



**SEVENTH FRAMEWORK PROGRAMME**  
**ICT-1-1.5**  
**Networked Media**

*Specific Targeted Research Project*

**My eDirector 2012**  
(FP7-215248)

**My eDirector 2012 - Real-Time Context-Aware and  
Personalised Media Streaming Environments for  
Large Scale Broadcasting Applications**

**D7.4 Technology Evaluation Report**

Due date of deliverable: 04-08-2010

Actual submission date: 13-05-2011

Start date of project: 01-02-2008

Duration: 36 months

### Summary of the document

<b>Code:</b>	<b>D7.4 Technology Evaluation Report-V.1.4</b>
<b>Last modification:</b>	12/05/2011
<b>State:</b>	<b>Final version</b>
<b>Participant Partner(s):</b>	ATOS, ICCS, AIT, BBC, INOV, QMUL, FBK
<b>Author(s):</b>	Ch. Patrikakis, L. Argyriou, N. Papaoulakis, Ch.Androulaki, David Salama, E. Garrido, Aristodemos Pnevmatikakis, Rui Cruz, Paul Chippendale, Hannah Fraser, Stefan Poslad
<b>Fragment:</b>	No
<b>Audience:</b>	<input checked="" type="checkbox"/> public <input type="checkbox"/> restricted <input type="checkbox"/> internal
<b>Abstract:</b>	<i>This document describes the technology aspects of My eDirector 2012. The final decisions on the technologies that were finally selected and evaluated are stated, along with other commercial relative solutions.</i>
<b>Keywords:</b>	<ul style="list-style-type: none"> <li>• <i>Networking</i></li> <li>• <i>Streaming</i></li> <li>• <i>Detection</i></li> <li>• <i>Annotation</i></li> <li>• <i>Personalisation</i></li> </ul>
<b>References:</b>	D2.3: "Network Platform requirements". D2.4: "System Architecture and Specifications". D4.4: "Configurable Camera Selection Tables and Algorithms". D6.2: "Personalised Media Stream over Radio Access Technologies".

## Change history

Version number	Date	Changed by	Changes made
0.1	22/10/2010	Lemonia Argyriou, Ch. Patrikakis, N. Papaoulakis, Christina Androulaki,	Creation of ToC for discussion with partners.
0.2	27/10/2010	Lemonia Argyriou, Ch. Patrikakis, N. Papaoulakis, Christina Androulaki,, Paul Chippendale, Aristodemos Pnevmatikakis, David Salama	Modified TOC including enhancements
0.3	05/11/2010	Aristodemos Pnevmatikakis	Section 4.3.2
0.4	05/12/2010	Aristodemos Pnevmatikakis	Sections 3, 3.1, 5.2
0.5	01/03/2011	Rui Cruz	Section 2
0.6	02/03/2011	Paul Chippendale	Additions and changes to section 3
0.7	02/03/2011	Hannah Fraser	Additions to section 3
0.8	03/03/2011	E. Garrido	Section 2, 5.15
0.9	24/03/2011	S.Poslad	Sections 4.1,4.2,6.1
1.0	27/04/2011	Rui Cruz	Section 4.1.2
1.1	29/04/2011	Lemonia Argyriou, Ch. Patrikakis, N. Papaoulakis, Christina Androulaki	Input to sections 2,4 & 5 Minor changes to the whole document Final editing
1.2	02/05/2011	Lemonia Argyriou	Minor corrections
1.3	12/05/2011	Emre Colakoglu,	QA
1.4	12/05/2011	David Salama	Close document.

## Table of Contents

1	Executive Summary .....	6
1.1	Scope .....	6
1.2	Audience .....	6
1.3	Summary.....	6
1.4	Structure.....	6
2	Networking and Streaming .....	7
2.1	State of the art.....	7
2.1.1	IP networking & media streaming technologies.....	7
2.1.2	Broadcasting technologies.....	10
2.2	Technology evaluation .....	12
2.2.1	Use of networking and broadcasting technologies .....	12
2.2.2	Use of protocols and media streaming technologies.....	12
2.3	Conclusions and Future Work.....	12
2.3.1	Internetworking issues .....	12
2.3.2	QoS support.....	13
2.3.3	Seamless mobility .....	13
3	Sensing athletes and action .....	14
3.1	State of the art.....	14
3.1.1	Visual 2D Tracking.....	14
3.1.2	Visual 3D Tracking.....	15
3.1.3	Face Detection, Tracking and Recognition.....	15
3.1.4	Text tracking and recognition.....	16
3.1.5	Motion analysis using optical flow.....	16
3.2	Technology evaluation .....	16
3.3	Conclusions and Future Work.....	17
4	Personalisation.....	18
4.1	State of the art.....	18
4.1.1	User profiling.....	18
4.1.2	Device and network Adaptation .....	18
4.2	Technology evaluation .....	19
4.2.1	Creating user profiles.....	19
4.2.2	Capturing user preferences (implicit/explicit).....	19
4.3	Conclusions and Future Work.....	19
4.3.1	User profiles in Internet based media access.....	19
4.3.2	Use of ontologies .....	20

4.3.3	Personalisation in broadcasting.....	21
5	State of the Art in Commercial Solutions .....	23
5.1	Http Adaptive Streaming Solutions .....	23
5.1.1	French Open Tournament '09.....	23
5.1.2	Sunday Night Football .....	25
5.1.3	Yinzcam .....	26
5.1.4	Beijing 2008 – NBC Sports .....	28
5.2	Sports Media Annotation Solutions.....	28
6	Implementation tools and development platforms .....	29
6.1	Front End Development .....	29
6.1.1	Silverlight .....	29
6.1.2	Silverlight and Smoothstreaming.....	29
6.1.3	Implementation Tools .....	29
7	Bibliography and References .....	30

## **1 Executive Summary**

### **1.1 Scope**

The scope of this deliverable is to evaluate the technologies that were selected and used for the implementation of the My eDirector 2012 platform. It starts with a full statement on the technologies that have been implemented till this moment to provide relative services to those of My eDirector 2012. Furthermore, after evaluating these technologies, a statement on the conclusions occurred and the future work planned is provided. Moreover, other commercial solutions on the related services are being presented. The deliverable ends up with a description of the implementation tools that were used in My eDirector 2012 and the development platforms that were finally selected.

### **1.2 Audience**

This deliverable is public. All the technical partners from My eDirector 2012 project should read and agree on the information, schedules and decisions presented in this document.

### **1.3 Summary**

In this deliverable it is provided a description on the state of the art that corresponds to each of the technologies used for the implementation of the project. This state of the art is followed by an evaluation of these technologies. At the end, the conclusions and the future work that relates to the presented technologies, methods and techniques are stated. Furthermore, commercial solutions regarding annotation and streaming techniques are being presented. And, finally, there is a description on the implementation tools and the development platforms that the whole project was based in order to complete its provided services.

### **1.4 Structure**

Chapter 1 is this introductory chapter, containing general information about the deliverable, including a summary of the document, the targeted audience and structure of the deliverable

Chapter 2 is a description of the state of the art on the networking and streaming technologies, the evaluation of the technologies used for the project and the conclusions and future work.

Chapter 3 is a description of the state of the art on the methods for sensing the athletes, the evaluation of the technologies used for the project and the conclusions and future work.

Chapter 4 is a description of the state of the art on the personalised services, the evaluation of the methods and techniques that were used in the project and the conclusion and future work.

Chapter 5 is a state of the art on Streaming and Annotation Commercial Solutions.

Chapter 6 is a statement on the implementation tools and the development platforms that were used for the project.

Chapter 7 is a statement on the relative bibliography and references to this deliverable.

## 2 Networking and Streaming

In the scope of My eDirector 2012, quality aware multimedia streaming adapted to network conditions is achieved using DVB-H in wireless broadcast, multicast over wireless (WiFi, WiMAX) and HTTP Adaptive Streaming for global Internet (unicast) over different access network technologies, including wired (xDSL, LAN-Ethernet), wireless (WiFi, WiMAX) and mobile (3G).

### 2.1 State of the art

The ability to offer end users quality aware multimedia streaming adapted to network conditions, together with the ability to support sophisticated mechanisms for user preferences capture (like preferred events, artists, athletes or particular times within an event) and the ability to manage the streams both in terms of space (e.g., camera angle) but also in time (pause, rewind, slow motion), has given the tools upon which enhanced user control over the offered services can be based.

Seamless bitrate switching, transparent transitions between access network technologies during media streaming and automatic, context-aware adaptation of transmission parameters are the challenges behind the development of new media streaming architectures.

#### 2.1.1 IP networking & media streaming technologies

There are several adaptation techniques that can be applied to multimedia contents that flow over IP networks. But the efficiency of these adaptation techniques, rely on information that needs to be known before the adaptation technique takes in:

- Information about the characteristics of the client device (size of display, colour depth, buffer size and the hardware and software types).
- Information about the content (content buffer size, the minimum streaming bitrate, frame rates and compression formats, hardware and software requirements).
- Information about the access network to which the client device is connected (mobile or wired, network technology, bandwidth, jitter, packet loss, delays and channel variations).

From the perspective of an adaptation system there are three components to consider: the adaptation of the content, the adaptation to network conditions and the adaptation for an adequate quality of service (QoS) and of experience (QoE) for the end-user.

The HTTP Adaptive Streaming technology [30] is a technique that allows the increase of video quality experienced on streamed contents over IP networks. The technique uses a pure web mechanism (HTTP based) for streaming, consisting on the request by the client application (media player) of a series of small file chunks that comprise the media stream.

But the most important innovation of this approach is the capacity to provide multiple encoded bitrates/qualities of the same media content, allowing the client player to seamlessly and dynamically switch between the available bitrates, not requiring any action from the HTTP Streaming Servers in the bitrate switching process [31][32]. It is a client-side code that implements the strategies for the selection of the most appropriate bitrate at any moment in time.

For the Apple HTTP Live Streaming method [64] or Microsoft Smooth Streaming IIS [63], the multimedia source is segmented into “chunks” of a few seconds long (typical values are 2 seconds for the Smooth Streaming of Microsoft and around 10 seconds for the Apple). These “chunks” are cut along a video group of pictures (GOP) boundary (starting with a key-frame), have no dependencies on past or future “chunks”/ GOPs and are encoded to the desired delivery format (H.264). Each “chunk” can be decoded and played in the client side independently of other “chunks”.

