APIDIS

Autonomous Production of Images based on Distributed and Intelligent Sensing

STREP Project, 1st FP7-216023

D2.1 End-user requirements, privacy issues, and architecture definition

Due date of deliverable: 30-06-2008
Actual submission date: 14-07-2008

Start date of project: 1st January, 2008 Duration: 36 months

Lead contractor for this deliverable: BM [Revision Final v1]

D2.1	End User requirements, privacy issues and architecture definition
Project Acronym :	APIDIS
Contract No :	FP7-216023
Due Date :	30-06-2008
Reply To:	Raquel Navarro Raquel.Navarro@barcelonamedia.org
Actual date of delivery	14-07-2008

Deliverable Identification Sheet

Project ref. no.	FP7-216023
Project acronym	APIDIS
Project full title	Autonomous Production of Images based on Distributed and Intelligent Sensing
Security (distribution level)	Confidential (CO)
Contractual date of delivery	Month 6, 06-30-2008
Actual date of delivery	Month 7, 07-14-2008
Deliverable number	D2.1
Deliverable name	End-user requirements, privacy issues, and architecture definition
Type	Report
Status & version	Submitted
Number of pages	71
WP / Task responsible	WP2 / BM
Other contributors	Mediapro, UCL
Author(s)	Raquel Navarro
EC Project Officer	Albert Gauthier
Abstract	Report of the work done in WP2 regarding the definition of the user cases, of user requirements, and of corresponding technological requirements. Specification of the goals and criteria that should guide the work in the rest of the work packages in APIDIS. Privacy issues are also considered. Future work is described.
Keywords	User Cases, User Group, Privacy, Technical Requirements
Sent to peer reviewer	EPFL
Peer review completed	Yes
Circulated to partners	Yes
Read by partners	Yes
Mgt. Board approval	Pending

Table of content

Executive Summary

The purpose of this document is to report the work done in WP2 regarding the definition of the user cases, of user requirements, and of corresponding technological requirements. Therefore, we have specified the goals and criteria that should guide the work in the rest of the work packages in APIDIS. The research reported in this document corresponds to the work done in Tasks 2.1, 2.3 and 2.6 during the first six months, in order to provide an integrated view of the vision of what APIDIS will deliver.

The document is organized as follows:

- Description of the user cases. In Section 2, we define three user cases and present relevant examples illustrating how APIDIS could help in each of the cases. The user cases are:
    - Automatic generation of content for Internet portals
    - Interactive and semantically-driven access to video surveillance
    - Application of APIDIS in the production room

- User group consultation and definition of the user requirements by user cases. Section 3 focuses on a qualitative approach in which we will gather in depth information about requirements from a reduced but informative group of users. We have used contextual observations and interviews to gather the data from the relevant actors. An interview guide has been created in order to increase homogeneity among the data collected by different partners. Relevant stakeholders ranged from TV producers, to end users like coaches, and experts in video surveillance areas.

- Privacy requirements. The legal and ethical issues that arose from the deployment of cameras and audio sensors in public areas are addressed in Section 4, together with their impact on APIDIS.

- Translation of user feedback into technical requirements. Section 5 presents the architecture envisioned to capture and process multi-camera content. It also defines the metadata that should support autonomous and personalized summarization of content.

General conclusions and future work are described in the last section of the report.

1. Introduction

The purpose of the document is to report the work done in WP2, regarding the definition of user cases, user requirements and corresponding technological requirements. Therefore, this document should guide the future research and development work in APIDIS. The research reported in this document corresponds to the work done in three of the tasks of WP2 (T2.1, T2.3 and T2.6) during the first six months in other to provide an integrated view of the vision of what APIDIS will deliver.

The goal of WP2 "**Requirements and functional specifications"** in APIDIS is to ensure that the project address real end-user needs, and can rely on experts' knowledge to fulfill these needs.
From the point of view of the User Centred Design, it is required to design technology based on a thorough study of the people who will or are expected to use it, because users will be the indicators of the success or failure of the solution. The best design of a product is the one in which the designers really appreciate user needs[1].
In this context, the objective of Task 2.1 was thus to conduct user studies and obtain significant information about people's needs, to ensure that the services and technologies that will be developed in APIDIS will not be the result of speculations but the result of the research on user expectations, problems, needs and opportunities.
Previous to gathering the user feedback, we have defined three user cases that allowed us to focus our investigation. We have then created a User Group that has been consulted to support the requirements definition. In this document we will describe the methodology used to gather this feedback and the results that we found. This is done in Section 2 (user cases) and 3 (user requirements). In Section 5, those learnings are exploited to drive the design of an architecture for autonomous content acquisition and personalized summarization. For completeness, Section 4 addresses legal and ethical issues related to the deployment of a network of sensors in public areas.

2. Overview of the user cases

The goal of the APIDIS project is the autonomous and/or interactive production of visual summaries from a network of audiovisual sensors. This document presents three use cases devoted to the 'automatic and personalized summarization of sport events', the 'interactive and semantically-driven browsing of video surveillance content', and the 'application of the APIDIS technologies in production rooms'. User cases have been respectively described by UCL, ACIC and MEDIAPRO.

The goal of user cases is twofold. First, it provides a pragmatic and federating initial understanding of the kind of applications that should be targeted by the project research,

[1] Beyer, H. R. and Holtzblatt, K. (1999) '*Contextual Design*'. Interactions / January and February, 1999.

which is useful to define a relevant development framework for APIDIS (see section 5). Second, it allows us to formulate the questions for the user group (see Section 3). Therefore, the description of the user cases, including their technical constraints and expected user interaction capabilities, provides a fundamental input to get a complete picture of what APIDIS would do, and how it would be useful for real users.

2.1. Use case: "Automatic generation of content for Internet portals"

2.1.1. Context and objective

The purpose here is to validate the federating idea of APIDIS, i.e. the automatic generation of personalized sport-event video summaries based on distributed and omnidirectional sensing.

In a typical application scenario, the APIDIS acquisition sensor network is composed of microphones and cameras, which for example cover a basket-ball field. Both conventional and (arrays of) omnidirectional cameras are considered. Distributed analysis and interpretation of the scene is exploited to decide what to show or not to show about the event, so as to produce a video composed of a valuable subset from the streams provided by each individual camera(s). In final, the system provides a solution to cover local (sport) events at low cost (no technical team or cameraman is involved anymore). The content could be sent or made available through a web-site or a mobile VoD service, to the basket-ball supporter, to the journalist or whoever.

By automating the production, APIDIS enables content access personalization. Generating a personalized summary simply consists in (re-)running the production process with input parameters corresponding to the specific constraints expressed by the user.

2.1.2. Infrastructure: functional and technical constraints

The *technical constraints* deal with the specifications of the acquisition sensor network, and with the infrastructure and mechanisms exploited to support the storage, the access and the manipulation of the content.

Regarding the acquisition network of sensors, we note that:

- the system is limited to fixed -potentially omnidirectional- cameras;
- the system captures side information such as the one driving the game control panel;
- the system also captures and locates the origin of key audio signals, typically referee whistle;
- cameras have a high resolution (>1Mpixels, >20fps);
- cameras have to be synchronized, so as to enable joint analysis of their images;
- camera calibration is a pre-request to joint processing of visual information. The way it is implemented remains open and subject to a cost/accuracy trade off studied during the project;

- camera field of views should be redundant enough to perform joint analysis of the scene, and to offer sufficient flexibility in terms of rendered views.

Regarding the storage and manipulation of content, we note that:

- the amount of captured content is huge (>Mbyte/s), so that compressed and hard drive storage is required;
- metadata associated to a game are likely to be split into annotations that are specific to a video stream or global to the session;
- metadata might be split into contextual (where and when) and scene-dependent (what happens in the scene) information;
- the user interface should allow for the introduction of manual annotations, so as to support inputs from individual end-users, e.g. a coach or a supporter who wants to identify key periods;
- metadata associated to a game should be partitioned into consensual information, and individual preferences.

The *functional constraints* correspond to the set of requests potentially expressed by the end-user about the summary (s)he wants to access. Those requests might include parameters like duration/resolution/bandwidth and some spatial (e.g. only the actions happening on the right side of the field) or temporal (e.g. only actions occurring during the last quarter time period) window of interest. They can also define more specific features like a player of interest or an action of interest, or a combination of both kinds of requests. A priori, the actions of interest can be defined based on shooting reports commonly used by basket-ball trainers. Typical actions of interest are shoots and free throws (+ indication of success or failure), ball losses, ball gains, faults, rebounds, assists. Duration and complexity (lots of passes, rapid counter-attack) of actions also provides useful hints to summarize the game. Allowing the end-user to select a player of interest implies that player recognition and tracking capabilities are supported by the analysis system (ideally, the project should reveal whether this is realistic or not).

The set of potential requests translates into (1) a set of events to detect from captured video streams; and (2) a set of requirements for annotation syntax and semantic.

2.1.3. Relevant examples of applicative scenarios

Content is captured by a network of sensors distributed around a basket-ball field. Information displayed on the game control panel is also recorded. A pre-defined set of events-of-interest are automatically detected based on video analysis. Events typically correspond to global behavior in the game (free throws, etc), but could also be related to one specific player (e.g. to monitor each individual player). The content is annotated and stored in a way that allows for personalized access, and for further manual annotation.

Two application scenarios are considered:

- First, the end-user defines his content access requirements, and content summaries are produced based on those requirements. This reflects a scenario where content is produced in a completely autonomous way.
- Second, incremental and manual annotation of the content is considered to further control subsequent summarization trials (e.g. by focusing on specific actions or

players). It corresponds to a semi-automatic scenario where some end-users (a coach, a player,…) further enrich the automatically generated content to highlight some of its segments. The access flexibility, the usability of the interface, and the fluency of the generated content will be evaluated as a function of the deployed infrastructure (number of sensors, type of sensors, etc.).

2.2. Use case: "Interactive and semantically-driven access to video surveillance content"

2.2.1. Context and objective

This use case considers the content acquired by a network of surveillance cameras, covering a building or a secured area. Typical examples of candidates for the APIDIS system are CCTV installations in confined areas such as metro/railway stations, airport, shops, shopping mall… However, APIDIS is not supposed to be integrated on any kind of CCTV network. There are some technical constraints at the system level and at a logical level. There are also some higher-level constraints at the functional level. These constraints are explained below.

2.2.2. Functional and technical constraints

From a functional viewpoint, the user takes a benefit from the APIDIS technologies when their usage results in a gain of time and/or a better accuracy in the search for visual events. It implies that the APIDIS technology should support efficient browsing of pre-recorded content, to visualize an event that is too complex to be detected completely automatically at recording time. An example of a counterexample where APIDIS does not help is the case where a very simple event is observed on a single camera. In this latter case, the summary is obvious and is a simple concatenation of video segments corresponding to an alarm. Such systems exist already.

APIDIS is thus expected to offer a real added value to security managers:

- By allowing the end-user to define some kind of event of interest a posteriori, typically based on the definition of a spatio-temporal area of interest, and a set of low level features captured about the scene;
- By allowing the end-user to interact with the summarization system to compensate and deal with the inaccuracies of the automatic analysis performed at recording time.

Two examples in which APIDIS may help are:

- A case for which a machine has been damaged in some area covered by the cameras. In that case, the user can point to that spatial area and tell that s(he) is interested in some activity in that area that has affected the rendering of the scene. The system can then rapidly converge to the instant at which the damage occurred based on an interactive bisection mechanism, because the user can interactively tell whether a given key frame is before or after the damage or event-of-interest.
- A multi-camera environment. In that case, the system might present to the end-user the main view of interest, together with a set of surrounding related views, so that the user can interactively decide to switch from one camera to another,

e.g. to follow one specific person in the scene. In a typical scenario, based on scene activity analysis, the system is able to assist the end-user in the selection of the most appropriate views to observe an event along the time, whatever the event is[2].

A key issue consists thus in the definition of the interactivity concept. Which kind of interactive commands should APIDIS support to be helpful? What kind of events would benefit from a semi-assisted interactive browsing system?
Those questions should be the central issues raised during discussions with the user group.

Technical constraints
Very obviously, there are some constraints on the nature and the position of the cameras. First the cameras must have a sufficient resolution and image quality. Recent CCTV networks most often verify this condition. The cameras must be fixed, or at least static. One can use PTZ cameras when they are not moving. Wide-angle and omnidirectional cameras are also suited for APIDIS purposes.

Concerning the position, if we suppose that all cameras were installed for optimal surveillance purposes. At least, their video streams must be accessible by the APIDIS system at the time of the summary computation. This means that all necessary metadata must already exist (produced automatically by the video analysis or annotated manually), including the camera ID, the time and date.

This is not enough, since it is nonsense to produce a summary with video extracts from cameras that capture non correlated scenes. The events detectable in their respective field of view must be similar on correlated. This constraint reduces the network to a set of cameras enabled to observe correlated events.

User's interactions with technology
In the CCTV use case, interactivity is not an essential functionality, or at least can be very trivial. In most cases the summaries produced automatically will be sufficient for a day to day rapid monitoring. However, in a more 'forensic' usage of the system, the user might want to influence the production of the summary to focus on specific area or events of interests. Imagine the day after an accident or a robbery in the facility of the company. The user knows a priori information he wants to input to the summary production system, like information about time, location or nature of the event. Of course, the main purpose of the project is not to develop a forensic solution, but to offer a service for gaining time by allowing the user to browse the content interactively and in an assisted manner (annotations defining global scene features may help to focus on spatio-temporal segments of interest). Those global scene features have to be defined by the end-user, to be translated in scene analysis tools and in annotation syntax and semantic.

Synthesis of the applicability conditions
- fixed cameras (or static PTZ) on the same network (can be omnidirectional)
- average to good quality and resolution of the videos

[2] Reinforcement learning could be envisioned here, for the system to learn which view to display (or to propose to the user) based on scene activity measurements. Learning could be based on automatic detection of pairs of identical objects in different views.

- events visible on camera streams used for the summary production must be correlated
- events of interests must not be too frequent
- events of interests must not be difficult to detect
- beyond specific events, APIDIS might be interested in scene features extraction
- metadata associated to video streams must be reliable, in the sense that they are not expected to be updated or refined a posteriori
- interactivity should be offered to support efficient browsing of the videos, in a way that improves a purely manual solution

2.2.3. Relevant examples of APIDIS technologies applied to CCTV networks

A good example where APIDIS helps is the surveillance of an industrial site. The automatic surveillance is especially helpful when the activity is low (after work or at night when the activity is limited). The events of interest concern an abnormal human activity at the perimeter of the facility or at very sensitive locations (e.g. reactors). In the morning, the security manager of the factory wants a very quick summary of what happened during the night, what the night watchman did not notice or did not consider as abnormal. If the criteria of interest are well defined and the APIDIS system is well trained, the summary will contain only the sequences of events that are of interest for the security manager (the entry of a person or vehicle, the trajectory of the latter, the grouping of workers etc.). This of course also applies to the director of an industrial SME trying to understand why the alarm of the security system did wake him up for the third time in one week and analysing the recording of its six cameras. This guy has no time to waste and wants an accurate and fast access to the information.
A second example considers a forensic-oriented scenario, which envisions the search for evidence (e.g. after a crime or an accident, on the road or in some public area) through a video surveillance data base. In that case, the captured video content is stored together with global scene characteristics, which are computed at recording time. Interactive requests from the end-user (e.g. the policy), formulated in terms of the global and pre-computed features, define the events of interest, which permits to accelerate the search.

2.3. Use case for application of APIDIS in the production room

Context

In APIDIS, we only aim at supporting professional content producers in the re-use of pre-recorded content. In this section we thus consider most of the content management issues raised by professionals when remixing content. This section is organized in subsections in which each one corresponds to a use case expressed by the professional users who work with MAM (Media Asset Management).

2.3.1. User's interactions with the technology

In this section we will describe a set of user tasks for which APIDIS could potentially provide some support. In the meantime, we will try to gain knowledge about which of these tasks would be of most interest for the professional content producers.

2.3.1.1. MAM cleaning

When you have a lot of indexed material, it is almost inevitable that different keywords refer to the same concept. This makes later searches difficult. It will be interesting to have a tool to detect these cases and present them to the user for correction.

