests include information centric network architectures, trustable network evaluation platforms and wireless networking. Dr. Turletti has been serving on the Editorial Boards of the following journals: Wireless Communications and Mobile Computing (2001-2010), Wireless Networks (since 2005) and Advance on Multimedia (since 2007).

MC-SDN: Introducing OpenFlow in the 3GPP Evolved Packet Core

Andreas Kunz[1], Surendran Palanimuthu[2], JaeSeung Song[1], Hans-Jörg Kolbe[1]
NEC Laboratories Europe, Heidelberg, Germany[1]
Nash Technologies, Nürnberg, Germany[2]
{kunz, song, kolbe}@neclab.eu[1], mprsuren@gmail.com[2]

1. Introduction

Software Defined Networking (SDN) recently gained a lot of momentum for flexible networking in both, the control and the user plane domain. Technologies, such as OpenFlow [1], provide full network traffic control on a per-flow basis. Having been mainly applied to data center networks, the time is right to look at its application to carrier networks.

The 3rd Generation Partnership Project (3GPP) specifies the most widely deployed mobile network architecture for the radio access and the network core. The latest developed Evolved Packet Core (EPC) in 3GPP introduced all-IP networks in the mobile domain. The underlying transport network is not considered by 3GPP and the EPC protocols just require IP connectivity between the nodes. While this allows having modular and independent network designs, it is not possible to benefit from the synergies that result from the optimization of both network layers to e.g., improve energy efficiency or optimize resource control and traffic routing.

A new approach of separating the control plane from the user plane in transport networks was developed by the latest research in the future internet area. The OpenFlow protocol was proposed as an innovation for campus networks [2]. The Open Networking Foundation (ONF) [1] is specifying the protocol and the related architectures for OpenFlow. The main concept is that a logically centralized OpenFlow Controller has full view on the network and controls all nodes in the user plane. The OpenFlow Switches behave like programmable switches and are routing the user plane traffic based on the instructions from the OpenFlow Controller.

While 3GPP already has defined the Policy and Charging Control Function (PCRF) [3] that can centrally control network and accounting/charging policies, the link to the transport network is missing. PCRF can control in user plane the QoS policies, but a mapping to the transport network needs to be done inside each node and the routing is currently out of scope. This opens the door for proprietary solutions to enable at least a basic combination of both principles, but best to our knowledge such solutions only exist for adding a new node to the network and configuring the network in a static way.

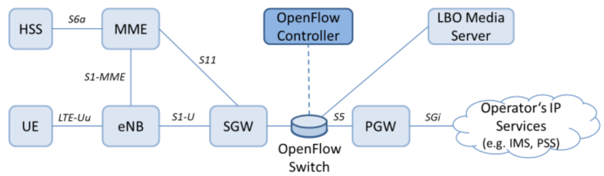


Figure 1: EPC Architecture with OpenFlow controlling the transport nodes (OpenFlow Switches)

This paper shows how OpenFlow can be also used in mobile networks in a first step using a non-disruptive approach. It integrates a transport network using a dynamic control plane with the EPC network architecture while minimizing the impact on the 3GPP architecture and protocols as far as possible. As first proof of concept, we describe a prototype on offloading data traffic within the EPC.

2. Gluing 3GPP and OpenFlow

In this section, we describe how a mobile network imposes on achieving a tight coupling with the transport network, i.e., making it aware of the mobile network tunnels' states. Once this is achieved, any routing functionality in the transport network is not needed any more as it can be completely taken over by the OpenFlow Controller.

Although there exist many different ways integrating the OpenFlow into the 3GPP system, in this paper, we are only considering the *S5* reference point between the Serving GateWay (SGW) and the Packet GateWay (PGW). Other scenarios, such as integrating OpenFlow switches on the *S1-U* reference point and connecting MME to the OpenFlow Controller, are not considered because of complexities and mobility management issues.

Figure 1 shows how OpenFlow Switches and the Controller can be integrated into the EPC on the *S5* reference point. The OpenFlow Switches provide the transport network in between the 3GPP nodes. The OpenFlow Controller is connected to the OpenFlow Switches. It is possible to easily integrate a dynamic policy routing between SGW and PGW, e.g., to route to a specific Media Server and breakout already in the EPC without going through the PGW. In this architecture, the OpenFlow Controller can learn all necessary information from the bearer setup procedure.

