APIDIS

Autonomous Production of Images based on Distributed and Intelligent Sensing

STREP Project, 1st FP7-216023

D2.4 End of project trials: specs and assessment methods

Due date of deliverable: 31-08-2009
Actual submission date: 25-09-2009

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

Lead contractor for this deliverable: [Revision Final v1]

D2.4	End of project trials: specs and assessment methods
Project Acronym :	APIDIS
Contract No :	FP7-216023
Due Date :	31-08-2009
Reply To:	Eric Martrou Email emartrou@mediapro.es
Actual date of delivery	25-09-2009

1 Executive Summary

This document defines the scenarios and the associated measures of success for the end of project trials scheduled between months 30 and 33. The trials deal with three scenarios; interactive and semantically–driven access to video surveillance content, sharing and remixing content, and automatic generation of content for Internet portals.

Deliverable Identification Sheet

Project ref. no.	FP7-216023
Project acronym	APIDIS
Project full title	FP7-216023
Security (distribution level)	Public (PU)
Contractual date of delivery	Month 20, 8 31, 2009
Actual date of delivery	Month 21, 9 25, 2009
Deliverable number	D2.4
Deliverable name	End of project trials: specs and assessment method
Type	Report
Status & version	Final Version
Number of pages	15
WP / Task responsible	WP2/ Mediapro
Other contributors	UCL, BM, ACIC
Author(s)	Eric Martrou
EC Project Officer	Marzio Morina
Abstract	This document defines the scenarios and the associated measures of success for the end of project trials
Keywords	Trials, specs, assessment
Sent to peer reviewer	*Yes*
Peer review completed	*Yes*
Circulated to partners	Yes
Read by partners	Yes
Mgt. Board approval	Pending

Table of contents

2 Introduction

This document defines the scenarios and the associated measures of success for the end of project trials scheduled between the months 30 and 33. The trials are run to prove the main concepts underlying the APIDIS framework.

The trials are divided in three sections:

- interactive and semantically–driven access to video surveillance content,
- sharing and remixing content,
- automatic generation of content for Internet portals.

For each trial, we define:

- the context and objective,
- the methodology,
- the trial scenario,
- the assessment.

3 Interactive and semantically–driven access to video surveillance content

Involved partner: ACIC (lead) with BM

Context and objective:

The aim of the intelligent video surveillance browsing tools developed in APIDIS is to ease the work of security guards operating in a control room when they have to browse a huge amount of video data.

We put ourselves in the general context where the video streams generated by surveillance cameras are recorded continuously, and for which efficient video browsing is needed to determine which event occurred during the recorded period.

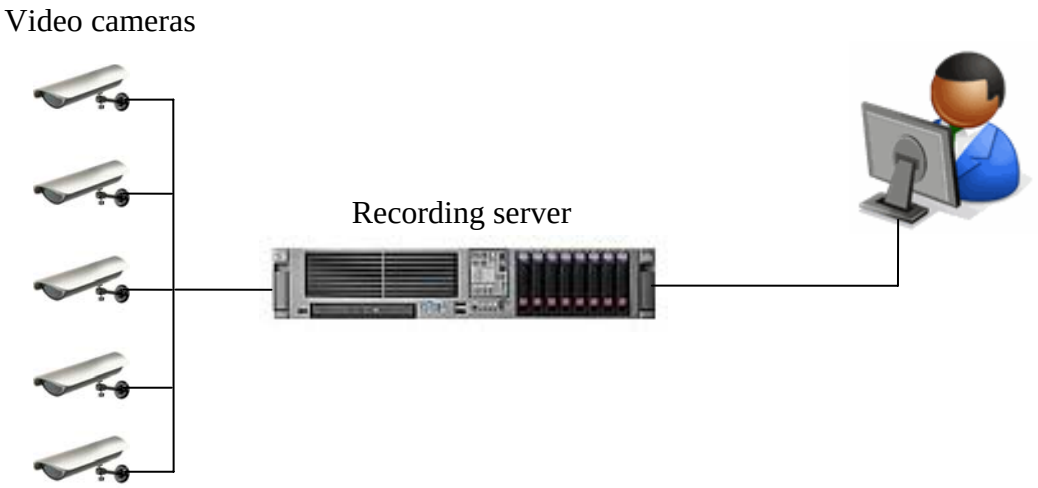

Figure 1: Video surveillance context

The video surveillance architecture proposed by APIDIS is the following:

1. Video cameras are continuously recorded by a server.
2. On-the-fly extraction of metadata associated to the videos, e.g. tracks of moving objects.
3. From the metadata, automatic learning of some relationships between the cameras, as well as automatic learning of frequent behaviours/trajectories.
4. Optimised browsing of the recorded data through the display of the most relevant cameras associated with what happens in a main view.

The system intelligence should help the video surveillance operator in the following cases:

- Browsing of all the events or activities that occurred when nobody was looking at the cameras, e.g. all the events of last Sunday.
- Supervised tracking of a person over several cameras by an operator that do not necessarily knows the topology of the deployed video surveillance system.
- Retrieval of the date at which a particular event has occurred, e.g. a car has left the parking.

The typical application field which is considered is the one of a building that is covered by a network of cameras.

Notes:

- Although the solution needs to extract some features from the videos, the APIDIS objective is not to develop complex video surveillance event detection algorithms. The task consists of exploiting easy-to-extract inputs in order to provide efficient techniques for browsing in the recorded videos.
- The system should never decide which camera to display by itself, otherwise explicitly defined in the parameters. The objective is to help the operator browsing the content despite the fact that low level features extraction might be subject to errors.
- The goal of APIDIS is not to produce the best interface in terms of design. The main role of the browsing prototype is to validate to which point we are able to improve the browsing experience through some automatic learning.

Methodology:

The trial will be based on pre-recorded video streams. However, the browsing interface will implicitly demonstrate the proof of concept for a live situation.

The datasets currently considered for the trial are:

- ACIC premises dataset: The video surveillance dataset acquired end of 2008. It is composed of 24 hours of recordings for eight cameras distributed around and in a building, as a real life deployed network of video surveillance cameras.
- PTZ dataset: New sequences that will be acquired with one or two of the cameras of the first dataset. A few other cameras will be installed close to each other but pointing towards different areas so as to simulate some static positions of a single PTZ camera. This dataset can be used to simulate how the APIDIS solution could automatically select a PTZ orientation to follow a target.
- End-users dataset. In case of need, we still have access to the network of cameras of the first dataset. We will give the opportunity to the end-users participating in the final trials to come and record the scenarios of their choice.

Note that only the two first datasets will be considered for system training purposes, whilst the third one will only be used to test and validate the system.

For completeness, note also that the Multiple Camera Tracking Scenario (MCTS) from i-LIDS has been considered as a candidate dataset by the APIDIS consortium. This dataset is composed of five cameras deployed in Gatwick airport. The context is however very challenging regarding the automatic tracking of people (crowded areas with lots of occlusions). As a consequence, the key issue has to do with scene feature extraction (which is not central to APIDIS), rather than with the learning of relationships between cameras. Therefore, this dataset will not be used for the final trial.

Video recording, metadata extraction and learning should run in real-time in an APIDIS deployed system. During the trial, no computational optimization was scheduled in the DoW. We will however rely on data that have been computed and processed off-line, and can be accessed from a storage device.

Trial scenario:

The people who will test the APIDIS browsing interface will be in front of a screen with a mouse and a keyboard.

In order to be as close as possible to a real situation, we will explain to the people who will test the prototype what the “product” aims to do.
Then, we will encourage them to browse the dataset freely so they can evaluate by themselves to which point the “product” fulfils its objective, and to which point it improves traditional video surveillance browsing.
Next, we will put them in front of a number of controlled scenarios, in order to get feedback about all the functionalities supported by the system. As an example, a pre-defined scenario may be: follow the person that is in camera *n* at time *d* in the network of cameras until (s)he leaves.