The streaming adaptation heuristics at the client, can continuously measure download times, buffer fullness, rendered frame rates, network bandwidth, window size, and other factors in order to seamlessly switch, without any pause in the video or “buffering” message, to a lower quality stream if needed or a higher quality stream if suitable, by requesting higher or lower bitrates of the same content from the HTTP Streaming Servers. The resulting end user experience is reliability and robustness, consistent playback without stutter, buffering or congestion, while the quality of the stream is dynamically adapted to the existing networking and computing context at the point of consumption, i.e., at the end user de-

vice.

Even though the client uses HTTP for media transport and control instead of Real-time Transport Protocol (RTP)/Real-Time Streaming Protocol (RTSP), it implements all of the most important functionalities available in traditional RTSP streaming, including the personal video recorder (PVR) trick functions like pause, stop, play, replay, fast-forward and Rewind but without requiring support from the Streaming Server. The player has the knowledge of the program timeline from the playlist received, and so, is able to request, at any moment, the corresponding “chunks” in the timeline.

The architecture incorporates all the necessary functionalities and architectures to support personalised media streaming for live events and the need for large scale media distribution (over heterogeneous network architectures and across domains) by its seamless integration with web oriented content delivery networks (CDNs), allowing the media to get easily cached along the network and much closer to the end users, not requiring special protocols or routers setup [33].

The architecture also uses HTTP for personalisation and program channel information (Channel Suggestions/Recommendations) and associated metadata, offering camera selection, fine control over presentation of closed captions and other timed text content.

The HTTP Adaptive Streaming supports a flexible user experience in a transparent manner, as long as the end user device provides adequate networking capabilities (multiple active interfaces like Wi-Fi/WiMax, 3G, etc.), in terms of a continuous uninterrupted playback by allowing retrieval of the media stream during the switch between multiple active wireless interfaces (network portability), avoiding session reconnection. As the connection relies entirely on HTTP, there is no need for requesting session reconnections to the Streaming Server [35].

Another flexible feature is the terminal portability allowing a user to suspend/resume session playback, either with the same terminal or with a different terminal (with the same or with different capabilities such as screen size/resolution or access network attachments) and in a different moment in time. When a session is suspended, both the content streaming and the suggestions metadata are halted, ceasing the communication between the client and the servers until the session is resumed [35].

**Adobe's** adaptive streaming solution is a web-based service available to all devices running a browser with Adobe Flash plugin. The server stores different streams of varying quality and size and switches among them during the playback adapting to user bandwidth and CPU. The service is provided using the RTMP streaming protocol. The supported video codecs are H.264 and VP6 which are available in Adobe Flash plugin. The advantage of AFDS is represented by the wide availability of Adobe Flash plugin at the client side. [37]

**Apple** has recently released a client-side HTTP adaptive live streaming solution. The server segments the video content into several pieces with configurable duration and video quality. The server exposes a playlist (.m3u8) containing all the available video segments. The client downloads consecutive video segments dynamically choosing the video quality using an undisclosed proprietary algorithm. The video codec employed is H264 using MPEG-2 TS container. Apple HTTP Live streaming is available on any device running iPhone OS 3.0 or later (including iPad), or any computer with QuickTime X or later installed, with support for encryption and authentication. [37]

Different vendors have implemented adaptive streaming technologies. In the table below, we compare the latest adaptive streaming technologies from Microsoft, Adobe, and Apple [36].

	<b>Microsoft IIS Smooth Streaming</b>	<b>Adobe Flash Dynamic Streaming</b>	<b>Apple HTTP Adaptive Bitrate Streaming</b>
<b>Cost per Streaming Server</b>	Free download <sup>1</sup> + \$469 for OS <sup>2</sup>	\$4500 <sup>3</sup> + OS cost <sup>4</sup>	Cost of Web server <sup>5</sup> + segmenter OS <sup>6</sup>
<b>On-demand &amp; Live Streaming</b>	✓	✓	✓
<b>Live Streaming DVR<sup>7</sup></b>	✓	Pause & Seek	✓
<b>Streaming Protocol</b>	HTTP	RTMP	HTTP
<b>Scalability via HTTP</b>	✓		✓
<b>Edge Caches</b>	✓		✓
<b>Stateless Server Connection<sup>8</sup></b>	✓		✓
<b>Supported Platforms</b>	Silverlight, Xbox 360, other Smooth Streaming-compatible players, and iPhone OS 3.0 <sup>9</sup>	Flash Player 10, AIR	iPhone OS 3.0, devices running QuickTime X
<b>DRM Support for Live, VOD</b>	PlayReady	None <sup>10</sup>	None
<b>DRM ASP &amp; Encoder ISV Support</b>	✓		
<b>Interoperable DRM (DECE Approved)<sup>11</sup></b>	✓		

<sup>1</sup> IIS Smooth Streaming is part of the free [IIS Media Services](#) download for Windows Server 2008 and Windows Server 2008 R2 .

<sup>2</sup> Runs on any edition of Windows Server® 2008 or Windows Server 2008 R2 including Windows® Web Server, which has a list price of \$469.

<sup>3</sup> Assumes use of Adobe Flash Media Interactive Server to support Pause, Seek, Authentication, and higher scalability

<sup>4</sup> Requires Windows Server 2003 SP2, Windows Server 2008, Red Hat® Enterprise Linux® 4, or Red Hat Enterprise Linux 5.2.

<sup>5</sup> Runs on any Web server. Also requires the Apple streaming segmenter – see next note.

<sup>6</sup> The Apple stream segmenter is a utility that receives encoded MPEG2-TS and breaks it into 10 second “chunks” for delivery. This free download requires an Intel-based Mac, with a Mac Pro or an XServe having two Ethernet network interfaces recommended.

<sup>7</sup> Full DVR features include Pause, Seek, Fast Forward (e.g., 2x, 5x playback speeds), Fast Rewind, Go To Live, Instant Replay, and Slow Motion.

<sup>8</sup> A stateless (non-persistent) connection between server and client increases scalability and allows seamless failover or rollover between load-balanced servers.

<sup>9</sup> Support for adaptive streaming to the Apple iPhone from IIS Media Services 4 was [announced](#) on October 18th, 2009.

<sup>10</sup> Future support was [announced](#) Sept. 10, 2009. Proposed availability is H1 CY10 for delivery to a future version of Adobe® Flash® Player and Adobe AIR™.

<sup>11</sup> Digital Entertainment Content Ecosystem (DECE, LLC) is a consortium of major Hollywood studios, consumer electronics manufacturers and retailers, network hardware vendors, systems integrators and Digital Rights Management (DRM) vendors chartered to develop a set of standards for the digital distribution of premium Hollywood content.

<b>Real-time Client and Server Logging</b>	✓ <sup>12</sup>		
<b>Programmable Client Side Switching Logic</b>	✓	✓	
<b>Live In-Stream Ad Integration</b>	✓		
<b>Built-in Analytics Framework</b>	✓		
<b>Delivery to Mobile Devices</b>	✓		✓
<b>Native 64-bit Server Support</b>	✓		✓
<b>Media Container</b>	MPEG 4 – Part 12 (Fragmented MP4)	MPEG 4 – Part 12 (MP4), FLV	MPEG-2 TS
<b>Supported Video Codecs</b>	Codec Agnostic (currently supports VC-1 Advanced Profile & H.264 Baseline, Main, and High)	H.264 Baseline, Main, and High; VP6	H.264 Baseline Level 3.0
<b>Supported Audio Codecs</b>	Codec Agnostic (currently supports WMA & AAC)	AAC, MP3	MP3, HE-AAC, AAC-LC
<b>Maximum Bit Rate</b>	No limit	No limit	1.6 Mbps
<b>Default Fragment Length</b>	2 seconds	n/a	10 seconds
<b>End-To-End Latency</b>	As low as 1.5 seconds (configurable)	6 seconds	30 seconds <sup>13</sup>
<b>File Type on Server</b>	Contiguous	Contiguous	Fragmented
<b>Client Programming Platform</b>	Microsoft .NET Framework	Adobe Action-Script	Objective-C

### 2.1.2 Broadcasting technologies

Though the use of unicast transmission is dominant in wired networks, the shared media in wireless and mobile access (air) make the use of multicast and broadcast more appealing and far more effective than in wired networks.

#### State of the art

The broadcast network infrastructure of My eDirector 2012 project utilises the DVB-T (EN 300 744) standard, as this technology was the state of art at the start of the project. On the other hand the DVB-T technology fulfils all the technical requirements of My eDirector 2012 [59] as such many features like different picture formats defined by the combination of size, aspect ratio even HD content like 1280 × 720 pixels in progressive scan mode (abbreviated 720p) or 1920 × 1080 pixels in interlace mode (1080i), each of these utilises a 16:9 aspect ratio. In H264/AVC format the 720p format demands a bandwidth of 8 Mbps and for the 1080i 10 Mbps. The maximum provided bandwidth from the DVB-T is 24.1 Mbps.

<sup>12</sup> Real-time logging for Silverlight applications using the [IIS Advanced Logging](#) extension.

<sup>13</sup> Encoded live streams are processed by the Apple stream segmenter. This intermediate step results in increased latency for live delivery.

