Multi-Camera Broadcasting Model with Automation of Optimal Scene Switching

D. Cymbalak, F. Jakab, M. Michalko, O. Kainz and R. Vapenik

Abstract— **Manuscript deals with introduction of model for automatic selection of the scene with the best position of the object of interest in a multi-camera live broadcasts. The novel metric for evaluation of object appearance in multi-camera scenes was proposed and designed following the deep analysis of relevant broadcasting technologies and object tracking methods. Evaluation of object appearance in scene comprises not only from the location and size of the object of interest in the actual frame but also from streaming transmission parameters and the subjective rating from the broadcast recipients. The proposed metrics serve as the basis for the establishment of a system for selecting the best scene with switching in the real-time. Model has been experimentally deployed in two alternative implementations using common and mobile devices. Results were compared with human based broadcast direction. Based on this comparison, the ability to respond to changes in the scene and also to capture the object of interest in the stream was observed. The resulting application of model should be adapted in different fields such as broadcast of conferences, sport events or security systems.**

Keywords*— *Broadcast, Multi-camera, Scene selection, Object tracking

I. INTRODUCTION

In recent years there has been a tremendous expansion of video technologies in terms of internet traffic. The amount of this traffic has grown exponentially and led to the formation of a new technologies dealing with the video processing, distribution and even recognition. Research presented in this paper is addressing the interconnection of the outputs from the late developments in streaming and computer vision to the area of multi-camera systems with automatic director.

We support research activities in Slovakia/This project is being 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, supported by the Research & Development Operational Programme funded by the ERDF.

D. Cymbalak is with University Centre for Innovation, Technology Transfer and Intellectual Property Protection on Technical University of Kosice, Slovakia (e-mail: david.cymbalak@cnl.sk).

F. Jakab is with University Centre for Innovation, Technology Transfer and Intellectual Property Protection and Department of Computer and Informatics on Technical University of Kosice, Slovakia (e-mail: frantisek.jakab@cnl.sk).

M. Michalko is with Department of Computer and Informatics on Technical University of Kosice, Slovakia (e-mail: miroslav.michalko@cnl.sk).

O. Kainz is with Department of Computer and Informatics on Technical University of Kosice, Slovakia (e-mail: ondrej.kainz@cnl.sk).

R. Vapenik is with Department of Computer and Informatics on Technical University of Kosice, Slovakia (e-mail: roman.vapenik@cnl.sk).

Tracking the object of interest in a scene from multiple cameras has become principal focus of the many researches, especially those dealing with security systems. Multi-camera broadcasts, however, exist also in other areas of life. Tracking of an object in video with subsequent shot evaluation is generally not the focal point of researches. Due to this reason, the ambition of this paper is to introduce a solution, which is to replace the manual switching of the output image in multi-camera streaming, i.e. provide the new approach to personalized broadcast, where each viewer receives a directed broadcast in real-time based on prior preferences and defined object of interest in the scene. Goal is to create an automatized model for streaming of exactly one video source, based on periodic evaluation of multiple video sources. Such source is to have the highest evaluation at given moment following the proposed metric. Idea of metric used for evaluation lies in rating of video sources based on actual position, size and other selected parameters of shot. Prior to design and implementation, the analysis of current streaming technologies is carried out as a theoretical basis for proposal of streaming management mechanisms and also another analysis of current detection of tracking algorithms resulting in a proposal of mechanism enabling evaluation of sources.

II. MULTI-CAMERA STREAM AND OBJECT TRACKING

A. Multi-cameras Systems

Views from individual cameras utilised in multi-camera system may overlap and continuously cover specific part of space or can be isolated without any overlapping, providing there is a greater distance between cameras.

Multi-camera system with overlapping views may be utilized to estimate the height of the tracked object in 3D space [1] or in creation of a global map of covered area, while providing the location of tracked objects on a map. In case of this configuration, the object is often detected concurrently by several cameras. For that reason, utilization of methods for definition of common areas in image and common points of interest is appropriate [2]. Field of view (FOV) lines are constructed based on fields of view [3], these lines are common for several views. Tracking of objects in case of multi-camera system with non-overlapping views from cameras is troublesome mainly due to separation in space and time during tracking of the object. In this case, the distance in time and space cannot be used as relevant information between particular nodes of system, unlike tracking approaches that utilize one camera. Example of utilization of methods for object tracking in multi-camera systems is selection of tracked person across the viewing areas, case of security systems [4]. Another field of utilization is highlighting the position of objects, e.g. players, ball, puck or other gear utilized at sports events.

Multi-camera surveillance systems commonly indicate the tracked object in a form of highlighted geometrical primitives surrounding the object across the outputs from all cameras or mark the position of object in global view on the map of monitored space. Due to this, the viewer has to monitor the output from all cameras and seek the highlighted square, circle or points of tracked object, i.e. coordinates and position in global map. Object in global map represents particular point, not the entire video. Useful solution is hence to stream the content from individual cameras of system, while utilizing streaming technologies and show only the view from one camera. One that detects the tracked object in scene and provides its best location.

B. Streaming Technology

Streaming technologies ensure the transfer of multimedia content over the computer network, in both audio and video along additional content, from the provider to the recipient [5].