Figure 2 shows the normal dedicated bearer setup procedure including the interaction with the OpenFlow Controller. Once the PGW gets triggered to setup a bearer, it sends a Create Bearer Request message to the SGW. This message is intercepted by the OpenFlow Switch and recognized as a new GTP-C [5] signaling message and send to the OpenFlow Controller for further action (steps 1 and 2). The OpenFlow Controller checks the packet and installs the default route for GTP-C packets towards the SGW and stores the PGW's TEID for later correlation (step 3). The OpenFlow Controller then sends a control message back to the OpenFlow Switch to install the default route to the SGW for this GTP-C flow (step 4). The OpenFlow Switch then forwards the packet accordingly to the SGW, and the intermediate nodes (i.e., SGW, MME and eNB) then handle the packet based on the regular procedure [3] for the bearer request message (step 5). The OpenFlow Switch that receives the incoming packet will also remove the GTP header from the packet and only route the IP packet based on the rule for the flow. The OpenFlow Switch at the destination will add back the GTP header as instructed by the OpenFlow Controller. This requires the support of the GTP protocol in OpenFlow. A response message to the Create Bearer Request message towards the PGW is then intercepted by the OpenFlow Switch and sent to the OpenFlow Controller (steps 5 and 6).

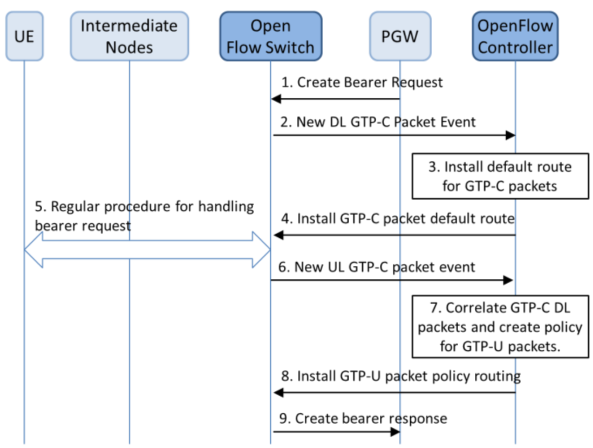


Figure 2: OpenFlow policy installation at bearer setup

The OpenFlow Controller can now correlate the message with the previous sent downlink GTP-C message based on the PGW TEID. It stores also the GTP-C SGW TEID for this flow and the corresponding GTP-U [6] TEIDs of SGW and PGW. The OpenFlow Controller constructs a policy for the GTP-U packets and sends a control message to the OpenFlow Switch to install the policy. The OpenFlow Switch installs the policy and routes the Create Bearer Response to the PGW.

With the introduced policy installation procedure at bearer setup, the OpenFlow Controller can instruct the Switch to do a local breakout to a media server already in the EPC. The dedicated bearer setup could be triggered, e.g., when the UE requested via the default bearer an IMS video service. Streaming services over Hypertext Transfer Protocol (HTTP) are very popular and got even enhanced to be adaptive to current network condition (DASH).

For example, let us consider that the OpenFlow Switch is instructed to forward packets that are destined to the address of web server A or asking specific contents to a local breakout in the EPC because the operator has a local content cache for these specific contents and wants to keep the traffic as local as possible. If a UE requests a certain video specified in the given contents from a web server, the OpenFlow Switch intercepts the packet from the SGW and checks the installed policy accordingly. In this case the policy foresees that packets towards this web server are rerouted and a local breakout in the EPC is performed towards a local media server. The local media server can directly process the packet since OpenFlow removed the GTP header for routing the packet in the OpenFlow network and will not add it back when routing towards the media server. A solution on how this can be implemented is described in the following section.

3. Prototype implementation

As a proof of concept we implemented a system called Mobile Core SDN (MC-SDN) for the local breakout scenario towards a media server inside the mobile core network. Instead of the EPC nodes, SGW and PGW, we used an implementation of the GERAN/UTRAN nodes, SGSN and the GPRS Gateway Support Node (GGSN), respectively. The SGSN and the GGSN are also using GTP on their direct interface, similar to SGW and PGW. The dedicated bearer setup procedure is a little different in GERAN/UTRAN and is called Packet Data Protocol (PDP) Context Activation procedure.