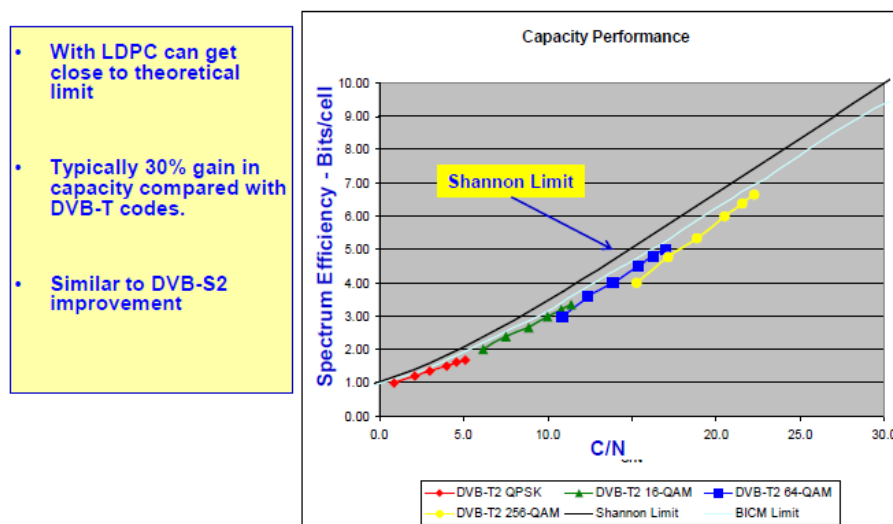
During the 3 years of project, there was a rapid evolution on the DVB technology which is the leading digital broadcasting technology within Europe and not only. Particularly the new standard is the DVB-T2 (EN 302 755) that promises enhanced features like the increased capacity and signal robustness. More specific the DVB-T2 could provide more services as it has bigger bandwidth capacity up to 40.2 Mbps, which is 66% more bandwidth than the legacy DVB-T with the same RF robustness.

In the following Table 1, we could see the difference of these two protocols and especially the bouquet capacity in Mbit/s.

	DVB-T	T2
Modulation	64QAM	256QAM
FFT size	2K	32K
Guard Interval	1/32	1/128
FEC	2/3 CC + RS (8%)	2/3 LDPC + BCH (0.3%)
Scattered Pilots	8%	1%
Continual Pilots	2.6%	0.35%
P1/P2 overhead	0%	0.7%
Bandwidth	Standard	Extended
Capacity	24.1 Mbit/s	40.2 Mbit/s

**Table 1: Comparison Capacity DVB-T / DVB-T2**

In the following diagram depicted in Figure 1 we could see that the DVB-T2 technology has reached the theoretical maximum capacity that determines the 2<sup>nd</sup> law of Shannon, which means that better radio coverage could be provided with less power RF transmitted power for the same number of carried TV channels.



**Figure 1: DVB-T2 Capacity Performance**

Moreover in contrast with the utilised content encoding of the DVB-T ( MPEG2 or MPEG4 and more specifically H264/AVC) the DVB-T2 technology will be based only on the H264/ AVC encoder. The transport stream remains the same and it is the MPEG2-TS (ETSI TS 101 154) as any change on this standard and method of streaming will change all the internal network and media equipment of the broadcasters.

**Benefits and changes**

However, in order to benefit from the best technical performance and the highest efficiency, countries could also make direct transition into DVB-T2. The considerations for choosing DVB-T or DVB-T2 can only be judged on a case-by-case basis, but generally depend on the actual ASO date as well as on the

services to be offered. DVB-T2 should be the system of choice for services with high data rate requirements like e.g. HDTV and 3DTV. In addition DVB-T2 is the ideal solution when different service characteristics need to be supported in one multiplex such as e.g. high data rate for HDTV via roof top antennas together with transmission to portable TVs with indoor antennas.

The alterations on any involved equipment are minor as from the side of the broadcasters, the only upgrade on the transmissions sites has to be, the change of the modulator to a correspondingly DVB-T2 capable. All the other components like RF power amplifiers and antennas systems remain the same. This means that the overall cost of the upgrade on the transmitting systems is around 5-10% to the overall CAPEX of the RF network infrastructure which is a minor cost in contrast with the capacity and radio coverage benefits that will have. Additionally the network infrastructure for transferring the media content from TV station master control to the transmission sites remain the same except the requirements for bigger throughput which is something obvious and already supported. From the side of the end user there is a need for new STBs, or install of DVB-T2 tuners on multi interface STBs. The cost of the CPE will not be high as according to the Freeview (leading STB manufacture) the sales of HD (DVB-T2) devices in the UK has passed the one million milestone and now stands at 1.2 million devices. This includes regular set-top boxes, PVR set-top boxes and integrated TV tuners (iDTV). Today, there are more than 150 models of Freeview HD boxes and TVs available to buy starting from £55.

### **Changes within My eDirector 2012**

Above we describe the demand for equipment changes either on the broadcast or end user side. Regarding the My eDirector 2012 platform from the side of the broadcasting system we have to replace the utilised modulator which is the DecTek DTA-110 [60], [61] to the DTA-115 which supports the new DVB-T2 standard. From the side of the end user and the My eDirector 2012 STB which is the Dream Box DM-800s the only change is the RF tuner with the DVB-T2 compatible.

## **2.2 Technology evaluation**

### **2.2.1 Use of networking and broadcasting technologies**

The use of broadcasting technology and infrastructures, though it minimises the use of the network resources in any case but on the other hand provides the lower level of personalisation capabilities as are based on unidirectional communication. Also it cannot be used for all the available streams but can be applied to specific technologies (i.e., DVB-H). The use of multicast, though it can be deployed for a larger number of streams than broadcasting, suffers from limitations imposed by network providers. The advantages of unicast when compared to the other transmission modes is counterbalanced by its limited efficiency in network resource utilisation, especially for large events that interest large audiences such as the Olympic Games [34].

### **2.2.2 Use of protocols and media streaming technologies**

The HTTP Adaptive Streaming architecture is capable of providing video streaming to end users, connected either directly or through content delivery networks (CDNs) over several access network technologies, enabling an uninterrupted streaming experience, guaranteeing, for each user, the best possible video quality reception. With this architecture, multimedia sources (video and audio) can be encoded in various bitrates, and a chunked HTTP progressive download method provides the capability to react to both bandwidth and local terminal conditions in order to seamlessly switch the video quality that the user is receiving, maximising the quality of experience (QoE).

## **2.3 Conclusions and Future Work**

### **2.3.1 Internetworking issues**

An important issue in media streaming is the transition between different transmission modes, namely unicast, multicast and broadcast. Since the streams in each mode may not be synchronised (even if they are, buffering can lead to significant time-shifting) the problems of jumping to another point in the video stream or interruption can still happen. Use of HTTP Adaptive Streaming does not solve the problem in this case, as it cannot be deployed for multicast or broadcast, making it harder to be considered a universal solution.

### **2.3.2 QoS support**

The HTTP Adaptive Streaming, in terms of QoS support, uses several strategies, namely a Network Quality adaptation strategy a Host Capabilities adaptation strategy and a Time Control strategy. The Network Quality strategy is used to determine the best bandwidth to use, maximising the network conditions at the client and keeping the buffer fullness within lower and upper boundaries. This strategy is re-frained by the Host Capabilities strategy that analyses the local conditions of the client device to limit the available bitrates to the ones that can be supported at any time. The Time Control strategy tries to achieve a rather smooth experience over time, minimising the changes between bitrates and avoiding oscillation conditions.

The appropriate buffering and the QoS-aware Adaptive Streaming approach, with content adaptation according to host environment and network characteristics, provides the optimum video quality experienced when compared to conventional streaming protocols, even under difficult network conditions like low throughput, big packet delays, jitter or big number of dropped packets.

### **2.3.3 Seamless mobility**

Two important features implemented in the architecture are the network portability and terminal portability capabilities, allowing a flexible user experience in terms of either a continuous uninterrupted playback during network interface transition with the same terminal or a suspend/resume session playback, either with the same terminal or with a different terminal (with same or with different capabilities like screen resolution or access network attachments).

### 3 Sensing athletes and action

In the scope of My eDirector 2012, athletes' presence and actions are sensed using video signal processing modules operating on the cameras feeds from broadcaster(s). There is an optional extra component of the My eDirector 2012 system of an active tracking system. This active tracking system could one of many technologies. GPS, Radar and Real Time Localisation Systems (RTLS) were considered during the period of the project. The RTLS system was investigated in more detail as this particular technology would be the best technology for the athletics scenario and many other Olympic events. As the final trials are track and field athletics the RTLS option was optimal for the project to investigate. The final trial did not implement this optional component, as despite being technically possible, acceptance of these tags at their current size will not be possible for track & field athletes. There is a trend for these tags to get smaller and for sporting bodies to become more accepting of technologies.

#### 3.1 State of the art

One of the main challenges faced in personalising an athletics broadcast, is the construction of video analysis modules to simultaneously process all of the raw-feeds, subsequently producing metadata that can enable an automated system to reason about important 'incidents', based on viewers' specific interests.

The automatic localisation and recognition of athletes [20] can provide users with a variety of features, such as smart-zoom or the delivering of optimum, un-occluded views of athletes. Understanding camera usage is equally important for extracting context, as a cameraman's actions can provide clues about content. For example, a static camera could indicate that a race/event is about to begin, whilst the onset of panning could indicate the beginning of the event.

##### 3.1.1 Visual 2D Tracking

For visual tracking, evidence is collected to support tracking hypotheses by means of measuring a wide variety of image properties. The different properties correspond to visual cues like colour, texture and shape. There are many different cues to measure in the video frames, each offering advantages and disadvantages. A perfect cue would be able to:

- **Discriminative foreground perfectly from the background**, to reduce ambiguity.
- **Robust to the forms in which a target is viewed**: The image of the target can change due to many factors: Illumination changes as the target moves relative to the light sources, or the sources themselves change. Pose changes as the target rotates relative to the cameras. Finally, shape changes as the target flexes (or in the case of human faces, changes expression).
- **Be persistent**, i.e. detectable in the majority frames in which a target should be visible.

Unfortunately there is not a single cue that possesses all of these qualities: **Colour** is persistent and quite robust to the appearance of the target, but may not be discriminative from the background. **Target outline** is discriminative and persistent (when the outline model describes the object adequately) but target flexing can render the learning of an outline model very difficult, and the application of the model very complicated. **Motion** is again discriminative, especially in fixed camera setups, but it is not persistent as the target can spend long periods almost immobile. Finally, **dedicated target detectors** (like face detectors) can offer varying degrees of discrimination, robustness and persistence, depending on their misses and false positives.

Multiple targets can be addressed either independently through the deployment of multiple trackers, or jointly using approaches like Multiple Hypotheses Tracking, Joint Probabilistic Data Association [14], or Approximate Bayesian. A successful example of the latter can be found in [9], where the update in a probabilistic framework (particle filtering) can be performed more efficiently than in previous approaches: the computational complexity grows at most quadratically with the number of targets rather than exponentially. This makes the system run robustly in real-time with 7 occluding targets.