2.3.1.2. Improve the annotation

At present, in professional content production the recorded materials are delivered to the content producer company as tapes, without the possibility to include extensive metadata. Typically, the annotation is done manually in a set of fixed fields (title, type, etc.) and another set of fields of free format (description, production notes etc.). APIDIS technology may be a great improvement by providing a set of initial descriptors already in the raw material or in the form of some 'twin' associated file. This will allow the annotators to improve not only their efficiency but provide information that is not captured in today's annotations like the importance of a shot automatically inferred through the capture of actions performed by the producer or the cameraman (e.g. a change of camera view, a zoom, a replay, etc) which might reflect the importance of a video segment. Open questions lie in the feasibility and relevance of deriving information based on those producer/cameraman actions.

2.3.1.3. Searching a shot for re-mixing

While APIDIS is not a search engine it will support efficient content browsing by the improvement of annotation. APIDIS will help to have metadata associated to shots and therefore it would make it possible to search through the metadata associated to frames and not only to the manually entered annotation in the text fields inside the MAM.

However, it is not clear at this point which criterion is used to select a shot in real time? Such knowledge would be critical for APIDIS to help the selection of shots of interest in post production. Therefore, Mediapro plans to further investigate how to monitor the actions from the producer, and how to exploit that information to infer the importance of a video segment. In particular, Mediapro will collect video data, and associated production actions, so as to provide reference material to demonstrate the usefulness of APIDIS content browsing facilities while remixing content in a production room. Part of this investigation will be reported while defining the specifications of the trial associated to production room (D2.3 and D2.4). Some preliminary ideas are:

- Some of the edited replies are not broadcast, a metadata could be added with the degree in which shot is interesting/good for "promotional material", sports programs; or mark a replay that has not been broadcasted.
- Detect insertion of advertisement, zoom-in, or replays in the broadcast content, so as to infer a level-of-interest for the corresponding video segment.

2.3.2. Technical and functional constraints

The technical constraints vary a lot from one MAM to another because production room infrastructures have very different characteristics, both in terms of how they capture and how they store the multimedia material. There is no standard set of requirements in previous literature as the requirements vary tremendously from a local television to a national or even international channel.

The functional constraints are derived from the type of interactions that are allowed for the users of the MAM. Any changes in the possibility allowed for the user has a big impact in the workflow of the user activity and the organization as a whole. Therefore, one of the major challenges is to understand how APIDIS can help the users integrating new functionalities inside their existing workflow. In addition, it is important that the technology to be developed in APIDIS does not force the users to create new content from the existing one, but rather provides them the information that allows them to take their decisions more quickly and effectively.

In order to gather more knowledge not only on the user requirements but on the technical constraints we will report the information gathered by two types of professional content providers: a national channel and a small local television.

2.3.3. Relevant examples of APIDIS technologies applied to production room

A good example where APIDIS technology could be useful for the production room is the creation of summaries. Actually, when the producer is working on the summaries of a round from a sport league, he receives the images of 10 games which represent a big amount of sequences to watch in order to prepare the summaries which are going to be sent to other international TV channels. All the process requires hours of work. If the annotation and indexation is well defined, the APIDIS system could support the autonomous extraction of the video segments that contain the events/actions that are useful for the producer, such as goals, cards, best plays etc. It will permit the producer to gain time. These developments also apply to a journalist who will have a direct access to the right shots.

In addition, the metadata associated to the content could also distinguish which sequences will be used for summaries for special week-end programs and the ones used for internal promotion of the TV channel.

3. User Requirements by user case

After we described the user cases, we have a much clearer idea of the specific APIDIS goals and the information that we needed to gather from real users. In particular we know the gaps in our knowledge that we need to gather. In addition, we will try to validate both the functionalities foreseen in the user cases and how the users would like to interact with the system. Logically, our next step was the creation of a User Group, which is a group of key informants that would provide the information that we need not only now but through out all the life cycle of the project. This section reports on the process to gather the information from this User Group and the main results from this task.

We need to consider this first requirements gathering task as the beginning of the collection of information about users. It will be complemented with the information collected when presenting proof-of-concept and final trials to the user group.

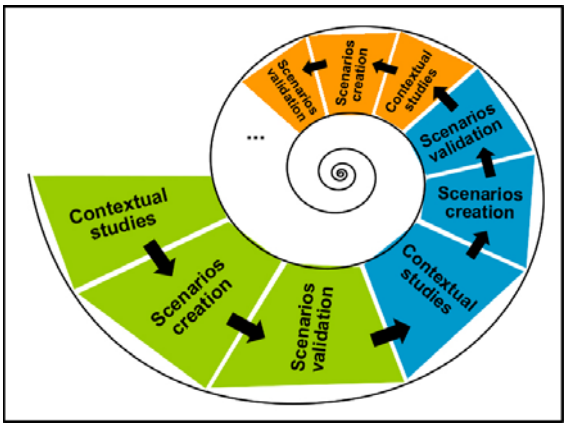


Figure 1:. *Methodological approach.*

3.1. User group creation

From the beginning, it was considered critical for APIDIS that the requirements would be based on real user's feedback about their day to day work, the problems they encounter at present, and their expectations about how a system like APIDIS may help them. Therefore, we needed to create a board of users which could give is the information required for each user case.

We consider this task to be so important that we decided that all the partners should take responsibility in using their industrial contacts to get the maximum possible number in the User Group, and to ensure that they were real experts that could provide us quality feedback. A plan was establish, coordinated by BM, to create all these contacts. This way we will also try to ensure that the user requirements for each user case is not based on informants from one country but spread as much as possible among all the countries represented in the project.

In order to help with the task of explaining APIDIS to the potential users, a User Group Flyer was created (see APPENDIX 1).
The result of this process is summarised in the following table.

Apidis partner who owns the contact	Contact Name and user case in which they are interested
From UCL	Alcatel-Lucent: • Currently develops an infrastructure for **personalized content delivery.** • Scenarios/Interest for APIDIS: At the source, they want to support multiple kinds of content providers going from the professional to the man in the street. At the sink, they consider **end-users accessing content through mobile handsets** or ADSL links. End-users get the capability to interactively skip some content within a pre-recorded list. We can also imagine that they interactively browse the content. In some scenarios, two contents provided by distinct sources might be merged before display (e.g. insertion of advertisement in a video).
From UCL	Canal C: - Local TV.==> **sport events summarization** - Interest for APIDIS: Canal C follows regional **sport events.** Besides, they have a weekly 30' program about research at university.
From UCL	RTBF/IAD, Vincent Housiaux: o Sport event summarization o Content producer
From UCL	Basket-ball coaches. o Browsing video for training purposes.
From UCL	Sport&Media journalism: Prof. Gerard Dereze (UCL/COMU). o Sport event reporting.
From QMUL	BBC The BBC would be interested in joining the APIDIS user group and to contribute to the **sport events scenarios** only (not the surveillance ones). It is a unit of the BBC that is looking at ways to manage multimedia user-generated content.
From MP	"Sexta" TV channel: - **Sport event** - Person responsible for MAM and visit to the annotation department - **Content producer** (productora): control room End users for user case 1: high performance school coaches
From ACIC	**Surveillance scenario:** - Hubert Lenoir from the Belgian Police (special forces) - Didier Nicholson from Thalès Communication in Colombes

Table 1. User Group for user requirements gathering

3.2. Methodology for user requirements gathering

We focus a qualitative approach in which we will gather in depth information about requirements from a reduced group of users, but all of them are key informants for us. As there are several methodologies to gather user feedback, we studied which one would be the most appropriate for the stage in the project cycle and the type of information that we needed to gather. A document was created that summarises the

main advantages and disadvantages of four ways to gather user feedback: questionnaires, interviews, observations, and scenario based **(APPENDIX 2)**.
We discarded questionnaires in spite of the fact that they could provide interesting quantitative information from a wider sample because the questions would inherently and inevitably restrict the nature of the collected information. In other words, questionnaires are not recommended at the very first stage of the requirements gathering because they could force the project to narrow the focus on a set of options at a time when we needed to understand which functionalities were more interesting and gather information about users' day to day work for understanding why.

During a project meeting it was agreed that we should use a combination of natural setting observations and interviews. Our goal was to have at least one observation and several interviews for each user case.
In order to make our results compatible across users, it was very important that we followed the same methodology and questions in each of the interviews and observations. Towards that goal we created an observation and interview guide for each of the user cases. That guide would serve as a check list to make sure that the partners collect all the required information from each user.

For the user case related to **automatic generation of content** we were interested in observing in a real context all steps of how journalists produce content summary in all its steps, from the preparation before the event to the long term storage of the content. It is indeed very important to understand what are the main actual problems and what would be the impact of being able to offer personalized summarization mechanisms.
Regarding the questions in the interview, we differentiate:

- Questions about the work-practices for companies that record sport events:
    - They describe step-by-step how they prepare for the recording of a sport event
    - Rules and constraints particular to a sport event
    - They describe step-by-step how they record a sport event
    - They describe step-by-step how they use the recorded footage
    - In which occasions would you find it useful to have an automatically generated summary?
- Questions about their expectations from an industry perspective about the major requirements and the how future users will personalise their own content.

For the **surveillance user case**, our main goals were to understand their work practices: how the users prepare for the recording of a CCTV (problems, barriers, expectations), when they record, and what they do with the recorded materials. In addition, we were interested in knowing about the main barriers they encounter while exploiting video recordings, and which kind of video summarization or browsing functionalities would help them in their everyday tasks.

Finally, for the **production room user case**, we wanted to understand how they handle:

- Annotation of a video: step-by-step, technology use, co-ordination with other people, work flow, problems, etc.
- Video re-use: what they do with the recorded material actually (problems, barriers, expectations)
- Video search: how they do it and major problems.

- Long term repository of the recorded materials: details on tools/procedures/ specialised roles/etc

3.3. User requirements for "Automatic generation of content for Internet portals"

In this section we summarize the main findings arising from the interviews and observations made with the key users in each one of the use case areas. In order to gather information from professional and from end users, we have interviewed expert users from TV channels and coaches that would be the end users of the system. We also contacted leaders in the design and deployment of content delivery infrastructures, to better understand how they envision the notion of personalized and potentially interactive access to content. A rough summary is provided in Table 2.

Table 2. Summary of the visits for the automatic content generation user case.

User and Role	**Area of investigation and main learnings**	**Methodology and APIDIS partner that made the contact**
Live observation of the retransmission of a football match of the Spanish league from an OB van. Interview with all the key members of the team.	Organisation and preparation of the cameras and logistics for the recording. Live retransmission of the most important information coming from several cameras. Best highlights selection. Creation of the event summary. Audio component management.	Contextual observation MP and BM
Marie-Pierre FIVET has been producing and commenting video summary reports for Canal C local television (www.**canalc**.be) for more than 5 years. She also relies on a solid experience as a semi-professional basket-ball player (National D1) to identify the main components of a game (tactical decisions, key players, representative actions, etc.), and help the viewer to understand them.	Identification of key principles governing the summarization of a basket-ball match.	Interview C. De Vleeschouwer, UCL.
Two coaches: - Jurgen VAN MEERBEECK (JVM) is the coach of DEXIA NAMUR, first national division. - Jean-Marc ALBERT (JMA) is the coach of a local non-	Getting feedback about potential usefulness of the APIDIS acquisition and summarization system.	Interview C. De Vleeschouwer, UCL

professional team (IATA, playing in the 3rd provincial division in Namur).		
Vincent Housiaux, RTBF/IAD, -leading the department guiding the artistic aspects of RTBF programs. He has been the producer of numerous television programs, including sports events, mainly football broadcasts and summarizations (p.e. sequences focusing on a player during EURO2000 in Belgium) - invited professor at IAD (www.iad-arts.be/v_british), an artistic postsecondary school in the domain of the performing arts, media and communication techniques.	Multi-camera television production mechanisms.	Interview C. De Vleeschouwer, UCL. 21/03/2008
Joan Amat. Joan is the trainer of 2 teams at a high performance school: - 1st team: boys 14-16 years old - 2nd team. girls same age	Feedback about the potential utility of the APIDIS acquisition and summarization system.	Interview Media Pro
Alcatel-Lucent. Benoit Quirynen is the head of the video department of Alcatel-Lucent Namur.	Information about expected future deployment of open IP (in contrast to the conventional walled garden paradigm) content distribution infrastructures. Personalized and interactive access to content.	Interview UCL

In the rest of the section, we only present the main conclusions drawn from those interviews and user group consultations, regarding the APIDIS project. For completeness, Appendix 4 details the main outcomes of a methodical analysis and integration of the data collected during the different consultations. A complete transcription of related interviews is also provided in Appendix 4.

3.3.1. Main conclusions

From the above discussion, we can draw a number of lessons for the APIDIS project:
- It is relevant to identify events of interest in the game, and derive a summary based on those events.
- It is important to drive the summarization process based on some underlying message about the game. In APIDIS, the user preferences introduced by the user about the actions he is interested in should partly fulfill that requirement.

- Interviews of key actors provide a natural support to the summarization. APIDIS should perhaps offer the capability to record some pre-formatted interviews. It raises an interesting open question.
- Deployment of multi-camera systems makes sense.
- Axis rule: Try to preserve consistency in players' movements and gazes. For that purpose, all cameras are located on the same side of an axis that is chosen in the main direction of the game.
- Implicit correspondence between the action and the way the scene is captured by a cameraman. Hence, if APIDIS recognize an action, it can rely on hard-coded rules to infer how to render it. Again, APIDIS should perhaps offer the capability to record some pre-formatted interviews.
- Many actions in a basket–ball game, therefore the camera can miss some events. To some extent, APIDIS can thus afford an incomplete coverage of the field.
- For the football match, several key actors intervene in the selection of the shots that have been broadcast.
- Next to the 'conventional' summarization process, coaches are interested in a manual or semi-automatic *browsing* of the content, based on the identification of events and actions of interest. It is close to the video surveillance scenario.
- APIDIS considers posterior summarization. Hence, we might exploit the fact that we know in advance who will score to select the image to show.

3.4. User requirements for "Interactive and semantically-driven access to video surveillance content"

Surveillance scenarios can be quite different from one application to the other. Therefore we have gathered information from different perspectives, from the police to the private enterprises, and from the end users to an important player in the software industry for surveillance. The expert consultations are summarized in the table below.

Conclusions:	**Area of Expertise**	**Methodology and APIDIS partner that made the contact**
	Surveillance for police	Interview ACIC 16/04/2008 – Mons, Belgium
From the above discussion, we can draw a number of lessons for the APIDIS project:	People detection	Interview ACIC
- It is relevant to identify events of interest in the game, and derive a summary based on those events.	Industry perspective on the needs for surveillance software in a variety of situations	Telephone interview 29 April 2008

Table 3. Summary of the visits for Interactive and semantically-driven access to video surveillance content user case.

In all our interviews we discussed with the users our two scenarios:

- Search within a video ➔ gain of time and/or a better accuracy while browsing content, searching for visual events, e.g. in a forensic context.
- Interactive navigation through a network of cameras ➔ the APIDIS system could support natural switching between camera views when following a specific object of interest.

3.4.1. Police

The police typically deal with the content acquired by a network of surveillance cameras, covering a building or a secured area. We first focused on their current work practices during four distinct stages:

- Recording request: Usually, the justice and/or the police (the investigators or interviewers) delegates CGSU for the close surveillance of a suspect. Typically they place 1 to 4 cameras to observe the activity in the house of the suspect (the entrances), who comes in, who goes out. Typically faces and license plates are extracted from the video. The investigator chooses the location.
- Recording settings: The cameras are hidden in a neighbor's house, behind the window. Their video stream is recorded on a portable DVR. This point is a major issue. Since the equipment is hidden, it is difficult to go often to download the videos and empty the HDD of the DVRs. Usually, the person hosting the system brings the HDD to the police. This puts the person at risk and the Police want a new system. Typically, there is a need to efficiently compress the video (by storing only the important moments) and send a summary of the day by radiocommunication (UMTS?).
- Creation of summaries: In practice, every time the HDD is sent back to the police, they have to watch all the recorded material and write a report for the viewed period. They use a laser telemeter pointing at the door to launch the recording when the door opens, but it is not enough. So, they have to reprocess manually afterwards. Sometimes, they use post-processing software to search backwards in the videosequences.
- Long term repository of the recorded materials: The repository is kept until the judgment, the archiving support is CD/DV, note that VHS tapes are still used at the Police. They just need a mandate to start the recordings. Then the confidentiality of the recordings must be granted.

