Extended Object Tracking and Stream Control Model Based on Predictive Evaluation Metric of Multiple-Angled Streams

Dávid Cymbalák, Ondrej Kainz, and František Jakab

***Abstract*—Paper describes the design of extended model for tracking and locating object in multi-cameras environment with controlling the streams from multiple angles in real time. Proposed and experimentally implemented solution is attempting to deliver via streaming technology of the only specific shot from scene that has the best view on tracked object in real time. Measurement of the best shot for the tracking object is based on proposed evaluation metric, consisting of dimensional, locational and reliability parameters gathered from object tracking methods. For latency optimization and accuracy of evaluation the prediction of evaluation metric was outlined. Solution is experimentally tested in standard and mobile environment as well. Approach described in this paper is innovative regarding the content related video stream delivery.**

***Index Terms*—Computer vision, live streaming, multi-cameras systems, object tracking.**

I. Introduction

In present the demand for using technologies with computer vision has increasing trends. Object detection, tracking, recognition are long known parts of intelligent systems not only in surveillance but also in shopping, education or entertainment. Using the computer vision technologies with mobile devices is more and more widespread because of computational capacity of mobile devices and also the quality of integrated capturing camera devices. With the increasing use of cloud computing and streaming services it is interesting to bring solutions which interconnect the computer vision techniques with streaming technologies and multimedia delivery issues.

The goal of our work is to outline the multi-cameras solution which is able to receive video in real time from various devices to find the chosen object in video using tracking methods and to bring only one video with the best shot on object in real time to the output for recipients. For successful implementation of this solution it is necessary to analyze mobile computer vision abilities, streaming platforms and also design own metric for evaluating each node of proposed system. Prediction of this metric is essential for achievement of the best results.

Manuscript received June 24, 2014; revised August 28, 2014. This paper/This project is co-financed by the European Union. Paper is the result of the Project implementation: University Science Park TECHNICOM for Innovation Applications Supported by Knowledge Technology, ITMS: 26220220182, supported by the Research & Development Operational Programme funded by the ERDF.

The authors are with the Computer Networks Laboratory at Technical University of Kosice, Slovakia (e-mail: {david.cymbalak, ondrej.kainz, frantisek.jakab}@cnl.sk).

II. Object Tracking in Real Time Video

Object tracking can be simply defined as the problem of estimating the trajectory of an object in the image as it moves around a scene [1]. The tracker is trying to assigns consistent labels to the tracked objects in different frames of a video. Depending on the tracking domain, the tracker can also provide locational or dimensional object information, such as coordinates, orientation, area, or shape. Primary goal of tracking methods is to build a model of what we want to track however we require the information about the object's position in the previous frames to achieve predictions about the current frame and restrict the search [1].

A. Object Tracking Algorithms

Nowadays there are various real-time tracking algorithms. Based on Wang's tracking algorithms comparisons [2], the Tracking-Learning-Detection algorithm (TLD) is currently considered to be the most effective and accurate in various types of videos. TLD simultaneously tracks the object, learns its properties in the following frames and detects and verifies the occurrence of the object in the image. The result is tracking of object in real time, where the accuracy and reliability of the tracking is improving related by time due to learning the new features of the observed object, e.g. changing the position, size, rotation or brightness. Tracking algorithm of TLD is based on the steps of monitoring recursive tracking in forward and backward direction, which is performed by Lucas-Canade algorithm in both directions by calculating of median at the end. Detection of the TLD uses dispersion filter, file classifier and nearest neighbor classifiers [3]. Learning in TLD is realized by PN learning, where the data can be classified with yielding the structure in the form of path of object and application of the positive (P) limitations and then the negative (N) limitation [4].