Many advantages emerge when we compare streaming multimedia formats to standard ones [6], the most significant are the following: possibility to control the bit rate of content, content protection against unauthorized use, possibility to select or filter the recipients of content and easy content management. Bearing in mind just stated and particularly due to the availability, these technologies are widely utilized in education systems, i.e. e-learning, and also in videoconferencing solutions [7]. Not only receiving of content but also broadcast directly from mobile devices is possible thanks to advanced development in the area of mobile devices [8].

Currently prevails a certain form of inconsistencies when it comes to usage of codecs and streaming formats in different browsers. H.264/AVC is an advanced codec developed by groups ITU-T VCEG and ISO/IEC MPEG. Codec consists of two layers: video coding layer (VCL) and network abstraction layer (NAL) [9]. VCL represents the video content itself and NAL is handling the structure of data, it is also carrier of information necessary for transfer. Such data is used by transport layer and storage media. [10].

Inherent component of streaming technologies are streaming protocols that ensure the transfer of packets with multimedia content. Streaming video server provides distribution of multimedia content into multiple packets and also initializes transfer to the end customer. Variant of client-server model is a content delivery network (CDN) model. Main streaming server, in the solutions based on CDN, initially sends multimedia content to a group of content delivery servers that are strategically deployed at the edge of a network. Another emerging model of service delivery related to streaming is distribution over P2P networks [11].

C. Objeckt Tracking in Stream

Following section is related to tracking itself or tracing the object in the sequence of images, i.e. in the video, considering one camera system. Analysis carried out for one camera system is to be beneficial in the process of proposing the evaluation metrics of multi-camera system. Resulting from this is the real model proposed in this paper.

Image detection methods use image segmentation, this technique divides the image into the area referred to as the foreground, where the tracked object is to be found. Second part is called background, this part is not relevant for further processing. The resulting foreground is consisted of so called image elements and based on predefined pattern the relevant element is detected. Foreground may be extracted based on certain parameters such as colour, shape or texture.

In terms of tracking the object in the video it is important to carry out object detection on the selected series of frames, separated by time, of the particular video. The location with respect to the captured view is calculated for each processed frame that contains the detected object. Series of location data subsequently define the nature of the object's movement, either in the form of a sequence of points with coordinates or in the form of a function.

Detection of moving objects in the video can be accomplished through the filtering out the background. Filtering of the background based on the chromaticity or gradient may be implemented on the level of the pixel, region or the entire image. Methods for tracking of deviations in the individual video frames face various issues that occur during capturing of a real environment, these are: gradual or sudden changes in lighting, low level of colour uniqueness of tracked object from the background, unwanted shadow of the object, change in the background of the object being tracked or the constant movement of the object when detection initializes. All just stated factors affect the recognition success rate. Hence it is only appropriate to use so-called learning algorithms to continuously monitor changes in the tracked object and also complement the model pattern used for object detection.

Common tracking system consists of three components: object representation, dynamic model and search mechanism. Object itself may be represented by either a holistic descriptor such as colour histogram and related brightness value of a pixel, or by local descriptor such as local histogram and chromaticity. Dynamic model is used to simplify the computational complexity in tracking of the object. Search mechanism is used to optimize the tracking of the real object and can utilize both deterministic and stochastic methods. Essential component in tracking methods is a motion model that can express, for example, translational movement, transformation based on similarities and the affinity transformation [12].

The existing methods used in object tracking in real-time include: incremental visual tracking (IVT) [13], variance ratio tracker (VRT) [14], fragments-based tracker (FragT) [15], online boosting tracker (BoostT) [16], semi-supervised tracker (SemiT) [17], extended semi-supervised tracker (BeSemiT) [18], Tracker (L1T) [19], multiple instance learning tracker (MIL) [20], visual tracking decomposition algorithm (VTD) [21] and track-learning-detection method (TLD). Reliability and functionality of these algorithms can be verified, for example, by measuring the success while keeping up the detection of the monitored object and by shift of central position detection. In this paper two algorithms were selected, reporting the best results for use in the proposed system: TDL and CMT algorithms.

TLD (Tracking-Learning-Detection) algorithm was developed for tracking and detection of objects in a video in a real-time. The object of interest is defined following the initialization of the square in one frame, such frame serves as a template. TLD simultaneously tracks the object, determines its features from the following frames and detects and verifies the occurrence of an object in the image. The result itself is tracking of the object in real-time. Accuracy and reliability is constantly improving as the duration of the tracking is getting longer, cause of this is learning of new features of tracked object, e.g. change of position, size, rotation, or luminosity [22].

Tracking algorithm used in the TLD is based on phase of recursive tracking in forward and reverse direction (Fig. 1), Canada-Lucas algorithm is executed in both directions and median is calculated as a final step. Detection in the TLD uses sparse filtering, file classifier and the nearest neighbour classifiers. TLD learning is performed in the form of P-N learning, extracted data is classified, while obtaining the its structure in the form of object path, positive (P) constraints are applied, with subsequent implementation of negative (N) constraints, then new data is generated and updated in the object classifier. P-type image segments represent objects with a high probability of correlation, when compared with the template, and vice versa, image segments of N-type represent objects with low probability of correlation [23].