Based on that feedback, we identified three major problems:

1. The difficulty to access the recorded material, as explained here above, and the risk for the police or the person hosting the equipment, when the transfer is done physically. Due to the amount of recorded material, it is impossible today to send the video streams through wireless infrastructure.
2. The requirement not to miss any event. Non-detections must be avoided, false alarms have to be minimized, but are less critical than false negatives!
3. The system must be portable and easy to install and must work day/night in all conditions

Regarding their **needs and requirements for video summaries**, our main conclusions are:

- The impact of having summaries would be enormous, from both efficiency and security perspectives, provided that the summary does not introduce missed detections.
- They will use the summaries once a day to once a week, depending on the cases
- They need to get all the faces and all the car license plates A very reliable system (no missed detections) and an efficient summarizer, which can be sent/viewed remotely through UMTS. 99,9% detection rates, recognition of the faces and car license plates.
- The system must be easy to use and install, portability, and the calibration of the system must be fast and discrete (the suspect can not see it).

3.4.2. Private Building Security

In the case of private security and surveillance, we have interviewed experts in two different environments:

- Nicolas point: User whose work is the creation of video summaries to monitor and control a big research building. In their case, the major problem is to know who the last person to leave the building was and who the first to arrive was. The building is equipped with an alarm system and they need to investigate who was responsible when a false alarm is raised (wrong code, doors left open). They currently have a script that produces a summary of the entrances and exits of the day. An Axis camera points towards the main door. They mainly rely on the motion detection function of the main door.
- Li-Qun Xu: User that is representative of companies producing software for video surveillance.

The main problems found are:

- The detection system is limited. This creates false alarms and missed detection. The produced summary is sometimes not accurate. Moreover, the program is not capable of filtering out the moments when a group of persons stay in front of the camera for some time. More globally, Nicolas Point would like a better image quality while at the same time keeping the volume of data reasonable.
- Illumination issues (when the person switched off the light before setting the alarms on/off), people staying in front of the camera, missed detection, false detections
- There is the need to define what a meaningful event is (depending on the application). For example, are moving objects of interest? Which ones? Or simply their direction of motion? Or if they are walking/running? Need to define ‘meaningful movements’ depending on scenario/ the application

While both users validated their interest in our user case, the interest on APIDIS technology depended on the type of expert user:

- For the person producing summaries, the interest is to gain time, but he would also be interested in the capability to send the summaries remotely (on the week-ends).
- For the software company, they thought that APIDIS technology would make them more competitive, especially if you can reveal to a customer some insights (business intelligence, statistics, etc) of the monitored scene that he did not know about.

The user needs and requirements for video summary are

- it would be of interest to summarize videos for a specific time slot (e.g., typical traffic situation on Monday mornings, without going through the whole video); or to provide video surveillance of a front door (motion trigger): 30 min video summarizing one day; the summary should be action/event driven, accompanied with the time stamps for each action. What has happened in front of your door during the day when I was away? (E.g. junk post advertise, charity collection …); or camera with dual use (summary of behaviour of birds in your garden during the day would be of interest for a keen bird watcher …)

- They would need to produce a video every day. The video is viewed everyday, but no archiving is done except when a major security problem happens, which is rarely the case.
- To be able not to miss entrances/exits, with no problem coming from the illumination. For one set of users, the ultimate goal is to know the last person exiting the building between 7PM and 6AM (when the alarm is set on) and the first person entering the building.
- Use IP-based systems, video can be stored in networked video recorder, locally or remotely, and accessed from anywhere in the network. Use of a standard video management system (e.g., for 20-30 cameras, use MPEG-4 or MJPEG).
- The importance of having the end-user in the loop and allow for on-line processing to refine/cluster events together.
- Regarding which attributes (time, etc) are used to browse the content, time is a standard attribute (that will always be needed). Specific objects like car or pedestrian are also relevant to allow for browsing/queries. Spatio-temporal features are also important.
- An important aspect is that the summarization system needs to work in real situations, on large datasets (real application scenarios) and to be reactive.

3.5. User requirements for "application of APIDIS in the production room"

For the production room scenario we were particularly interested in supporting professional content producers in the re-use of pre-recorded content. In this section we thus consider most of the content management issues raised by professional when remixing content

User and Role	Area of Expertise	Methodology and APIDIS partner that did the contact
Vincent Housiaux, RTBF/IAD, -leading the department guiding the artistic aspects of RTBF programs. He has been the producer of numerous television programs, including sports events, mainly football broadcasts and summarizations (i.e. sequences focusing on a player during EURO2000 in Belgium) - invited professor at IAD (www.iad-arts.be/v_british), an artistic postsecondary school in the domain of the performing arts, media and communication techniques.	Multi-camera television production mechanisms.	Interview UCL. 21/03/2008
Thierry Delrue, TDE@rtbf.be,	Strong experience in multi-camera sport event production.	
Team of people in the "La Sexta" TV national channel responsible for: - supervision of all the content annotation and storage - annotation of the content - providing pre recorded video to other parts of the company - maintenance of repository of video material in the MAM	MAM, video summarisation, video annotation, video searching	MP Field observation and field interviews
Person responsible for creating summaries in La Sexta Oscar Tomas	Expert on creating summaries in post production	MP

Table 4. Summary of the visits for the application of APIDIS in the production room user case.

3.5.1. Summarization of multi-camera sport-events

According to Vincent Housiaux, the **summarization of a sport event currently** implies that a journalist surveys the whole match, and keeps tracks of the time stamp associated to an action of interest together with its level of interest (via a number of stars: one star means 'worthwhile action'-four stars corresponds to a 'crucial action' such as a goal). We conclude that the approach is close to the salient segment identification envisioned by APIDIS.

Current approach to **content personalization** is pragmatic and does not really involve viewer's feedback. It mainly consists in making programs available on the web portal at multiple rates.

3.5.2. Post production in a national TV channel

In this section we will try to answer some of the open questions we asked in the a priori descriptions of the user cases, where we envision a number of user interactions that would support content remixing. We have gathered the information from a national TV station in Spain.
In particular we have studied:

a. Annotation of video
b. Video reuse
c. Video search
d. Long term repository and MAM cleaning

a. Annotation of a video: step-by-step, technology use, co-ordination with other people, work flow, problems, etc.

Our main findings about how they annotate is that there are two main processes of annotation:

- Pre-catalog: they create number of identifiers of the clip (ID, journalist, copyrights, time, date, duration, international/agency or national), all them added manually ➔ they would like an automatic ID
- Annotation of the news and the frames, which are kept in two different records:
  - Annotation of the frames: which describes the frames based on several criteria: type shot, if it would be useful for promos,
  - New cataloguing: which keeps the main info from the new agency

The annotation of the frames and selection of the storyboard is based on two main types of knowledge:

– Knowledge of the broadcaster, that is to say inhouse knowledge, of what may have been valuable before
– Active collaboration with the journalist that tells them what is the most relevant information

The tools that are now in use to help in the annotation:

– Annotation guidelines from direction
– Internet: APTN, Reuters, EFE, etc. In other words, external data sources.
– Thesaurus (keep by the director only)
– Wikipedia and R.A.E

The type of descriptors of the frames that we found are**:**

- type of shot
- "promo material"
- "interesting shot", "supergol" "person talking"

The criteria to select frames for the storyboard is subjective, depending on the annotators expertise. It is a bit more clear what the automatisation for football is because they have clear guidelines about the type of shots they will need for the summaries and promos.

b. **Video re-use: what they do with the recorded material actually (problems, barriers, expectations)**

We have observed how they select the content to keep and store in archives. This decision depends on where that content comes from. If the content is National content the journalist makes a summary and sends it to archive department. On the other hand, for International content the decision is taken by the archiver.
The usage and type of material that gets re-used depend on the TV section , for instance "promos" or "sport summaries" ask for content, for instance if it is for a documentary or promotional material.

When the users want to access stored content, this requires several tools (AVID; Unitecnic, Insctinct (iNews (only text) & Media Manager), TDIAL, Newscutter, Flow monitor). Each content has an ID and the mark of the disk where it is stored, typically in WM-MPEG2. If the content is less than 2 days old it will be very quickly accessible, if not it is in disks. Because of the number of simultaneous requests of content in the MAM it is important to be able to prioritise the requests.

c. **Video search**

The Descriptors used at present are:

- ID
- Time code (the most important)
- Place of the news
- "Visual Material"
- "Hidden Material"
- Visual entity (i.e. government, a company, etc)
- Hidden entity (i.e. government, a company, etc)
- Onomástico (name of the place, i.e. Barcelona) visual
- Onomástico (name of the place, i.e. Barcelona)hidden

Thesaurus to make sure they search with the right word.

d. **Long term repository of the recorded materials: details on tools/procedures/ specialised roles/etc**
   - All the programs are duplicated in Barcelona and in Madrid.
   - The most recent material is online and the rest on LT03 disks which are physically stored on shelves
   - Madrid uses DV and in Barcelona BP

User requirements conclusions for production room applicative context:

Based on the above arguments, we conclude the following:

- **Annotation: improved (not only words) annotation is needed**
  - Voice and face automatic recognition
  - Automatic metadata in pre-catalogue (specially date)
  - For News: they would like Interplay, just one tool
  - For News: iNews is too basic
  - iNews to prevent errors in ID
  - Automatic ID is needed
  - Need help with the language for the annotation
- **Video reuse: Automatisation and MAM cleaning**
  - Automatisation would be very helpful specially for football storyboard where the change in shot is more important (it is seen as too complicated for other news)
- **Video search: advance search is needed**
  - Be able to add operators to the search (plus, minus, no, etc)
  - Highlight of the search text or image

In addition, we consider very important:

- To improve the way in which the user interacts with the **User interface**, because we have seen examples where the actual interfaces for the management of video is too textual and not intuitive.
- Connections with other sources of information (e.g. Google) are done constantly but not implemented in a structured way in the MAM.

4. Privacy

Sport events video footages and video surveillance raise both legal and ethical questions. This section will not provide an in-depth essay on the matter but discuss some of the main issues an APIDIS system should be concerned with.
Section 4.1 lists the main applicable rules in the European Union. In section 4.2, we discuss the main issues to be addressed especially in the APIDIS context.

4.1. Main applicable rules

4.1.1. International level

The **European Convention for the Protection of Human Rights and Fundamental Freedoms** (Rome Treaty, 1950) states:

- Article 8: Right to respect for private and family life
    - Article 8.1: Everyone has the right to respect for his private and family life, his home and his correspondence.
    - Article 8.2: There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others.

The **International Covenant on Civil and Political Rights** (New York Pact, 1966) states:

- Article 17
    - Article 17.1: No one shall be subjected to arbitrary or unlawful interference with his privacy, family, or correspondence, nor to unlawful attacks on his honour and reputation.
    - Article 17.2: Everyone has the right to the protection of the law against such interference or attacks.

4.1.2. European directives and laws

In order to remove the obstacles to the free movement of data without diminishing the protection of personal data, Directive 95/46/EC (the data protection Directive) was developed to harmonise national provisions in this field. As a result, the personal data of all citizens should have equivalent protection across the Union. The fifteen Member States of the EU were required to bring their national legislation in line with the provisions of the Directive by 24th October 1998.

Let us recall the content of the Charter of Fundamental Rights of the European Union.

The rules are:

- Data must be processed fairly and lawfully.
- They must be collected for explicit and legitimate purposes and used accordingly.
- Data must be relevant and not excessive in relation to the purpose for which they are processed.
- Data must be accurate and where necessary, kept up to date.

- Data controllers are required to provide reasonable measures for data subjects to rectify, erase or block incorrect data about them.
- Data that identifies individuals must not be kept longer than necessary.
- The Directive states that each Member State must provide one or more supervisory authorities to monitor the application of the Directive. One responsibility of the supervisory authority is to maintain an updated public register so that the general public has access to the names of all data controllers and the type of processing they do.
- In principle, all data controllers must notify supervisory authorities when they process data.
  Member States may provide for simplification or exemption from notification for specific types of processing which do not entail particular risks. Exception and simplification can also be granted when, in conformity with national law, an independent officer in charge of data protection has been appointed by the controller.
  Member States may require prior checking, to be carried out by the supervisory authority, before data processing operations that involve particular risks may be undertaken. Which types of processing operations involve particular risks is for the member states to determine.

When can personal data be processed ?

Personal data can only be processed (e.g. collected and further used) if:

- The data subject has unambiguously given his or her consent, i.e. if he or she as agreed freely and specifically after being adequately informed;
- Data processing is necessary for the performance of a contract involving the data subject or in order to enter into a contract requested by the data subject, e.g. processing of data for billing purposes or processing of data relating to an applicant for a job or for a loan;
- Processing is required by a legal obligation;
- Processing of data is necessary to protect an interest that is essential for the data subjects life. An example is in the case of a car accident and the data subject is unconscious, emergency paramedics are allowed to give blood tests if it is deemed essential to save the data subject's life;
- Processing is necessary to perform tasks of public interests or tasks carried out by official authorities (such as the government, the tax authorities, the police etc.);
- Finally data can be processed whenever the controller or a third party has a legitimate interest in doing so. However, this interest cannot override the interests **or** fundamental rights of the data subject, particularly the right to privacy. This provision establishes the need to strike a reasonable balance, in practice, between the business interest of the data controllers and the privacy of data subjects. This balance is first evaluated by the data controllers under the supervision of the data protection authorities, although if required, the courts have the final decision.

Sensitive data

Very stringent rules apply to processing sensitive data: data relating to racial or ethnic origin, political opinions, religious or philosophical beliefs trade union membership, data concerning health or sexual preference. In principle, such data cannot be processed. Derogation is tolerated under very specific circumstances. These circumstances include

the data subject's explicit consent to process sensitive data, the processing of data mandated by employment law, where it may be impossible for the data subject to consent (e.g. blood test to the victim of a road accident), processing of data has been publicly announced by the data subject or processing of data about members by trade unions, political parties or churches. Member states may provide for additional exceptions for reasons of substantial public interest.

The rights as a data subject

You have the right to be informed of any data processing when you are the data subject.

Data controllers are required to inform you whenever they collect personal data concerning you, unless you have previously been informed. You have the right to be informed of: the identity of the controller, the purposes for the processing and any further information such as the recipients of the data and the specific rights that you, as data subject, are entitled to. You have the right to receive this information whether the data was obtained directly from or indirectly from third parties.

Derogation may be allowed in the latter case if giving this information proves impossible or extremely difficult, or if it's required by law.

You have the right of access data about you.

You are entitled to approach any data controller to know whether or not he is processing personal data that concern you to receive a copy of the data in an intelligible form and to be given any available information about their sources. If the personal data are inaccurate, or if they were unlawfully processed, you are entitled to ask for the correction, block or erasure of the data. In such cases, the data subject may also require the data controller to notify third parties who had previously seen the incorrect data, unless this proves impossible. A reasonable fee for providing access may be charged in some cases.

You must also have access to the logic on which automated decisions are based.

Decisions, which significantly affect the data subject, such as the decision to grant a loan or issue insurance, might be taken on the sole basis of automated data processing. Therefore, the data controller must adopt suitable safeguards, such as giving the data subject the opportunity to discuss the rationale behind the data collected or to contest decisions based on inaccurate data.

Exceptions and limitations

The right to privacy may sometimes conflict with freedom of expression and in particular, freedom of the press and media. It is therefore up to the Member States to establish exceptions in their data protection laws in order to strike a balance between these different but equally fundamental rights.

National law might allow other exceptions to provisions of the Directive. (These include the obligation to inform the data subject; the publicising of data processing operations; the obligation to respect the basic principles of good data management practice.) Such exceptions are permitted if, among other things, it is necessary on grounds of national security, defence, crime detection, enforcement of criminal law, or to protect data subjects or the rights and freedom of others.