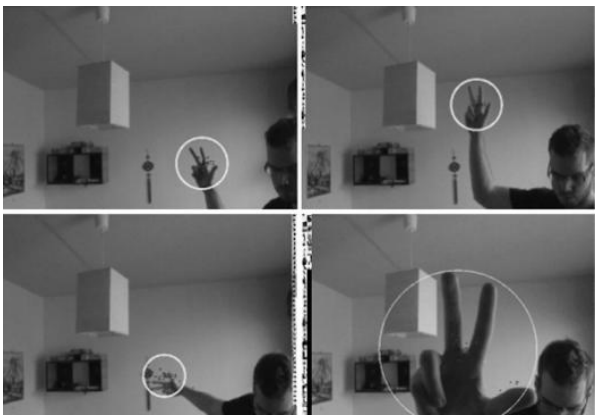

Fig. 1. C++ implementation of OpenTLD used by testing to track hand in various angles and shapes.

DOI: 10.7763/IJCTE.2015.V7.983

There are various open implementations of TLD algorithm, which were tested for proposed solution. OpenTLD C++ application was deployed for tracking on desktop platform with camera (Fig. 1). Couples of implementation were carried out on mobile platform while using different computer vision libraries: TLD based on OpenCV, TLD based on Fast CV, TLD based on BoofCV.

B. Object Tracking on Mobile Platforms and CV Libraries

The most popular computer vision libraries with object tracking ability on mobile platforms are OpenCV, FastCV and BoofCV.

FastCV originally developed by Qualcomm supports Android and Windows mobile platform and provides a clean processor-agnostic hardware acceleration API under which chipset vendors can hardware accelerate FastCV functions on their hardware [5]. BoofCV is Java based open source library for real time computer vision and robotics applications. Easily usable and with high performance, it often outperforms even native libraries with specific situations [6]. OpenCV is the most widespread computer vision library with huge support of the user community. Yet it embodies C, C++, Python and Java interfaces and supports Windows, Linux, Mac OS and mobile platform iOS and Android. OpenCV was designed for computational efficiency with a strong focus on real-time applications. The library is written in optimized C/C++, and also takes advantage of multi-core processing [7].

Fig. 2. Various tracking techniques in different computer vision libraries on android mobile device.

Tracking efficiency was tested and deployed using different computer vision library as experimental applications on Android mobile device (Fig. 2). Android application based on BoofCV offers object tracking based on Circulant, MeanShift, Sparse Flow or TLD method. Based on experiments, the TLD method based on BoofCV was the most accurate and stable; nonetheless it causes high CPU load and the dropping of FPS. Another implementation of TLD using FastCV library had the same results in accuracy and stability of tracking but the efficiency of using system resources was more efficient, but in specific cases the FastCV library register a compatibility problem with new version of Android 4.4 with quad-core CPUs.

III. Live Streaming and Content Delivery

Previously discussed tracking method and computers vision application shows resulting image on local devices, where also the actual coordinates of object are evaluated. However for the purposes of our concept we need to deliver

image in real time to another recipients without layer of tracking lines or squares. This is to contain the information about position of tracked object in background.

A. Streaming Servers and Protocols

In this solution we need to consider utilization of suitable streaming protocol to deliver real time video from system nodes without undesirable latency or high load. Streaming protocols for transmitting multimedia contents can vary in implementation details that divide them into two categories: PUSH protocols - establish a connection between the server and the client, connection is maintained and packets are send to the client until the connection is interrupted until the expiration of limit or disruption on side of client, PULL protocols - make the client to be an active element that establishes a connection in the form of requests for streaming media content from a server [8]. Based on approach from [9] were identified differences between RTMP (PUSH type) and HTTP (PULL type), these differences were mainly in duration of request time on server. In this case the efficiency of RTMP protocol is several times higher than HTTP (Fig. 3) although the results may vary depending on the type of application.

