PORTIVITY: OBJECT BASED INTERACTIVE MOBILE TV SYSTEM

J. Deigmöller[1], G. Fernàndez[2], A. López[2], B. Mérialdo[3], H. Neuschmied[4], F. Pinyol[2], R.Trichet[3], P. Wolf[5], R. Salgado[6], F. Milagaia[6], S.Glaser[7], A. Duffy[7]

[1]Institut für Rundfunktechnik, Munich, DE, [2]GTAM - Enginyeria La Salle, Universitat Ramon Llull, Barcelona, SP, [3]Institut Eurecom, Sophia Antipolis, FR, [4]Institute of Information Systems, Joanneum Research, Graz, AT, [5]Fraunhofer-Institute for Secure Information Technology, Darmstadt, DE, [6]Portugal Telecom Inovação, Aveiro, PT, [7]Rundfunk Berlin-Brandenburg, Potsdam, DE

deigmoeller@irt.de, gabrielf@salle.url.edu, alexl@salle.url.edu, bernard.merialdo@eurecom.fr, helmut.neuschmied@joanneum.at, fpinyol@salle.url.edu, remi.trichet@eurecom.fr, patrick.wolf@sit.fraunhofer.de, roger-a-salgado@ptinovacao.pt, fernando-j-milagaia@ptinovacao.pt , Sven.Glaser@rbb-online.de, Annette.Duffy@rbb-online.de

Abstract: **The porTiVity project is developing a converged rich media iTV system, which integrates broadcast and mobile broadband delivery to portables and mobiles and which will enable the end-user to act on moving objects within TV programmes. porTiVity has developed tools which enable producers to link rich media information with objects in the TV programme. The developments of the project include the playout of portable rich media iTV and the middleware, data and presentation engine in the handheld receiver. For middleware, metadata and the application layer, the project makes use of international standards such as MPEG-4 A/V, MPEG-4 LASeR, MPEG-7 Metadata and DVB-H. On the distribution side, the open standard MXF is used as material exchange format to deliver A/V content to the authoring suite and to distribute the authored rich media information between authoring suite and playout centre. In order to test the overall system and to show its feasibility, two different services were implemented: »Spur & Partner« and »iSports«.**

Keywords: Mobile TV, Interactive TV, LASeR, DVB-H, MXF.

1 INTRODUCTION

Personal devices like mobile phones or PDAs have acquired astonishing multimedia capacities in recent years. Now, broadcasters, telecommunication and Internet companies alike are pushing mobile *tele*vision. The syllable tele- is of Greek origin and means *far away,* which stands in blatant discrepancy to the distant of personal devices to their users – they are hand distance. They are thus perfectly suited for (manual) interaction. This already hints that mobile TV needs to be something more than a scaled down version of traditional TV. But what kind of interaction is suitable for such devices?

1.1 Motivation

Previous projects have shown that interaction with video content at object level is highly attractive to users (GMF4iTV,[1]). Others taught us that broadcast content can be enriched by interactive means (SAMBITS, [2]) and even be in sync with additional content delivered over alternative distribution channels (SAVANT, [3]).

With DVB-H [4] or eDAB effective mobile broadcast standards have been established and UMTS allows point-to-point connections to mobile devices with a bandwidth that makes multimedia streaming possible.

The multimedia capabilities of modern personal devices together with the broadcast possibilities to personal receivers allow us now to combine the visions of these projects to realise something new - direct interaction with video objects in content broadcast to mobile devices. In other words: an object-based interactive mobile TV system.

1.2 System Overview

The architectural diagram of the porTiVity end-to-end system is shown in Figure 1. porTiVity [5] is developing tools which enable producers to link rich media information (additional content) with objects in the TV programme (main content). The result is an interactive scene (realised in MPEG-4 LASeR [6]) that allows end-users to click on objects highlighted in the video in order to retrieve additional content. The main content plus the synchronised LASeR scene are broadcast over a mobile broadcast channel (DVB-H) while the additional content is requested by end-users over a mobile broadband channel (UMTS).

1.3 Structure of this work

The remainder of this paper introduces individual components of the porTiVity system and discusses major challenges in their realisation. The Authoring Suite discussed in chapter 2 is responsible for tracking of objects that are going to be highlighted and creation of the interactive LASeR scene. All of this is wrapped into an MXF [7] container and given the playout module (chapter 3). Here the content is adapted to the various distribution channels and it is made sure that the tight synchronisation between individual content parts is kept on its way to the terminals. The terminals (chapter 4) render the LASeR scene and implement vital functions like time shift viewing. Finally, chapter 5 describes first services that are possible through porTiVity technology.

