A Multi-view E-learning System for Remote Education

Ján Valiska*, Lukáš Sendrei†, Ľudmila Maceková ‡ and Stanislav Marchevský§
Department of Electronics and Multimedia Communications
Technical University of Košice
Košice, Slovak Republic
*jan.valiska@tuke.sk, †lukas.sendrei@tuke.sk, ‡ludmila.macekova@tuke.sk, §stanislav.marchevsky@tuke.sk

***Abstract*—In this paper, a multi-view e-learning system for supporting remote education is introduced. Since video is the most natural way to perceive information and of interaction between people, it also has inherent capabilities when using in education environment. The system, we propose, should encourage students more use the opportunities offered by remote learning to foster their training on the university. To enable multi-view video (MVV) streaming a system with three cameras is used. A central server is used to capture the partial streams and it also serves as the control unit to manage the whole system. The student will have the choice to decide which particular part of the captured scene he/she wants to watch. Another option will be a guided stream, where the streams are switched automatically, based on the area where the presentation is currently focused. We believe that this model might contribute to the educational process.**

***Keywords*—Distance education, E-learning, Multimedia, System**

I. Introduction

Multimedia is becoming a part of our everyday live. The rapid development of information technologies changed the way we communicate, work, buy things, and how we entertain. And even education is heavily affected by the possibilities that multimedia offers in terms of interactivity as well as seamless connection of people. In this manner, various on-line learning application emerged in the past and the distance between the lecturer and the student is no more an issue [1]. Current rapid development of using computers and various types of networks has not only increased the volume of multimedia content, but also the requirements for its sharing. As part of fulfilling these requirements, a so-called e-learning, like a method for supplement and improving direct education (i. e., face-to-face) or, respectively, for its full replacement with remote access to various forms of electronic educational content, is developed [2]. E-learning technologies include video conferencing, audio conferencing, electronic whiteboards, application sharing, chat, chat engines, virtual classroom and learning with computer support. Unique technologies, known as services for streaming media, are widely used for delivering multimedia educational materials in e-learning.

Media streaming is the simultaneous transmission of digital media (i. e., video, audio, data) so that they are received as a continuous data stream in real time. The data are transmitted by a server application and are received and displayed using a client application. These applications can trigger video displaying or audio playing once sufficient amount of data in the receiving station is received and stored in the cache. Multimedia content for e-learning is delivered by the streaming server. According to this, the performance of the server is a key factor which influences the stream quality. Students attitude to education may be indirectly affected by this fact. Because when the capacity and the performance or bandwidth are not satisfying the long waiting for data can decrease his/her motivation to study. During video streaming system design, increased emphasis on reliable and fast network and high-performance data storage devices, to offer good video streaming services, is needed.

MVV is a set of 2D pictures of the same 3D scene captured synchronously by a camera field placed close together with different angles of view. By an interactive MVV streaming the client can play the video from one view, he/she can stop or randomly switch to output from any camera. Interactive video streaming technology is embedded in many modern applications for mobile communication devices, e.g.: smartphones or tablets. An example of this kind of service is the ClassX Mobile application for Android operating system, developed and already used in education at Stanford University [3]. ClassX is an open source interactive platform enabling to receive video, multitouching (i.e., interactive zooming of selected area), etc.

The detailed description of the multi-view streaming system, which will be used as a supporting tool for educational purposes, is the main goal of this paper. The proposed system captures the area where an lecture will take place with using three video cameras, with high definition. Each camera is recording the lecture from different viewpoints. In this manner, a more complex and detailed scene representation can be obtained. This has an enormous meaning in technical lectures where an specific area of interest with greater importance can be emphasized. Enabling different end user device, e.g.: mobile phones, tablets, notebooks, etc., must be taken into account when designing such type of multimedia services. It is a fact, or at least high probable, that this application must be able to cope with multi-user streaming. Therefore, also the capability of the control server must be high enough to enabling reliable delivery of high quality multimedia content to the end user.