Additionally, derogation from the right to access data may be granted for data processed for scientific or statistical purposes.

Data transfers to non-EU countries

In cases of transfer of data to countries that are not members of the European Union, it may be necessary to take special precautions if the level of data protection in the third country is inconsistent with that provided by European Law. Without such rules, the

high standards of data protection established by the Directive would quickly be undermined, given the ease with which data can be moved around in international networks.
The principle of the Directive is that personal data can only be transferred to countries outside the EU that guarantee an "adequate" level of protection. Analysis of data protection laws and dialogues with the EU's more important trading partners is underway in order to decide which countries can be seen as offering adequate protection.
Where a non-EU country does not ensure an adequate level of protection, the Directive requires the blocking of specific transfers. Member States must inform the Commission of any such blocking measures, and this triggers a Community procedure to ensure that any Member State's decision to block a particular transfer is either extended to the EU as a whole, or reversed.

4.2. Consequences for APIDIS

Since human beings and their belongings (vehicles, houses or cars in the surveillance context) are filmed and recorded, APIDIS does deal with such personal data. Intelligent surveillance or sport event automatic summarization systems must thus protect this data appropriately.
In the video surveillance context, privacy is threatened if the systems or the stored personal data are misused, i.e. if they are used with any other goal than crime or incident investigation, such as social control. The problem gets worse when cameras are combined with other sensors and personal information sources, such as credit cards, RFID's, mobile phones, etc. Interestingly, though anyone can imagine the potential abuses of cameras, it seems not to be the case for such more subtle information sources. In short, protection of privacy is a global concern and should be addressed, not only when speaking about video surveillance.
From this perspective, it is important to note that any interactive information retrieval system can potentially threaten the privacy of the users. Indeed, if user feedback and behavioural information are stored and can be associated with his identity, e.g. using login, it represents sensitive personal data, that should be protected as well.
Ignoring privacy would be a mistake since it could prevent deploying an APIDIS system given the governing legislation. In short the legal approaches put the emphasis on the minimisation of the data acquisition and the fairness of the processing.
Last, it is worth mentioning that the protection of privacy in surveillance represents a market opportunity in itself. As an example, companies like Emitall (http://www.emitall.com) are focusing on the problematic of hiding private areas in video surveillance streams.

Although specific legal rules may differ in each country, we can draw a minimal list of recommendations for an APIDIS system installation:

- Warn the public about the recording of personal data such as images or sound as soon as they enter an area covered by APIDIS sensors (video cameras, microphones...).
- Explain the goal and policy of the data recording, storage and diffusion (e.g. at which extent the videos will be published: only for the coach, for the players, for the players and their family, directly on Internet...)
- Permit access to and consultation of the data to the individuals whom personal data are recorded and processed.

- In the context of sport events, obtain an explicitly written permission from the people (or their legal entities) who will be the main subject from the data processing point of view, i.e. the players, the coaches... Those people must explicitly agree on the fact that both their image and other personal data could be used in the system (e.g. if a request for a summary is possible for a given player name). In order to minimize problems, they have to officially agree on the APIDIS system processing of their personal data as well as with the modalities of the APIDIS system outputs (at which level the data captured by APIDIS and the features of APIDIS will be made public).

It should be mentioned that the entity legally responsible of the law application may change from the APIDIS system reseller (if APIDIS is sold as a service) or the buyer of such a system (if APIDIS is sold as a product). Nevertheless, all these legal issues can only be addressed when the system will be operational and with regard to each country legislation.

Bibliography:

Council of Europe: “European Convention on Human Rights,” available at http://conventions.coe.int/treaty/en/Treaties/Html/005.htm, 1950.

United Nations :“International Covenant on Civil and Political Rights,” available at http://www2.ohchr.org/english/law/ccpr.htm, 1966.

Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, available at http://ec.europa.eu/justice_home/fsj/privacy/docs/95-46-ce/dir1995-46_part1_en.pdf

5. Personalized and (semi-)autonomous summarization: acquisition and processing architecture definition, and content annotation.

Figure 2 presents the architecture considered by APIDIS to access content in a personalized manner. It reflects the practical constraints imposed by the video acquisition setting, and integrates most of the components required to address the requirements and expectations formulated by the user group during interviews. In particular, with respect to the end-user, the envisioned architecture supports:

- manual definition of (emotionally) interesting time periods by the end-users,
- exploitation of individual semantic preferences (player or kind-of action) based on rich and semantically relevant metadata,
- exploitation of the narrative context associated to the browsing or summarization process,
- compliance with individual technical requirements (resolution, bandwidth, duration).

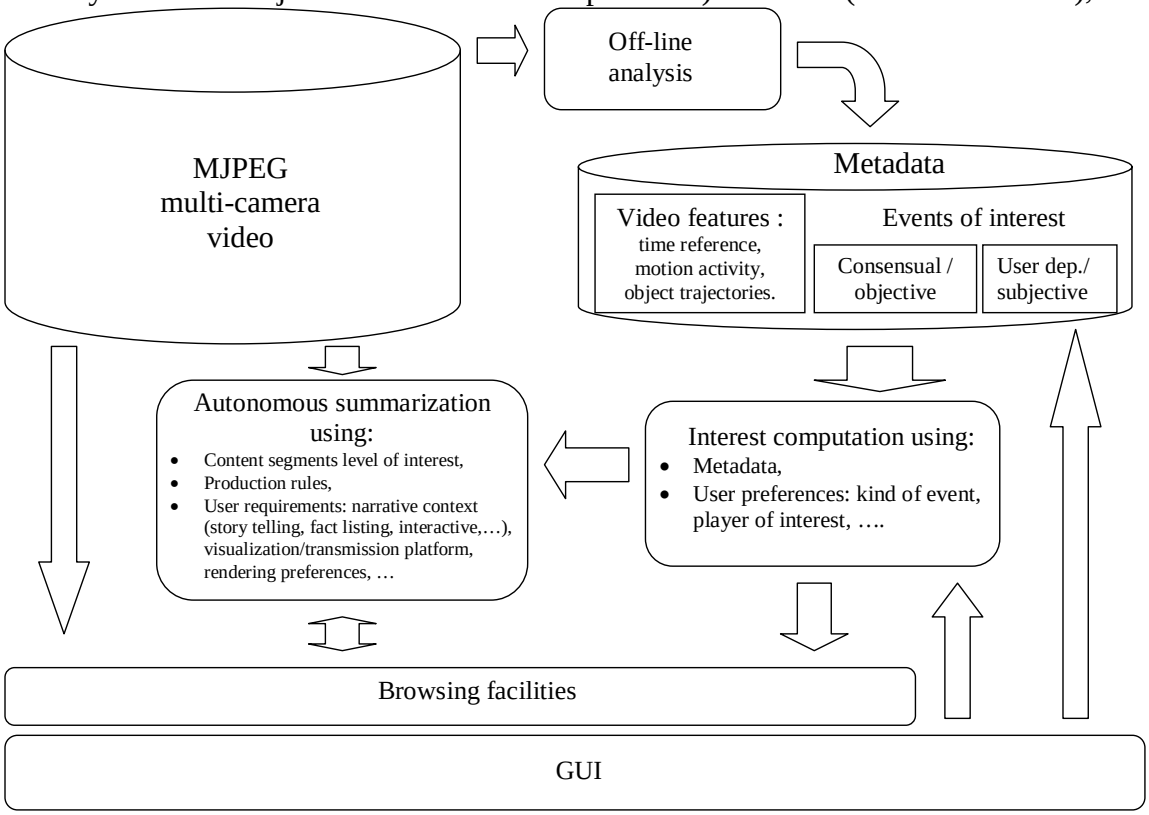
Specifically, in Figure 2, we observe that:

- The stored information is composed of raw multi-camera video content, and of metadata. For practical reasons, the raw content has been encoded in MJPEG format by the IP cameras exploited during the acquisition campaign. Metadata include low- and mid-level generic video features, together with semantically meaningful annotations of events of interest. In more details:
  - Low-level features refer to the (motion vector fields, foreground masks) features extracted in a particular camera view.
  - Mid-level features typically correspond to object trajectories, computed based on joint processing of multiple views. Trajectories are defined in absolute terms, independently of camera index.
  - Events of interest refer to the scene at hand, and not to the stream captured by a specific sensor. They can be objective or subjective. In the first case they correspond to typical actions (e.g. throw or fault in a basket-ball match), and should be defined automatically, through off-line analysis of the video or audio content. In the second case, they correspond to subjective opinions or interest expressed by the end-users about any period of time. This information is expected to be introduced manually through the GUI, and recorded to support subsequent content access sessions.

- The level of interest or relevance of a scene segment depends on (1) the information collected through metadata, and (2) the preferences defined by the client in terms of player or action of interest. Computation of scene segment relevance is used for autonomous summarization purposes, or to support efficient interactive browsing of the content (e.g. by displaying most relevant segments along a time-line).

- The autonomous summarization module integrates content relevance, production rules, and user requirements to generate the content to display to end-users. Among user requirements, we distinguished between the preferences related to the kind of

production (e.g. pragmatic style devoted to players and coaches vs. story telling kind-of-style devoted to journalists or common spectators) or access (interactive or not),



and the constraints related to the device or channel used to access the content.

Figure 2: APIDIS framework.

In the rest of the section, we first present the common test-bed envisioned to unify the way to access and process content within the APIDIS consortium (Section 5.1). We then focus on the sport event video summarization scenario, and detail the acquisition settings (Section 5.2) and the metadata collection (Section 5.3) in a basket-ball context.

5.1. Test-bed for unified data access and processing

This section describes the skeleton of the platform considered for demonstrating and developing the APIDIS prototype. This common platform has been defined to guarantee coherence between the modules developed by each individual APIDIS partner, so that the final prototype should be obtained by simple plug-in of the GUI, browsing, and summarization tools to the skeleton.

5.1.1. Hardware

For both maintenance and economic efficiency, an APIDIS system should use only one processing unit. This server is assumed to perform the following tasks:

- Acquisition and storage of the video streams and metadata.
- Offline processing for the extraction of the low-, mid- and high-level features.

- On-demand summarization of video surveillance or sport events acquisitions with respect to end-users parameterised requests.

Depending on the scenario, it is likely that this central processing unit will be assisted by several other units for the real-time acquisition of all video streams. Indeed, when deployed in a large scale environment (e.g. hundreds of video surveillance cameras), the amount of video data bandwidth must be shared between several recording units.
The APIDIS main server will be a standard 32-bits architecture PC machine.

5.1.2. Operating system

The developments will be held under Linux. This operating system offers a large number of free but good quality development tools and libraries. Furthermore, Linux can be used for free in a commercial product, which reduces potential exploitation costs.

5.1.3. Software

Although we decided to use Linux as our operating system, the APIDIS software prototype developments must be as generic and re-usable as possible. There are two reasons for that:

- If the project (partially) succeeds in its objectives, APIDIS software could be sold independently of any hardware. Supporting several architectures and operating system is then welcome.
- In order for each partner to be able to develop its components for the prototype and be able to use them with minimum modifications in its usual working environment.

APIDIS partners agreed on the following development principles:

- C++ programming language.
- wxWidgets for the Graphical User Interface components. This library provides the advantages of being supported in a large number of operating systems and forwards the requests to the native functions of each operating system rather than emulating them. Furthermore, wxWidgets implements a C++ API.

Several open source libraries will help us in our developments, e.g. ffmpeg, libxml, etc.

5.1.4. Unified development of APIDIS modules

As a solution for efficient integration of each partial development, a VMware virtual machine will be provided, to be used by all partners**.** Figure 3 shows this virtual machine running with VMware. It is foreseen that the final demonstrator will use exactly the same operating system and libraries versions than the ones installed in this virtual machine.

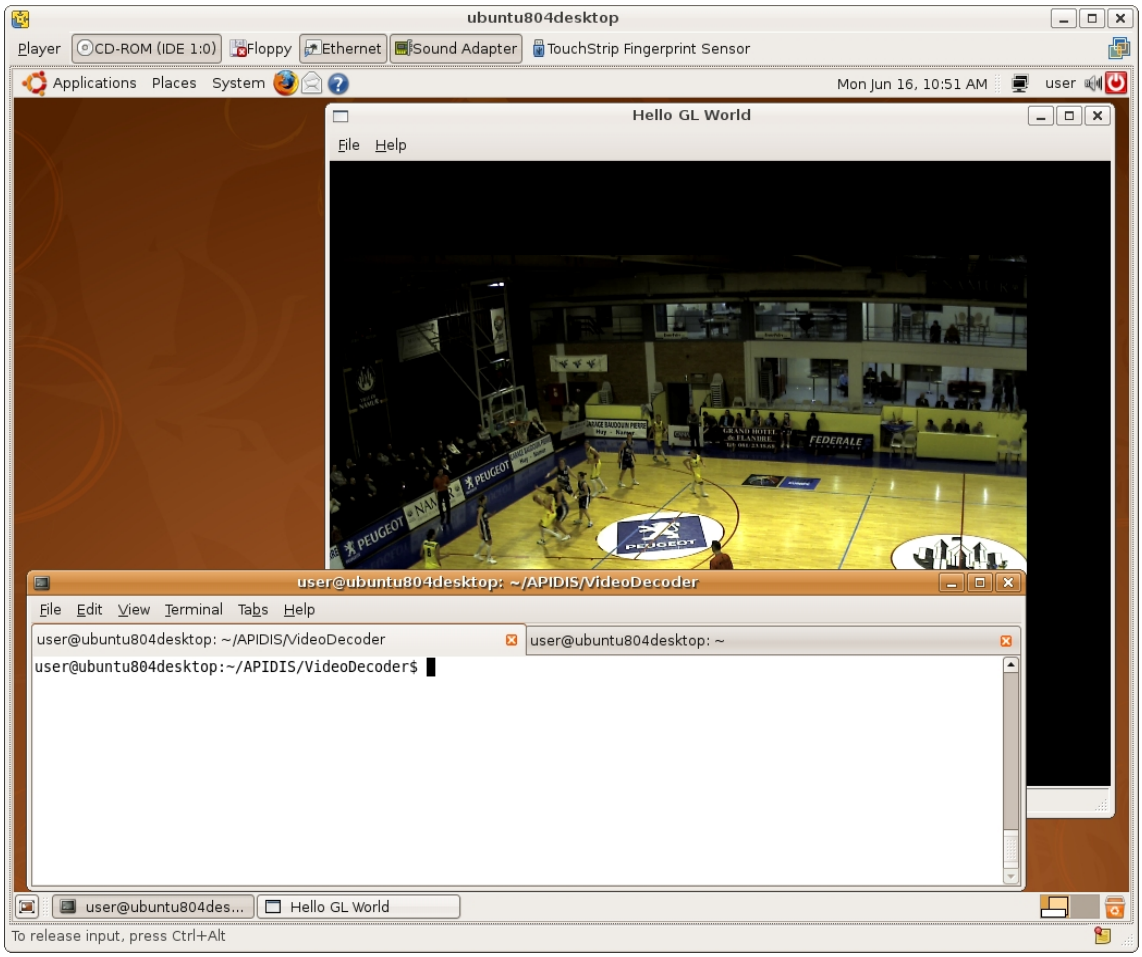


Figure 3: VMware Ubuntu 8.04 (32 bits) virtual machine

For efficient and coherent input/output management, the following rules are proposed to name media and metadata files:

- Filename convention is *MediaName.Date.Extension*. For example, camera1.20080405T194800Z.mjpeg contains the video stream starting at 7:48 PM the fifth of May 2008 for camera 1.
- The directory tree structure provides part of the date and time of day of the file timestamp. Assuming that the acquisition generates one file every minute, the previous file path could have be e.g.: (root directory)/2008/04/05/19/48/ camera1.20080405T194800Z.mjpeg or (root directory)/20080405/19/48/ camera1.20080405T194800Z.mjpeg or (root directory)/20080405 camera1.20080405T194800Z.mjpeg, etc.
- It is assumed that the timestamps of the content of a file are greater or equal to the timestamp provided in the filename. It is also assumed that the timestamps of the content of a file are all included in the time period the directory path represents.