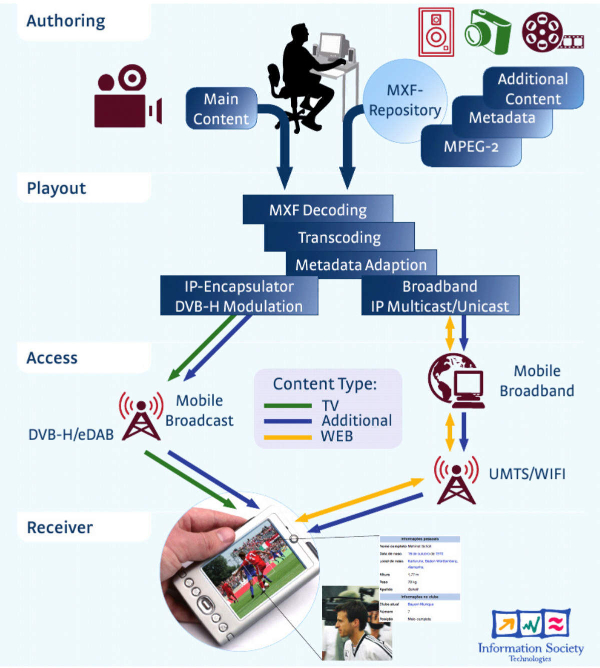

Figure 1 System overview

2 AUTHORING

The Annotation Suite includes all necessary tools for the post-production of the interactive LASeR application (see Figure 2) that makes object appearing on the video clickable. The annotation is done in several steps.

The video is automatically preprocessed by the *Media Analyze Tool*. It extracts metadata for navigation and structuring such as shots and key-frames and metadata which enables a fast semi-automatic annotation such as object tracking and shot/object re-detection. The analysis process can be started for several videos. The metadata is saved in MPEG-7 [8].

The extracted metadata can be viewed and edited by the *MPEG-7 Video Annotation Tool*. Furthermore, it supports manual structuring of video content, adding textual and semantic annotation, and the annotation of objects (image regions). The annotation work can be done very efficiently through object redetection and tracking functionalities and the possibility of multiple assignments of textual annotations [9]. The produced metadata is also saved in MPEG-7 format.

Based on the MPEG-7 annotation the LASeR service is generated by the *Interactive Application Tool*. Interactivity can be provided by inserting predefined menus and buttons in the video and by assigning additional content to moving objects. The appearance time of predefined objects can be specified using the annotated video structure. The additional content is a link to an image, video or audio data, a HTML page, or a LASeR scene. It can also be a LASeR script which enables functionalities such as personalisation of the additional content or games. The authoring suite is connected to the playout by means of MXF files which contain multiplexed AV and additional content as well as metadata in a synchronous manner.

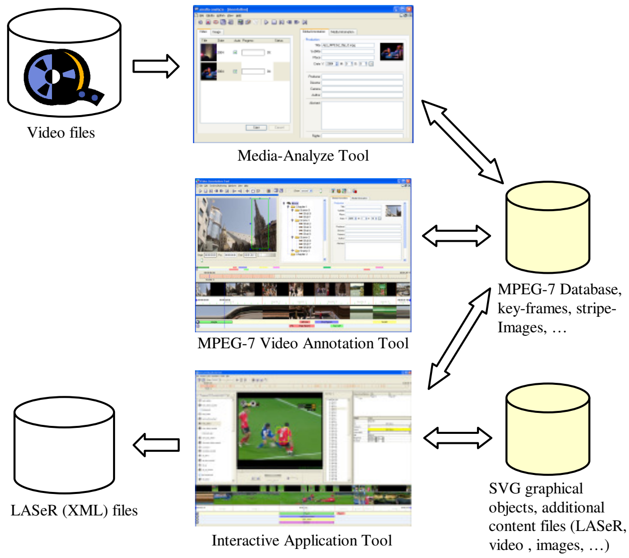

Figure 2: Authoring Suite

2.1 Object Tracking

The *MPEG-7 Video Annotation Tool* enables an efficient specification of interactive regions (objects). Object tracking allows the user to identify a video object on one frame, and automatically get all positions of this object throughout the entire shot.

For the annotation process to be ergonomic, the tracking has, first, to be as fast as possible. Moreover, in order to limit the user intervention, the object needs to be coarsely located with a bounding box. And finally, our tracker has to be generic, handling all types of videos and thus, all kinds of difficulties. In order to fulfil these specific requirements, we have build up a keypoint based tracking algorithm. Keypoints are located at key positions (usually corner or extrema of a given function), making them easy to recover. Moreover, they are enriched by local descriptors in order to increase their robustness to usual transformations (scale changes, illumination changes, rotations, affine transformations, …). Thus, keypoints are a reliable tool for many tasks in the computer vision domain and have been the subject of growing interest from the scientific community in recent years.