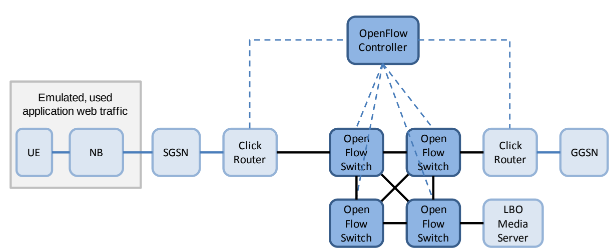


Figure 3: Prototype Setup of the MC-SDN system

Figure 3 shows the setup of MC-SDN using four OpenFlow Switches and a Media Server. The UE and the base station (NB) are emulated using a web application. The user plane packets trigger at the SGSN directly a PDP Context Activation procedure. Since OpenFlow does currently not support the GTP protocol, we use an implementation of the Click Modular Router [7] to process the GTP packets and send them to the OpenFlow Controller via an enhanced interface. Using the Click Router allows us to migrate the developed code later to tiny, highly scalable software routers controlled via OpenFlow [8].

The OpenFlow control framework used in the prototype setup is TREMA [4]. The implemented controller learns the TEIDs and all necessary information during the PDP Activation procedure. Compared to the proposed procedure in Figure 2, the difference here is that the Click Router is a separate entity and does a first processing before the OpenFlow Switch can handle the packet. When the first GTP-C packet is sent from the SGSN to the Click Router, the Click Router forwards it to the OpenFlow Controller. The OpenFlow Controller then detects the GTP-C packet and stores the necessary identifiers. Furthermore the OpenFlow Controller instructs the Click Router what to do with a default route for the GTP-C packets installed in the switches, i.e., removing and adding the GTP header. The Click Router then forwards the IP packet without GTP header to the OpenFlow Switch where the packet will be handled according to the installed route. When the Click Router at the terminating side receives the packet, it adds back the GTP header and forwards the packet to the GGSN. When the GGSN responds to the PDP Context Activation request, the Click Router again forwards the packet to the OpenFlow Controller. The OpenFlow Controller learns all identifiers for GTP-C and GTP-U packets and can install a route also for the GTP-U packets. In our prototype implementation we installed a simple policy for the local breakout based on the destination IP address in the user plane packet from the UE.

4. Conclusion

We studied how to make the transport network below the 3GPP EPC aware of the mobile network's bearers using the OpenFlow protocol and architecture while having minimal impact on the EPC itself. This allows for selective routing in emergency cases, for energy efficient routing and for traffic offload without involving any enhanced 3GPP node. Furthermore, redundant virtualized EPC nodes can be supported without any impact on the attached other EPC nodes. We have implemented a basic prototype for a proof of concept. Next steps are load tests with enhanced tiny software routers and the inclusion of the PCRF into the architecture to form a coherent system, also including fixed line broadband access.

References

[1] OpenFlow Switch Specification Version 1.1.0

[2] N. McKeown, T. Anderson, H. Balakrishnan, G. Parulkar, L. Peterson, J. Rexford, S. Shenker, J. Turner: OpenFlow: Enabling Innovation in Campus Networks, *OpenFlow white paper*.

[3] 3GPP TS 23.401: "GPRS Enhancements for E-UTRAN Access".

[4] H. Shimonishi, Y. Takamiya, K. Sugyo, "OpenFlow Framework Trema", CHANGE & OFELIA Summer School, 2011

[5] 3GPP TS 29.274: "Evolved General Packet Radio Service (GPRS) Tunnelling Protocol for Control plane (GTPv2-C)"

[6] 3GPP TS 29.281: "General Packet Radio System (GPRS) Tunnelling Protocol User Plane" (GTPv1-U)

[7] Eddie Kohler, "The Click Modular Router", Massachusetts Institute of Technology, February 2001

[8] M. Ahmed, F. Huici, A. Jahanpanah. "Enabling Dynamic Network Processing with ClickOS". SIGCOMM 2012 Poster Session, August 2012, Helsinki, Finland

Andreas Kunz received his diploma degree and his Ph.D. in Electrical Engineering from the University of Siegen, Germany. He is working for NEC Laboratories Europe with focus on 3GPP standardization, mainly in the system architecture working group.