A library for automatically retrieving a file given a media name, an extension and a timestamp will be developed. This library will also provide the filename (and path) to use when it is required to write a new media file (e.g. low-level features metadata).
A library providing decoded video frames starting at a given timestamp for one specific camera will also be developed so that video processing modules don't need to deal directly with the recorded video files.

5.2. Basket-ball event acquisition settings

This section describes the system considered to capture a set of complementary images around a basket-ball field. The objective was to capture images that could support both efficient analysis and nice-looking rendering of the event. Only still cameras have been considered. They were connected to a central and unique server.

5.2.1. Camera and server specifications

The APIDIS acquisition system is composed of several high-resolution cameras distributed around the sport field. The main criterions for the selection of the cameras were the resolution, the frame rate, the sensibility and the cost of the overall system. The main features of the selected Arecont Vision AV2100 cameras depicted in Figure 4 are: 2 Mpixel IP-cameras sensitive to 0.1 lux at F1.4 providing 24 fps at 1600x1200 and featured with a captor size of ½ inch.
For two cameras providing a top view of the basket-ball field, Fujinon wide angle lenses have been used, see Figure 5. When installed with the AV2100 cameras, they provide horizontal and vertical view angles of 136°18'x102°19' for a focal length of 2.7 mm. The other cameras use more standard lenses.
The server that collects all the video streams is a Hewlett Packard DL380 G5 with Intel(R) Xeon(R) CPU E5420 at 2.5 GHz with 2 GB of memory and two 73 GB disks. In the course of the APIDIS project, this machine will also be used to serve images and metadata to the algorithms in charge of the automatic generation of sport events summaries.

Figure 4: Arecont Vision AV2100 color, 2 Mpixels.

Figure 5:: Fujinon FE185C086HA-1 super wide angle lens.

5.2.2. Camera positioning

To preserve consistency between the movements rendered in all views, cameras have all been distributed on the same side of the axis joining the two baskets. In order to limit the number of cameras to deploy, some asymmetry has been introduced and more cameras have been placed on the left-hand side of the field. Figure 6 displays all seven camera views at the same time. The first two super wide angle pictures provide top views (one for each side) of the basket-ball field. Each one of the two pictures below

shows half of the field with an incidence angle of about 45°. The last three pictures provide zoom in views on the left-hand side of the field[3]. Those three zoom in views are particularly interesting for rendering, while large angle views allows for consistent tracking along the game.

Figure 6: Overview of the 7 cameras at the same time

[3] In a complete system, it is foreseen that the same three cameras are installed for the other side of the game court.

Figure 7: Top view with one of the super wide angle lenses

5.2.3. Storage format

The seven video streams are recorded on the server in their native MJPEG format. The bandwidth and required disk space is about 300 MB per camera per minute. The files are organized with a directory for each camera and a file for each minute of video sequence. A file containing the timestamp of each frame is also recorded with each video file.
The metadata associated to specific cameras (e.g. foreground mask, as defined in the next section) are also distributed in the respective directories, while metadata relative to the game (e.g. clock and non-clock events, as defined in the next section) are stored in the upper directory. All filenames describe the source, the timestamp of the first element they contain and a media type extension, e.g. *cam1.20080405T131900Z.mjpeg*.
This directory tree structure and filenames convention allow fast search of files for browsing.

5.3. Basket-ball metadata definition and generation

This section defines the set of complementary metadata that have been considered to identify salient segments within captured video content, thereby supporting user-driven and personalized summarization of the basket-ball game.

5.3.1. Input data: synchronized multi-views video streams

The main assumption underlying the annotation format is the existence of a common temporal and spatial reference for all camera views, so that all information defined relatively to one camera can be mapped to the absolute spatial and temporal coordinates of the scene at hand.
A *common and single* ***time reference*** *for all camera views,* could obviously be obtained by synchronizing the instants at which frames are captured by the cameras. However, such synchronization capability requires expensive professional firewire cameras. In contrast, each IP camera from the setting described in Section 2 captures as much frames as possible, and sends them to a common server. When it receives a frame, the server stores it, and labels it with a timestamp, corresponding to the instant of arrival. Hence the timestamp refers to the server clock, which is common to all cameras, but corresponds to the storage instant rather than to the instant of capture. In this context, we propose to build a common reference time line for all camera views as follows. We sample the server clock at 20 Hz, and for each time sample and each camera view, we select the frame with the larger timestamp below the instant of interest. Doing so, we generate a regular stream of multiple frames, each frame being a reasonable approximation of the signal that would be captured by the corresponding camera view at the instant of interest. This multi-view stream is the one considered for subsequent processing and annotation.
Regarding the spatial correspondence between views, we rely on calibration parameters and define an *homography* to link the points of the basket-ball ground in every camera views, thereby defining a *single* ***reference position*** for all objects on the field at hand. Note that the annotation does not assume any kind of overlapping between views, nor a complete coverage of the captured scene.

5.3.2. Metadata : event, trajectories, and low-level features distribution

The annotation process consists in the collection of three kinds of complementary metadata:

- First, events of interest identify specific actions of the game, and are characterized by a number of attributes (type of event, instant, position, player involved).
- Second, visual object trajectories define the position of a given object (a player or the ball) within a finite period of time.
- Third, efficient representations of the spatial (and potentially temporal) distribution of low-level features, typically foreground mask or motion vectors fields, are considered to support scene rendering.

Each kind of metadata is further defined in the rest of the section, together with a description of the mechanisms that have been implemented to collect them.

5.3.3. Clock- and Non-clock- events

In the following, an event refers to a game action or a player/spectator(s) behaviour that can be inferred based on the analysis of the signal captured by the sensor network, mainly composed of cameras in our case.

The events potentially relevant to identify salient segments of a basket-ball game have been listed and defined through the XML hierarchical syntax depicted in Figure 8.

Two categories of events have been distinguished:

- The first one is composed of events that have a direct impact on the 24'' clock of the basket-ball game. Therefore, we name them the *clock-events*. They correspond to the events associated to the starting, stopping or re-initialization of the 24'' clock, i.e. to instants at which the game is interrupted or at which the ball hits the basket or is gained by the opponent's team.
- The second category of events encompasses all events that do not cause any specific action on the 24'' clock. They typically have to do with displacements and interactions of players during the game, or with some subjective interest expressed by spectators/viewers about the game.

We decided to differentiate the clock and non-clock events because the 24'' clock is easy to monitor in an automatic way and is closely linked to the semantic of the game, or at least to all objective statistics that are generally collected by coaches and players about the game (scored points, fault, etc…). Hence, for clock-events, the objective of the annotation tools reduces in recognizing and characterizing the event associated to the change of clock state.
In contrast, non-clock events refer to global or individual behaviour in the game, and do not have specific and objective time anchors. They are thus more challenging to detect, and are mainly considered to support the manual introduction of complementary information about the game. The APIDIS objective will thus not be to detect and characterize those non-clock events in an exhaustive and automatic way. Rather, only partial hints about some of those events are provided.

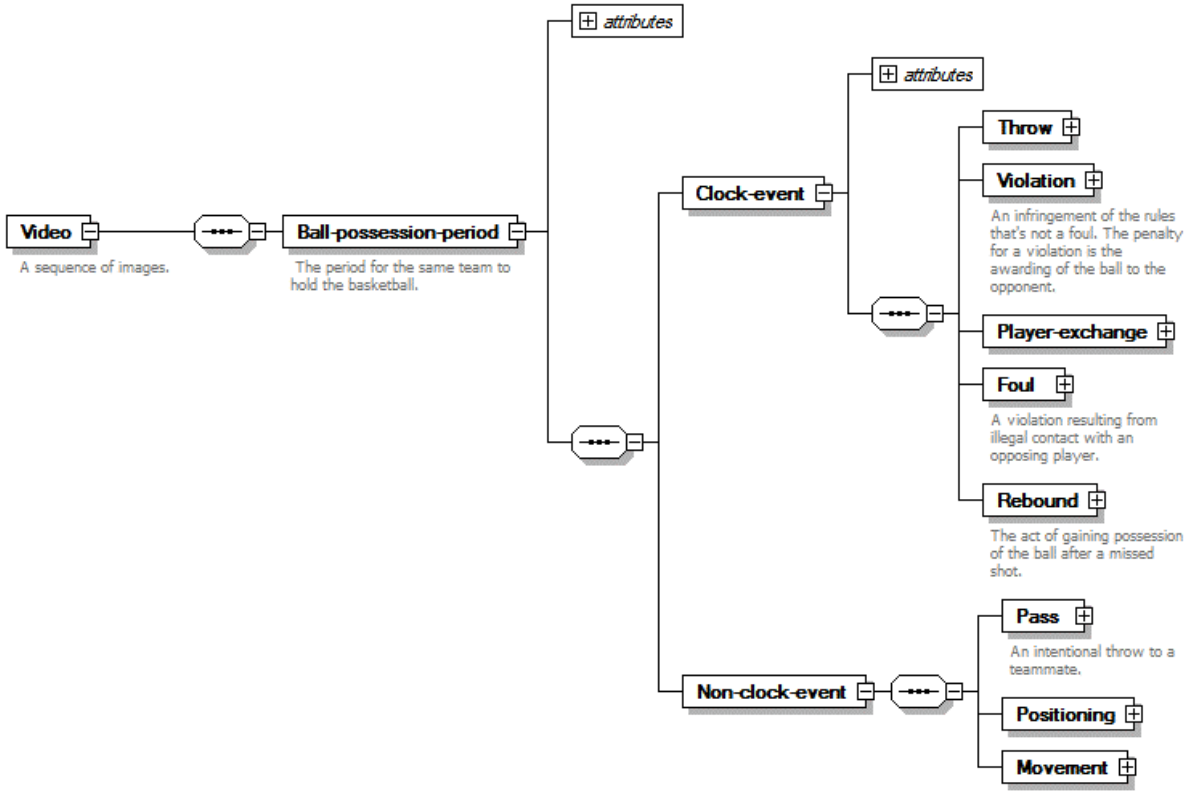


Figure 8: XML file format for basket-ball event annotation.

We observe in Figure 8 that events are grouped in subsets corresponding to ball possession periods, which are expected to be meaningful in a summarization context. In practice, since the clock is initialized each time a team gains the ball, the period of time between two clock events is characterized by a single attacking and defending team. A ball possession period can thus easily be defined by merging adjacent periods between clock-events for which the same team is attacking/defending.

In more details, the attributes of the ball possession period typically include start- and end-times, team label, and optional information about the behaviour of offending and/or defending teams (fast-attack, zone defence, press, etc.).

In contrast, the attributes of a clock or non-clock event include its timestamp in the camera time and game time[4] references, plus a set of attributes that are specific to the event type (e.g. throw, fault, etc). As an example, Figure 9 presents the list of attributes associated to a fault event. An exhaustive description of the attributes associated to each type of event is provided on the APIDIS web site. Typically, the attributes identify the players involved in the event, define a time frame for the event, and refines the nature of the action at hand through a number of options (e.g. foul-type, throw-type,…).

[4] Here, the game time refers to the clock aggregating the actual time played in the game. The game time can be computed based on the timestamp associated to clock events, since the game clock is stopped/started each time the 24'' clock is stopped/started.

At the current state of the project, an interface has been developed to support manual definition of clock- and non-clock-events. The annotation XML file should be made publicly available before September 2008, together with the video streams and calibration parameters. Within APIDIS, it will be used to initiate the automatic personalized summarization mechanisms, and will serve as a ground truth reference to validate the video analysis/recognition tools that will be developed to generate those metadata automatically.

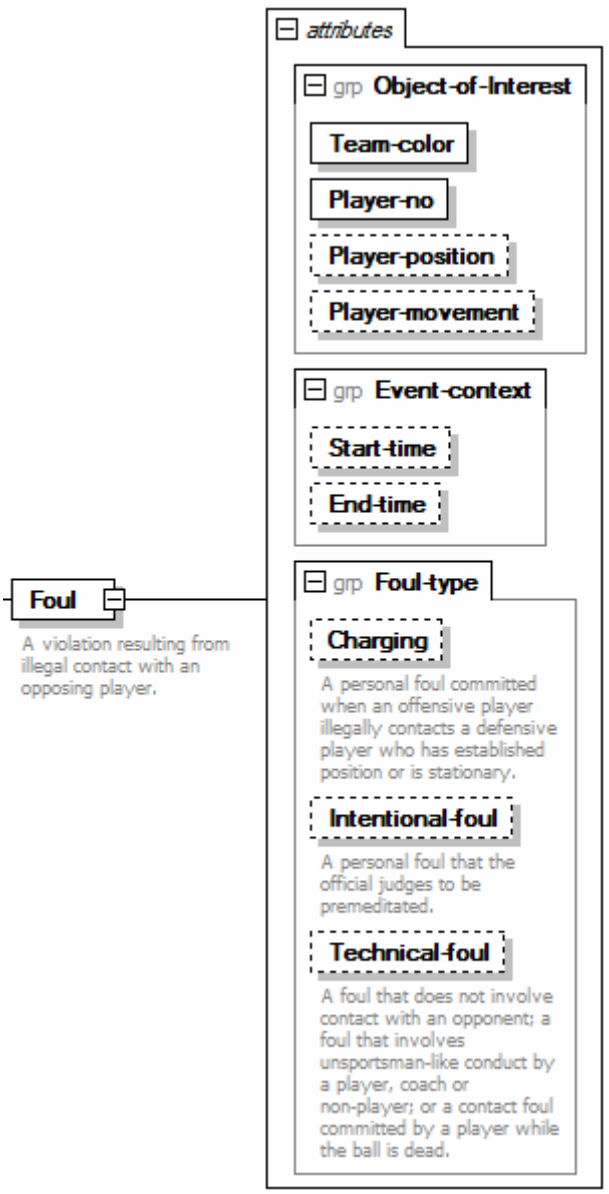


Figure 9: List of attributes associated to a fault event.

5.3.4. Object trajectories

In our sport-event analysis framework, object trajectories define the position of a given object (a player or the ball) within a period of time. It is worth mentioning that the trajectory does not necessarily track the object in a continuous manner all along the

game. Instead, the file collecting the trajectories of an object consists in a set of potentially interrupted tracks. Hence, those metadata files can be progressively incremented, based on automatic detection and tracking of objects within the scene.
In a first stage, an object is detected and tracked without being identified. In a second stage, a recognition engine parses the object bounding boxes, so as to recognize it automatically, e.g. by recognizing the number of the player from the image captured at a well-chosen instant and from a well-chosen viewpoint along the trajectory. Joint processing of multiple views is exploited to detect objects and define their trajectories. The first release of object trajectories metadata should be based on conventional particle filters, and on manual identification of the player associated to most relevant trajectories. Advanced tracking methods and automatic player recognition modules will be developed in the course of the project, and should result in updated and completed release of those trajectory files.

5.3.5. Foreground mask and motion vector distribution

Here, we consider the characterization of low-level features within each individual camera view. The envisioned features define the location of foreground objects within the scene, and could potentially be augmented based on motion vector fields computation.
In APIDIS, those features are considered to fill in the lack of information resulting from the potential interruptions of object trajectories, when selecting rendering parameters. In other terms, those features are expected to support the selection of appropriate camera view and cropping parameters when rendering a given period of the game.
At the current stage of the project, only the extraction of foreground objects has been considered. A sub-sampled mask of foreground areas has been defined automatically for each frame of each camera view, based on the subtraction of a background Gaussian mixture model[5].

5.4. Conclusion

This section has presented the general framework envisioned for interactive browsing and personalized summarization of multi-camera content within the FP7 APIDIS project, together with the acquisition setting and annotation methodology implemented for basket-ball events.
We identify two practical and short-term[6] outcomes associated to the above defined requirements. The first one is the release of a project content processing test-bed (see Task 6.1 in DoW). The second one is the public release of multi-camera basket-ball video content and of its corresponding metadata, ranging from low-level scene descriptors to ground-truth high-level semantic concepts.

