

# DIRECTING YOUR OWN LIVE AND INTERACTIVE SPORTS CHANNEL

*Stefan Poslad<sup>1</sup>, Aristodemos Pnevmatikakis<sup>2</sup>, Mario Nunes<sup>3</sup>, Elena Garrido Ostermann<sup>4</sup>, Paul Chippendale<sup>5</sup>, Peter Brightwell<sup>6</sup>, Charalampos Patrikakis<sup>7</sup>*

<sup>1</sup>Queen Mary University of London, <sup>2</sup>Athens Information Technology, <sup>3</sup>INESC Inovação Instituto De Novas Tecnologias, <sup>4</sup>Atos Origin, <sup>5</sup>Fondazione Bruno Kessler, <sup>6</sup>British Broadcasting Corporation, <sup>7</sup>National Technical University of Athens

## ABSTRACT

The ability to mark-up live sports event content, viewed from multiple camera angles, such that athletes and other objects of interest can be tracked, facilitates an exciting new personalised and interactive viewing experience for spectators, enabling spectators to act as directors of their own customised live sports videos. In this paper, such an approach is described as part of the My-e-Director 2012 project. The design of this platform is described here and a discussion of a prototype system is given.

## 1. INTRODUCTION AND MOTIVATION

Large Scale Broadcasting events such as the Olympic Games provide a first class forum for research, development and demonstration of novel large scale live broadcast audio-video-information services across multi-venue, multi-event heterogeneous environments with many actors and high-activity scenes. Athens 2004 and Beijing 2008 are prominent manifestations of this trend, and the London 2012 Olympics is expected to continue this trend.

Professional types of live event video content delivery often tend to record multiple viewpoints of live-events but to only broadcast one at any one time. Only basic user interaction such as pause and resume of video delivery is supported. The EU FP7 My-e-Director 2012 project (<http://www.myedirector2012.eu/>) is creating a video delivery platform to offer viewers much greater choice and freedom to interact with live sports video content delivery. It proposes to do this in several innovative ways based upon multi-camera selection, person and object tracking, video content distribution over heterogeneous ICT infrastructures and personalisation.

My-e-Director 2012 puts the spectator in the director's seat and provides the power to enable them to choose amongst a wide range of camera streams, tracked objects and athletes, to use personal preferences, and for content to be delivered over a range of network infrastructures and access devices.

The outline of this paper is as follows. Next, an analysis of the state of the art is given. Third the design and

implementation of the platform is presented. Finally, conclusions and further work are presented.

## 2. SURVEY OF RELATED WORK

This survey of related work is partitioned according to the main areas of research and development in the project: multi-camera selection, person and object tracking, video content distribution and personalisation

### 2.1. Multi-Camera Selection

The Automated Coverage project [1] developed technologies to capture appropriate content from a scene by detecting and responding to events in the real world. The PolyOmaton, Greek for "Many Eyes", System [2] is an ambient multi-camera director which constantly monitors an indoor environment. The main innovation of My-e-Director 2012 is to provide a more dynamic mechanism for camera selection attuned to person tracking, including face detection and identification, that can be effective in much more challenging environments, including out-door venues, high activity scenes, with more complex sensor set-up and involving many actors.

### 2.2. Person and Object Tracking

State of the art sports tracking can be divided into two main dimensions: active versus passive, supported by GPS or RFID based tracking versus vision based systems respectively; the type of objects identified and tracked, e.g., body, face or text labels. My-e-director 2012 will focus on vision based approaches. Tracking is based upon detecting and filtering movement across frames. Algorithms are based upon Kalman filters, particle filters [3], CONDENSATION [4] and modelling of image pixels [5]. Particle filters seem promising for use with sports events and have been initially deployed more for indoor sports [6]. Face detection is a solved problem for still images but for real-time outdoor sports use, one needs to resort to a tracker to narrow down the search region [7] and then to use skin colour models to detect the face and eyes and to geometrically normalize this [8].

The My-e-Director 2012 innovation is to develop and apply person, including face tracking algorithms, for use in outdoor sports events which have more variable background environments and less ability to optimise camera positions for tracking. In addition, My-e-Director 2012 tracking algorithms must deal with the challenging problem of transient body and face rotation, distortion and occlusion of athletes.

## 2.2. Content Distribution and User Access

User access to video broadcast streams will be based upon DVB-T (Digital Video Broadcasting – Terrestrial), a DVB European consortium standard for broadcast transmission of digital terrestrial television. A variant of this (DVB - Handheld) uses MPEG-2 and H.264 encoding for accessing broadcast video on handheld receivers. There are several projects that enhance multimedia content through adaptive and interactive content distribution and access. The DANAE project has investigated dynamic and distributed adaptation of multimedia content [9]. mCDN focuses on Multimedia Content Discovery and Delivery [10]. The MEDIANET Project [11] concerns a common open delivery platform for multimedia content. The PHOENIX project concerns advanced multimedia services for mobile devices [12]. The main innovation of My-e-Director 2012 is to support scalable, real-time, personalised, interaction based upon tracked objects and multi-camera selection of multimedia content.