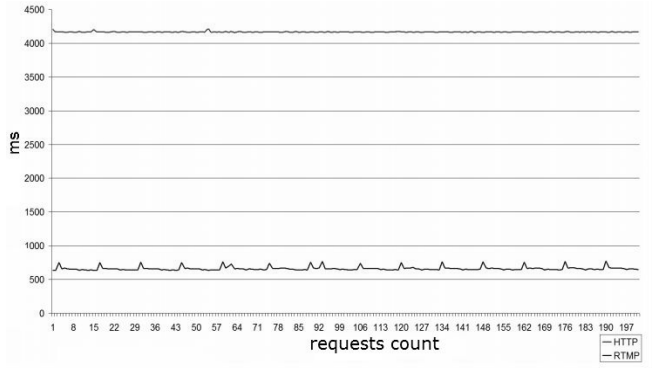

Fig. 3. Efficiency in time between answer and request on server between streaming protocol HTTP vs RTMP [9].

B. Streaming Source Control

Another need is to redistribute, clone and switch video stream sources to resulting streaming output. It is crucial to select one stream from many streams incoming from devices - these devices are simultaneously tracking the object to one resulting output stream in real time with the best shot focused on tracked object. To achieve this we used the vMix API of software video mixing solution. The vMix API provides access to common functions such as switching output based on parameters through the HTTP protocol. The parameter related to input source is in our solution chosen by the results of evaluation metric on each tracking device.

IV. Design of Extended Model

Overall proposed extended model deals with object tracking process, evaluation metric, stream switching process, prediction of metric, redistribution of object tracking pattern etc. Focal point of paper is primarily to introduce outlined predictive evaluation metric.

A. Evaluation Metric

Evaluation metric (1) of streaming source *n* at time *t* consists of the components: M_p^t (positional component of metric), M_v^t (dimensional component), M_d^t (detection reliability component). It can be expressed in the form as:

$$M_n^t = (M_p^t + M_v^t) * M_d^t \quad (1)$$

Individual components of proposed metric are evaluated from information gathered from TLD tracking algorithm where one line contains coordinates, dimensions and reliability for each frame in video (Table I).

TABLE I: Output for 6 Frames of Video from OpenTLD

F	*X*	*Y*	*W*	*H*	*D*
489	*NaN*	*NaN*	*NaN*	*NaN*	*0.000000*
490	*63*	*176*	*75*	*59*	*0.748394*
491	*64*	*186*	*75*	*59*	*0.651912*
492	*64*	*186*	*75*	*59*	*0.620531*
493	*64*	*184*	*75*	*59*	*0.613741*
494	*63*	*182*	*75*	*59*	*0.608190*

The positional component of proposed metric is based on zonal division of image based on ideal composition. The image captured by each node in system is divided to 5 zones with values from 0 to 4. The division is made by golden cut ideal composition rule based of Fibonacci spiral (Fig. 4).The object gets the value of the zone in which its current coordinates of the middle of object belong. If the object is outside of the screen, object gets zero value in positional component of evaluation metrics.

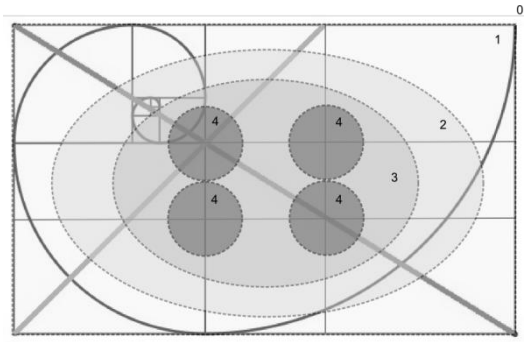

Fig. 4. Zonal division of shot based on fibonacci spiral.

B. Global Prediction of Evaluation Metric

Very appropriate is to predict evaluation metric which is important to achieve optimal latency between discovery of the source with the highest rating and the final appearance of the output video source. Such prediction could initialize switching of video source at time *t* to source which should have the highest ratings of metric at time *t*+1.Under ideal conditions, this approach should provide a video source switching before or in the right moment, i.e. when the tracked object is arriving to the capturing scene.