**Target initialisation** is a difficult issue in visual tracking. This refers both to how the initial model of the person is acquired, and to how it is efficiently updated throughout the lifetime of the track. Typically adaptive background segmentation techniques and face detectors are used for initialisation, both being quite problematic for crowded scenes.

### 3.1.2 Visual 3D Tracking

The 3D tracking of people in a room with multiple cameras typically follows one of two distinct approaches, as illustrated in Figure 2:

- A model-based approach:** A 3D model of the tracked object is maintained by rendering it onto the camera views, searching for supporting evidence per view, and based on that, updating its parameters. An important advantage of the model-based approach is that rendering can be implemented in a way that mimics the real image formation process, including effects like perspective distortion and scaling, lens distortion, etc. In the context of multi-body tracking this is particularly advantageous, since occlusions can be handled at the rendering level. In this way, updates are achieved through the search for supporting evidence only in the image parts where the different models are visible, thus occlusions are handled in a systematic manner [9]. The disadvantage, of course, is that the person models have to be initialised and occasionally updated, which in some situations may be tricky (e.g. multiple people entering simultaneously into the monitored space).
- A data-driven approach:** 2D trackers operate independently on the separate camera views; then the 2D tracks belonging to a same target are collected into a 3D one. The handling of occlusions and the association of (possibly split or merged) tracks are the main drawbacks of the data-driven approach. There is not enough information in the independent camera views to efficiently address them. The work-around in this case is to work with faces and or heads instead of bodies. In this case the initialisation problems of the direct approach are diminished as initialisation is handled by face detectors.

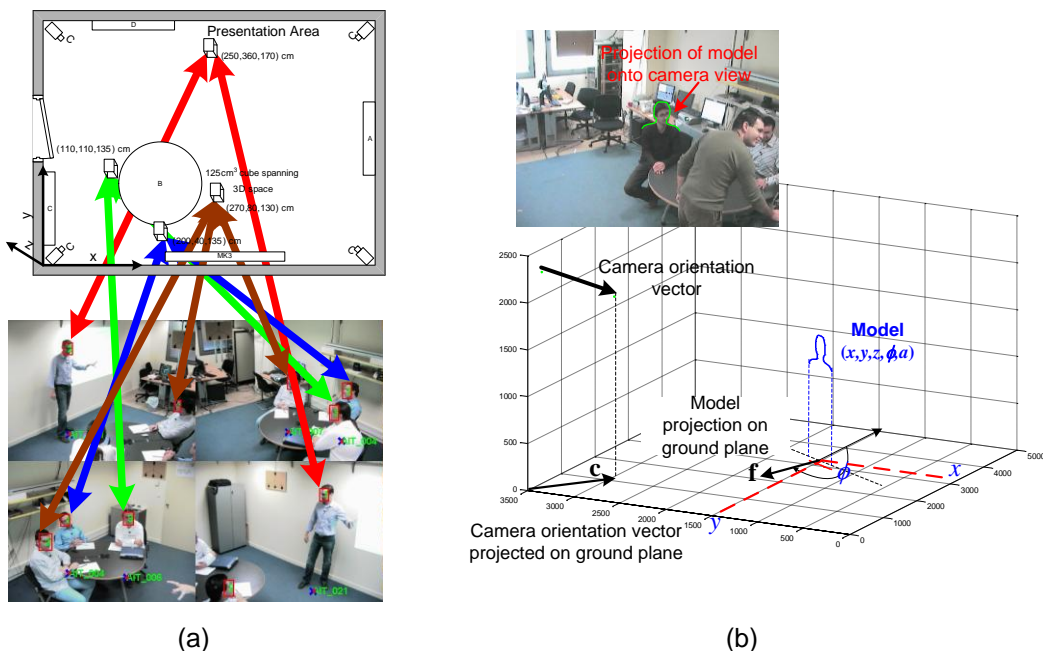


Figure 2: Data driven (a) and model-based (b) approaches towards 3D tracking.

### 3.1.3 Face Detection, Tracking and Recognition

'Non-attentive face recognition' refers to the recognition of faces in a natural (unconstrained) environments under arbitrary pose, expression or illumination (Ruiz 2009). Accurate face detection is a crucial first step in the detection of the tracking process. Different approaches for face detection were proposed based on the boosted cascades of simple classifiers [6,7,13,16], that achieved high accuracy paired with low computational complexity. To improve face alignment accuracy, successful approaches considering models for facial features, e.g. based on Active Appearance Models [3], Constrained Local Models [2] or Boosted Ranking Models [17] were proposed. When dealing with video sequences, different approaches have been proposed for face tracking, e.g. [10,19], or the Multiple Instance Learning (MIL) paradigm [1]. Although great improvements in the field of face recognition have been achieved in the last decade, complex scenarios remain an unsolved problem. The recently released challenging

dataset Labelled Faces in the Wild [5] features “real world” conditions. The best performing algorithms ([11], [4], [15]) focus on local features (SIFT [8], Gabor, Local Binary Patterns [18], etc.) followed by complex decision engines (SVM, MLP, metric learning, etc.).

### 3.1.4 Text tracking and recognition

Using detected text in the scene can provide a very robust means of athlete identification. Providing that the text string of an athlete is sufficiently different to any other possible text in the scene, the detection of such text can almost guarantee that an athlete is present and ‘visible’.

The robust identification of athletes in an unconstrained environment, like an outdoors stadium, is problematic as faces and text (located on the athletes’ bib) are often seen as blurry, occluded or too small to identify, hence accurate identification cannot be guaranteed in 100% of the frames.

The extraction of textual information from scenes (still images or videos) can be divided into sub-stages, not all necessary present:

- analysis of the image in order to detect the presence of text,
- localisation,
- tracking,
- extraction and enhancement of text,
- interpretation as words, also known as text recognition.

An interesting and detailed survey of these sub-stages, except for recognition, is provided by Jung et al. [55]. They categorise text localisation methods into two main types:

- region-based
- texture-based,

according to the utilised features. The first group is further divided into connected component and edge based approaches. The method [56] to extract text from images that we use in the present work, with some adaptations, is also reviewed in their connected component-based sections. In some recent approaches, scene text segmentation relies on graphical models and belief propagation [57], methods that are interestingly applied also to text recognition [58].

### 3.1.5 Motion analysis using optical flow

Key-point extraction can be done in several ways. The simplest defines a ‘point of interest’ as a corner, and utilises the Harris corner detector [21]. More recent and complicated approaches involve the use of SIFT [8] [22] and its variants like SURF [23]. Once key-points have been located, they can be matched in successive frames, thus providing an estimate of camera motion in the form of the projective transformation that matches one frame to the next.

## 3.2 Technology evaluation

Tracking athletes is a very complicated task due to many factors; primarily, moving cameras, adverse and changing lighting conditions and adverse poses of people in extreme action. To succeed in this task, we employed a variety of image processing techniques for athlete analysis, including:

- Face detection. This searches the distinctive facial features in the gray-scale images. If the faces that exist in the camera view are approximately frontal and fully visible, i.e. without extreme shadowing or expression, then they can be very accurately framed using face detection algorithms, such as the Viola-Jones classifier [16].
- Colour analysis. Human skin has a distinctive gamut of colours, no matter the race. While a generic human skin colour model like the one of Jones and Rehg [24] can serve for most people, it easily mistakes background regions as skin. If a colour model is carefully trained and updated on a per-target basis, then colour can prove a very powerful cue no matter the expression and the pose of the head, and is also capable of coping with gradual illumination changes.
- Outline location. Whilst a full-body outline is very difficult to model, the head and shoulders part of an athlete is fairly constant, hence more easily modelled. The only sources of flexing

of this part of the human body are the neck and the shoulders. The head stays mostly upright in many instances, as it only occasionally bends and head rotation does not change the outline significantly.

- Motion analysis. In cases where the cameras are immobile, the motion of people and objects appearing in the video can be accurately localised, creating potential athlete targets. However, if the target stops moving, the adaptive background estimation algorithm used for segmenting the image into foreground/background regions will fade as the person blends into the background. In order to reduce this effect, the learning rate of the image region under consideration can be adjusted based on the tracker results.
- Both stochastic and deterministic tracking algorithms. The stochastic algorithm employed for 3D person tracking was based on particle filtering, and utilised non-linear measurement models. The deterministic approach selected for 2D person tracking was based on Mean-Shift.
- Text tracking and recognition. An algorithm was implemented to detect, extract and interpret text on a per-frame basis. It is based on the previously published method [8], regarding the extraction, filtering and clustering of connected components, using prior knowledge of the applicative scenario. This method is then followed by a restoring of components, taking into consideration (during a second analysis), filtered components in the candidate zones. Each candidate text line is fed into an OCR system, which is used to both filter false alarms detected as text and also to obtain a string of characters from pictorial text.
- Projective transformation between frames has also been investigated to provide an estimate of camera motion, in terms of translation and zoom.
- Frame differencing between projectively transformed frames has been used to yield motion outliers and effectively separate the far-field scene from the bodies of the athletes. In such scenes, obtaining the athlete bodies as motion outliers is of paramount importance, since the lack of adequate resolution makes the application of the other tracking options discussed very difficult. The same goal has also been achieved through optical flow analysis of the motion.

### 3.3 Conclusions and Future Work

The majority of graphics that are applied to TV broadcast that highlight athletes is done manually or in post production. The illusion of tracking can be achieved without actually directly tracking.

There are many companies [38][39][40][41] that use graphics to aid the understanding of how athletes move.

The athletes in athletics track and field events are currently not tracked directly in TV broadcasts. There are many sports where the athletes are tracked live. GPS is used for sports such as:

- Sailing
- Motor Racing such as Formula 1 and Nascar
- Red Bull Air Races
- Horse Racing

Broadcasters have used other tracking technologies such as the use of mobile phones and GPS to track athletes. Occlusion is able to be overcome by use of non- image based tracking.

## 4 Personalisation

### 4.1 State of the art

#### 4.1.1 User profiling

A user profile is a collection of personal data associated to a specific user. When it is exploited by a computer system, a user profile can be considered as the computer representation of a user model. There are different choices for user profiling.