## 2.3 Personalised Interactive Viewing

Many projects have researched and developed personalised and interactive broadcast multimedia content delivery and access. ELIN focuses on personalisation of newspaper content. ePerSpace focuses on building unified virtual personal environments and mCDN [10] focuses on personalisation to enhance the content retrieval process. IST-LIVE [14] supports the concept of a future Video Conductor which supports user participation in real-time by allowing users to feedback to a director their choices for which parallel Olympic sports event they wish to view, enabling these to form ‘non-linear stories’ of events. Although IST-LIVE uses a similar concept to My-E-Director 2012, that of a director, the director concepts is used very differently. IST-Live seeks to support manual viewer feedback to schedule future events whereas My-e-Director 2012 focuses more on automated personalised viewing with respect to tracked objects of interest and multi-cameras during live sports events that match personal profiles, and on the use of multicasts in addition to broadcasts.

## 3. MY-E-DIRECTOR 2012 FRAMEWORK

My-e-Director 2012 adopts a concurrent engineering development process model [16], a variant of the spiral development model. The spiral model supports rapid iterative prototyping of a complex system as a linear sequence from initial system model to final system, e.g., the terminal application component is developed incrementally from a basic standalone PC version AV player to a customisable interactive AV player version, to a networked version, to a Web version, to a mobile version.

The concurrent development model consists of activities, tasks, and their associated states. Some of these activities can occur in parallel with states being used to synchronise sequential and parallel activities. For example, several engineering tasks occur in parallel, including athlete tracking, camera selection, streaming server and personalised access.

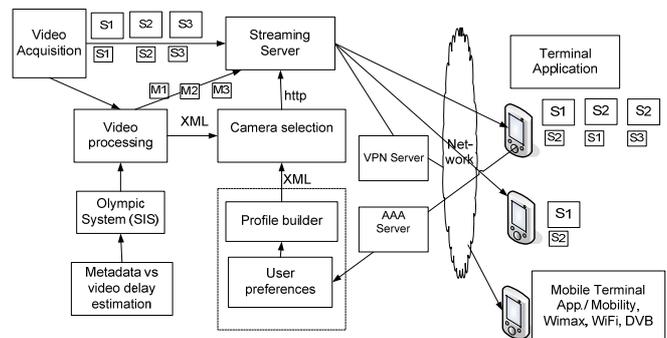


Fig. 1 My-e-Director 2012 System Architecture

The system (Figure 1) consists of the following main components: video acquisition, video processing, camera selection, security, metadata generation and streaming, personalisation and terminal applications. Each of these is discussed in turn. The video acquisition receives the camera inputs and generates uncompressed video. Using, video encoders, the raw video is encoded in different formats and bitrates, e.g., suitable for SDTV and mobile devices. Video processing supports tracking analysis and synchronisation of video streams with metadata streams which describe events. The video streaming server generates the video streams for different users, both in unicast and multicast modes. Camera selection is triggered by detected sports incidents and by group personal profiles. The security sub-system supports both VPN and AAA Service based access. Metadata behaves as the system glue to support interoperability. In the personalisation sub-system, user preferences are gathered from access terminals and aggregated to form group profiles. The core user terminal application is based upon a custom built audio-video player

which supports sport event metadata and personal profile driven interaction.

### 3.1 Athlete and Objects Tracking

Although there are many problems associated with the extraction of useful content from video images, nevertheless video provides us with a rich source of information ranging from low-level features such as environmental context, e.g. track, field, crowds, etc., to more complex content such as which athletes are visible, where they are in 3D space and in which direction a camera is pointing. It is often necessary to analyse video feeds purely from television cameras that cover an event. These may not always be best placed or suited for video processing, e.g. they only offer a wide-angle view, are situated on the shoulder of a moving cameraman or are moving on a rail. The un-calibrated and moving nature of such cameras makes 3D video tracking difficult. The automatic estimation of camera motion and orientation parameters can help us in this respect. Such a rough estimate of camera usage can be sufficient for more complex 3D understanding. Person identification can be based upon the 2D detection of faces, text found on clothing, simple shape and colour models of athletes, etc, or in 3D space using the information from multiple cameras covering the same scene from multiple points of view.

### 3.2. Multi-Camera Selection

This subsystem utilizes information from low-level video processing to understand the scenes in the various cameras, extracting events of interest for the viewers. These incidents relate to: the beginning and end of sports events, individual attempts, level of competition and to the presence of specific athletes. Incidents are mapped to user preferences, resulting in automated proposals for camera switching, between simultaneous sports events, or within an event, to better cover the incidents of interest.