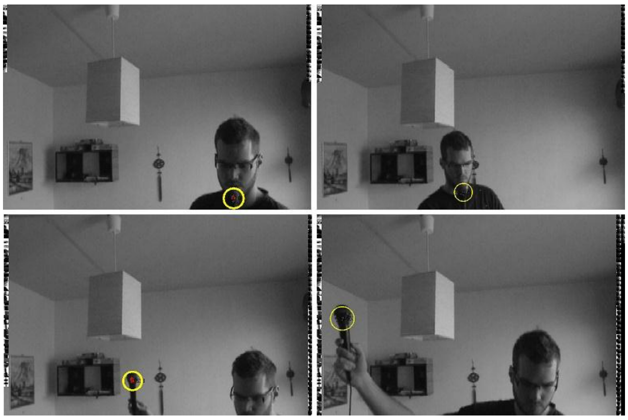

Fig. 1. Experimental testing of Tracking-Learning-Detection algorithm for tracking the microphone in the video stream.

CMT algorithm - Consensus-based matching and tracking is based on the keypoints method in combination with a matching-and-tracking framework (Fig. 2). CMT algorithm and his pythom implementation (pyCMT) expects input to be a sequence of frames and initialization region as pattern for detection in the very first frame. The goal of this algorithm is to renew the position of the pattern in each following frame and thus obtain the position of the tracked object. The output is the information of tracked object's position. Operation of CMT algorithm includes three essential steps: comparison of the correlation and keypoints tracking, voting and approval. To localize the object in each frame, the voting on position of central point from each keypoint is acquired. In order to detect the other points, the system based on approval of correlation in the voting is used. Changes in size, location and rotation of detection region are visible as the transformation of votes follows the specific position of keypoints. Utilization of the fast keypoint detectors and binary descriptors introduces suitable domain for the application of the algorithm in terms of video processing in a real-time [24].

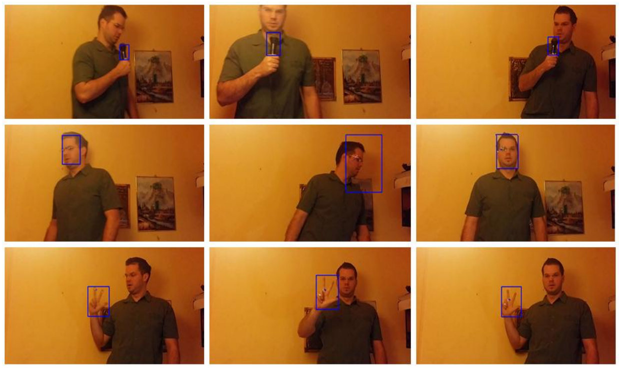

Fig. 2. Experimantal realization of tracking the head, microphone and hand based on CMT algorithm.

III. DESIGN OF MODEL FOR OPTIMAL SHOT SELECTION

The principal goal is to propose a metric that is to combine rating based on parameters with regard to the presence of the object in the shot as gained by the tracking algorithm, having the parameters of picture and voice. As a result, each node of multi-camera system is to be evaluated every second, based on this in each second the source of video stream is estimated and provides the best view and quality on the object of interest. Subsequently such shot becomes the primary and the only broadcasted view.

A. *Metric for evaluation of video sequence*

The principal metric in the proposed model will evaluate video sequence of each video source connected to the multi-camera system. Based on this metric the system will be able to determine which shot is best designed to visually track the object of interest in real-time. The metric (M_n^t) for evaluation of video sequence has been designed as the sum of the five components of the metrics, which are multiplied by a constants defined in the input vector (Equation 1).

$$M_n^t = (c_p \times Mp_n^t + c_v \times Mv_n^t) \times Md_n^t + c_h \times Mh_n^t + c_q \times Mq_n^t \quad (1)$$

Depending on the type of broadcast in which the proposed model is to be deployed, the utilization of c constants allows adjustment of weights of individual metric components, or exclusion of selected components from the evaluation.

Positional component of metrics (Mp_n^t), used for the evaluation of video sequence, is based on data as extracted from tracking algorithm deployed in every stream of multi-camera broadcasts. A significant effect on metrics component are the x and y coordinates, representing the occurrence of the object's central point contained within the input vector for a particular source of the stream. The positional component itself has been designed to act as the optimal composition of the golden ratio, i.e. the centre of the tracked object of interest should be located as close as possible to the nearest golden section of the shot (Fig. 3).

In this case, it is also possible to seek a composition where the object of interest is approaching the best location with respect to the termination of Fibonacci spiral in the four different shots. In this case, zones are designed to keep the most highly rated areas closests to the intersections of the golden rations of shot (see Equations 2).

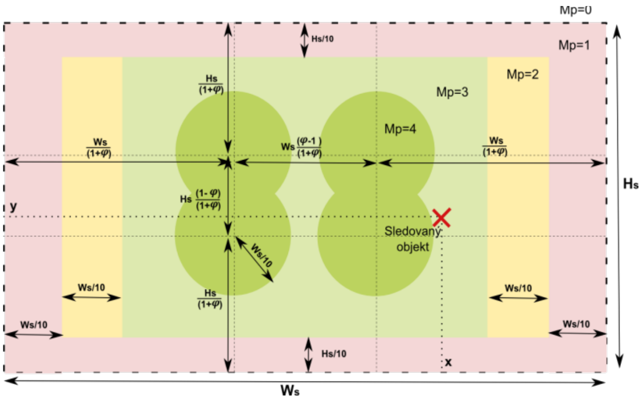

Fig. 3. Zonal division of scene based on golden ratio with positional component of metric calculation.