Prediction mechanisms to estimate of the subsequent evaluation of streaming sources can be integrated to the overall design on global level or on local level. First option considers the global prediction mechanism located between the mechanisms for calculation the evaluation metrics and streaming control mechanism (Fig. 5). In this case the prediction mechanism uses PH_k matrix (2) containing H_t

vectors of previous metrics values (M_n^{t-k}) for each of *n* sources:

$$PH_k = \begin{pmatrix} M_1^t & M_2^t & \cdots & M_n^t \\ M_1^{t-1} & M_2^{t-1} & & M_n^{t-1} \\ \vdots & \vdots & \ddots & \vdots \\ M_1^{t-k} & M_2^{t-k} & \cdots & M_n^{t-k} \end{pmatrix} \quad (2)$$

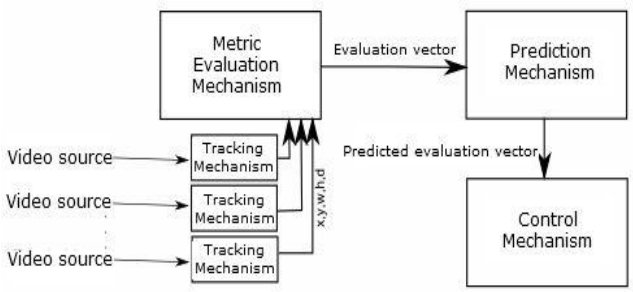

Fig. 5. Design of integration of the global prediction mechanism.

Based on the PH_k matrix the global prediction mechanism creates a new predicted evaluated vector (3) containing metric's values for *n* sources in *t*+1 time:

$$H_{t+1} = [M_1^{t+1}, M_2^{t+1}, \ldots M_n^{t+1}] \quad (3)$$

After calculating the maximum $max_{0<k\leq n} M_k^t$ from n nodes at time t will be obtained the index of source with best evaluation metric in actual time. This index determines which video source should be broadcasted in specific time (Fig. 6).

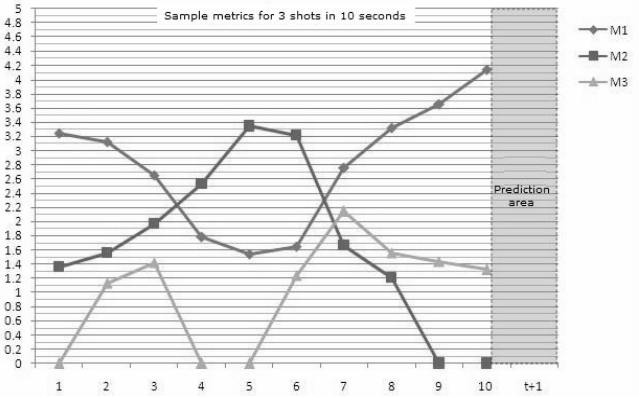

Fig. 6. Balancing of evaluation metric in real time for 3 camera sources and finding the predicted part.

Prediction mechanism on global level could be based on utilization of Active Lezi algorithm which has the advantage of preserving a variable length window. After assembling the tree using Active LeZi it is possible to predict the next evaluation metric. This calculation should also utilize PPM (Prediction by Partial Match) algorithm, yet in our solution is considered to use simplified customized version of PPM algorithm.

C. Local Prediction of Evaluation Metric

Another outlined approach of prediction is to integrate local mechanisms for prediction between the detection mechanism and the mechanism for calculating an evaluation metric (Fig. 7). In this method, each local prediction mechanism uses the PO_k matrix (4) containing the O_n^t vectors of *k* previous properties of tracked object (coordinates *x* and *y*, dimensions *w* and *h* and credibility *d*):

$$PO_k = \begin{pmatrix} x^t & y^t & w^t & h^t & d^t \\ x^{t-1} & y^{t-1} & w^{t-1} & h^{t-1} & d^{t-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x^{t-k} & y^{t-k} & w^{t-k} & h^{t-k} & d^{t-k} \end{pmatrix} \quad (4)$$