Hans-Jörg Kolbe holds a Ph.D. in Physics from Universität Marburg. He received his diploma degree from the same university in 1996. He now leads the Converged Networks Research and Standards group at NEC Laboratories Europe in Heidelberg. Prior to joining NEC in 2007, he was responsible for the broadband network design at Arcor AG & Co KG, which later became part of Vodafone.

Surendran Palanimuthu received his MSc in Business Information System from ESIGELEC, France (2012) and BE in Computer Science from CEG, Anna University, India (2007). Currently, he is working as Software Test Engineer for UTRAN at Nash Technologies in Nürnberg. He was working as Student Research Intern at NEC, Heidelberg and was working as a Software Engineer for packet core networks in Starent Network.

JaeSeung Song is currently working as a senior researcher and oneM2M standardization engineer at NEC Europe Ltd, Heidelberg, Germany. Previously, he worked for LG Electronics as a senior research engineer from 2002 to 2008. He received a PhD at Imperial College London in the Department of Computing, United Kingdom and BS and MS degrees from Sogang University.

KANDOO: A FRAMEWORK FOR EFFICIENT AND SCALABLE OFFLOADING OF CONTROL APPLICATIONS

SOHEIL HASSAS YEGANEH AND YASHAR GANJALI
UNIVERSITY OF TORONTO
{SOHEIL, YGANJALI}@CS.TORONTO.EDU

1. Introduction

Frequent network events can easily stress the control plane and hinder the scalability of Software-Defined Networks (SDNs) [**1**]. For instance, network-wide statistic collection and flow arrivals at a high frequency can over-consume the bandwidth allocated for the control channels with severe impact on other parts of the control plane. Thus, limiting the overhead of frequent events on the control plane is essential for scalable Software-Defined Networking.

Existing solutions try to address this problem by processing frequent events in the data plane. For instance, DIFANE [**3**] offloads part forwarding decisions to special switches, called authority switches, and DevoFlow [**1**] introduce new functionalities in the data plane to suppress frequent events and to reduce the load on the control plane.

To limit the load on the controller, an effective option is to handle frequent events close to the datapath, but preferably without modifying switches. Adding new primitives to switches comes at the cost of visibility in the control plane and necessitates changes to the southbound protocols (e.g., OpenFlow).

Kandoo. Taking an alternative route, we have proposed Kandoo [**4**], a framework for preserving scalability without modifying switches. Kandoo is a novel distributed control plane that offloads *local* control applications (i.e., applications that do not require the network-wide state), over available resources in the network to process frequent events at scale.

2. Design

As illustrated in Figure 1, a network controlled by Kandoo has multiple *local controllers* and a logically centralized *root controller* structured in a two layer hierarchy: (i) at the bottom layer, we have a set of controllers with no interconnection, and no knowledge of the network-wide state, and (ii) at the top layer, we have a *logically centralized* controller that maintains the network-wide state. Controllers at the bottom layer run only local control applications near datapath. These controllers handle most of the frequent events and effectively shield the top layer.

Kandoo is OpenFlow-compatible in a sense that it does not introduce any new data plane functionality in switches, and, as long as they support OpenFlow, Kandoo supports them, as well. To that end, all network functionalities are implemented as control applications that are automatically distributed by Kandoo without any manual intervention. In other words, Kandoo control applications are not aware of how they are deployed in the network, and application developers can assume their applications would be run on a centralized OpenFlow controller. The only extra meta-data Kandoo requires is a flag showing whether a control application is local or not.

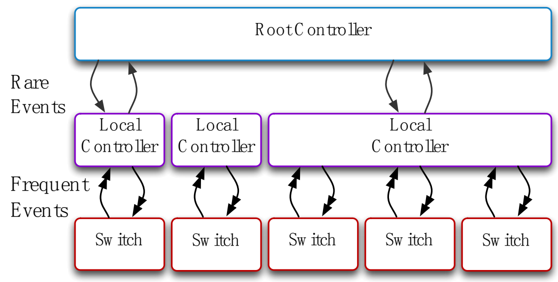


Figure 1. Two layers of controllers in Kandoo.