[5] See ‘C. Stauffer and W.E.L. Grimson. Adaptive background mixture models for real-time tracking. In *IEEE Conference on Computer Vision and Pattern Recognition*, volume 2, pages 246–252, June 1999’, or ‘X. Desurmont, C. Chaudy, A. Bastide, C. Parisot, J.F. Delaigle and B. Macq. Image analysis architectures and techniques for intelligent systems. In *IEE proc. on Vision, Image and Signal Processing, Special issue on Intelligent Distributed Surveillance Systems*, 2005’.

[6] Short term means by September 2008.

6. General conclusions and further work

In this document we have described the work done in WP2 by all the partners in APIDIS in their effort to define and guide the future development of APIDIS. The information gathered in this deliverable will be used by the other workpackages in the project to develop the technologies.

In section 2, we first defined a set of potential and foreseen use cases for APIDIS. Section 3 has then collected and analyzed the user requirements associated to those cases. The interviews and observation have been very useful in order to understand professional work-practices, actual problems, and constraints and expectations. In addition, we have asked end users, in particular coaches already using video, how APIDIS could help them. The requirements gathering is expected to be an iterative and continuous process where we will go on collecting information from the users during all the stages of the technology development. Therefore, as soon as we have a prototype, we will gather more information from both professionals and end users. Scenario and prototype-based requirements gathering will be particularly useful regarding scenarios for which potential end users do not using video at present, like in the internet portal scenario.

To complement user feedback, Section 4 has addressed the main legal and ethical issues related to the deployment of multi-camera networks in public areas.

Section 5 has presented the general framework envisioned for interactive browsing and personalized summarization of multi-camera content within the usage context defined in earlier sections. It has also defined the acquisition setting and annotation methodology implemented for basket-ball events.

7. Appendix 1 User Group Flyer

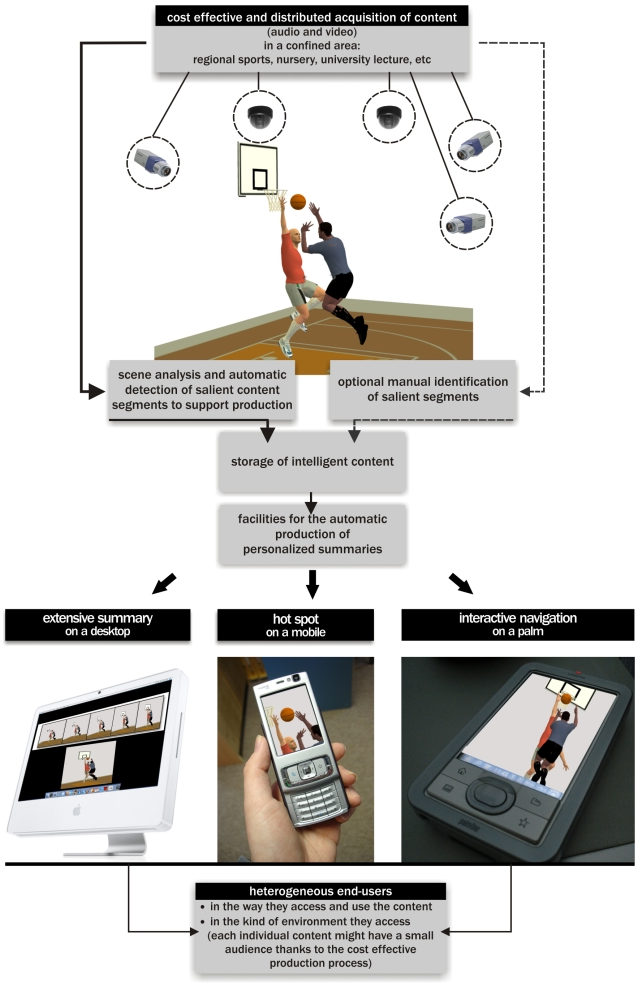
APIDIS will investigate the automatic extraction of intelligent content from networks of multi-modal sensors. It will exploit the knowledge to automate the production of video content for controlled scenarios (sports events or surveillance). The project will consider personalised and potentially interactive content summarization mechanisms to address heterogeneous user needs and access conditions. **APIDIS will develop applications that are cost-effective and fully automated for production of content dedicated to small-audiences.**

The APIDIS project, which started in January 2008, will respond to the exciting **challenge of personalized production.** APIDIS will propose a framework to automate the collection and production of digital content. Based on this framework, APIDIS will develop **application solutions** that are cost-effective and fully automated. These **content provider solutions** will **lower costs of local event coverage**, as no technical team or camera crew involvement is needed.

APIDIS also promotes the collection and supply of **intelligent digital content**. Intelligence refers here to the identification of salient segments within the content, and to the exploitation of that knowledge to adapt and personalize content summarization according to the individual user needs.

Participate in APIDIS Breakthrough Innovations as a *USER GROUP MEMBER*

The integrated technology and methodologies developed within APIDIS have the potential to be applied to **numerous user scenarios**. In a typical application scenario, the APIDIS acquisition sensor network is composed of microphones and still cameras, which cover a location, like a basketball court in a community center where the local team plays. Both conventional and (arrays of) omnidirectional cameras are considered. Distributed analysis and interpretation of the scene is exploited to decide what to show about the event, so as to produce a video composed of a valuable subset from the camera stream(s).



An important component of APIDIS is the User Group that will be consulted to support the user scenarios and requirements definition and will be invited to evaluate both mid-project proof-of-concepts and final user-field trials. **If you interested in participating in APIDIS** as a **techno-user** - exploit APIDIS

technology to offer novel services and to tap new markets for your customers or an **end-user** - watch the images produced by the APIDIS technology, let us know!
APIDIS is an ambitious and challenging project with the potential to **benefit national and European markets with a range of exploitable results**. The **user-centred design approach** adopted by APIDIS builds on a thorough understanding of the context of use, and relies on participation of users through an iterative analysis-design evaluation process. Help us to understand how you exploit visual content and what your requirements are. Tell us your feelings and interests regarding the potential services offered by the APIDIS technology. **Help to shape and validate the breakthrough innovation and benefit from the APIDIS achievements**. Participate in questionnaires and/or semi-structured interviews to help define user requirements and create user scenarios.

Participate in APIDIS Breakthrough Innovations as a *USER GROUP MEMBER*
PIDIS CONTACT: Christophe De Vleeschouwer, devless@tele.ucl.ac.be 32 10 472543

APIDIS Facts

APIDIS will investigate the automatic extraction of intelligent content from networks of multi-modal sensors. It will exploit the knowledge to automate the production of video content for controlled scenarios (sports events or surveillance). The project will consider personalised and potentially interactive content summarization mechanisms to address heterogeneous user needs and access conditions. APIDIS will develop applications that are cost-effective and fully automated for production of content dedicated to small-audiences. The potential applications of the integrated technology and methodologies that will be developed within APIDIS are numerous, ranging from:

- personalized access to digital media related to local sport events through a web portal or a mobile hand-set;
- cost-effective and fully automated production of content dedicated to small-audience, e.g. souvenir DVDs, university lectures, etc;
- automated summarization for video surveillance.

The APIDIS Consortium is made up of leading experts from academia and industry:

iversité catholique de Louvain	CL	lgium
ueen Mary& Westfield College, University of London	MUL	ited Kingdom
ole Polytechnique Fédérale de Lausanne	FL	vitzerland
ndació Barcelona Media Universitat Pompeu Fabra	M	ain
CIC sa	CIC	lgium
ediapro	P	ain

Project coordinator: Université Catholique de Louvain (UCL), Belgium
Project type: STREP, FP7 Grant Agreement Number 216023
Start Date: 2008-01-01 Duration: 36 Months
Project Cost: 2615541 Project Funding: 1,925,000
APIDIS CONTACT: Christophe De Vleeschouwer, devless@tele.ucl.ac.be 32 10 472543 **APIDIS** is co-funded by the European Union under the seventh framework programme**.**

8. Appendix 2 Summary table of methodologies

Methodologies	**What information do we obtain**	**Usefulness for APIDIS**
Questionnaires	Quantitative data, we need to know in advance the options that we want to present to them i.e. about: - list of possible problem they may have at work - what they would like to have that they do not have now - etc	Not recommended at the very first stage because it would be too narrow focus
Interviews – open ended questions	We can ask about: - main role in the organisation -they can describe how they do their work (not a complete workpractice gathering but an overview) - what they see as the major current problems at work	Could complement easily the observations Requires less experience from the interviewer about how to gather information from users The information gathered depends a lot on the interviewer.
Observation (Shadowing)	-Current workpractices -UI already used ➔ input for APIDIS UI -Real work problems and the impact in their day to day work -Coordination among member of a team	Is the most informative Requires experience from the researchers
Scenario-based interviews: we show the users a presentation with APIDIS scenarios and asked them to describe how they will used it in their work	-it already guides the user around our scenarios so we may miss anything outside the scenario that may have been useful -good to gather specific feedback about he scenarios	Probably better for a second round of user feedback gathering Good for gathering information from non-professional users
Possible methodologies to gather feedback about UI ➔ paper prototypes Present sketches done quickly in paper (only boxes with names of functionalities, representing bottoms/taps/etc) ➔ give the users set of cards with the name of the functionalities that we thinks should have the interface, as it was a puzzle, and let the user to place them in a piece of paper	Gather their expectations about the UI	

9. Appendix 3 Observation and interview guide by user case

1. Automatic generation of content for Internet portal:

Observation guide (for end users)

Goal: gather user requirements for personalized sport-event video summaries based on distributed and omnidirectional sensing.

Steps/ what to observe:

- how the users prepare for the recording of a sport event (problems, barriers, expectations)
    - who records it
    - how is it decided what to record
    - materials needed
- how the users now record a sport event (problems, barriers, expectations)
- what they do with the recorded material actually (problems, barriers, expectations)
- long term repository of the recorded materials

Questions:

- what do they see as their major problems barriers (complementing the observation)
- would be the impact of being able to personalise content➔ real day to day examples about how they would use it
- What does "personalisation" mean for them➔ examples of different ways in which they would like to have the content from different "perspectives"
    - Temporal? Spatial? By actors?
    - It is any similar to the *shooting reports??*
- Use of internet and internet portals (related to the sport/team) or in general

Interview guide (for industry perspective)

Goal: gather user requirements for personalized sport-event video summaries based on distributed and omnidirectional sensing from an industry perspective.

Questions:

- Question about the work-practices for companies that record sport events:

  - They describe step-by-step how they prepare the recording of a sport event
  - Rules and constrains special for a particular sport event
  - They describe step-by-step how they record a sport event
  - They describe step-by-step how they use the recorded footage

- Questions about their industrial requirements: After the description of what APIDIS technology goal is:

  - For your daily work, how often/when/why do you need to generate a
  - What do you do with the video (how to store it, publish, send it to other people, etc)

  - How do you create your summaries currently?

  - Do they have any technology that facilitates the creation of summaries?

  - In which occasions would you find it useful to have an automatic generated summary?

  - Would it save you time /make your work more efficient?

  - What do they see as the major technical constraints to it?

  - What do they see as the major functional constraints to it?

- Questions about their perspective
  - What are your expectations about when/how/why non-professional users will use their self recorded videos
    - Any expected trends
    - Do they have data about it

  - What is the company plans to create technology that facilitates the generation of summaries?

  - How they see that facilitating the UGC would influence their business?

Any other suggestions or information that they would like to give

2. Interactive and semantically-driven access to video surveillance content- Observation and interview guide

Goals of the technology for which we are gathering requirements:
This use case considers the content acquired by a network of surveillance cameras, covering a building or a secured area.
Search within a video ➔ gain of time and/or a better accuracy in the search for visual events.
Creating summaries ➔ 'forensic' usage of the system, the user might want to influence the production of the summary to focus on specific area or events of interest.

Observation guide (for end users)

Goal of the observation: to gather user requirements from end users

Steps/ what to observe (if not possible to observe, to ask in an interview):

- how the users prepare for the recording of a CCTV (problems, barriers, expectations)
    - why and how they record it
    - how what to record is decided
    - location of the cameras and which information they try to get with the location of the cameras

- how the users now record ➔ observations rooms? Person responsible?
- what they do with the recorded material actually
    - in which situations do they go back to the recorded material
    - how they do it
- long term repository of the recorded materials

Questions:

- what they see as their major problems barriers (complementing the observation)
- what would be the impact of being able to do have summaries➔ real day to day examples about what they would use it for. In which occasions would you find it useful to have an automatic generated summary? Would it save you time /make your work more efficient?
- For your daily work, how often/when/why do you need to generate a summary of a video
- What do you do with the video (how to store it, publish, send it to other people, etc)
- How do you create your summaries currently?
- Do they have any technology that facilitates the creation of summaries?

- What does it mean for them to be able to choose interactively and semantically in the content➔ examples of different ways in which they would like to have the content form different "perspectives"
    - A particular event? A suspect? Groups of people? Anything strange (what does it mean?)
- What they would like to do that they cannot do now: examples from their daily activity
- Legislation constraints in what they can do with the video material and how they can/can not use it.
- Questions about their industrial requirements: After the description of what APIDIS technology goal is:
    - What do they see as the major technical constraints to it?
    - What do they see as the major functional constraints to it?

Interview guide (for industry perspective)

Goals: to understand industry perspective

- Questions about their industrial requirements: After the description of what APIDIS technology goal is:
    - In your experience in the area of surveillance:
        - how useful do you think that it would be to generate a summary of a video of a surveillance video? In which occasions
        - What do you think that companies want to do with the video (how to store it, send it to other people, etc)
        - Do they provide any technology that facilitates the creation of summaries?
        - Would it make your company more competitive if it has this kind of technology to offer to customers?
        - What do they see as the major technical constraints to it?
        - What do they see as the major functional constraints to it?

- Questions about their perspective
    - What are your expectations about when/how/why companies/police/etc will use their CCTV recorded videos
        - Any expected trends
        - Do they have data about it
    - Does your company have any plans to create technology that facilitates the generation of summaries?
    - How do they see that facilitating this technology would influence their business?

Any other suggestions or information that they would like to give

3. Production room User case -Observation and interview guide

Technology goal:
Supporting professional content producers in the re-use of pre-recorded content. In this section we thus consider most of the content management issues raised by professionals when remixing content, with the main advantage of:

- MAM cleaning of repeated videos and keywords to designate the same concept
- Language of the annotation
- Improve the annotation (looking for not only words)
- Approximated textual search
- Semantic search
- Learning search
- User interface
- Connections with other data sources

Observation guide (for end users)

Goal: gather user requirements from potential end user of the system (professional content providers, and specifically producers)

Steps/ what to observe:

- Annotation of a video: step-by-step, technology use, co-ordination with other people, work flow, problems, etc.

- Video re-use: what they do with the recorded material actually (problems, barriers, expectations)
    - usage of the video by him/her/other groups in the company
    - tools used for accessing the video

- Video search
    - how they do it now: details on tools/procedures/ specialised roles/etc
    - in which context, for which programs, with how much anticipation?

- long term repository of the recorded materials: details on tools/procedures/ specialised roles/etc
    - who is responsible for it?

Questions:

- what do they see as their major problems barriers (complementing the observation) in:
    - annotation
    - video reuse
    - video search

- what would be the impact of having some technology that would help you with annotation, video reuse, video search or any other problem observed before➔ real day to day examples about how they would use it

- expectations about how the interfaces should be for annotation, video reuse, video search

- Which other data source do you use in your job?

Interview guide (for industry perspective)

Goal: gather user requirements from potential end user of the system (professional content providers, and specifically producers)

Questions:

➔ All the previous questions from the observation / interview if it is not possible to observe, plus

- Would it save you time /make your work more efficient to have help with annotation, video reuse, video search?

  - What do they see as the major technical constraints to it?
  - What do they see as the major functional constraints to it?
  - What do they see as the major organisational constraints to it?

- Questions about their perspective
  - What are the company plans to create technology that facilitates annotation, video reuse, video search?
  - How they see that the UGC would influence your business?