- Implicit vs. Explicit model: implicit profiling means indirect acquisition of the user profile, e.g. by observing the users. Explicit profiling done via direct user interactions, e.g. query profile from UI
- User instance modeling vs. user type modeling: user stereotypes can be derived through data mining and clustering. Individuals can then be associated with that user stereotype.
- Static vs. dynamic user models: static profile is personal information tend to be invariant, such as user's prints, gender, birthplace etc. Dynamic user profile varies across user tasks, time or space. Dynamic user profile can be enabled through machine-learning algorithms which can cause the adaptation of the system to user profile in different context.
- Generic vs. Application-specific models: the latter applies to specific task or application domain whereas the former can be used across tasks and application domain.
- Content-based vs. collaborative user models: content-based user profile depends on prior characters of the content and matches them to the user's individual preferences. Collaborative user profile depends on matching individual's preferences to a stereotype user's preferences and then using the latter to help complete the preferences of the individual.

Hybrid user models may also be used. Stereotype user models may be used in conjunction with explicit user modeling. Stereotype user models are also often used in conjunction with collaborative user models.

#### 4.1.2 Device and network Adaptation

The device adaptation allows the services to be accessed via heterogeneous computers and terminals. The most focus for presentation adaptation is the adapting content designed for decimeter sized screens and for small centimeter size displays. Adaptation to different heterogeneous terminals should preferably automatically adapt content to the terminal capabilities. There are two main approaches to this:

- Lowest common denominator approach: content is created that can be used on a few categories of devices that can cover a large number of devices. Each device in the category supports a lowest common denominator profile for that category.
- Transcoding of content to adapt it to specific types of access devices: this transforms content from one from to another via clearly defined mapping functions.

These strategies (presentation adaptation and device type adaptation) were implemented by means of heterogeneous network and terminal capabilities functions (supporting different interfaces, resolution screens, and CPU capabilities, maintaining usability and all functionalities).

The quality of the streams is adapted to the end user's device and network conditions, and its change is transparent to the user. The adaptation process bases its decisions on conditions from the network, from the host environment and from the context of the program being watched, as these are the factors that will most probably cause impact on the quality experienced by the user, in order to determine the boundaries for the media bitrates that shall be used.

A specialised module in the client is responsible for the monitoring of the quality of the network connection and of the host conditions in which it is running. In the host it monitors the CPU conditions, checks for dropped frames and screen resolution. Even if the bandwidth available is enough for the maximum

quality stream, a slow or loaded CPU may not be able to render the stream, thus dropping frames. In this case, a lower quality stream is requested to meet the bottleneck requirements.

A Terminal Portability feature together with a Total CPU Limiter, a Player CPU Limiter, a Screen Size Adaptation and a Content Adaptation component bring the adaptation enhancements to the solution.

## **4.2 Technology evaluation**

### **4.2.1 Creating user profiles**

User profiles can be directly asked from users. But the drawback of this is it assumed user's profile is invariant which might only apply to demographic information. In practice, only a few questions should be asked such as demographics and other indicators, which allow the user to be defined. Such information can also be used to classify user stereotypes in collaborative recommender systems.

User profile can also be created indirectly. User's interaction can offer a great amount of user information in terms of his input. The input will be afterwards used to represent user profile. This approach is more suitable for a 'dynamic' problem domain or task in which the user profile changes more frequently.

User profile can be maintained either on client side or service provider side. The latter is often preferable to the former as it allows a cross-device usage of these profiles.

### **4.2.2 Capturing user preferences (implicit/explicit)**

User preference can be created either via explicit user input or implicit user input.

In the former case, the user's input is seen as a single-shot feedback, which would be used as invariant user information. The apparent advantage of this approach is that the answer will be more precise. Nevertheless, the disadvantages of this approach are also obvious as it might disrupt user's current task and might be time-consuming and annoying. One typical application of this approach is a one-off interview or questionnaire. However, the potential risk of using a full interview or questionnaire is that user may not fully complete it.

In the latter case, the user's current and historical interaction will be considered as the input to form his profile. The obtained profile can be used to personalise user input to filter user selections of content and services. Such indirect approach can be improved by combining several context values such as the activity, time and etc. and may generate a more complete understanding of the current situation.

## **4.3 Conclusions and Future Work**

### **4.3.1 User profiles in Internet based media access**

User profiles are critical in enabling the context-aware systems. In the Internet based media access domain, the user profiles are used to maintain user information and serve as a clue for future use pattern. However, coupling defined user profiles with the Internet based media access system could be a challenge in some perspectives.

First, the mean of retrieving of the profile is constrained by many other aspects such as accuracy and expected retrieving frequency. Explicit user profile can be accurate in a one-time shot but may not be accurate enough in a long term. Higher retrieving frequency also requires less disturbing UI support.

Second, maintaining these profiles could be a dilemma. If the profiles are maintained centrally, it requires good network condition especially when the amount of each user's profile data increasingly grows. Maintaining the user profile locally may lose the advantage of group profiling, such as tackling the cold start problem.

In the context of the My eDirector 2012 system, the user profiles are defined in terms of different user tasks so that the user profile management can be best fit for different user requirements and interactive tasks. The existing user profiles are mainly consisted of user device screen information, user demographic information and user usage information, in the future we also wish to exploit other information such as physical location and physical environment context so that the profiles can be used to enable

other personalisation services.

User profiles are able to be automatically updated by exploiting a set of on-line learning algorithms as well as user's feedback. In this way, user profiles are dynamically designed in a context aware framework. In addition, ontological structures are able to be exploited in order to enhance the degree of personalisation to the end users. Ontologies describe the knowledge of the sport events and provide relationships among the detected incidents and the sport events. In this way, we are able to construct more complex queries enhancing the level of personalisation to the end users.

Of particular interest for future work is the examination of social networking aspects in the construction and evolution of user profiles as well as the ontologies associated with them. In particular, research should be performed towards this direction of update the relationships between the objects of the ontology by incorporating either the feedback of the user or the social aspects of other "similar" users.

#### **4.3.2 Use of ontologies**

Ontological descriptions of world concepts can greatly assist intelligence extraction. The role of ontologies and semantic web technologies in general is:

- To formally define a domain of discourse, by defining the concepts of interest and the relationships between them
- To infer knowledge based on the existing one.

However, employing ontological description does not constitute a panacea for formal knowledge definition.

Using ontology, one can create more sophisticated queries according to his/her information needs. In this way, we increase the degree of personalisation achieved in the framework of My eDirector 2012 project [62]. Ontology can be applied either to real-time video events or to stored, annotated video archives. For real-time sport events, the ontology can be exploited as interfaces that enable users to create more sophisticated queries regarding selection of the most important video segments, according to his/her information needs. On the other hand, in case of stored video archives, ontology can be used for detecting the most interested video segments according to user profiles.

The ontology is described using the interoperable format of the RDF/SML framework. This provides the users an interface for constructing sophisticated queries and interoperable encoding the knowledge and the queries under the RDF/XML framework. The interoperable RDF/XML description allows for third party enterprises to establish interfaces for parsing the ontology structure and implementing tools for providing queries by exploiting both the knowledge, described through the ontology, and the proper reasoner tools.

##### **4.3.2.1 Computational burden**

The usage of semantic web technologies is an additional computational burden that is deemed unnecessary (or even wrong) in cases when:

- There are limited computing resources. Inference procedures tend to require numerous calculations. In a system already performing complex calculations for face tracking, reasoning would bring it to its limits.
- Information processing needs to be conducted in real-time. In this case, the upper bound in the time needed to process newly generated information is kept relatively low. For instance, a higher frame rate in a camera imposes even stricter time limits for the real-time information processing.
- Information is intended for internal consumption only; not to be made publicly available. In the scope of a project, arbitrary development of a vocabulary is not a bad choice as long as the vocabulary covers its needs and there are no plans for integration with third party sources and repositories.
- The complexity in the information (meta)model is low to the extent that usage of ontologies is

not justified and would be an overkill. Inclusion of ontological descriptions entails a computational burden that needs to be justified by the benefits it will yield.

One could argue that usage of semantic web technologies could be kept to a minimum (i.e. use of RDF graphs with sparsely interconnected concepts). In this case however, the needs of an application with such a simple world model can be covered by more mature (and often more lightweight) technologies compared to semantic technologies, for instance programming languages structures, relational database models, XML, etc.

#### 4.3.2.2 A matter of usefulness

Ontologies do not have to be developed from scratch. Nowadays there are numerous ontologies available on the web, covering all needs of concept definitions: from high-level concepts to application specific ones. One could therefore wonder why an existing ontology was not reused in the scope of My eDirector 2012. The answer, however, is not as simple as it sounds.

Take for instance an individual property: *hasComponent*. This property can be defined in an ontology lying in an ontology repository filled with robust, experience-based and well tested ontologies, in a way analogous to code libraries. There are two options in including *hasComponent* in our application description. These options include:

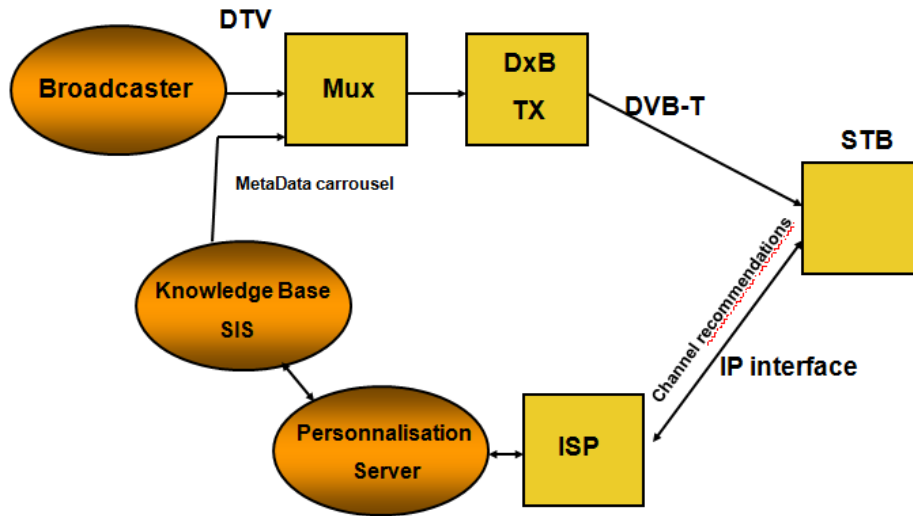
- Merely include a reference to the public ontology URI. This means that the ontology developed for the needs of My eDirector 2012 would not benefit by the hard work put into the public ontology developers. This happens because none of the concepts will be made available for use.
- Import the whole ontology in our application description. This is a justifiable approach when the import chain is relatively small. But, more typically, one may import any number of ontologies, of lower, middle and upper level. In this case, the local ontology will contain a lot of useless information. Therefore, the additional burden will be disproportional to the problem it is intended to solve, since it will cause the inference engine to behave much slower.