The rest of the paper is organized as follows. Section two defines the term of multi-view video streaming. The proposed multi-view streaming system is introduced in section three. In this section, the network topology, respectively, each part of the system is described with its particular functionality as well as hardware and software components. Finally, conclusions and future research intentions are given in the last section.

978-1-4799-7740-6/14/$31.00 ©2014 IEEE

II. MULTI-VIEW VIDEO STREAMING

One can find many definitions of describe MVV in the literature. We define multi-view as recording of a scene from different positions. It means that multiple video cameras, at least two of them, are placed around a captured scene in different angles. The view direction and the viewpoint can be changed interactively by the user. Other meanings of multi-view refer not only to different positions in space but also to different time instances and other things. Based on this, we recognize various types of MVV described in the following subsection.

A. Multi-view video

The weakness of classic 2D video representation is notable in some cases, e.g. interactive or real-time entertainment-oriented applications. As an example, we can imagine a live music concert where the viewer might want to control the viewpoint. In the traditional 2D, the viewer has no tool to influence the viewpoint, which is fixed. Hence, interaction is not possible and the viewer is just a passive participant and can watch the stream based on non-user selected video sequence (i. e. single-view video). On the other hand, when the scene is synchronously recorded from different viewpoints, we call this MVV [4]. The user (i.e. viewer) has in this case the choice between several views and can easily switch to the desired one, based on his preference. In MVV streaming, the viewer interaction is considered as a most important feature.

1) 3D Video: 3D video is a type of media where the scene is captured using mostly two cameras. By applying suitable video progressing techniques a depth can be experienced in the video representation. The ability to perceive depth scenes from a 2D representation format is a quite old technique. Sir Charles Wheatstone is a pioneer in 3D video and described the general principle of stereoscopic imaging already in 1838 in [5]. When applying the stereoscopic principle, the viewer sees two separate images, recorded in subtly different angles of view, one for the right eye and one for the left. Further, the viewer must use polarized glasses which filters out the images for each eye. Assuming this requirements, an illusion of a depth in the video representation can be achieved. This procedure is already well known and widely used e.g in IMAX 3D theaters. However, the introduction of 3D broadcast television or other home platform has not been successful yet. One reason for its low acceptance is the need of special filtering glasses, which can act as a annoying and hindering matter This might be the argument why this platform failed. A solution to this particular problem might be the usage of autostereoscopic displays which do not require 3D glasses [6].

Another drawback of traditional stereoscopic video is the absence of interactivity. No matter in which direction for the display the viewer moves, he/she gets the same 3D image. This is caused by head motion parallax [7]. This can be solved by installing a head tracker [8], [9], and by creating a new pair of stereoscopic images which is based on the new position of the head. This allows the view of a so-called ”look around objects” effect . The number of views in an autostereoscopic displaying can vary in the range from 2-1000. With the crowing number of views the complexity of the video system the usage of resources increases as well.

Although the illusion of dept can be created by MVV, there exist other technologies which are also capable to this such as Zcam [8]. To get depth in the video sequence, Zcam uses a high-speed pulsed infrared light source. This is done by measuring the time of the emitted and reflected light.

2) Free Viewpoint Video: Free viewpoint video (FVV) is an another example of application of MVV representation. Also in this case, a scene is captured by multiple cameras. With techniques from computer vision, these synchronized video sequences can be transformed, with the help of computer vision techniques, into a data representation that allows the viewer to freely choose viewpoint and direction, respectively [10]. Many of the same principles known from 3D computer graphics (e.g. rendering) are used in FVV as well. FVV is captured from real world objects, whereas 3D computer graphics is usually built upon pure imaginary graphical structures. However, FVV is for broadcast services really unpractical because it requires high data rates and computer computation power. On the other hand, this disadvantage can be relatively applicable for storage applications. One should note, that FVV and 3D video does not exclude each other. It is possible to make a FVV with a 3D video scene representation using the right processing tools. Although, this will be very resource-demanding. An illustration of FVV capturing is shown in Fig. 1.