$$\boldsymbol{Mp_n^t} = \mathbf{0} \Leftrightarrow x = null \;\; OR \;\; y = null$$

$$\boldsymbol{Mp_n^t} = \mathbf{1} \Leftrightarrow x < \frac{W_s}{10} \;\; OR \;\; x > W_s - \frac{W_s}{10}$$

$$OR \;\; y < \frac{H_s}{10} \;\; OR \;\; y > H_s - \frac{H_s}{10}$$

$$\boldsymbol{Mp_n^t} = \mathbf{2} \Leftrightarrow x < \frac{W_s}{5} \;\; OR \;\; x > W_s - \frac{W_s}{5}$$

$$\boldsymbol{Mp_n^t} = \mathbf{4} \Leftrightarrow$$

$$\left(x - \frac{W_s}{(1+\varphi)}\right)^2 + \left(y - \frac{H_s}{(1+\varphi)}\right)^2 < \left(\frac{W_s}{10}\right)^2$$

$$OR \left(x - \frac{W_s}{(1+\varphi)}\right)^2 + \left(y - \frac{\varphi \times H_s}{(1+\varphi)}\right)^2 < \left(\frac{W_s}{10}\right)^2$$

$$OR \left(x - \frac{\varphi \times W_s}{(1+\varphi)}\right)^2 + \left(y - \frac{H_s}{(1+\varphi)}\right)^2 < \left(\frac{W_s}{10}\right)^2$$

$$OR \left(x - \frac{\varphi \times W_s}{(1+\varphi)}\right)^2 + \left(y - \frac{\varphi \times H_s}{(1+\varphi)}\right)^2 < \left(\frac{W_s}{10}\right)^2$$

$$\boldsymbol{Mp_n^t} = \mathbf{3} \Leftrightarrow else \tag{2}$$

Dimensional component (Mv_n^t) (Equation 3) of the proposed metrics is based on the parameters as derived by tracking algorithm for estimation of the actual height h and width w of tracked object in the shot with dimensions W and H. In this situation, the evaluation of this component is to be increased with increasing size of the reference object.

$$Mv_n^t = \sqrt{\frac{w^2 + h^2}{{W_s}^2 + {H_s}^2}} \tag{3}$$

Component concerning the reliability (Md_n^t) (Equation 4) of the object detection is directly dependent on the input data. This data represents reliability of detection d of tracking algorithm at time t contained within the input vector. Consequently, the component is to be in the form of percentage correlation, indicating the match of detected object in the image at time t to the pattern used for tracking.

$$Md_n^t = d \tag{4}$$

The proposed metric also includes a component enabling the control of the resulting metric through the subjective evaluation (Mh_n^t) by the recipients of stream from multi-camera broadcasts. In case of specific broadcast, each recipient would be able to evaluate each stream source at time t by a percentage value, i.e. change own evaluation dynamically throughout the broadcast. In order to guarantee the relevant processing of the evaluation of individual stream sources it is essential to introduce the calculation of evaluation based on Bayesian estimation – bearing in mind the stream is dependent on the number of recipients who are able to evaluate it, not only on the average evaluation (Equation 5).

$$Mh_n^t = \frac{\left(\sum_{k=1}^{k=j} r_k\right) + (c_{min} \times r_{av})}{j + c_{min}} \tag{5}$$

Following the analysis of evaluation methods of video quality, the calculation of objective video quality component of metric (Mq_n^t) (Equation 6) dependent on three fundamental parameters of picture was designed. These image parameters are the following: resolution of video stream (width W and height H), the actual transmission bandwidth B^t (bit-rate) and frame rate F^t compared to optimal bandwidth B_{op} and optimal frame rate F_{op}.

$$Mq_n^t = \frac{\sqrt{{B_{op}}^2 - \left(B_{op} - B^t\right)^2}}{2 \times \left(\sqrt{{W_s}^2 + {H_s}^2}\right) \times 10^{-3}} * \frac{\sqrt{{F_{op}}^2 - \left(F_{op} - F^t\right)^2}}{F_{op}} \tag{6}$$

B. Processing the results of evaluation

Processing requires individual results of partial evaluation metrics of image to be stored in the matrix (Equation 7) containing other sub-metrics for all sources of stream n at time t:

$$Rcomp^t = \begin{pmatrix} Mp_1^t & Mv_1^t & Md_1^t & Mh_1^t & Mq_1^t \\ Mp_2^t & Mv_2^t & Md_2^t & Mh_2^t & Mq_2^t \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ Mp_n^t & Mv_n^t & Md_n^t & Mh_n^t & Mq_n^t \end{pmatrix} \tag{7}$$

Subsequently, the results of overall metrics R^t for each source at time t assemble the image metrics vector (Equation 8) of results for the entire multi-camera system in time t:

$$R^t = (M_1^t \;\; M_2^t \;\; \dots \;\; M_n^t) \tag{8}$$

The final step is to select the stream source in real-time following the calculations of chosen metrics. Such source provides, in the specific broadcast the highest result of the overall evaluation. Basis of selection is to calculate the maximum (Equation 9) of vector R, i.e. maxima of resulting numerical evaluations of all currently connected broadcasting sources:

$$\max_{0<k\leq n} M_k^t \tag{9}$$

Each image capturing device generates a certain latency until the image gets to the output. In the terms of multi-camera heterogeneous system, e.g. consisted from standard SDI cameras, mobile phone with enabled 3G transfer, sports camera with Wi-Fi connectivity or IP camera via fixed line, the delay between video input and the output has to be on the same level. The ambition is to ensure that the object of interest in the image at time t is present in the same global position for all the input video sources during processing. This can be done indirectly by processing the audio track of stream, where the part of audio track from reference source with the lowest latency is selected at regular intervals and compared with other sources. The system would therefore be able re-evaluate how many *ms* is the shift from detected audio sequence of specific source (Fig. 4).