In the scope of My eDirector 2012, our investigation into ways of incorporating semantic technologies led us to the conclusion that the expressiveness, intelligence, and standards-compliance potentially offered by semantic web technologies have to be sacrificed to the altar of performance: Real-time, low-latency behaviour, given the hunger for resources of the perceptual component, leaves no room for semantics in the suggested approach. Moreover, regarding the description of the world concepts, there is no clear payoff, in terms of inferred knowledge, to justify the expense.

#### 4.3.3 Personalisation in broadcasting

The personalisation within the broadcasting DVB-T technology in My eDirector 2012 is based on the usage of the DVB-T My eDirector 2012 STB. More specific this STB provides two interfaces with the My eDirector 2012. The 1<sup>st</sup> one is the RF interface (tuner) that can receive the broadcasted content that is transmitted on the UHF frequencies and simultaneously serves a bidirectional communication interface over IP with the core My eDirector 2012 platform. Within this return channel the end user can submit to the system the preferences and his profile in order for the core platform to have the adequate data and provide the automatic channel alteration according to the submitted requirements, The STB is based on a market available device the DM-800S DreamBox in which we have installed the My eDirector 2012 plugin. This plugin can also be downloaded and installed via the DreamBox plugin repository. This plugin provides the GUI in order for the end user to edit and submit his preferences and requirements to the core My eDirector 2012 system. Particularly the STB manages to download from the UHF the program and channel description from the My eDirector 2012 DVB-T transmitted bouquet. On the other hand it submits the end user profile via HTTP interface to the Profile Builder [59] and Figure 3. Before this stage there is the login and authentication stage to the SP/AAA module and the assignment of unique userid within the My eDirector 2012 system. Additionally, the personalisation information regarding the automatic channel and recommendations and alterations are received over the same http interface and from the knowledge base module. Finally the My eDirector 2012 STB gives to the end user enhanced personalisation features like the dynamic selection on metadata regarding the content and also the dynamic presentation for the channel change in PiP format. This means that the end user

could see in a small window in the upper side of the TV screen the proposed new channel and accept the change. In the case of the real time personalised metadata the My eDirector 2012 STB receives the metadata stream from Olympic Game SIS system and accordingly, the initially user preferences are being displayed.



**Figure 3: My eDirector 2012 STB interfaces**

Regarding the state of the art in commercial solutions there are alternative methods for the distribution of interactive multimedia services over broadcasting technologies, including wire and wireless telecommunication networks. In particular, non-real-time and real-time programme services that are downloaded and stored in the receiver (STB) are expected to be offered via these networks. However, compared to broadcasting networks, the downstream path lacks the attractive cost/capacity ratio. There are also alternatives for upstream transmissions in relation to broadcasting networks; for example the DVB-RCT system: an in-band terrestrial return channel. But it is unlikely that the DVB-RCT system will achieve wide application because of frequency usage limitations. In contrast, the My eDirector 2012 selected technology for the interactive return channel is the most flexible and popular as it based on the IP protocol and most specific on the most popular application protocol which is the HTTP.

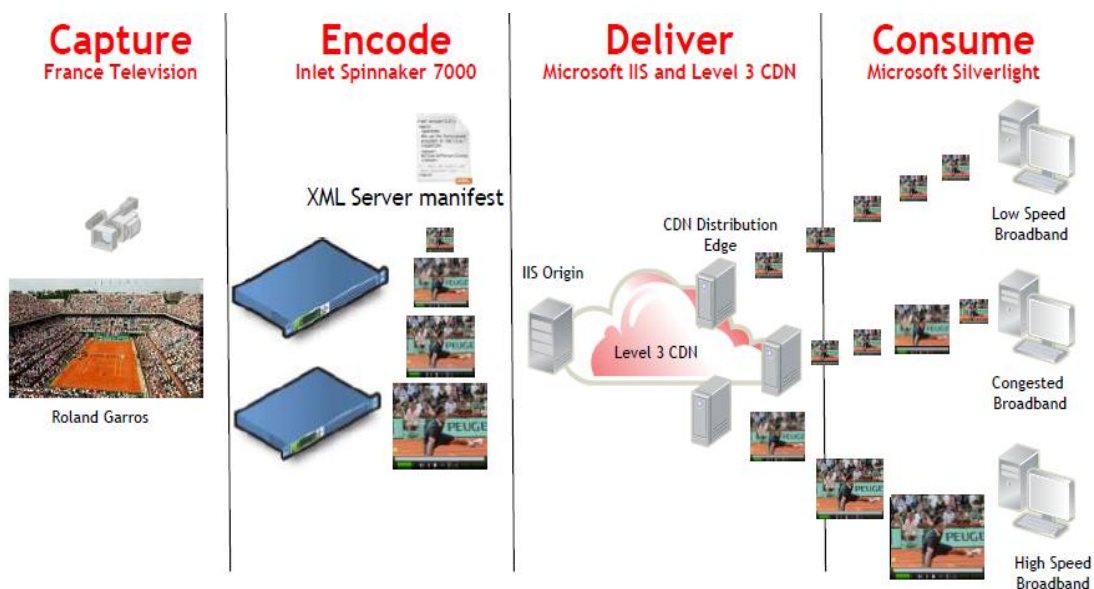
## 5 State of the Art in Commercial Solutions

### 5.1 Http Adaptive Streaming Solutions

#### 5.1.1 French Open Tournament '09

French Open Tournament is an international tennis tournament also known as Roland Garros. In June 2009, the France Televisions TV channel, wanted to offer to its viewers a unique real-time web experience of HD (720p+) quality using its websites France2 0, France3 0, France4 0. For that reason, France Televisions chose to cooperate with Microsoft, Inlet Technologies 0 and Level3 Communications 0, in order to manage finally to transmit the tournament in HD quality using the Smooth Streaming technique. This was the first time that such an event was transmitted live using the Smooth Streaming.

As shown at the following Figure 4, the tournament was at first captured by camera and then it was modified in HD format 1080i50.



**Figure 4: Capturing and Broadcasting method of French Open 09**

Moreover, the signal that occurred was passed through Spinnaker encoders of Intel Technologies through HD\_SDI connections. The synchronisation of the HD signal was completed through two Spinnaker encoders reaching rates from 350 Kbps to 3Mbps, offering 7 streams of video. Those 7 streams were sent to the IIS Microsoft servers and were stored for DVR future purposes. IIS servers were acting like primal servers for the transport network of Level3. Therefore, the requests of the end users for content were addressed to the closest servers of Level3 through technology based to DNS. Therefore, the network situation is always checked to assure that the final content receivers are connected to the closest web-server to them. That Level3 server gets the requested content from the local web caches, which are powered by the Microsoft IIS servers and delivers the requested content via http protocol to consumers, taking advantage of the mass distribution of Level3. Finally, the Microsoft Silverlight player dynamically tracked the circumstances of the user's machine and changed the incoming video pieces.

Consumers with high speed internet connections combined with a powerful terminal had the opportunity to experience the True HD (720p) resolution, while those with lower speeds and lower power terminals were the most suitable video bit rate stream at these conditions (Figure 5).

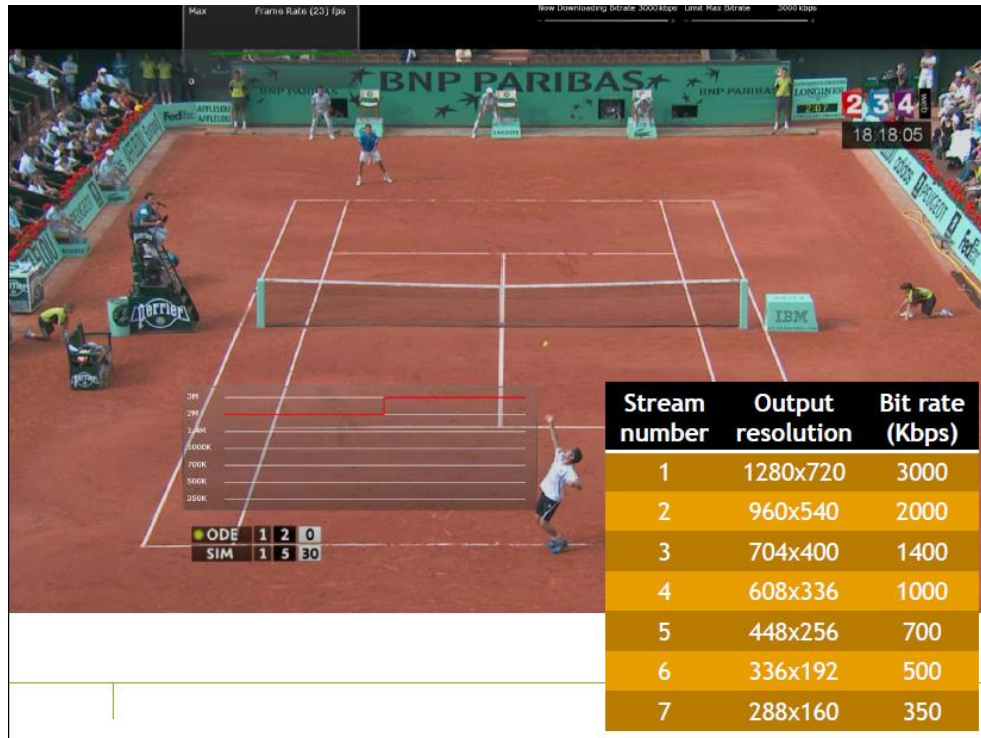


Figure 5: Snapshot of stream changing on the Silverlight player

It is worth mentioning that the user could take advantage during live broadcasting of the network DVR capabilities that were offered, ie pause, instant replay and even rewind a few hours (Figure 6) without breaks on the picture or buffering.