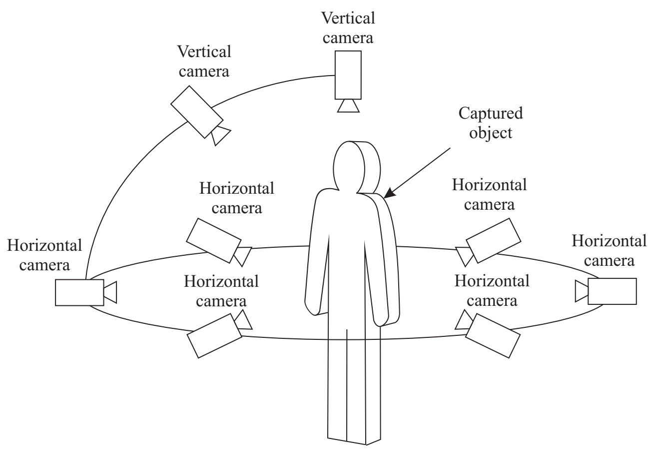

Fig. 1. Camera set-up illustration by capturing video sequences using FVV.

3) Omnidirectional Video: In omnidirectional video (also known as immersive video) [11] the scene is captured by multiple cameras, where the individual views represent spherical or cylindrical fields of the captured environment. In other words the surrounding is captured at different position, in every direction. The position can move over time. With the appropriate software, the viewer can easily zoom and rotate around the different captured images. The final video representation provides the user with the feeling of being in the scene. However, in contrast to FVV, the user is not able to change the position of the viewpoint interactively. One way how to change the viewpoint is that the cameras are moved during capturing, but it might be a challenging task. One can describe omnidirectinal video as a special case of interactive MVV [12]. An example of omnidirectional video capture is shown in Fig. 2.

Fig. 2. An example of omnidirectional video captures [13].

B. Multi-view video coding

It is obvious that MVV requires large portion of the bandwidth available in todays data networks, and also storage capacities. Hence, video compression is an vital part of the video system design. Naturally, the simplest solution to solve this problem, is to encode each individual video sequence independently using a state-of-the-art codec such as H.264/AVC [14] or the recently introduced H.265/HEVC [15]. The first one can be with advantage used in wireless networks as shown in [16]. Recent investigations done by the Moving Picture Experts Group (MPEG) has shown that further improvements can be done by exploiting statistical dependencies within MVV Coding (MVC). Large amount of inter-view statistical dependencies can be observed in a MVV system, because all of the cameras are set up to capture the same scene[17]. We can this dependencies divided into two types, temporal similarity and inter-view similarity. Inter-view similarity is predicted from corresponding images in adjacent views, whereas temporal similarity is determined by temporal neighboring images. Therefore, an approach that combines inter-view/temporal prediction algorithms can lead to better coding efficiency. Numerous researches have shown (e.g., [18], [19]) that specialized MVC techniques can under specific circumstances outperform, in terms of objective and subjective measurements, the independent encoding of individual streams from each camera. However, this performance gain requires its price, it increases computational complexity, memory requirements as well as delay. Fig. 3 shows a simplified block diagram for MVC. Based on the recent activities in standardization organizations and various experts group from industry, one can clearly see that MVV is becoming an integral part of modern entertainment-oriented multimedia platforms.

III. Proposal of multiview e-learning system

The goal of this paper is to propose the e-learning system with multiview capabilities and interactive communications between system and attenders of university courses.

We decided to develop a system that will have client-server architecture. The database and stream server (DBSS) is as heart of proposed system. It is responsible for storing the incomming streams from multiple cameras and serving this streams, based on client capabilities (e. g. screen resolution of tablet, smartphone etc.). Switching input sources by tracking of speaker or next slide events and storing this timeline information of presentation is also very important, because the attenders can have interactive experience even when they watch presentations from recorded streams. The server is controlled by speakers control point in realtime. That means that the speaker can explicity define what video source is served to the client [20].