Our keypoint object model [10] offers two main advantages in the context of a tracker dedicated to annotation. First of all, the robustness of the associated descriptors makes them particularly efficient in dealing with the various difficulties that could be encountered by a generic tracker. Moreover, their computation is object location independent. Hence, the keypoints and their descriptors could be calculated off-line (for each frame) leaving only their matching [11] and the consequent global object motion estimation [10] to be performed by the algorithm during the annotation session, thereby drastically reducing the in-line computation.

3 VALIDATION OF AUTHORING

In the course of the project the Authoring Suite tools were validated by professional users and benchmarking was conducted. Validation by professional users conducted at an early stage in the project ensured that proposed improvements in the usability of the tools were implemented.

The 'Spur & Partner' service was used as the demo scenario to generate the required benchmarking figures. For this final test and benchmarking, an off-the-shelf PC (Core2Duo - 2x 2.66GHz, 2 GB RAM, 1280x1024 pixels resolution) was used and was set up with the Offline Annotation Suite version 2. The broadcast TV programme has duration of 08:40 minutes and in the tested version sixteen objects were annotated.

As a first step the *Media Analyse Tool* was started and the video to be analysed was then loaded. The analysis was conducted and saved as an XML file. The second step involved importing the analysis data into the *Semantic Video Annotation Tool*. With this tool, the objects to be annotated were located and marked with bounding boxes[1)]. Object tracking based on keyframe results was carried out, and any necessary corrections made before the results were saved[1)]. Finally, additional content for each annotated object was either imported or referenced in the *Interactive Application Tool* [2)].

For this 08:40 min episode the whole process, as described above, lasted 2:59:56h.

Process\Tool	MAT	SVAT	IAT
Start	00:00:05 h	00:00:04 h	00:00:07 h
Load content	00:00:02 h	00:00:08 h	00:00:07 h
Processing analysis:	01:44:25 h	-	-
Annotate one object		∅ 00:03:54 h[1)]	∅ 00:00:47 h[2)]
Annotate all objects		01:02:30 h	00:12:28 h

Figure 3: Benchmarking Offline Annotation Suite

4 PLAYOUT

The playout module consists of several modules: MXF Demuxer, Transcoder Module, Metadata Processing Module, Additional Content Processing Module, Additional Content Server and DVB-H Module.

4.1 MXF Demuxer

The input for this module is an MXF file containing the main content, as well as the additional content. The module extracts all the tracks, and sends them to the next modules:

- Audio and video from the main content are sent, over UDP, to the Transcoder Module.
- LASeR from the main content is sent, over UDP, to the Metadata Processing Module.
- Additional content is put on a directory shared by the Additional Content Server.

4.2 Transcoder Module

This module accepts the main audio and video from the MXF Demuxer, transcodes them from MPEG-2 to MPEG-4, encapsulates them over RTP, and sends them to the DVB-H Module. Audio is converted from MPEG-2 layer 2 to AAC. Video is converted from MPEG-2 video at 720x576 resolution, 25 fps, to H.264 at 320x240 resolution, 12.5 fps and 160kbps, in order to adapt it to handheld terminal requirements.

4.3 Metadata Processor Module

This module receives LASeR access units from the MXF Demuxer, it adapts them (constructing absolute URL addresses for the links in them), encapsulates them over RTP, and sends them to the DVB-H module.

4.4 Additional Content Processing Module

This module is the responsible for adapting the additional content, usually in MPEG-2 formats, to MPEG-4.

4.5 Additional Content Server

This server accepts requests from the broadband channel, and is able to serve the additional content for downloading (http) or streaming (rtsp). The later may include a reflector module which allows several clients to receive the same additional content synchronously, e.g. multi-cameras feature. Also, a combination of a reflector with a Policy Based Selection Module can be used to automatically re-route the stream from the broadband channel to the broadcast channel, according to the amount of clients connected to a given content.

4.6 DVB-H Module

This module takes all the IP packets sent to it, from the Transcoder Module (audio and video) and from Metadata Processing Module (LASeR) and encapsulates them for DVB-H (Multi Protocol Encapsulation over MPEG-2 Transport Stream). Finally, this Transport Stream is modulated and broadcast.