Figure 6: DVR capabilities during live broadcasting

### 5.1.2 Sunday Night Football

The American football league, known as the NFL (National Football League), is one of the sports events with the largest audience in the U.S. For this reason, the NBC [47], the channel which had the rights to broadcast the games in America at the time of the 2009 season, wanted to give viewers a unique, interactive viewing experience of the Games through their computers. This was managed in collaboration with Microsoft, Vertigo [48], Akamai [49], iStreamPlanet [50] and was based on the technology of Smooth Streaming.

So every Sunday during the course of the live broadcast of the game via the sports page of NBC [51], the user was able to "play" with various interactive features [12], including four different camera angles of the camera and a video player with full DVR functionality, as shown in Figure 7 below. Furthermore, the transmission was of great interest to advertising agencies as they were given the opportunity in co-operation with DoubleClick [52] for advertisement insertion in real time

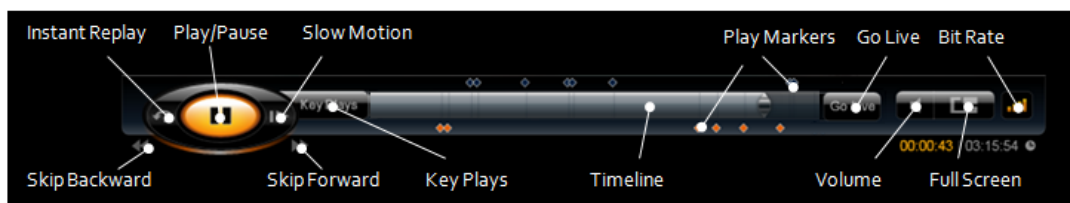


Figure 7: The "Sunday Night Football" video player

The viewers who started watching the race in the middle of the duration, had the opportunity by rewind, to see what was lost by the time they tuned into the game. An important added feature was the bar that allowed the user to navigate to the main points of the game until that moment and through an index to see in which point of the video there was a touchdown, for example. Also, the viewer could send comments to the show-host of live broadcasting.

The video player was implemented by the company Vertigo Software Inc including the Silverlight 3.0 plug in. The iStreamPlanet Co was given the task of encoding the video in which the bit rate reached up to 3.45 Mbps, which gave to the user the feature of HD (720p) quality. Though, for delivering the content to the terminals, the distribution network and the servers of Akamai were used.

The optional features that the user could enjoy through the offered video player implemented by the Verizon [53] are presented in Figure 8 below.



**Figure 8: Features of the video player**

*Play/Pause:* Starts or stops playback of the video respectively

*Time Bar:* Displays the total duration of the match, giving the user the ability to navigate to the time-point he wants by selecting the corresponding point on the timeline bar.

*Instant Replay:* Move back 5 seconds, in order for the user to be able to watch the last snapshot. Then by selecting the Go Live button, the user turns back to live transmission.

*Slow Motion:* Slow retransmission.

*Skip Backward/Forward:* Skip 5 seconds back / forward, respectively.

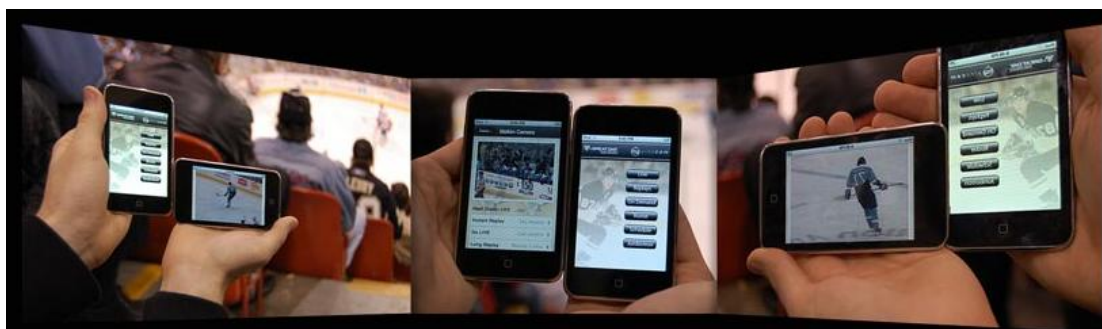
*Key Plays:* By clicking this button the user can watch through the menu that pops-up, important time-stamps of the game.

*Play Markers:* Representation using indicators for the important timestamps of the game including a brief caption.

*Bit Rate:* Figure depicting the signal strength, which reflects the quality of the image received by the user. The quality is affected by the prevailing conditions in the network but also from those of the terminal and may be changed during the transmission.

### 5.1.3 Yinzcam

YinzCam[54], is a mobile video technology for broadcasting sports events that are free to users. Through the use of YinzCam, the users can choose to watch in the mobile device (Figure 9) and while they are at the stadium, real-time video through unique camera angles. In particular, they also have the ability to choose a camera that follows the athlete in the race. Of course, their mobile devices should be able to connect to the stadium's wireless network. Moreover, they have the ability to manage the video content and create their own instant replay from multiple angles, with duration of 20 seconds or 2 minutes, and also they can access to statistics related to the event taking place at that time or biographical information of the players. All these options are shown in Figure 10 through the initial menu of the YinzCam.



**Figure 9 : Capture from different angles**



**Figure 10 : YinzCam Menu**

It should also be noted that the YinzCam offers a solution for desktop systems (Figure 11) via a touch screen, offering the same opportunities to the users. Nevertheless, all these options are available at this moment for viewers that are present at the stadium and not for remote users.



**Figure 11 : Touch screen for desktop systems**

#### **5.1.4 Beijing 2008 – NBC Sports**

NBC Sports had ambitious plans for the Web component of its 2008 Olympics coverage and sought help from strategic alliances with Microsoft, the MSN® network, and others to realise those plans.

Developers created NBCOlympics.com on MSN, which broadcast more than 5,500 hours of live and on-demand streaming content and tallied more than 1.3 billion page views, 50 million unique visitors, 70 million viewer-initiated video streams, and 10 million hours of viewer-watched video.

Developers on that project team used the Microsoft® Silverlight™ browser plug-in version 2.0, the Microsoft .NET Framework version 3.5, the Visual Studio® 2008 Team System development system, Expression Blend™ design software, and other Microsoft technologies for the next-generation Web to create the advanced features and capabilities essential for the kind of experience they wanted to give Olympics viewers.

Schematic developers also worked closely with deltatre, a partner of NBC Universal (the parent company of NBC Sports), to synchronise a rich set of real-time commentary and game statistics. The developers also used Silverlight and related Microsoft Web technologies to synchronise Olympics content with advertising and to integrate and match live commentary from Beijing with the event videos.

### **5.2 Sports Media Annotation Solutions**

Metadata offered to viewers is nowadays an integral part of athletic events coverage. A few systems exist that provide real-time metadata, such as the Sports Information System (SIS) of Atos Origin [25] [26] tailored towards the Olympic Games, or the Amisco multiple-camera football match analysis system [27]. Other systems generate off-line metadata, suitable for analysis by team coaches, like the computerised football analysis system by Mathball [28] and the sports video analysis systems for football or handball by SportVAS [29]. Of the online systems, only Amisco is semi-automatic, requiring some human intervention and it utilises dedicated cameras instead of those of the broadcaster.

## 6 Implementation tools and development platforms

### 6.1 Front End Development

#### 6.1.1 Silverlight

Silverlight is a platform used for building Rich Internet Application (RIA). Its primary focus is on rich media, interactivity, and animation. Compared to another famous RIA platform Flash, Silverlight adds more tools to the tool belts of web designers and developers everywhere. Coming from different perspectives on RIA, Silverlight introduces more advanced concepts including separation of presentation and logic and dependency properties, which allows the developer to literally create any conceivable applications. As a front-end RIA platform, Silverlight is independent of any backend system which allows it to be flexibly integrated with any backend systems.

#### 6.1.2 Silverlight and Smoothstreaming

Smooth Streaming, an [IIS Media Services](#) extension, enables adaptive streaming of media to Silverlight and other clients including iPhone, windows phone 7 over HTTP. Smooth Streaming provides a high-quality viewing experience that scales massively on content distribution networks, making true HD 1080p media experiences a reality.

#### 6.1.3 Implementation Tools

Visual Studio 2010 and Expression Blend 4 are the major tools used to create the Silverlight Applications.

Visual Studio 2010 has everything needed to develop Silverlight applications, with a visual designer for Silverlight pages. Using this designer, the user interface can be dragged, dropped, drawn into existence and a live preview can be done.

Blend 4 provides the rich support for creating Silverlight user interface, with visual tools that surpass Visual Studio. For certain types of user interface grunt work (for example, creating a nice gradient fill), it's a tremendous help. Expression Blend also supports a fun application prototyping tool called SketchFlow and includes a decent coding editor that's designed to look like Visual Studio. However, it lacks many advanced and important development tools, such as debugging, code refactoring, and project source control.