Any other suggestions or information that they would like to give

10. Appendix 4 Analysis of the interviews and full interview transcripts

10.1. Analysis of the data: Requirements for professional and end users

In summary, **for the professional content creators**, we have gathered information about the workflow and the way in which several professionals do their work in all the main phases that emerged:

a. Preparation of the match

The full information about the how the cameras are mounted would be presented in D2.2. It is interesting to notice that we have gathered very detailed information for premium football matches and for basketball.

b. During the match

The second stage of the production consists in selecting the right image at the right time, among the ones captured by the multiple cameras. The exploitation of a priori knowledge about the players supports that second stage in a real-time broadcasting context. For example, knowing that player 'Fellaini' is tall and strong is an incentive to focus on him during a corner.

During the game, the summary production is prepared by keeping track of the key events occurring during the game. Some of those events are defined in an objective way, while others are subjective. A typical example of objective event corresponds to a player scoring some points at some time instant. A subjective event corresponds to a representative/spectacular/intelligent/… action, performed by player X or team W at some time instant. It is worth mentioning that two parameters are used to define a time instant. First, the clock associated to the game defines the temporal location of the event within the game. Note that this clock is stopped every time the game is interrupted (fault, ball leaving the field, etc.), and is thus not synchronized with the actual time. Second, the timestamp T defines the time at which images are captured by the camera. Hence, it permits to rapidly retrieve the images corresponding to an event.

In relationship with the interaction between the journalist and the cameraman, we have investigated two different cases of professional data gathering:

- The case of only one cameraman or a small coverage of cameras for gathering the video from the match in with case we found a close interaction between the journalist and the cameraman during the acquisition process. Here the journalist is the producer, and might ask the cameraman for some specific images, e.g. the board displaying the game status at the end of every period. In any case, there are many actions in a basket–ball game which gives some freedom to the cameraman. Hence, he can move around the field and capture images from different point of views.
- On the other hand, when the match is covered with a quite complete set of cameras it is the producer who control them, and there is no direct interaction between the journalist and the camera people.

In addition to the journalist, and obviously the producer, we have identified other important roles for the selection of scenes in real time and the replays that are broadcast during the football match.

- Editor: the person in contact with the journalist via specific microphones when in an important football match, who sits in the OB van. He is the person responsible for the images that are broadcast agreeing with what the journalist is saying. In addition the journalist informs the editor of interesting things that may be happening in the field that have not been captured at this moment. The editor communicates this information to the producer.
- The producer assistant and the video team that he leads. The video team consist of 2 to 3 people that are making the decision about which images form all the video coming live from each camera is been kept and preparing in real time the replays that are broadcast for instance just after a goal. In many occasions, they have different views of a particular action and they themselves decide with one is the best angle and propose it to the producer and his assistant. The producer assistant is constantly "shouting" to the video team to pay attention to particular actions.

Nice looking production is a mixture of long and close views. Close views are important to render the emotions.

Some production rules:

- Avoid switching between cameras. In general, a switch occurs when an action ends.
- As long as the ball is in the game, no views on the public and no replays.
- Avoid two consecutive large view shots in a summary
- Explain the action by alternating 'intelligently' between long and close views.
- Take care with audio connections!

Regarding when other images are inserted during the live broadcast, like advertising, the producer is the person responsible for deciding when to insert the banner inside some of the replays ("sponsored by"). The team in charge of the replays edits some videos that they think are interesting to be used with the banner.

c. Immediately after the match:

Between the end of the game and the production process, Marie-Pierre Fivet writes down her feeling about the game, and decides about the message that should be carried by the summary, i.e. what should be explained to the viewers. Most often at this stage, Marie-Pierre already writes a draft of the text to read as an audio comment, next to the visual summary. At this stage, she might also run interviews of the main actors (players, coach, etc.), to support her message.

d. Long term storage and views on summaries

For the football match, after the match all the information selected by the video operators is saved on tapes and the team brings the tapes with them to Madrid. We are not clear about the tape format, but in Madrid they told us that they use LTO3. We think that they copy the videos in same format (SD, HD) in which it is broadcasted because they do not transcode the material before saving it on the tape.

They reported that they do not use any kind of metadata, because they use tapes to record the matches. They do not seem to have specific tags for the tapes. They told us that when, in the future, they will change and not use tapes; usage of metadata seems interesting.
Vincent Housiaux's feeling towards the APIDIS summarization objectives:

- Enthusiastic about the challenge and the perspectives that it could open, but quite sceptical towards the capability to create nice summaries without the intervention of a cameraman to zoom-in on players of interest.
- Interested in viewing intermediate proof-of-trials results.

Special effort will be made in the identification of the impact of the annotation process in the generation of summary, and the efficiency of potential search. New metadata will be identified according to the needs of the professionals.

The information about the creation of summaries in the production room would be reported in the user case related to the production room. Here we will mention that when summarizing a football match, the cameraman has to continuously capture the action from a given position, as he can not afford missing a goal.

On the other hand for the **end users**, in our case coaches, we have validated the scenarios that we proposed in the user cases. At present, the systems they have allow them to show to the players if they did the right or wrong movement, if they respect the system that was chosen for the attack or defence.
In our interviews, we found two major interests of the coaches related to APIDIS technology:

- Interest in getting the capability to browse the content by himself, so as to analyze the behaviour of players in the situations he is interested in. Getting some support for efficient content browsing (eg through the automatic detection of actions and recognition of players) sounds thus interesting in that context.
- Interest in the fun and pedagogical dimension associated to APIDIS. According to him, such system might increases the interest of young players towards basket-ball. In this context, getting the ability to focus on a given player, or to provide a representative and meaningful summary of the game, sounds thus relevant. However, as a coach, he also notes that the system might help to improve players by showing them their actual weaknesses of some players in specific situations. In that context, getting the possibility to access specific actions sounds thus interesting to him. Examples of actions are rebound and initiation of following action, shoots, counterattack, etc. Right now when they are looking a tape of the last game, they always lose time in searching for the right sequence, "always doing Forward and Rewind". If they can have access to a summary where the actions are separated, for example attack or defence or player A in defence or Player B in attack, this technology will be very useful for his work.
- One coach was interested by the deployment of the cameras used by ACIC and UCL during the game in Belgium. As their current system didn't cover perfectly the field. Joan was very interested by the cameras located on the top of the field. He said that this view is the best view that you can get.

10.1.1. Analysis of the data: Content distribution infrastructure

The information in this area has been gathered in the interview with Alcatel-Lucent: Open

- Internet could potentially overcome the current walled garden paradigm. Hence, generating content automatically makes sense.
- Interactivity is not yet the main path considered personalizing content. They are more interested in advertisement insertion, and content adaptation to the user needs (based on a profile map).

10.2. Analysis of the data: Key rules for sport event video capture using multi-cameras

The rules and key principles for sport event production have significantly evolved since the 1998 Football World Cup in France, under the impulse of a French producer who wanted to *explain* the actions as much as possible. Since that period, explaining to the viewer what is happening in the game has become the fundamental and federating principle behind all artistic and technical production innovations. Among the key elements that support a modern sport event production, we note:

- The use of more and more cameras, which are close to the action. Typically 15 cameras are used to cover a football match (on the side: cameras with large- and medium view angles + offside cameras, in corners: ground-level and higher-level cameras for slow motion replays + cameras in the game axis). In Munich there is a camera fixed on a cable that capture images from above the game. It provides a novel and deep insight into the game. It encourages us to deploy omnidirectional cameras on the ceiling of the basket-ball court.
- The 'axis rule'. When distributing multiple cameras around the sport field, one should always take care to display consistent players' movements and gazes. For that purpose, all cameras are located on the same side of an axis that is chosen in the main direction of the game. Typically, the axis joins the two goals in a football game.
- The first stage in the production consists in capturing the right image at each camera-end. For that purpose, an implicit reference 'bible' is exploited, which defines for each kind of actions what should be done by each cameraman. In other words, it defines what to capture depending on what happens in the scene. For example, during a penalty, the camera behind the goal should zoom-in on the player kicking the ball, while another camera should focus on the referee, etc…

Conclusions:

- Key objective for the producer: explain the action to the user.
- Some production rules:
  - Avoid switching between cameras. In general, a switch occurs when an action ends.
  - As long as the ball is in the game, no views on the public and no replays.
  - Avoid two consecutive long view shots in a summary
  - Explain the action by alternating 'intelligently' between large and close views. Close views are important to render the emotions.
  - Take care with audio connections!

10.3. Interview transcripts

10.3.1. Report interview Vincent Housiaux, RTBF/IAD, 21/03/08.

Author: C. De Vleeschouwer, UCL.

About Vincent Housiaux:

Vincent Housiaux is now leading the department guiding the artistic aspects of RTBF programs. He has been the producer of numerous television programs, including sports events, mainly football broadcasts and summarizations. For example, he was in charge of production of short sequences focusing on a player during EURO2000 in Belgium.
Besides, Vincent Housiaux is also invited professor at IAD (www.iad-arts.be/v_british), an artistic postsecondary school in the domain of the performing arts, media and communication techniques. He teaches students about (multi-camera) television production mechanisms.

Introduction:

This document reports a fruitful feedback received from Vincent Housiaux at RTBF. The main objective was to identify and understand the key principles governing the production and summarization of multi-camera sport-events. Envisioned use cases are the automatic generation of summaries for Internet portal, and assistance to the production room.

The questions that have been raised during the discussions include:

- Question about the work-practices for companies that record and summarize sport events:
  - **Rules and constraints special to the production** of a particular sport event.
  - How do they create their summaries currently? Do they have any technology that facilitates the creation of summaries?

- Questions about automatic generation of summaries:
  - In which occasions would you find useful to have an **automatic generated summary**?
  - Would it save you time /make your work more efficient?

- Questions about salient segment identification:
  - what do they see as their major problems/barriers when dealing with:
    - annotation
    - video reuse
    - video search

  - Which attributes/annotations/keys would be desirable to efficiently search and access content?
  - Does video shots prioritization make sense? Is it realistic to monitor the producer actions so as to infer a level of importance for each shot?

- Questions about the perspectives opened by APIDIS:
  - What does "**personalisation**" mean for them➔ examples of different ways in which they would like to have the content form different "perspectives"

  - Use of internet and **internet portals** (related to the sport/team) or in general

Outcomes of the discussion:

Rules and key principles for sport event production have significantly evolved since the 1998 Football World Cup in France, under the impulsion of a French producer who wanted to *explain* the actions as much as possible. Since that period, explaining to the viewer what is happening in the game has become the fundamental and federating principle behind all artistic and technical production innovations. Among the key elements that support a modern sport event production, we note:

- The use of more and more cameras, which are close to the action. Typically 15 cameras are used to cover a football match (on the side: cameras with large- and medium view angles + offside cameras, in corners: ground-level and higher-level cameras for slow motion replays + cameras in the game axis). In Munich there is a camera fixed on a cable that capture images from above the game. It provides a novel and deep insight into the game. It encourages us to deploy omnidirectional cameras on the ceiling of the basket-ball room.
- The 'axis rule'. When distributing multiple cameras around the sport field, one should always take care to display consistent players' movements and gazes. For that purpose, all cameras are located on the same side of an axis that is chosen in the main direction of the game. Typically, the axis joins the two goals in a football game.
- The first stage in the production consists in capturing the right image at each camera-end. For that purpose, an implicit reference 'bible' is exploited, which defines for each kind of actions what should be done by each cameraman. In other words, it defines what to capture depending on what happens in the scene. For example, during a penalty, the camera behind the goal should zoom-in on the player kicking the ball, while another camera should focus on the referee, etc…
- The second stage of the production consists in selecting the right image at the right time, among the ones captured by the multiple cameras. The exploitation of a priori knowledge about the players supports that second stage in a real-time broadcasting context. For example, knowing that player 'Fellaini' is tall and strong is an incentive to focus on him during a corner. Note: APIDIS considers posterior summarization. Hence, we might exploit the fact that we know in advance who will score to select the image to show.
- Nice looking production is a mixture of large and close views. Close views are important to render the emotions.
- Some production rules:
  - Avoid switching between cameras. In general, a switch occurs when an action ends.

- As long as the ball is in the game, no views on the public and no replays.
- Avoid two consecutive large view shots in a summary
- Explain the action by alternating 'intelligently' between large and close views.
- Take care to audio connections!

Contact point related to multi-camera sport-event production: Thierry Delrue, TDE@rtbf.be, strong experience in multi-camera sport event production.

Currently, the **summarization of a sport event** implies that a journalist surveys the whole match, and keeps tracks of the time stamp associated to an action of interest together with its level of interest (via a number of stars: one star means 'worthwhile action'-four stars corresponds to a 'crucial action' such as a goal). We conclude that the approach is close to the salient segment identification envisioned by APIDIS.

Current approach to **content personalization** is pragmatic and does not really involve viewer's feedback. It mainly consists in making programs available on the web portal at multiple rates.

Vincent Housiaux's feeling towards the APIDIS summarization objectives:

- Enthusiastic about the challenge and the perspectives that it could open, but quite sceptical towards the capability to create nice summaries without the intervention of a cameraman to zoom-in on players of interest.
- Interested in visioning intermediate proof-of-trials results.

10.3.2. Report interview Marie-Pierre FIVET, Canal C, 05/04/2008.

Author: C. De Vleeschouwer, UCL.

About Marie-Pierre FIVET:

Marie-Pierre FIVET has been producing and commenting video summary reports for Canal C local television (www.canalc.be) for more than 5 years. She relies on a solid experience as a semi-professional basket-ball player (National D1) to identify the main components of a game (tactical decisions, key players, representative actions, etc.), and help the viewer to understand them.

Introduction:

This document reports the feedback received from Marie-Pierre FIVET about her way to summarize a basket-ball match. The main objective was to identify some of the key principles governing the summarization of a basket-ball match. Relevant APIDIS use cases are the automatic generation of summaries for Internet portal, and assistance to the production room. The questions that have been raised during the discussions include questions about the work-practices, and questions about salient segment identification.

Outcomes of the discussion:

Marie-Pierre Fivet generally attends the games she has to summarize. The production process can be split into three consecutive and complementary stages.

During the game, the summary production is prepared by keeping track of the key events occurring during the game. Some of those events are defined in an objective way, while others are subjective. A typical example of objective event corresponds to a player scoring some points at some time instant. A subjective event corresponds to a representative/spectacular/intelligent/… action, performed by player X or team W at some time instant. It is worth mentioning that two parameters are used to define a time instant. First, the clock associated to the game defines the temporal location of the event within the game. Note that this clock is stopped every time the game is interrupted (fault, ball leaving the field, etc.), and is thus not synchronized with the actual time. Second, the timestamp T defines the time at which images are captured by the camera. Hence, it permits to rapidly retrieve the images corresponding to an event.

Between the end of the game and the production process, Marie-Pierre Fivet writes down her feeling about the game, and decides about the message that should be carried by the summary, i.e. what should be explained to the viewers. Most often at this stage, Marie-Pierre already writes a draft of the text to read as an audio comment, next to the visual summary. At this stage, she might also run interviews of the main actors (players, coach, etc.), to support her message.

In the production room, the work of the journalist mainly consists of finding the images supporting and illustrating the message to transmit. Particular attention is paid to the balance between interviews and images of the game.

Some additional interesting observations:

- there is a close interaction between the journalist and the cameraman during the acquisition process. Here the journalist is the producer, and might ask the cameraman for some specific images to, e.g. the board displaying the game status at the end of every period;
- the fact that there are many actions in a basket–ball game gives some freedom to the cameraman. Hence, he can move around the field and capture images from different point of views. In contrast, when summarizing a football match, the cameraman has to continuously capture the action from a given position, as he can not afford missing a goal.

Conclusions:

From the above discussion, we can draw a number of lessons for the APIDIS project:
- It is relevant to identify events of interest along the game, and derive a summary based on those events.
- It is important to drive the summarization process based on some underlying message about the game. In APIDIS, the user preferences introduced by the user about the actions he is interested in should partly fullfill that requirement.
- Interviews of key actors provide a natural support to the summarization. APIDIS should perhaps offer the capability to record some pre-formatted interviews. It raises an interesting open question.

10.3.3. Report interview basket-ball coaches, 04/2008.

Author: C. De Vleeschouwer, UCL.