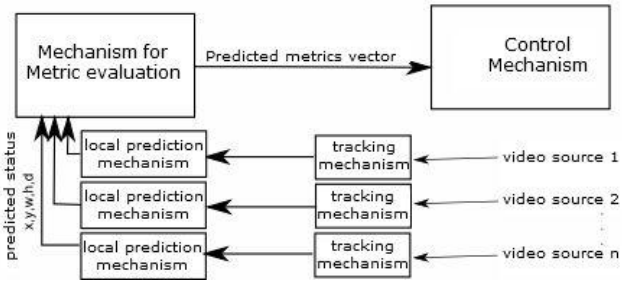

Fig. 7. Design of integration of the local prediction mechanisms.

Based on PO_k is locally created the new vector (5) containing predicted object properties (coordinates, dimensions, credibility) for each *n* source in time *t*+1:

$$O_n^{t+1} = [x^{t+1}, y^{t+1}, w^{t+1}, h^{t+1}, d^{t+1}] \quad (5)$$

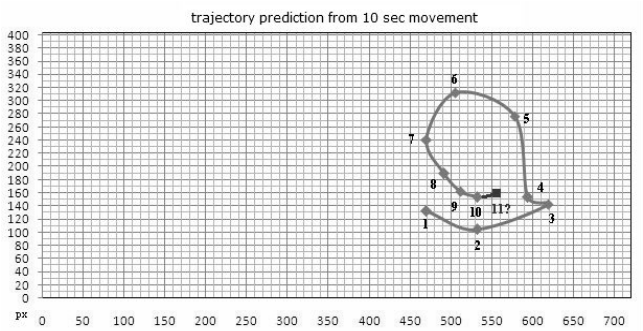

Fig. 8. Movement of tracked object on one camera screen with prediction of trajectory.

Prediction mechanisms at the local level could be based also upon the methods of predicting of trajectory of movement based on previous coordinates such are Kalman filter or extended Kalman filter (Fig. 8). Kalman filter (KF) is used for prediction of the next position based on the movement description of the previous motion section. Model Kalman filter requires learning the representation of the tracked object in the status perception [10]. An alternative approach to increase the accuracy of prediction is to use the extended Kalman filter (EKF). Here the status function may take also in non-linear form. The accuracy of the EKF is particularly useful in difficult movements with sudden change of direction.

D. Implementation of Solution

Currently the proposed system is in experimental operation, ready to implement and test changes in metrics definitions and use the prediction methods. Solution itself consists of interconnection between software video mixing program created with the API, which provides switching of streaming sources that is depended on the information as delivered by the mechanism for calculating evaluation metric that are gathered from the local instances of openTLD applications (Fig. 9). The experimental set-up allows us to integrate various kind of capture input into the multi-cameras

system. Support of direct connection with standard cameras with low-latency, HTTP and RTP stream from IP cameras and also mobile streaming applications is as well enabled. Utilization of hardware HDMI input cards with the standard cameras allows handling very low latency between appearing the object on screen and reality. Streaming server based on WSE modified modules was also deployed all this in cooperation with media encoder application. WSE server is handling the delivery of resulting stream with best shot on tracked object to the web interface with video player.

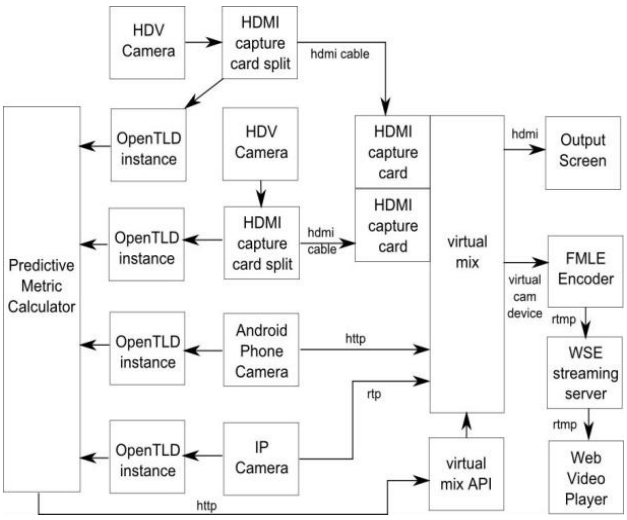

Fig. 9. Interconnection between used components in experimental implementation of solution.