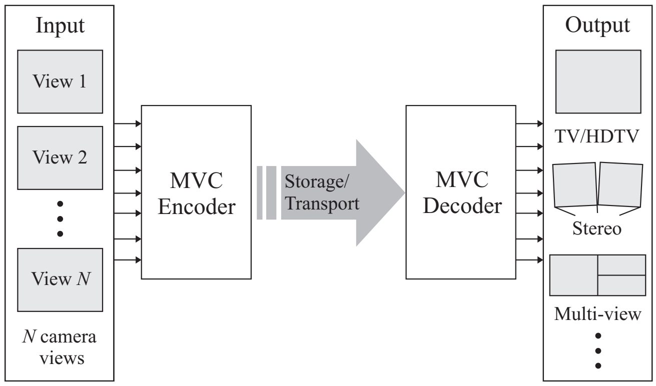

Fig. 3. Block scheme of a multi-view video coding system.

A. Topology

The topology of the proposed system is shown in Fig. 4. As you can see, all components are bounded together using network switch with *power on ethernet (PoE)* ports. This decision simplifies the connection of cameras.

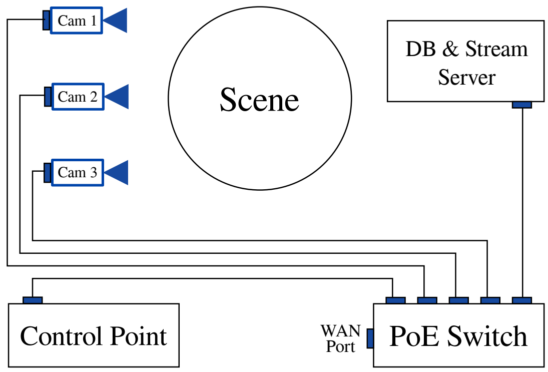

Fig. 4. Topology of proposed e-learning system

Remote users can access to database and streaming server through WAN port of switch, which is connected to university LAN network and Internet.

Multi-view approach is very flexible and every camera can record one point or area of interest in audience hall. For example during the lecture of Ultra Wide Band (UWB) radar system in our laboratory, one camera can be oriented to UWB radar, second camera on speaker, and another camera on whole audience room (e.g. from top of room). See figure 5.

The slideshow has a separate channel and is streamed from control point (CP) to DBSS server and next to clients.

Audio signal is captured by every camera and is transmitted to DBSS server. At this point audio signals are mixed together and transmitted through main transmission channel to clients. It is posibble to use dedicated wireless audio microphones situated far of cameras for discussion question recording.