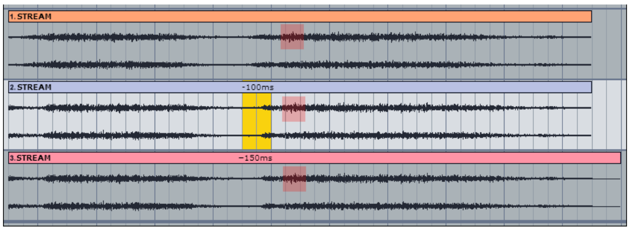

Fig. 4. Synchornization of delay via audio track of multiple streams

IV. Implementation of Model

Building a system following the proposed model is possible on multiple levels according to the character of use. Implementation diagram of the system, based on the model using the external video sources, is depicted in Fig. 5. Note that distribution mechanism is employed prior to the evaluation.

Individual sources of live broadcasts from different angles are transmitted through the streaming distribution device, i.e. streaming server. The transferred stream of each video source, acquired from the distribution mechanism, is received and then processed by the tracking algorithm. Input is in the form of single instance per each video source. Results of tracing are sent, in a real-time, by each instance of object tracking mechanism. Evaluation is performed by the calculation mechanism, which receives the initial parameters of metrics as the input, these are: constants for subjective assessment, vector containing the constants of compression ratio and optimal frame rate, and vector encompassing the constants of the individual components of the global metrics. At the same time, the vector comprising of scene parameters and stream quality is provided to the calculation mechanism. Such vector is based on the encoding used in distribution mechanism or input encoder of video source. Evaluation mechanism of recipient associated with the content playback mechanism collects the ratings from recipients and provides them as the input to the calculation mechanism. Calculation mechanism continuously collects the input parameters. Output of this mechanism, a vector of ratings from all the input sources, is sent directly to the control mechanism. The control mechanism determines the best rating and source to be set as the output of the current broadcast, both is done in real-time. Control itself is thus connected directly to the content playback mechanism, where the distribution mechanism provides only the stream with the best rating to the recipient.

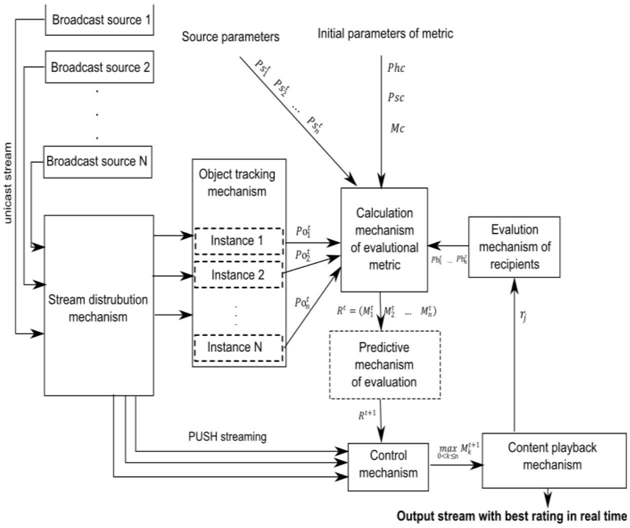

Fig. 5. Implementation scheme with stream distribution mechanism

Several system implementations were carried out as a part of this project, all were following the proposed model for the automatic selection of the optimal shot while utilizing specific streaming and computer vision technologies. One of such system configuration was the implementation encompassing the mobile devices with build-in camera, experimentally linked to the evaluation mechanisms located behind the streaming server (Fig. 6).

In this configuration, two Android mobile devices, one iOS mobile device and sports GoPro camera were used as the sources of the image. Image extracted from Android devices was received directly by WSE streaming server, in this case by means of created RTMP encoder. iOS devices utilized Wowza GoCoder and stream from GoPro camera was distributed via FFmpeg to WSE server. As a result, all the image sources were distributed through streaming server. Subsequently, each distributed RTMP stream was processed by FFmpeg in separate instances of pyCMT tracking algorithm. Individual pyCMT instances passed on results of the position and size of the tracked object for further calculation of the final metrics, along with the WSE server stream and calculation of recipients rating parameters. The final metrics values were send, containing the information of maxima and identifier of best evaluated via vmix API. Then the output stream was passed to the output in the virtual mix, which received each stream separately from WSE server. Subsequently, the output stream encoded the required format through FMLE and forwarded it to a recipient's FlowPlayer player and to a control output.