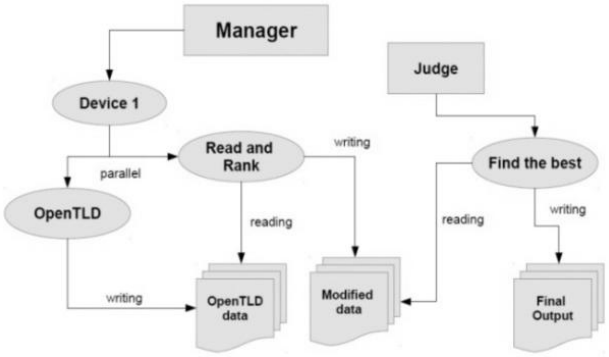

Fig. 10. Practical usage of proposed experimental implementation.

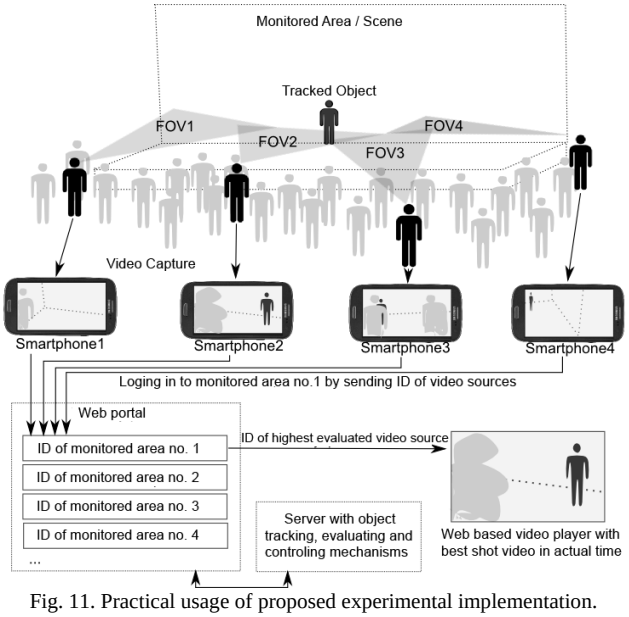

Fig. 11. Practical usage of proposed experimental implementation.

Detailed process of TLD object tracking for one specific capturing device with ranking in evaluation process is depicted in Fig. 10.

Issue of sharing pattern for object tracking is in our experimental implementation solved by providing the access to one object tracking pattern for all devices during capturing of event. Initial object tracking pattern could be defined by primary device in system by cropping from specific video frame or could be imported from external file. Other devices are using this defined pattern and extend it with additional information about object's appearance, such as brightness changes, rotation or different angles view. Extended patterns for object detection and tracking from all nodes of system could be exported for next usage.

Overall solution could be also utilized in the real environment as system which enables sharing of live events based on location – by streaming from mobiles, smart glasses and cameras to the web. This system is to choose the best shot from the streams at the same location in real time automatically (Fig. 11). In this way a new experience of consuming live video for distant audience is provided.

V. Conclusion

The goal of this paper was to introduce system with extended model for object tracking in multiple-angled streams based on evaluation metric. The overall solution enables capturing the image from the set of cameras or the set of mobile devices in real time – further capabilities enable tracking of selected object for each source, calculate the proposed evaluation metric and switch the source with best shot to final streaming output. We have experimentally created prototypes for conditional switching of resulted live streams, streaming from mobile phones, object tracking via mobile phones and optimization of streaming server for live streaming delivery in form of experimental system for evaluation the best shot from multiple streams in real time.