4.7 Synchronisation between audio, video and LASeR tracks

One usage of LASeR in this project (amongst others) is highlighting objects that appear on the video, which have a link to additional content. This highlighting must be tightly synchronised with the video, with frame accuracy. In order to achieve this precision, the RTP streaming of the three main streams (audio, video and LASeR) uses the RTCP (Realtime Control Protocol) mechanism [12], which synchronises the packet timestamps with a common clock. The terminal uses this mechanism to align the received tracks.

5 TERMINAL

The terminals receive the DVB-H stream sent out by the playout through external modules (SDIO and USB). The DVB-H service, consisting of an audio, video and LASeR MPEG-4 Elementary Stream encapsulated in RTP (Real-Time Transport Protocol), is synchronised on terminal side by RTCP. The player for rendering the porTiVity service is based on the open source player Osmo4, which is integrated in the GPAC framework developed at ENST (École Nationale Supérieure des Télécommunications). For presentations a QTek 9000 (Windows Mobile 5) and a Samsung NP-Q1 (Windows XP) was used. Two major challenges for the project have been the optimisation of the source code for video performance on Windows Mobile devices and the full integration of LASeR into the player. LASeR serves here as light-weight sort of middleware ensuring that porTiVity services can be played on any device [13].

The additional content – which is linked by interactive graphical elements in the LASeR DVB-H stream – can be accessed by an HTTP or Streaming Server via any broadband connection (here WiFi or UMTS). As both DVB-H and the broadband connection are based on IP layers, the navigation between both can be achieved by simply linking the services with the correct IP addresses. This also facilitates streaming of the porTiVity service over any IP based network beside DVB-H, for example WiFi.

An important feature needed on the terminal side is PVR (Personal Video Recorder) or at least time shift capabilities. Viewers do not want to miss parts of the broadcast while they are viewing additional content. The time shift module allows recording of the whole porTiVity service (video, audio, LASeR) so that when viewers go back to the main content, the programme resumes where they left to view the additional content.

For GSM enabled devices, the porTiVity player also implements a SMS message sending mechanism, which can be used to design quiz like services and to provide real-time feedback, allowing the design of community conversation channels.

6 PORTIVITY SERVICE EXAMPLES

Two exemplary services use porTiVity technology to enhance and transform traditional TV broadcasting and show the innovative potential of this project.

6.1 Spur & Partner

»Spur & Partner« is based on a successful interactive TV crime series for children produced by German public broadcaster ARD. The mobile variant shows porTiVity's simple and intuitive interaction (see Figure 4). Clicking the highlighted objects, children can collect pieces of evidence, reply to questions, and finally find the culprit. They can also retrieve a solution video on demand or access additional web content, like an interactive detective magazine.

6.2 iSports

»iSports« on the other hand is an all-round mobile football service. By clicking on the football players during a match, users can access statistic information on each player (see Figure 4). A click on the field brings information on the match, the team selection or the stadium. Exciting scenes can be bookmarked and a re-play function allows for repeating highlights. The service incorporates personalisation features into a normal TV programme. In this case, each user has its own profile (customisable) and the service will take advantage of this to deliver specific content according to the user preferences. The result is a further customisable and enhanced user's watching experience. Other features like real-time updated MMS banner and SMS sending mechanism are also included in the service.

7. VALIDATION OF SERVICES

End user validation of porTiVity services was carried out by Rundfunk Berlin-Brandenburg in Berlin, Germany and Universitat Ramon Llull in Barcelona, Spain. The purpose of the tests was to determine the usability of the applications and the suitability of the devices.

Using representative groups of three boys and three girls for 'Spur & Partner' and six young adult males for 'iSports', the results show clearly that both services transfer well to handheld devices.

7.1 Spur & Partner

This scenario has already been presented at IFA (*Internationale Funkausstellung*[1], Berlin), *Medientage*[2] (Munich) and IBC (International Broadcast Congress, Amsterdam). The set-up then was a DVB-H transmitter, a WiFi connection and an HTTP server to present the services on the terminals. The Annotation Suite was shown in offline mode.

Figure 4: Spur & Partner screen captures.

During the conducted end user validation the majority of test users singled out the interactive possibilities of 'Spur & Partner' as one of the strong points of the handheld variant of this well-known childrens' programme.

The service requires active involvement on the part of the users in order to explore the full range of interactive possibilities, and users' confidence and skill level rises significantly as they become more familiar with the device and the game. For example, recognition of the bounding box increases from 50% of users to almost 85% during the test.

On-screen feedback contributes greatly to users' confidence and involvement and conversely, where feedback is not forthcoming, users' ability and interest does not develop at a comparable rate. Unfamiliarity with the device was sometimes interpreted by users as a technical limitation in the tested version of the application but, with improved feedback, 'Spur & Partner' will be a compelling application for the development of multiplatform interaction between users and television.