## 7 Bibliography and References

- [1] Boris Babenko, Ming-Hsuan Yang, Serge J. Belongie: Visual tracking with online Multiple Instance Learning. CVPR 2009: 983-990
- [2] D. Cristinacce and T.F.Cootes, "Feature Detection and Tracking with Constrained Local Models", Proc. British Machine Vision Conference, Vol. 3, pp.929-938, 2006.
- [3] T.F.Cootes, G.J. Edwards and C.J.Taylor. "Active Appearance Models", IEEE PAMI, Vol.23, No.6, pp.681-685, 2001
- [4] Matthieu Guillaumin, Jakob Verbeek, and Cordelia Schmid., 'Is that you? Metric Learning Approaches for Face Identification.' International Conference on Computer Vision (ICCV), 2009.
- [5] Gary B. Huang, Marwan Mattar, Tamara Berg, and Erik Learned-Miller, 'Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments', Faces in Real-Life Images Workshop in European Conference on Computer Vision (ECCV), 2008.
- [6] S.-Z. Li and J. Lu: Face Detection, Alignment and Recognition, in G. Medioni and S. Kang (eds.), Emerging Topics in Computer Vision, (2004).
- [7] S.-Z. Li and Z.Q. Zhang: FloatBoost Learning and Statistical Face Detection. IEEE Transactions on Pattern Analysis and Machine Intelligence, 26 (2004), 1112-1123.
- [8] David G. Lowe, "Distinctive image features from scale-invariant keypoints," International Journal of Computer Vision, 60, 2 (2004), pp. 91-110.
- [9] O. Lanz, "Approximate Bayesian Multibody Tracking," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 28, No. 9, Sep 2006 , pp 1436-1449
- [10] M. Nechyba and H. Schneiderman: PittPatt Face Detection and Tracking for the CLEAR 2006 Evaluation, in R. Stiefelwagen and J. Garofolo (eds.) CLEAR 2006, Lecture Notes in Computer Science, 4122 (2007), 161-170.
- [11] Nicolas Pinto, James J. DiCarlo, and David D. Cox, 'How far can you get with a modern face recognition test set using only simple features?', Computer Vision and Pattern Recognition (CVPR), 2009.
- [12] Javier Ruiz-del-Solar, Rodrigo Verschae, and Mauricio Correa, 'Recognition of Faces in Unconstrained Environments: A Comparative Study.', EURASIP Journal on Advances in Signal Processing (Recent Advances in Biometric Systems: A Signal Processing Perspective), Vol. 2009.
- [13] H. Schneiderman: Feature-Centric Evaluation for Efficient Cascaded Object Detection, in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, (June 2004).
- [14] Lawrence D. Stone, Carl A. Barlow, Thomas L. Corwin, "Bayesian Multiple Target Tracking," Artech House Radar Library
- [15] Yaniv Taigman, Lior Wolf, and Tal Hassner, 'Multiple One-Shots for Utilising Class Label Information.', British Machine Vision Conference (BMVC), 2009.
- [16] P. Viola and M. Jones: Rapid Object Detection using a Boosted Cascade of Simple Features, IEEE Conf. on Computer Vision and Pattern Recognition, (Dec 2001), 511.
- [17] Wu, H., Liu, X., and Doretto, G., 'Face alignment using boosted ranking models', In Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 1-8, 2008.
- [18] Hong Yang and Yiding Wang, 'A LBP-based Face Recognition Method with Hamming Distance Constraint', Proceedings of the Fourth International Conference on Image and Graphics, pages 645-649, 2007.
- [19] S. Zhou, R. Chellappa and B. Moghaddam, "Visual tracking and recognition using appearance-adaptive models in particle filters," IEEE Transactions on Image Processing, Vol. 13, No. 11, pp. 1491-1506, 2004.
- [20] Q. Ye, Q. Huang, S. Jiang, Y. Liu, and W. Gao. Jersey Number Detection in Sports Video for Athlete Identification. In Visual Communications and Image Processing, volume SPIE 5960, pages 1599-1606, Beijing, China, July 2005.
- [21] C. Harris and M. Stephens, "A combined corner and edge detector," in 4th Alvey Vision Conference, 1988, pp. 147-151.
- [22] Mikolajczyk, K., and Schmid, C., "A performance evaluation of local descriptors", IEEE Transactions on Pattern Analysis and Machine Intelligence, 10, 27, pp 1615--1630, 2005
- [23] Bay, H., Tuytelaars, T., Gool, L.V., "SURF: Speeded Up Robust Features", Proceedings of the

- ninth European Conference on Computer Vision, May 2006.
- [24] M. Jones and J. Rehg: Statistical colour models with application to skin detection, Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, (1999), 274–280.
- [25] S. Poslad, A. Pnevmatikakis, M. Nunes, E. G. Ostermann, P.I Chippendale, P. Brightwell, C. Z. Patrikakis, “Directing Your Own Live and Interactive Sports Channel”, 2009 International Workshop on Image Analysis for Multimedia Interactive Services, London U.K., 2009.
- [26] “Information diffusion systems,” [http://www.atosorigin.com/enus/olympic/games/services/solutions/information diffusion systems/](http://www.atosorigin.com/enus/olympic/games/services/solutions/information%20diffusion%20systems/), 2009.
- [27] Amisco Live Analysis Service, URL: [http:// www.sport-universal.com/services/matchanalysis/matchanalysis.html](http://www.sport-universal.com/services/matchanalysis/matchanalysis.html), last accessed on 22/9/2010.
- [28] Computerised Football Match Analysis Systems, URL: <http://www.mathball.com>, last accessed on 22/9/2010.
- [29] Sports Video Analysis Systems, URL: <http://www.sportvas.com>, last accessed on 22/9/2010.
- [30] R. S. Cruz, J. Espadanal, and M. S. Nunes, “A Personalised HTTP Adaptive Streaming WebTV,” in *Proceedings of the First International Conference on User Centric Media Workshop, UCMedia '09*, (Venice, Italy), Dec. 2009.
- [31] R. Pantos, “HTTP Live Streaming,” Internet-Draft draft-pantos-http-live-streaming-04, Internet Engineering Task Force, Nov. 2010. Work in progress.
- [32] A. Zambelli, *IIS Smooth Streaming Technical Overview*. Microsoft Corporation, Mar. 2009.
- [33] C. Patrikakis, N. Papaoulakis, C. Stefanoudaki, and M. S. Nunes, “Streaming content wars: Download and Play Strikes Back,” in *Proceedings of the First International Conference on User Centric Media, UCMedia '09*, (Venice, Italy), Dec. 2009.
- [34] N. Papaoulakis, C. Z. Patrikakis, M. S. Nunes, and R. S. Cruz, “ROBUST MULTIMEDIA TRANSMISSION OVER WIRELESS AND MOBILE NETWORKS,” in Proceedings of the International Conference on Wireless Information Networks and Systems, WINSYS'10, Jul. 2010.
- [35] J. E. Gonçalves, R. S. Cruz, and M. S. Nunes, “Implementing a QoS-Aware, Context-Aware and Personalised WebTV,” in *Proceedings of the 2010 6th EURO-NF Conference on Next Generation Internet (NGI)*, (Paris, France), pp. 1 –8, Jun. 2010.
- [36] Adaptive Streaming Comparison. Available at: <http://learn.iis.net/page.aspx/792/adaptive-streaming-comparison>, last accessed on 03/03/2011
- [37] Adaptive Streaming: a comparison. Available at: <http://www.quavlive.com/adaptive-streaming>, last accessed on 03/03/2011
- [38] RBM, “Red Bee Media Website”. Available at: <http://www.redbeemedia.com/>, last accessed on 02/03/2011.
- [39] ORAD “ORAD Website”. Available at: <http://www.orad.co.il/>, last accessed on 02/03/2011.
- [40] VZRT, “Vzrt Website”. Available at: <http://www.vizrt.com>, last accessed on 02/03/2011.
- [41] SportVision, “SportVision Website”. Available at <http://www.sportvision.com>, last accessed on 03/03/2011.
- [42] Inlet Technologies, “Inlet Technologies Website”, Available at: <http://www.inlethd.com> , last accessed on 06/04/2011.
- [43] Level3 Communications, “Level3 Communications Website”, Available at: <http://www.level3.com> , last accessed on 06/04/2011.
- [44] France2 TV Channel, “France2 TV Channel Website”, Available at: <http://www.france2.fr> , last accessed on 05/04/2011.
- [45] France3 TV Channel, “France3 TV Channel Website”, Available at: <http://www.france3.fr> , last accessed on 05/04/2011.
- [46] France4 TV Channel, “France4 TV Channel Website”, Available at: <http://www.france4.fr> , last accessed on 05/04/2011.
- [47] NBC, “NBC Website”, Available at: <http://www.nbc.com> , last accessed on 07/04/2011.
- [48] Vertigo, “Vertigo Website”, Available at: <http://www.vertigo.com> , last accessed on 07/04/2011.
- [49] Akamai, “Akamai Website”, Available at: <http://www.akamai.com> , last accessed on

- 07/04/2011.
- [50] iStreamPlanet, "iStreamPlanet Website" Available at: <http://www.istreamplanet.com> , last accessed on 08/04/2011.
- [51] NBCSports, "NBCSports Website", Available at: <http://nbcsports.msnbc.com/> , last accessed on 07/04/2011.
- [52] DoubleClick, "DoubleClick Website" Available at: <http://www.doubleclick.com> , last accessed on 08/04/2011.
- [53] Verizon, "Verizon Website" Available at: <http://www.verizon.com> , last accessed on 09/04/2011.
- [54] YinzCam, "YinzCam Website" Available at: <http://www.yinzcam.com>, last accessed on 10/04/2011.
- [55] K. Jung, K.I. Kim, and A.K. Jain. Text Information Extraction in Images and Video: a Survey. *Pattern Recognition*, 37(5):977–997, May 2004.
- [56] S. Messelodi and C.M. Modena. Automatic Identification and Skew Estimation of Text Lines in Real Scene Images. *Pattern Recognition*, 32:791–810, 1992.
- [57] H. Shen and J. Coughlan. Finding Text in Natural Scenes by Figure-Ground Segmentation. In *International Conference on Pattern Recognition*, Hong Kong, August 2006.
- [58] J.J. Weinman, E. Learned-Miller, and A.R. Hanson. Scene Text Recognition Using Similarity and a Lexicon with Sparse Belief Propagation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(10):1733–1746, February 2009.
- [59] My eDirector 2012- Real-Time Context-Aware and Personalised Media Streaming Environments for Large Scale Broadcasting Applications project, Deliverable D2.3: "Network Platform requirements".
- [60] My eDirector 2012- Real-Time Context-Aware and Personalised Media Streaming Environments for Large Scale Broadcasting Applications project, Deliverable D2.4: "System Architecture and Specifications".
- [61] My eDirector 2012- Real-Time Context-Aware and Personalised Media Streaming Environments for Large Scale Broadcasting Applications project, Deliverable D6.2: "Personalised Media Stream over Radio Access Technologies".
- [62] My eDirector 2012- Real-Time Context-Aware and Personalised Media Streaming Environments for Large Scale Broadcasting Applications project, Deliverable D4.4: "Configurable Camera Selection Tables and Algorithms".
- [63] IIS "Microsoft Smooth Streaming Website". Available at: <http://www.iis.net/overview>, last accessed on 15/04/2011.
- [64] Http Live Streaming "Apple Http Live Streaming Website". Available at: <http://developer.apple.com/library/ios/#documentation/networkinginternet/conceptual/streamingmediaguide/Introduction/Introduction.html> , last accessed on 15/04/2011.