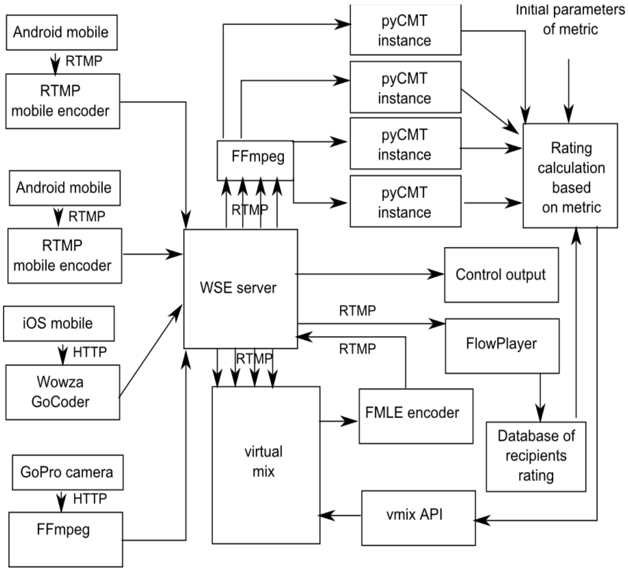

Fig. 6. System configuration with using pyCMT mechanism

Several experiments were carried out in the simulation environment. The moving object of interest was captured simultaneously using multiple cameras, both in different environments and different image dynamics. Then, the feedback, whether the object of interest in sequence *t* is visible, was retrieved from the recipients. Subjective evaluation of each sequence was done to provide the evaluation of the composition quality of the tracked object and also its distinctiveness when compared to the expectations of the viewer. Comparison included the manual director as is present in the standard system of video mix, this provided the information about the video stream selected as the output in time t. In case of same real-time video sequence, the table for comparison was continuously filled with the results as gathered from both, the automatic director based on the proposed model and manual director in standard video mix. Results included in the table create a map of the object visibility in the video sequence from both perspectives - comparison of the selection of output image and evaluation of composition. This allows us to express the correlation of automatic director to manual director, i.e. when it runs faster or, vice versa, when the lag is present.

Proposed model was experimentally tested on extensive sequences of live broadcasts throughout each development and optimization phase of relevant metrics. Experimental multi-camera broadcast was selected as a sample of the model variability. This included combination of standard cameras and mobile devices build-in cameras. Pattern for the initialization box used for tracking of the object was set to copy the figure of a lecturer in front of blackboard (Fig. 7).

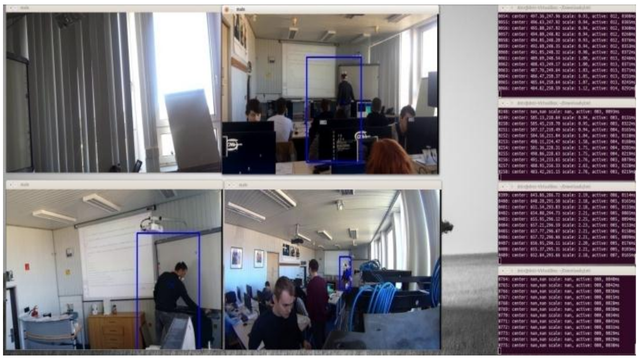

Fig. 7. Experiment of tracking the lecturer in 4-camera stream

Simultaneously, the video streams assigned to a streaming mix were processed and combination of parameters from the evaluation mechanism and another combination of qualitative parameters allowed final evaluation metrics to provide the information of current source. Such source was set, at time *t*, as the output by streaming mix (Fig. 8).

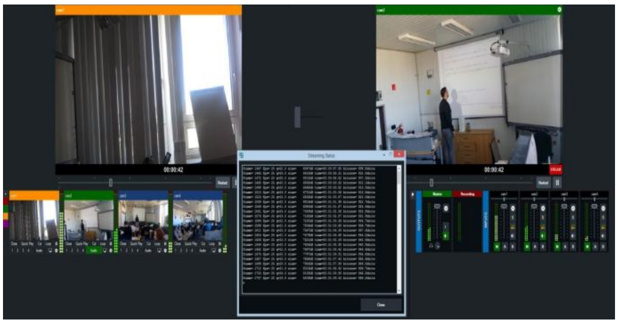

Fig. 8. Experiment of choosing the best rated stream based on tracking

The results of the final evaluation of individual sources for the selected broadcast sequence is depicted on the graph, see fig 9.

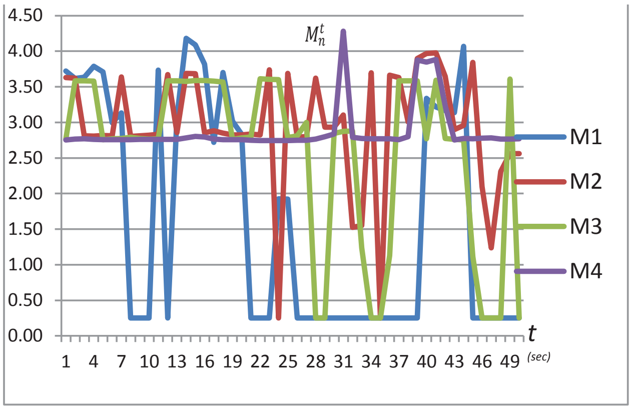

Fig. 9. Results of 4-camera stream evaluation by proposed automatic system

Comparison of the results served for the evaluation of object composition in the selected sequence of a broadcast, both automatized system using proposed model (blue) and manual director (red) were deployed (see Fig. 10).