An Example. To shed more light on Kandoo's design, we show how Kandoo can be used to reroute elephant flows in a simple network of three switches (Figure 2). Our example has two applications: (i) $\text{App}_{\text{detect}}$ that detects elephant flows, and (ii) $\text{App}_{\text{reroute}}$ that reroute the detected elephant flows. To detect elephant flows, $\text{App}_{\text{detect}}$ constantly queries each switch. Once an elephant flow is detected, $\text{App}_{\text{detect}}$ notifies $\text{App}_{\text{reroute}}$, which in turn may install or update flow-entries on network switches. Without modifying switches, it is extremely challenging, if not impossible, to implement this application in current OpenFlow controllers [**1**]. That is to say, collecting network-wide statistics from a (logically) centralized control would place a considerable load on control channels that results in profound degradation in the quality of service.

As shown in Figure 2, Kandoo replicates $\text{App}_{\text{detect}}$ on processing resources close to the switches; hence, each application instance can frequently query each switch to detect an elephant flow without affecting the other parts of the control plane. Once an elephant flow is detected, $\text{App}_{\text{detect}}$ notifies $\text{App}_{\text{reroute}}$ residing on the root controller. Since these events are significantly less frequent than statistic queries, Kandoo can scale considerably better than a normal OpenFlow network.

Kandoo local controllers alongside with the logically centralized root controller collectively form Kandoo's distributed control plane. Each local controller can control multiple switches, but each switch is controlled by one and only one local controller. If the root controller needs to communicate with a switch, it delegates the request to the respective local controller. For high availability, the root controller can register itself as a slave controller in switches supporting OpenFlow 1.2 or higher.

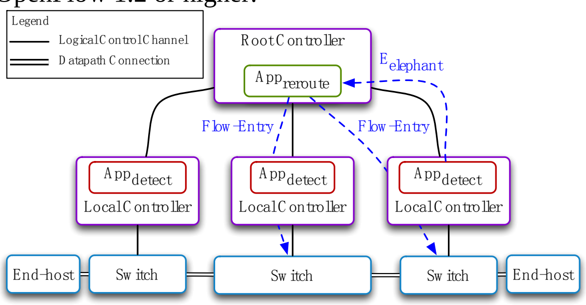


Figure 2. Rerouting elephant flows in Kandoo.

Deployment Model. Kandoo has an elastic architecture, and thus can be tailored based on the characteristics of a network. For in-software, virtual switches, local controllers can be directly deployed on the same end-host. Similarly, for programmable switches equipped with general purpose co-processors, we can deploy Kandoo directly on the switch. Otherwise, we deploy Kandoo local controllers on the processing resources closest to the switches (e.g., on an end-host directly connected to the switch). In such a setting, one should provision the number of local controllers based on the workload and available processing resources. We note that one can utilize a hybrid model in real settings.

3. Concluding Remarks

Kandoo is a highly configurable and scalable control plane, with a simple yet effective approach: it processes frequent events in highly replicated local control applications and rare events in a logically centralized controller. As confirmed by experiment [**4**], Kandoo scales considerably better than a normal OpenFlow implementation.

Although distinctive, Kandoo's approach is orthogonal to other distributed control plane. HyperFlow [**2**] and Onix [**5**] try to distribute the control plane while maintaining logically centralized, eventually consistent network state. These approaches can be used to realize a scalable *root* controller; the controller that runs non-local applications in Kandoo.

Moving forward, we are extending Kandoo's abstraction to include control applications that are not necessarily local but that have a limited scope and can operate by having access to the events generated by a part of the data plane. We note that this necessitates using a hierarchy of controllers (as opposed to the presented two-level hierarchy). We have also started porting Kandoo to programmable switches. When such switches are equipped with Kandoo, they can natively run local control applications.

References

[1] Jeffrey C. Mogul et al., "DevoFlow: cost-effective flow management for high performance enterprise networks," in *Proceedings of the 9th ACM SIGCOMM Workshop on Hot Topics in Networks*, 2010, pp. 1:1--1:6.

[2] Amin Tootoonchian and Yashar Ganjali, "HyperFlow: a distributed control plane for OpenFlow," in *Proceedings of the 2010 internet network management conference on Research on enterprise networking*, 2010, pp. 3-3.

[3] Minlan Yu, Jennifer Rexford, Michael J. Freedman, and Jia Wang, "Scalable flow-based networking with DIFANE," in *Proceedings of the ACM SIGCOMM conference*, 2010, pp. 351-362.

[4] Soheil Hassas Yeganeh and Yashar Ganjali, "Kandoo: a framework for efficient and scalable offloading of control applications," in *Proceedings of the first workshop on Hot Topics in Software Defined Networks (HotSDN'12)*, 2012, pp. 19-24.