### 3.2. Video Content Distribution

In [17] different architectures were proposed in order to support personalized and user centred media streaming, which were named terminal centric, network centric and hybrid scenarios. In the terminal centric scenario the best camera selection is performed at users' terminals. This approach has the advantage of having a simple and scalable architecture that could use a standard video streaming server already deployed in many video streaming solutions. Furthermore, the user has full control over the selection of the camera, which can be done manually and upon demand. On the other hand, it requires special applications to be deployed in users' devices and there is the possibility of high stream and camera switching delays since the

selections of users need to be sent to servers in order to perform stream switching.

In the network centric scenario, the terminal performs the minimum functionality since the best camera selection is performed centrally at the content provider servers. This approach has the advantage that simple standard terminals can be used for accessing the service, without the need for special software, while controlled advertising (e.g. between clips) can be easily implemented. On the other hand, it is not so scalable as the decisions for stream switching are performed centrally, resulting in excessive demands for processing power for large numbers of users. Furthermore this centrally processing of user preferences and best camera selection leads to requirements for a special design of streaming server.

The hybrid scenario distributes the processing to enable stream personalisation and best camera selection between the server and terminals. In this way, they combine some of the main advantages of both scenario groups, offering a scalable solution that introduces fewer requirements for the introduction of new servers and development of new applications on terminals. It should be noted however, that the hybrid scenario does not offer the same level of functionality and platform capabilities as regards user group type personalisation.

The design decision of which of the three aforementioned architectures is best, should take into account the complexity of the final solution, the amount of investment that can be made and the importance of personalisation level that would be required by users. Following this approach, in the context of my-eDirector2012 project, after a comprehensive study of user requirements based on previous surveys by broadcasters[18][19] and a project initiated survey[20], it was concluded that the hybrid scenario (Figure 1) presented the best trade-off in order to simultaneously achieve seamless stream switching, user preference flexibility and network scalability, matching the needs of the my-eDirector2012 platform requirements. Consequently this was the type of architecture selected for implementation in the My-eDirector demonstrators.

### 3.3. User Terminal and Personalisation

The initial system interface is driven by a task analysis of: general AV player interaction such as pause and resume live video streaming; generic metadata driven interaction such as displaying and clearing sports venue information; specific incidents in sports events, e.g., by a triple jumper overshooting the start board in that event.

User interface task models are implemented using Microsoft's Windows Presentation Foundation (WPF) Application Programmer's interface (API). One of the reasons for using WPF is that it supports declarative user interface expressed in Extensible Application Markup

Language (XAML) enabling user interfaces to be more easily integrated into cross-platform Web interface style applications, e.g. Microsoft Silverlight. Although, this must be balanced against some restrictions in the use of real-time video streaming support that Web applications may offer. Applications use a Microsoft Windows operating system platform, video streams transmitted using the real time streaming protocol (RTSP) and encoded in MPEG-4 format. Figure 2 shows a prototype system UI for the objects of interest selection task.



**Fig. 2** Screenshot of objects of interest selection

Individual personal profiles are created and maintained via a mix of both explicit user input and through observations of user interaction. For example, users can explicitly rank preferences for which nationalities of athletes and types of sports they are interested in. Sports specific preferences may also be related to user preferences for viewing of incidents in sports events. In addition, in the initial prototype, three video streams will be offered to the user terminal, one on the main screen and two on smaller side streams. The side streams will be 'recommendations' output from the profile builder part of the personalisation sub-system and may be sports incident related. Depending on how users react to the recommended side streams, either by choosing to view a side stream in the mainstream, or to signal fixing a side-stream, or ignoring it, the individual personalisation model will be updated.

### 3.4. Metadata Generation and Streaming

Within the My-e-Director 2012 system, metadata enables sub-systems to interoperate, to be dynamically configurable by users and for processes to be orchestrated. Several major types of metadata can be specified about sports events, sports incidences, athletes, objects of interest, cameras, video frames and streams, user interface layouts and personal profiles. For example, a Sport Information System (SIS) creates and gathers metadata about sports incidents which is synchronised to the video streams. This information will be gleaned from commercially available services such as an Internet Data Feed Service [21] provided by the Organizing Olympic Committee to distribute the results and statistical data from the games to the official website, broadcasters and to other Internet customers. This

is enriched by combining it with the tracking outputs from the visual processing and the camera selections.

## 4. CONCLUSION AND FUTURE WORK

The My-e-Director project is developing a framework to enable much more personalised and interactive viewing of live, multicast and broadcast, video streams of sports events with respect to tracked objects of interests, incidents and camera angles. At the end of its first year of its three year span, My-e-Director has completed a comprehensive user requirements analysis, analysed the core design choices, specified metadata interfaces to support component and process interoperability and developed and unit tested simple sub-system prototypes. The next phase will focus on advancing the prototype by integrating the core components into a complete system than can be tested and validated with users.