This experiment shows that automatic selection of the best view does not match the manual director, dissimilarity rate is estimated to be two thirds. Object of interest (lecturer) was in the case of automatic selection visible throughout the entire broadcast of segment. Manual director achieved error rate of 2% - it means the object was not in the shot in 2% of duration of the broadcast segment. The automatic system scored worst composition rating then manual direction in 36% of duration of broadcast segment. However in 54% of duration of broadcast segment the composition of object of interest (figure of lecturer) with automatic system has better rating then manual direction. In this experiment, the model succeeded especially in case of object arrival and disappearance from the image, response time was in the case of proposed model faster when compared to the manual mode.

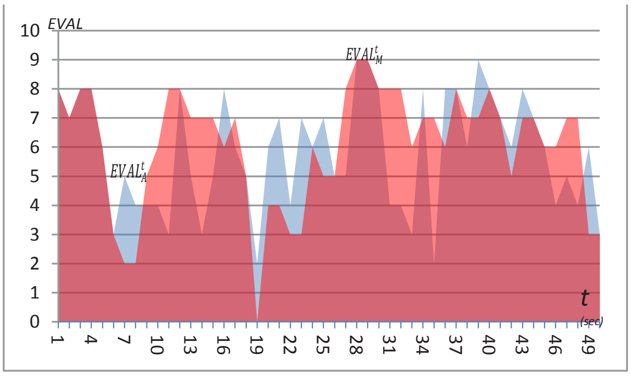

Fig. 10. Comparison of evaluation of composition between proposed model (blue) and manual direction (red) in 4-camera stream experiment

The overall model for the selection of the best image in a multi-camera system proposed in this paper is adaptable for usage in different types of broadcasts. The primary objective was to develop a model for the selection of optimal shot that would be useful for heterogeneous multi-camera system, e.g. broadcasting of the events from multiple angles using different types of devices such as classic camera, mobile phone, sports camera, drone, IP camera, etc. (Fig. 11)

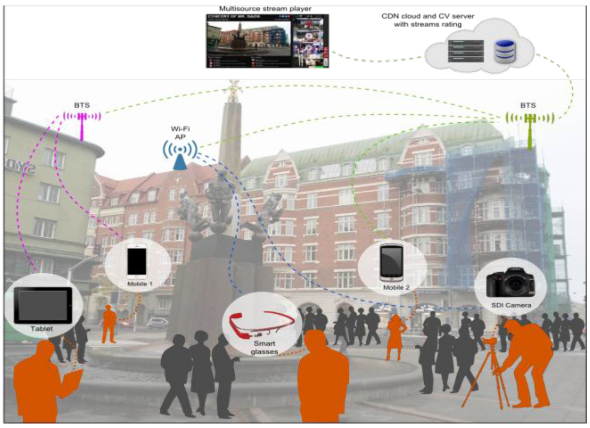

Fig. 11. Example of use case of proposed model

CONCLUSION

In this paper, the innovative model for automatic selection of the most suitable shot regarding the object of interest in the multi-camera system was proposed. Depth analysis of the current streaming and computer vision technology lead to a proposal of evaluation metrics for individual broadcast sources. The proposed evaluation metric was adapted to a different types of uses through changing the value of weights of individual metric components. Overall, the evaluation of multi-camera system sources is derived from several factors: the current position and size of the reference object in the shot, the parameters of the video broadcast and audio track, lastly also the objective and subjective evaluation by viewers. Described evaluation metrics became the basis for the establishment of a system for the selection of the most suitable shot and related switching of the output streaming source in a real-time. The created model was experimentally verified in a number of alternative implementations using standard and mobile devices. The results compared with the standard manual director showed the advantages of model in the form of faster response time to changes in a shot and a tendency to capture the object of interest in the image when being closest to the ideal composition. Experimental testing has shown the ability of adaptation the model for different types of applications, such as broadcast sporting events, security systems, industrial monitoring or conference broadcasts.

REFERENCES

[1] CHU C. et al..: Tracking Across Multiple Cameras with Overlapping Views Based on Brightness and Tangent Transfer Functions, Distributed Smart Cameras (ICDSC), 2011 Fifth ACM/IEEE International Conference on. IEEE, 2011.

[2] ZHU L. – HWANG J. – CHENG H.: Tracking of multiple objects across multiple cameras with overlapping and non- overlapping view, IEEE Intl. Symposium on Circuits and Systems, s. 1056-10360, Taipei, 2009

[3] KHAN S. – SHAH M.: Consistent labeling of tracked objects in multiple cameras with overlapping fields of view, IEEE Trans. Pattern Analysisand Machine Intelligence, vol. 25, no. 10, s. 1355-1360, 2003

[4] ZAMBANINI, S. - BLAUENSTEINER, P. - KAMPEL, M.: Automated multi-camera surveillance for the prevention and investigation of bank robberies in Austria: A case study Crime Detection and Prevention ICDP, 2009

[5] O'DRISCOLL G. : Next generation IPTV services and Technologies, Hoboken: Wiley, 2008, ISBN:978-0470163726

[6] HAYAKAWA T. et al..: Management of Multimedia Data for Streaming on a Distributed e-Learning System, 26th International Conference on Advanced Information Networking and Applications Workshops, 2012