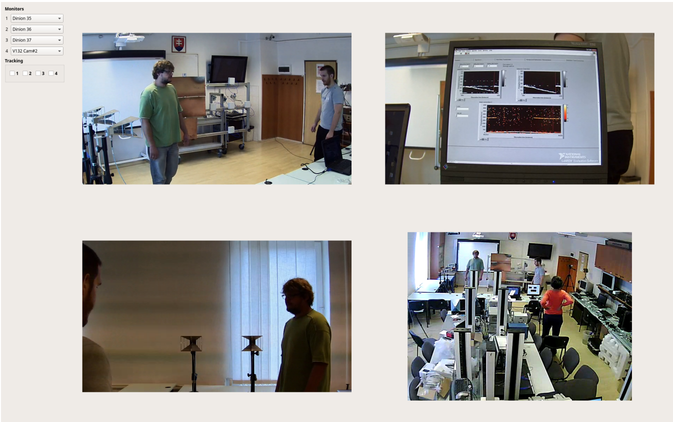

Fig. 5. Example of four camera topology of MVV system

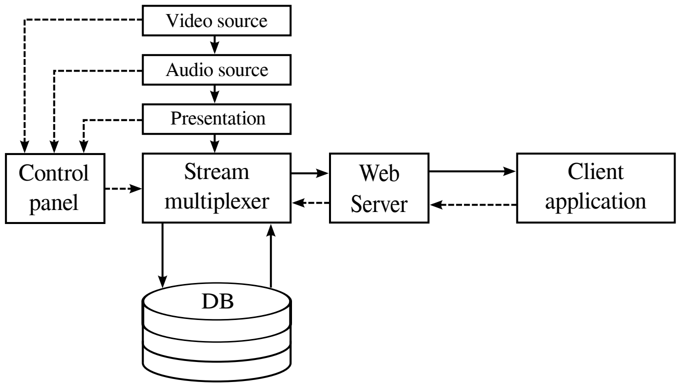

Fig. 6. The architecture of the MVV system

B. Components

In this section particular components of proposed system are described.

For capturing of object in scene we decided to use Full HD surveillance cameras (Bosh DINION HD 1080p HDR IVA). This camcorders have very good quality in poor illuminated areas. Every camera is connected and powered through PoE ethernet ports of switch with only single ethernet cable that simplifies wiring of system. We have connected three cammeras to a switch. Camera lenses are adjustable by software geometric calibration [21].

PoE switch have enough power for powering peripherals(cameras, positioners…). Every port of ethernet switch (Cisco WS-C2960C-8PC-L) has fast 1Gbps specification, what is good for transmitting of huge amount of video data from camera to DBSS server and consequently to clients.

Every PC with Windows or Linux operating system can be control point. In our proposed system CP is laptop with DVI link connected to LCD projector, and ethernet port connected to ethernet switch. From this place, the whole course of presentation can be controlled (e.g. change of slides, camera switching, speaker tracking).

Database and streaming server is midpoint between audience hall and remote students. It is key part of proposed system and therefore it must be very reliable. We use dedicated HP server (ProLiant) with fast harddrives in hardware RAID 1 configuration. If one HDD fails, it can be simply changed.

C. Architecture

In this section the architecture of the system is described. Basic architecture is shown in figure 6.

Operational data are obtained in separated blocks.

Video source is array consisted of multiple cameras. In our test enviroments there is three full HD cameras. Every camera is configured in one configuration XML file. There are described IP adresses of cameras, symbolic names (usefull for identification), description and also information about cameras positioner (if included in camera).

Audio sources are group of microphones (e.g. embedded in camera or dedicated).

Presentation is source of slides or other video or picture examples. Originally they are played or presented from second workspace of control point and are streamed as presentation video track.

Control panel (CP) is block used for controlling progress of whole lecture or talk. In this block director of presentation can switch camera source which is intended to be streamed to clients. Also, there is option for automatic camera source switching based on tracking of speaker using bayesian filtering (Particle filters) [22].

Stream multiplexer (SM) is controlled by control point, and from many video, audio or presentation tracks the one track is created and then streamed to clients. Stream multiplexer can get data directly from sources or from database (DB). If client wants to view old archived presentation, then stream multiplexer reads control information and source data from database and streams composed track to client. In other way if client wants to view live presentation, SM gets actual data from sources and multiplexes them together using actual control information from CP [23].

The main interface between multi-view stream system and client is web server and web application. Web server provides graphical user interface (GUI) to clients. In this interface clients can view list of all lectures or presentations from database, and if there is also live presentation, the users are informed about it.

For live experience, web application is implemented in event-driven javascript library. This approach is best for resolution of this problem, because server can automaticaly (based on events) control video player in user interface.

D. Software framework

Software for controlling, capturing and streaming of audio and video signals is very important. There are also software products for multi-view streaming like ClassX [3]. One disadvantage of this software is that the multi-view is proposed as point of interest from one big video picture. Student is able to zoom in or zoom out to this region what is interested for.

For capturing and streaming of video streams the Video Lan Client (VLC) is one of the biggest open source candidate.