## 5. ACKNOWLEDGMENT

The My-e-Director 2012 Project is funded by the European Community, FP7. The authors wish to thank the members of the project consortium for their support and guidance.

## 6. REFERENCES

- [1] (BBC, 2006) Automated Coverage, BBC research:<http://www.bbc.co.uk/rd/projects/virtual/automated-coverage/index.shtml>
- [2] N. Azodolmolky, V. Dimakis, G. Mylonakis et al., 'Middleware for In-door Ambient Intelligence: The PolyOmaton System. Proc. 2nd NGNM workshop, Networking Waterloo, Canada, 2005.
- [3] D. Fox, J. Hightower, L. Liao et al, "Bayesian Filtering for Location Estimation". *IEEE Pervasive Computing*, vol. 2, no. 3, pp. 10-19, 2003.
- [4] M. Isard, A. Blake, "CONDENSATION - Conditional density propagation for visual tracking", *Int. J. Computer Vision*, Vol. 29, No. 1, pp. 5-28, 1998.
- [5] C. Stauffer, W.E.L. Grimson, Learning patterns of activity using real-time tracking. *IEEE Trans. PAMI*. Vol. 22, No. 8, pp. 747-757, 2000.
- [6] R. Brunelli, A. Brutti, P. Chippendale et al, "A Generative Approach to Audio-Visual Person Tracking," in CLEAR'06 Evaluation Workshop, Southampton, UK, 2006.
- [7] A. Pnevmatikakis, L. Polymenakos, V. Mylonakis, "The AIT Outdoors Tracking System for Pedestrians and Vehicles," in CLEAR'06 Evaluation Workshop, Southampton, UK, 2006.
- [8] R. Stiefelhagen, J. Yang, A. Waibel, "Tracking eyes and monitoring eye gaze," in *Workshop on Perceptual User Interfaces*, Banff, Canada, 1997.
- [9] A. Hutter, P. Amon, G. Panis et al, "Automatic adaptation of streaming multimedia content in a dynamic

- and distributed environment,” in *Int. Conf. Image Processing, ICIP 2005*, vol. III, pp. 716-719.
- [10] J. Sokol, K.-P. Eckert, “MCDN: Multimedia Content Discovery and Delivery,” in *Proc. 8<sup>th</sup> Int. Symp. on Autonomous Decentralized Systems*, 2007, pp 411-420.
- [11] S. Travert, M. Lemonier, “The MEDIANET Project,” in *Proc. WIAMIS*, 2004, Lisbon, Portugal.
- [12] E. Roddolo, G. Panza, C. Lamy-Bergot et al, “Joint source and channel (de)coding in 4G networks: the PHOENIX project,” in *Proc. WPMC 2004*, Padova, Italy.
- [13] J. Casademont, F. Perdrix, M. Einhoff et al, “Elin: A newspaper universal multimedia access platform based on MPEG standards,” in *Proc. 5<sup>th</sup> IASTED*, 2005, pp.27-32.
- [14] P-Y. Danet, K. Madani, A. Kapovits, “A Global Platform for Seamless Personalised Services at Home & Elsewhere - the ePerSpace approach,” in *16th Biennial Conf. ITS*, 2006, Beijing, China.
- [15] M. Williams, R. Wages, “Video Conducting the Olympic Games 2008: The iTV Field Trial of the EU-IST Project LIVE,” in *Proc. 3<sup>rd</sup> ACM Int Conf. DIMEA*, 2008, Athens, Greece.
- [16] A.M. Davis, P. Sitaram, “A concurrent process model of software development,” *ACM SIGSOFT Software Engineering Notes*, 1994, Vol. 19, No. 2, pp. 38 - 51.
- [17] M. Nunes, C. Patrikakis, N. Papaoulakis. “A Network Oriented Perspective on the Personalization of Media Streaming,” in *2nd IEEE Workshop on Enabling the Future Service-Oriented Internet, GLOBECOM*, New Orleans, 2008.
- [18] ESPN, Entertainment and Sports Programming Network, official web site, URL: <http://www.espn.com>.
- [19] BBC internal Audience Research Reports from Olympic Games Athens 2004: Audience Summary, Interest by Sport, and Interactive TV Review
- [20] My-e-Director 2012 project, “Deliverable D2.1- End-User Requirements”, 7/11/2008, URL: <http://www.myedirector2012.eu>.
- [21] AtosOrigin, Olympic Games Info Difussion System from Major Events:[http://www.atosorigin.com/en-us/olympic\\_games/services\\_solutions/information\\_diffusion\\_systems/](http://www.atosorigin.com/en-us/olympic_games/services_solutions/information_diffusion_systems/)