[5] Teemu Koponen et al., "Onix: a distributed control platform for large-scale production networks," in *Proceedings of the 9th USENIX OSDI conference*, 2010, pp. 1-6.

Soheil Hassas Yeganeh is PhD student at the University of Toronto. He received his bachelors and masters from Sharif University of Technology. His research interests include software-defined networking, network virtualization, and congestion control.

Yashar Ganjali is an associate professor of computer science at the University of Toronto. His research interests include packet switching architectures/algorithms, software defined networks, congestion control, network measurements, and online social networks. Dr. Ganjali has received several awards including an Early Researcher Award, Cisco Research Award, best paper award in Internet Measurement Conference 2008, best paper runner up in IEEE INFOCOM 2003, best demo runner up in ACM SIGCOMM 2008, and first and second prizes in NetFPGA Design Competition 2010.

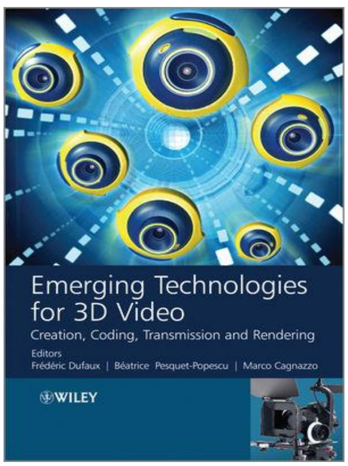
NEW BOOK ANNOUNCEMENT: Emerging Technologies for 3D Video: Creation, Coding, Transmission and Rendering

Wiley, publication date: May 2013, ISBN: 978-1-1183-5511-4

Editors: Frederic Dufaux, Béatrice Pesquet-Popescu, Marco Cagnazzo,
Télécom ParisTech/CNRS LTCI, Paris, France
{dufaux,pesquet,cagnazzo}@telecom-paristech.fr

With the expectation of greatly enhanced user experience, 3D video is widely perceived as the next major advancement in video technology. In order to fulfill the expectation of enhanced user experience, 3D video calls for new technologies addressing efficient content creation, representation/coding, transmission and display.

The underlying principles of stereopsis have been known for a long time. Stereoscopes to see photographs in 3D appeared and became popular in the 19th century. First demonstrations of 3D movies took place in the first half of the 20th century, initially using anaglyph glasses, and then with polarization based projection. Hollywood experienced a first short-lived golden era of 3D movies in the 1950's. In the last ten years, 3D has regained significant interests and 3D movies are becoming ubiquitous. Numerous major productions are now released in 3D, culminating with 'Avatar', the highest grossing film of all time.



In parallel with the recent growth of 3D movies, 3DTV is attracting significant interests from manufactures and services providers. This is obvious by the multiplication of new 3D product announcements and services. Beyond entertainments, 3D imaging technology is also seen as instrumental in other application areas such as video games, immersive video conference, medicine, video surveillance, and engineering.
With this growing interest, 3D video is often considered as one of the major upcoming innovation in video technology, with the expectation of greatly enhanced user experience.

This book intends to provide an overview of key technologies for 3D video applications. More specifically, it covers the state-of-the-art and explores new research directions, with the objective to tackle all aspects involved in 3D video systems and services. Topics addressed include content acquisition and creation, data representation and coding, transmission, view synthesis, rendering, display technologies, human perception of depth, and quality assessment. Relevant standardization efforts are reviewed. Finally, applications and implementation issues are also described.
More specifically, the book is composed of six parts:

1. **Part1 : acquisition and creation**
2. **Part 2: representation, compression and transmission**
3. **Part 3: view synthesis and rendering**
4. **Part 4: 3D display technologies**
5. **Part 5: human perception and perceptual quality assessment**
6. **Part 6: Applications of 3D video**

By covering general and advanced topics, providing at the same time a broad and deep analysis, the book has the ambition to become a reference for those involved or interested in 3D video systems and services. Assuming fundamental knowledge in image/video processing as well as basic understanding in mathematics, this book should be of interest to a broad readership with different backgrounds and expectations, including professors, graduate and undergraduate students, researchers, engineers, practitioners, and managers making technological decisions about 3D video.

MMTC OFFICERS

E-LETTER BOARD MEMBERS