It have many options like capturing from various sources (e.g. web cams, USB connected video devices, files etc.). Also every stream can be transcoded to desired image resolution and required bitrate. Streaming of this video and audio streams is possible through many output modules. Convenient way of streaming is streaming through HTTP protocol and RTSP protocol (for internet streaming) and RTP/UDP multicast streaming in local area networks (like our university network). Disadvantage of VLC is that there is no option for live switching between input sources, what is our most wanted ability [24].

One piece of software that meets our needs is FreeJ (Free Vision Mixer). It is totally free instrument for realtime video manipulation. It consists of a commandline application on GNU/Linux, a graphical application on Apple/OSX, a C++ multimedia framework library and a javascript engine for efficient live video manipulation. It can be controlled in different ways: Joystick, MIDI, OSC, sensors, keyboard. It can be controlled locally or remotely from multiple places.

FreeJ is an asynchronous video rendering engine, it can be scripted using javascript syntax in an object oriented way, to control its operations through a procedural list of commands and actions [25].

Developers of FreeJ also provides the core of freej software as C++ library with API for a multimedia framework. And this is main reason for our decision using this software. We can build our streaming server on top of this software using bundled library and Python or Perl bindings.

Other softwares are also proposed but they are not open software. Open software means, that everybody can view, edit and share source code. And this is our preferred way. We want software that is open and shared over the Internet because we want support also from other developers and clients.

IV. Conclusion

The aim of this paper is to introduce a MVV system for e-learning purposes to support the education of university students. The detailed description of the system architecture has shown that a multi-view camera configuration can be effectively used to display a technical lesson with detailed view to a particular area of interest. This could be not possible with classical video representation techniques, where the camera can capture just the whole scene and the view is limited to this concrete image. However, many technical problems occur with the introduction of this type of service. The first issue is the need of high bandwidth, caused by the requirement to transport multiple independent video streams. This can be partially solved with the assistance of MVC techniques. The second open issue is the multi-platform and multi-device support. Assuming that the user, in this case the student, will use various operating systems as well as devices, then the platform independence is an important aspect of the system design. These, and probably many other, open questions will be the object of study of future research. Regardless the efforts taken in the system design it is up to the student if they accept these new multimedia content type. It is our best hope that this type of platform might ameliorate and enrich the classical education process.

Acknowledgment

This work was supported by the Slovak Cultural and Educational Grant Agency (KEGA) under contract No. 062TUKE-4/2014 and by the Scientific Grant Agency of the Ministry of Education, Science, Research and Sport of the Slovak Republic and the Slovak Academy of Sciences (VEGA) under contract No. 1/0563/13.

References

[1] S. Malik and A. Agarwa, "Use of multimedia as a new educational technology tool - a study," *International Journal of Information and Education Technology*, vol. 2, no. 5, pp. 468–471, Oct 2012.

[2] J. Vassileva, "Toward social learning environments," *Learning Technologies, IEEE Transactions on*, vol. 1, no. 4, pp. 199–214, Oct 2008.

[3] D. Pang, S. Halawa, N.-M. Cheung, and B. Girod, "Classx mobile: Region-of-interest video streaming to mobile devices with multi-touch interaction," in *Proceedings of the 19th ACM International Conference on Multimedia*, ser. MM '11. New York, NY, USA: ACM, 2011, pp. 787–788. [Online]. Available: http://doi.acm.org/10.1145/2072298.2072457

[4] J.-G. Lou, H. Cai, and J. Li, "A real-time interactive multi-view video system," in *Proceedings of the 13th Annual ACM International Conference on Multimedia*, ser. MULTIMEDIA '05. New York, NY, USA: ACM, 2005, pp. 161–170. [Online]. Available: http://doi.acm.org/10.1145/1101149.1101173

[5] C. Wheatstone, "On some remarkable, and hitherto unobserved, phenomena of binocular vision." *Philos. Trans. R. Soc. Lond*, 2007.