[7] ANDERSON T. : The theory and Practice of Online LEarning, second edition, AU Press : Edmonton, 2011

[8] CHEN D. M. et al..: Streaming Mobile Augmented Reality on Mobile Phones, ISMAR, 2009

[9] DHONSALE K.V. : Overview, implementation and comparison of Audio Video Standard (AVS) China and H.264/MPEG -4 part 10 or Advanced Video Coding Standard, 2012

[10] WIEGAND T. et al..: Overview of the H.264 / AVC Video Coding Standard, IEE Transactions on Circuits and Systems for Video Technology, 2003

[11] LIU Y - GUO Y - LIANG C.: A survey on peer-to-peer video streaming systems, Peer-to-Peer Netw Appl, 2008

[12] WANG Q. et al..: An Experimental Comparison of Online Object Tracking Algorithms, Proceedings of SPIE: Image and Signal Processing Track, 2011

[13] ROSS D. et al..: Incremental learning for robust visual tracking, International Journal of Computer Vision 77(1-3), s. 125–141, 2008

[14] COLLINS R. – LIU T. – LEORDEANU M.: Online selection of discriminative tracking features, IEEE Transactions on Pattern Analysis and Machine Intelligence 27, s. 1631–1643, 2005.

[15] ADAM A. – RIVLIN E. – SHIMSHONI I.: Robust fragments-based tracking using the integral histogram, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, s. 798–805, 2006.

[16] GRABNER H. – BISCHOF H.: On-line boosting and vision, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, s. 260–267, 2006.

[17] GRABNER H. – LEISTNER C. – BISCHOF H.: Semi-supervised on-line boosting for robust tracking, Proceedings of European Conference on Computer Vision, s. 234–247, 2008.

[18] STALDER S. – GRABNER H. - VAN GOOL A.: Beyond semi-supervised tracking: Tracking should be as simple as detection, but not simpler than recognition, Proceedings of IEEE Workshop on Online Learning for Computer Vision,2009.

[19] MEI X. - LING H.: Robust visual tracking using l1 minimization, Proceedings of the IEEE International Conference on Computer Vision, s. 1436–1443, 2009.

[20] BABENKO B. – YANG M. – BELONGIE S.: Visual tracking with online multiple instance learning, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, s. 983–990, 2009

[21] KWON J. - LEE K.: Visual tracking decomposition, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 1269–1276, 2010.

[22] KALAL Z. – MATAS J. – Mikolajczyk K.: P-N learning: Bootstrapping binary classifiers by structural constraints, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 49–56, 2010.

[23] KALAL Z. - MIKOLAJCZYK K. - MATAS J.: Forward-Backward Error: Automatic Detection of Tracking Failures, International Conference on Pattern Recognition, Istanbul, Turkey, ICPR 2010

[24] NEBEHAY G. - PFLUGFELDER, R: Clustering of Correspondences for Deformable Object Tracking, Computer Vision and Pattern Recognition, IEEE, 2015

D. Cymbalak received his M.Sc. and Ph.D. degree in Informatics from Technical University in Kosice (Slovakia). He is a member of University Centre for Innovation, Technology Transfer and Intellectual Property Protection on Technical University of Kosice. He is also member research institution - Computer Networks Laboratory at Technical University of Kosice. His research includes telepresence technologies, video streaming solutions, web and cloud services and computer vision.

F. Jakab (IEEE member) graduated from the Faculty of Computer Science and Electrical Engineering at the St. Petersburg Institute of Electrical Engineering in the field of System Engineering (Russian Federation). He established Computer Networks Laboratory (www.cnl.sk, in 1995) at TUKE. Main areas of his research activities: computer networks, new form of multimedia based communication (videoconferences, IP streaming). He coordinated important educational initiatives and projects linked to cooperation with practice, international projects in the Leonardo, Tempus, the European Commission Framework Programmes. He was at the birth of a global educational initiative Cisco Networking Academy. He stands behind the success of the Cisco Networking Academy global initiative in the region of the Eastern Europe. He is a head of the Application Section of the Communication Technology Forum Association in Slovakia, director of the University Centre for Innovation, Technology Transfer and Intelectual Property Protection at Technical university of Kosice. In 2006 he was awarded a prestigious award "IT Personality of the Year" in Slovakia. He has participated in the establishment of The Kosice IT Valley Association.

M. Michalko received his M.Sc. and Ph.D. degree in Informatics from Technical University in Kosice (Slovakia). For more than 10 years he is member of well recognized research institution - Computer Networks Laboratory at Department of Computers and Informatics at Technical University of Kosice (DCI TUKE). Now he is an Assistant Professor at DCI TUKE and his lectures are focused on Computer Networks. His research includes multimedia content delivery, video streaming services, web and cloud services, innovative teaching&learning techniques and IoE/IoT solutions.

O. Kainz was born in 1988. In 2013 he graduated (MSc. in Applied informatics) from the Technical university in Kosice, Slovakia. 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 computer vision, e-learning, human-computer interfaces, computer graphics, computer networks, biological engineering and body area network.

R. Vapenik graduated (MSc. in Informatics) from the Technical university in Kosice, Slovakia. 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 streaming technologies, computer networks and multimedia content delivery.