7.2 iSports

This scenario has already been presented at the GSMA Mobile World Congress 2008 (former 3GSM, Barcelona). The set-up then included two QTek 9000 and a HTC Shift. The service was shown in offline mode. Latest implementations are the time shift module and the RTCP which are currently being tested.

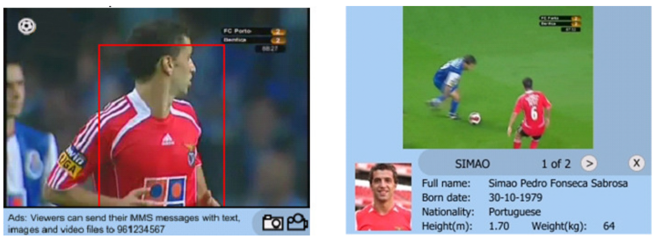

Figure 5: iSports screen captures.

[1] International consumer electronics trade fair

[2] German media congress and media exhibition

During the conducted end user validation iSports users were enthusiastic about many features of this application, with interactivity and access to additional content about the football match and players being particularly well received by all test users. An updated 'View the match' function allowed users to watch the live match while doing other iSports activities, and this was also highly rated.

All users accessed match information and found it to be 'one of the best features in form and content'. All users were able to view an image sent as MMS and all easily accessed such non-menu functionalities as MMS, the quiz, match info, etc. Users also efficiently carried out tasks involving the 'bounding box'. Other popular features were the 'Help' function and replays, and users also rated highly the ability to configure their own user profile.

Comments and recommendation were made about various aspects of the application, including the design of the bounding box. Once these developments are implemented, 'iSports' will serve as a convincing demonstration of the possibilities for personalised rich media mobile TV.

Devices

In their hardware preferences, users of both scenarios identified the 'Origami' device as the preferred option due to its size and portability.

8. SUMMARY

In porTiVity, rich content is linked to objects seen in a TV programme on mobile TV devices. User interaction, i.e. clicking on objects on the screen, allows the reception of this content. In this paper, the whole chain has been covered: tracking of the objects in the video; authoring of the content and association to the object; generation and processing of interactive information in MPEG-4 LASeR format; synchronisation and transmission of all the streams through DVB-H; and reception in a mobile personal terminal.

In a first stage, the system only deals with off-line content, but future work foresees on-the-fly live content annotation and interactivity generation.

9 AKNOWLEDGEMENTS

This work was developed within porTiVity, an European Project (www.portivity.org), co-funded under the European Commission IST FP6 programme.

10 REFERENCES

[1] B. Cardoso et al, "Hyperlinked Video with Moving Objects in Digital Television", ICME 2005

[2] Healey, P. et al., "Integrating Internet and Digital Video Broadcast Data", 4th World Multiconference on Systemics, Cybernetics and Informatics (SCI 2000), volume I, pages 624– 627, Orlando (USA), July 2000

[3] U. Rauschenbach et al, "A Scalable Interactive TV Service Supporting Synchronized Delivery over Broadcast and

Broadband Networks", Proc. IBC 2004 conference, Sept. 9-13, 2004, Amsterdam, The Netherlands
[4] DVB-H: ETSI EN 302 304 "Transmisión for handheld terminals (DVB-H)"
[5] Gerhard Stoll et al.. "porTiVity: New Rich Media iTV Services for Handheld TV". 2nd European Workshop on the Integration of Knowledge, Semantic and Digital Media Technologies. London, 2005
[6] MPEG-4: ISO/IEC FDIS 14496-20:2006(E)
[7] Material eXchange Format (MXF): SMPTE 377M
[8] MPEG-7: ISO/IEC 15938:2001
[9] H. Neuschmied, R. Trichet, and B. Mérialdo, "Fast annotation of video objects for interactive TV", ACM Multimedia 2007, 158-159.
[10] R. Trichet and B. Mérialdo, "Generic Object Tracking for Fast Video Annotation", *VISAPP*, Barcelona, Spain, 2007.
[11] R. Trichet and B. Mérialdo, "Probabilistic Matching Algorithm for Keypoint Based Object Tracking Using a Delaunay Triangulation", *WIAMIS*, Santorini, Greece, 2007.
[12] RTP, A transport protocol for real-time applications. RFC 3550.
[13] Björn Stockleben, "Approaches towards a Mobile Interactive Platform", Euro ITV Workshop on Interactive Applications for Mobile TV, Amsterdam, the Netherlands, May 24-25, 2007