About the coaches:

Two coaches have provided some feedback about the APIDIS acquisition and summarization system:
- Jurgen VAN MEERBEECK (JVM) is the coach of DEXIA NAMUR, first national division.
- Jean-Marc ALBERT (JMA) is the coach of a local non-professional team (IATA, playing in the 3rd provincial division in Namur).

Outcomes of the discussion:

Both coaches have somewhat different interests in the APIDIS summarization process:

- JVM is definitely interested in the acquisition system and in the raw data content, especially in the images captured by the cameras located at the roof. Those images provide clear and non-ambiguous views of the position of players over the field, which in turn permits him to identify the strong/weak points of each team. JVM is not that much interested in a TV-like summary of the game. He seems more interested in getting the capability to browse the content by himself, so as to analyze the behaviour of players in the situations he is interested in. Getting some support for efficient content browsing (eg through the automatic detection of actions and recognition of players) sounds thus interesting in that context.

- JMA is more interested in the fun and pedagogical dimension associated to APIDIS. According to him, such system might increases the interest of young players towards basket-ball. In this context, getting the ability to focus on a given player, or to provide a representative and meaningful summary of the game, sounds thus relevant. However, as a coach, he also notes that the system might help to improve players by showing them their actual weaknesses of some players in specific situations. In that context, getting the possibility to access specific actions sounds

thus interesting to him. Examples of actions are rebound and initiation of following action, shoots, counterattack, etc.

Conclusion:

Next to the 'conventional' summarization process, coaches are interested in a manual or semi-automatic browsing of the content, based on the identification of events and actions of interest.

10.3.4. Report interview basket-ball Coach high performance school, 06/2008.

Author: Eric Martrou, Mediapro

About the coach:

Joan Amat is the trainer of 2 teams at a high performance school in Esplugues de Llobregat, Barcelona:

- 1st team: boys 14-16 years old
- 2nd team. girls same age

Actually, Joan and his assistants used a system of 3 cameras (Handycam Sony mini DV). 2 cameras are mounted on tripods (around 4m high) on each side of the field. The other camera is used by one of his assistants along the field.

This system allows Joan to see, after each training or game, the weaknesses of his teams and players. This allows him to show to the players if they did the good or wrong movement, if they respect the system that was chosen for the attack or defence.

Outcomes of the discussion:

Regarding the APIDIS technology, Joan was interested by the deployment of the cameras used by ACIC and UCL during the game in Belgium. As their current system do not cover perfectly the field. Joan was very interested by the cameras located on the top of the field. He said that this view is the best view that you can get.

Another interesting point is that Joan seems to be curious about the automatic summarization of content. Right now when he and his players are looking a tape of the last game, he said that he always lose time in searching the right sequence, "always doing Forward and Reward". If he can have access to a summary where the actions are separated, for example attack or defence or player A in defence or Player B in attack, this technology will be very useful for his work.

Conclusion:

The coach shows a positive interest in the cameras deployment and the automatic summarization process.

10.3.5. Alcatel-Lucent vision about content access personalization.

Author: C. De Vleeschouwer, UCL.

Contact and context:

Benoit Quirynen is the head of the video group in Alcatel-Lucent, Namur. UCL and Alcatel-Lucent meet on a monthly basis, in the context of the project WalCoMo, funded by the Walloon Region. The vision expressed in this report results from these meetings and discussions. It is expected to be completed in the future by Alcatel-Lucent public slideshows.

The questions that have been raised during the discussions include:

- What does "**personalisation**" mean for them➔ examples of different ways in which they would like to have the content form different "perspectives"
- Constraints associated to the deployment of infrastructures for personalized access to content?
- Use of internet and **internet portals** (related to the sport/team) or in general

Alcatel-Lucent's vision:

The main idea underlying personalized access to content is driven by the will to break the walled garden principle underlying current video distribution mechanisms. The objective is to offer an infrastructure that allows to make the content produced by anybody (amateur web portal, professional content provider, …) accessible by any end-viewer, through open Internet connections.

Such architecture builds on video servers that can handle thousands of connections at a time. To make it more profitable, the infrastructure is enriched to allow for content mixing, insertion of advertisement into the content, and so forth. Obviously, the technologies envisioned by APIDIS will contribute to increase the mass of content made accessible through such kind of infrastructures, thereby increasing their return on investments.

A major concern regarding the economic viability of the system lies in the way the transmission cost will evolve on the Internet. Currently, high bandwidth connections are cheap, but if services with added values are deployed through those nearly free connections, Internet providers might decide to increase the cost of the bit. This might kill the 'open garden' video distribution model.

The opportunity to deploy interactive functionalities to further personalize the content depends on the application context, but does not constitute a fundamental aspect of the personalization for the following reasons:

- When considering television, a set top box can be used to provide some interactive functionality. However, really personalized content can only be envisioned in a VoD context, since multicast is exploited for television programs.
- When dealing with wireless mobile access, personalization appears to be more natural to deploy since each user receives an individual access. In such context, however, the main interest of Alcatel-Lucent lies in the possibility to give the end user the capability to access any content available on the web at any time, rather than thinking about interactive personalization of the content. This is because the mobile phone interface is quite limited and already largely used to control playlists.

10.3.6. Nicolas Point is responsible for the security of Multitel building.

Multitel is a research center of 80 persons managing their time with a great diversity. There cleaning staff arrive very early and some researchers leave very late. The problem of N Point is to know who was the last person to leave the building and who was the first to arrive. The building is equipped with an alarm system. N Point has to investigate who was responsible when a false alarm is raised (wrong code, doors left open). N Point being a computer scientist, he has developed a script that produces a summary of the entrances and exits of the day. An Axis camera points towards the main door. The motion detection function is on. When a motion occurs in from of the door, a recording session is launched. The system records 10 seconds before and 10s after the event at 5frames per second. At the end of the day, his script puts all videos sequences together to form one video clip, that he watches every morning.

Interactive and semantically-driven access to video surveillance content- Observation and interview guide

Goals of the technology for which we are gathering requirements:
This use case considers the content acquired by a network of surveillance cameras, covering a building or a secured area.
Search within a video ➔ gain of time and/or a better accuracy in the search for visual events.
Creating summaries ➔ 'forensic' usage of the system, the user might want to influence the production of the summary to focus on specific area or events of interests.

Observation guide (for end users)

Goal of the observation: to gather user requirements from end users

Steps/ what to observe (if not possible to observe, to ask in an interview):

- how the users prepare for the recording of a CCTV (problems, barriers, expectations)
    - why and how they record it
    - how is decided what to record
    - location of the cameras and which information they try to get with the location of the cameras

- how the users now record ➔ observations rooms? Person responsible?
- what they do with the recorded material actually
    - in which situations they go back to the recorded material
    - how they do it
- long term repository of the recorded materials

Questions:

- what are they see as their major problems barriers (complementing the observation)

The detection system is limited. This makes false alarms and missed detection. The produced summary is sometimes not accurate. Moreover, the program is not capable to filter out the moments when a group of persons stay in from of the camera for some time. More globally, N Point would like a better image quality while at the same time keeping the volume of data reasonable.

- What would be the impact of being able to do have summaries➔ real day to day examples about they would use it In which occasions would you find useful to have an automatic generated summary? Would it save you time /make your work more efficient?

Basically, the interest for N Point is to gain time, but he would also be interested in the capability to send the summaries remotely (on the week-ends).

- For your daily work, how often/when/why do you need to generate a summary of a video

Every day

- What do you do with the video (how to store it, publish, send it to other people, etc)

The video is viewed everyday, but no archiving is done except when a major security problem happens, which is rarely the case.

- How do you create your summaries currently?

An Axis camera points towards the main door. The motion detection function is on. When a motion occurs in from of the door, a recording session is launched. The system records 10 seconds before and 10s after the event at 5frames per second. At the end of the day, his script puts all videos sequences together to form one video clip, that he watches every morning.

- Do they have any technology that facilitates the creation of summaries?

See above

- What does it mean for them to be able to choose interactively and semantically in the content➔ examples of different ways in which they would like to have the content form different “perspectives”
    - A particular event? A suspect? Groups of people? Anything strange (what does it mean?)

To be able not to miss entrances/exits. The ultimate goal is to know the last person exiting the building between 7PM and 6AM (when the alarms are set on) and the first person entering the building.

- What they would like to do that they cannot do now: examples from their daily activity
- Legislation constraints in what they can do with the video material and how they can/can not use it.

He has to respect private life in the context of the working laws. People are warned that a camera is on. No archiving is done.

- Questions about their industrial requirements: After the description of what APIDIS technology goal is:
    - What do they see as the major technical constraints to it?
    - What do they see as the major functional constraints to it?

Illumination issues (when the person switched on the light before setting the alarms on/off), people staying in front of the camera, missed detection, false detections

10.3.7. M Hubert Lenoir, Federal Police, CGSU (Investigation special unit)

is the equivalent of the GIGN in Belgium and is well-known throughout Europe
16/04/2008 – Mons, Belgium

Interactive and semantically-driven access to video surveillance content-
Observation and interview guide

Goals of the technology for which we are gathering requirements:
This use case considers the content acquired by a network of surveillance cameras, covering a building or a secured area.
Search within a video ➔ gain of time and/or a better accuracy in the search for visual events.
Creating summaries ➔ 'forensic' usage of the system, the user might want to influence the production of the summary to focus on specific area or events of interests.

Observation guide (for end users)

Goal of the observation: to gather user requirements from end users

Steps/ what to observe (if not possible to observe, to ask in an interview):

- how the users prepare for the recording of a CCTV (problems, barriers, expectations)
    - why and how they record it
    - how is decided what to record

  Usually, the justice and/or the police (the investigators or interviewers) delegates CGSU for the close surveillance of a suspect.

    - location of the cameras and which information they try to get with the location of the cameras

      Typically they place 1 to 4 cameras to observe the activity in the house of the suspect (the entrances), who comes in, who goes out. Typically faces and

licence plate are extracted from the video. The investigator chooses the location.

- how the users now record ➔ observations rooms? Person responsible?

The cameras are hidden in a neighbour's house, behind the window. Their video stream is recorded in a portable DVR.

- what they do with the recorded material actually
  - in which situations they go back to the recorded material
  - how they do it

This point is a major issue. Since the equipment is hidden, it is difficult to go often to download the videos and empty the HDD of the DVRs. Usually, the person hosting the system brings the HDD to the police. This puts the person at risk and the Police want a new system Typically there is a need to efficiently compress the video (by storing only the important moments) and send a summary of the day by radiocommunication (UMTS?). In practice, every time the HDD is back to the police, they have to watch all the recorded material and write a report for the viewed period.

- long term repository of the recorded materials

The repository is done until the judgement, the archiving support is CD/DVD.

Questions:

- what are they see as their major problems barriers (complementing the observation)

There three major problems:

4. The difficulty to access the recorded material, as explained here above, and the risk for the police or the person hosting the equipment, when the transfer is done physically. Due to the amount of recorded material, it is impossible today to send the videostreams trough wireless
5. The requirement not to miss any event. Non-detections must be avoided, false alarms have to be minimized, but are less critical then false negatives!
6. The system must be portable and easy to install and must work day/might in all conditions

- What would be the impact of being able to do have summaries➔ real day to day examples about they would use it In which occasions would you find useful to have an automatic generated summary? Would it save you time /make your work more efficient?

As explained above, the impact would be enormous, from both an efficiency and a security perspective, provided that the summary does not introduce missed detections.

- For your daily work, how often/when/what for do you need to generate a summary of a video

Once a day to once a week, depending on the cases

- What do you do with the video (how to store it, publish, send it to other people, etc)

Burnt DVDs (note that VHS tapes are still used at the Police)

- How do you create your summaries currently?
- Do they have any technology that facilitates the creation of summaries?

The use a laser telemeter pointing on the door to launch the recording when the door opens, but it is not enough. So, they have to reprocess manually afterwards. Sometimes, they use a postprocessing software to search backwards in the videosequences.

- What does it mean for them to be able to choose interactively and semantically in the content➔ examples of different ways in which they would like to have the content form different "perspectives"
    - A particular event? A suspect? Groups of people? Anything estrange (what does it mean?)

  They need to get all the faces and all the car licence plates
- What they would like to do that they cannot do now: examples from their daily activity

A very reliable system (no missed detections) and an efficient summarizer, which can be sent/viewed remotely through UMTS. The system must be easy to use and install and the calibration of the system must be fast and discrete (the suspect can not see it).

- Legislation constraints in what they can do with the video material and how they can/can not use it.

They just need a mandate to start the recordings. Then the confidentiality of the recordings must be granted. The VHS or DVD must be kept until the judgment.

- Questions about their industrial requirements: After the description of what APIDIS technology goal is:
    - What do they see as the major technical constraints to it?
    - What do they see as the major functional constraints to it?

See above:

99,9% detection rates, recognition of the faces and car licence plates, portability,

easy to use and easy to install

10.3.8. Visit to "La Sexta" TV channel

1. **Annotation of a video: step-by-step, technology use, co-ordination with other people, work flow, problems, etc.**

Main findings:
6 has two main processes of annotation:

- pre-catalog: they create number of identifiers of the clip (ID, journalist, copyrights, time, date, duration, international/agency or national), all them added manually ➔ they would like an automatic ID
- Annotation of the news and the frames, which are keep it two different records:

- Annotation of the frames: which describes the frames based on several criteria: type shot, if it would be useful for promos,
- New cataloguing: which keeps the main info from the new agency

The annotation of the frames and selection of the storyboard is based on two main type of knowledge:

- Knowledge of the 6, inhouse knowledge, of what it may be valuable before
- Active collaboration with the journalist that tells them what is more relevant information

Tools that help in the annotation:

- Annotation guidelines from direction
- Internet: APTN, Reuters, EFE, etc.
- Thesaurus (keep by the director only)
- Wikipedia adn R.a.e

Type of descriptors of the frames: type of shot, "promo material", "interesting shot", "supergol" "person talking"
The criteria to select frames for the storyboard is subjective, depend on the annotators expertise ➔ more clear the automatisation for football because they have clear guidelines about the type of shots they will need for the summaries and promos

2. Video search

- how they do it now: details on tools/procedures/ specialised roles/etc
- Thesaurus to make sure they search by the right word
- Descriptors:
  - ID
  - Time code (the most important)
  - Place of the news
  - "Visual Material"
  - "Materia oculta"
  - Visual entity
  - Hidden entity
  - Onomástico visual
  - Onomástico hidden
- in which context, for which programs, with how much anticipation? ➔ for a variety of programs but we did not observe that in detail

3. Video re-use: what they do with the recorded material actually (problems, barriers, expectations)

- **How they select the content to keep and document**
  - National content: the journalist makes a summary and send it to documentation department
  - International content: the decision is take by the archiver
- **usage of the video by him/her/other groups in the company**

the usage and type of material depend on the section inside the TV, for instance "promos" o "sport summaries"

- **tools used for accessing the content**
  - It is important to be able to prioritise the request
  - If the content is less than 2 days old it will be very quickly accessible, if not it is in discs

- Each content has an ID and the mark of the disc where it is stored
- WM-MPEG2
- AVID; Unitecnic, Insctinct (iNews (only text) & Media Manager), TDIAL, Newscutter, Flow monitor

4. Long term repository of the recorded materials: details on tools/procedures/ specialised roles/etc

- **who is responsible for it?**
- 2 MAM:
  - Madrid→brutos informativos
  - Barcelona→emisión
- All the programs are duplicated in Barcelona y en Madrid.
- The most recent material is online and the rest in discs LT03 which are physically stored in shelves
- Madrid uses DV and in Barcelona BP

5. Design suggestions

- **Annotation**
- Voice and face automatic recognition
- Automatic metadata in pre-catalog (specially date)
- Informativos: they would like Interplay, just one tool
- Informativos: iNews is too basic
- iNews to prevent errors in ID
- Automatic ID
- **Video reuse**
- Automatisation of football storyboard where the change in shot is more important (it is seem as too complicated for other news)
- **Video search**
- Be able to add operators to the search (plus, minus, no, etc)
- Highlighted of the search cue