[6] N. Dodgson, "Autostereoscopic 3d displays," *Computer*, vol. 38, no. 8, pp. 31–36, Aug 2005.

[7] K. Lackner, A. Boev, and A. Gotchev, "Binocular depth perception: Does head parallax help people see better in depth?" in *3DTV-Conference: The True Vision - Capture, Transmission and Display of 3D Video (3DTV-CON), 2014*, July 2014, pp. 1–4.

[8] C. Fehn, R. de la Barre, and S. Pastoor, "Interactive 3-dtv-concepts and key technologies," *Proceedings of the IEEE*, vol. 94, no. 3, pp. 524–538, March 2006.

[9] M. Lopez, J. Hannuksela, O. Silven, and L. Fan, "Head-tracking virtual 3-d display for mobile devices," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2012 IEEE Computer Society Conference on*, June 2012, pp. 27–34.

[10] A. Smolic, K. Mueller, P. Merkle, C. Fehn, P. Kauff, P. Eisert, and T. Wiegand, "3d video and free viewpoint video - technologies, applications and mpeg standards," in *Multimedia and Expo, 2006 IEEE International Conference on*, July 2006, pp. 2161–2164.

[11] A. Smolic and P. Kauff, "Interactive 3-d video representation and coding technologies," *Proceedings of the IEEE*, vol. 93, no. 1, pp. 98–110, Jan 2005.

[12] A. Smolic and D. McCutchen, "3dav exploration of video-based rendering technology in mpeg," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 14, no. 3, pp. 348–356, March 2004.

[13] J. Fermoso, "Omni directional pictures," 2008, [Online; accessed 09-September-2014]. [Online]. Available: \url{http://www.wired.com/images_blogs/photos/uncategorized/2008/07/01/omni_directional_pictures_with_new_.png}

[14] T. Wiegand, G. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the h.264/avc video coding standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 13, no. 7, pp. 560–576, July 2003.

[15] G. Sullivan, J. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (hevc) standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 22, no. 12, pp. 1649–1668, Dec 2012.

[16] L. Sendrei, J. Valiska, and S. Marchevský, "H.264 video transmission over wlan in opnet modeller," *Journal of Electrical Engineering*, vol. 64, no. 2, pp. 112–117, 2013. [Online]. Available: www.scopus.com

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

[18] M. Flierl, A. Mavlankar, and B. Girod, "Motion and disparity compensated coding for multiview video," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 17, no. 11, pp. 1474–1484, Nov 2007.

[19] P. Merkle, A. Smolic, K. Muller, and T. Wiegand, "Efficient prediction structures for multiview video coding," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 17, no. 11, pp. 1461–1473, Nov 2007.

[20] Ľ. Maceková, D. Levický, and S. Marchevský, *Progresívne technológie v DVB-T*. TUKE, 2011, ch. DIGITÁLNA TELEVÍZIA, TEÓRIA A PRAX, pp. 18–47.

[21] O. Kováč and J. Mihalík, "Teoretická analýza geometrickej kalibrácie kamery," *Posterus*, vol. 7, no. 7, pp. 1–9, 2014.

[22] J. Valiska, S. Marchevský, and R. Kokoška, "Object tracking by color-based particle filter techniques in videosequences," in *Radioelektronika (RADIOELEKTRONIKA), 2014 24th International Conference*, April 2014, pp. 1–4.

[23] S. Marchevský, D. Levický, Ľ. Maceková, S. Pillár, B. Hrušovský, and Ľ. Holováč, *Progresívne technológie v DVB-T*. TUKE, 2011, ch. IPTV: ZÁKLADNÉ TECHNOLOGICKÉ RIEŠENIE, pp. 150–199.

[24] B. Al-Madani, A. Al-Roubaiey, and Z. A Baig, "Real-time qos-aware video streaming: A comparative and experimental study," *Advances in Multimedia*, vol. 2014, p. 11, 2014.

[25] "Freej :: Free vision mixer," http://www.dyne.org/software/freej.