The next step is definition and methods optimization of evaluation metric prediction for individual sources. This is to include the definition of the recursive neural networks, implementing Bayes classifier and finally comparison of the prediction reliability in compiled solutions, respectively. The future work could be also focused on extending the solution with techniques of augmented reality and utilization of wearable smart devices with cameras connected to network.

References

[1] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *Acm Computing Surveys (CSUR)*, vol. 38, issue 4, 2006.
[2] Q. Wang *et al.*, "An experimental comparison of online object tracking algorithms," in *Proc. SPIE Conf. Image and Signal Processing Track*, 2011.
[3] Z. Kalal, J. Matas, and K. Mikolajczyk, "P-N learning: Bootstrapping binary classifiers by structural constraints," in *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, 2010, pp. 49-56.
[4] Z. Kalal, K. Mikolajczyk, and J. Matas, "Forward-backward error: automatic detection of tracking failures," in *Proc. International Conf. Pattern Recognition*, Istanbul, Turkey, 2010.
[5] Fast CV Library 1.1.1. Qualcomm Inc. (2014). [Online]. Available: https://developer.qualcomm.com/docs/fastcv/api/index.htm
[6] P. Abeles, "Speeding up SURF," *Advances in Visual Computing*, Springer Berlin Heidelberg, pp. 454-464, 2013.
[7] G. Bradski and A. Kaehler, *Learning OpenCV: Computer Vision with the OpenCV Library*, O'Reilly Media Inc, 2008.

[8] A. Begen, T. Akgul, and M. Baugher, "Watching video over the web, part I, streaming protocols," *IEEE Internet Computing*, vol. 15, issue 2, pp. 54-63, 2011.
[9] T. MizerÁK, "Protocols for clients communication in flash," MUNI Brno Patent, 2010.
[10] N. Funk, "A study of the kalman filter applied to visual tracking," *Project for CMPUT 652*, 2003.

David Cymbalak was born in 1987. In 2011 he graduated and got the MSc degree from the Department of Computers and Informatics, Faculty of Electrical Engineering and Informatics at Technical University in Kosice. Since 2011 he is a PhD student at the Department of Computers and Informatics. His scientific researchs are focusing on streaming technologies, e-learning solutions or object tracking methods. In addition, he also investigates questions related with the optimization of mobile access to multimedia sources.

Ondrej Kainz was born in 1988. In 2013 he graduated and got the MSc degree in applied informatics from the Technical University in Kosice, Slovakia. Since the very same year he is a PhD student at the Department of Computers and Informatics of the Faculty of Electrical Engineering and Informatics at the Technical University of Kosice. His scientific research interests include e-learning, human-computer interfaces, computer graphics, computer networks, biological engineering and body area network.

František Jakab was born in 1959. He graduated from St. Petersburg Electro Technical University and got the MSc degree in system engineering in Russia in 1984, the PhD degree from Technical University of Kosice, Slovakia in 2005. He has been an associated professor since 2008. He has extensive experience in networking and utilization of ICT in education where he established well known research centre - Computer Networks Laboratory. He has been an coordinator of several large international projects financed by EC, an coordinator of national wide ICT projects and research grants; the chair of many international symposiums and conferences, the editor of conference proceedings. Since 1999, he involved into Cisco Networking Academy Program in Slovakia as an instructor (CCAI certification) and since 2001 in the position of coordinator of the Program in Slovakia, the regional lead for Russia, Ukraine and CIS from 2008 to 2014. He is also the head of the Application Section of the Communication Technology Forum Association in Slovakia, the head of Committee on business – academic cooperation, American Chambers of Comers in Slovakia and general manager of University Centre for Innovation, Technology Transfer and Intelectual Property Protection at Technical university of Kosice.