The current objectives are that the browsing interface supports the following functionalities:

- The video of one camera is played in the main view. According to what happens in it, a set of views that are expected to become relevant in the future are presented.
- The video of one camera is played in the main view with at least two moving objects. By clicking on one of the bounding boxes, the list of relevant future cameras is adapted to focus on this object.
- Interactive search of when a particular object has disappeared (e.g. been stolen) from the scene. Given a time interval, the interface allows us to efficiently retrieve the time when the car has left the parking.
- High level description of events. Similar events are grouped for each camera, e.g. all the tracks that start at the same position and end at another same position are grouped. This allows us either to browse only a group of events (to search for someone, knowing his/her path) or to browse ungrouped events (that could be considered as suspicious).
- Same feature as the previous one but for events that are described in multiple views.
- One of the small views is used to show the best view of the current scene (possibly with zoom on the mobile object).

- If a PTZ camera with a set of pre-defined positions is available, the system is able to select the most relevant position according to the activity observed in the other cameras.

Assessment:

Subjective evaluation:

The people who will evaluate the prototype will have to test different browsing scenarios and give their feedback on how the prototype has helped them in each of them. For each of the functionalities, they will rate the prototype and optionally give some comments.
The questionnaires will be filled in with additional information like: personal background on video surveillance browsing tools, initial knowledge of the topology of the network of cameras, etc.

Objective evaluation:

An objective assessment of a browsing tool is not easy to implement. However, here are some tracks:

- Logging of the user actions in order to get some metrics, e.g. the number of frames they have seen for a particular search.
- Storage space required by the metadata compared to the video data.
- Processing power required to generate the metadata and automatic learning.

4 Sharing and remixing content

Involved partner: MEDIAPRO (lead), and UCL

Context and objective:

The purpose of this trial is to demonstrate the novel form of content consumption enabled by the automatic summarization capabilities of APIDIS.

The content considered as input to our trial corresponds to the content broadcasted by TV channels, augmented by metadata (e.g. which camera is used for the view, type of shot, player and his position. etc.) that can be either provided in a natural way by professionals working in the production room (such potential inputs are derived from the ethnographic studies performed in Task 2.1 and 2.2) or computed automatically based on simple audio and/or image processing tools (e.g. computing the view type, the audio intensity, etc.).

Optionally, additional semantic information can be collected through manual annotation, e.g. to identify and characterize the events-of-interests within the content.

Given those inputs, the APIDIS solution is expected to build length-constrained summaries that can be personalized in terms of semantic interest or narrative preferences of the users.

During the trial, we want to demonstrate how this capability can be exploited:
1. to give a second life to the content, by offering the access to personalized summaries through web or mobile access channels;
2. to help the (professional) user to search for the segments to include in a re-mixed content, e.g. by locating all actions of a specific type or involving a specific player in one or several games. In this case, a summary of key actions is provided, and the user selects on the fly the segments (s) he wants to further process to produce his/her remixed content.

Methodology:

The content and the metadata directly collected from the production room will be provided by MP. UCL and MP signal processing tools will then augment those metadata based on generic image and audio analysis tools. As explained in D6.2 and D7.1, those metadata are sufficient to segment the video, and build local stories for each segment.

In addition, through manual annotation mechanisms, the content provider gets the opportunity to define which parts of the content are important or not, thereby embedding some form of semantic knowledge in the stored content. The APIDIS summarization technology exploits that information to automatically adapt the

content to the user needs (in terms of duration, semantic interest, etc.), while preserving the information considered as fundamental by the content provider.

However, for both exploitation scenarios described above, the richer the semantic annotation, the better the personalization of the summary. To circumvent the human load associated to accurate manual annotation of the content, we plan to investigate a (social media) content sharing framework. In this framework, the annotation tool would be made interactive and available online. The tool would invite and encourage[1] users to fill-in annotation forms for those video segments that are expected to be of particular interest, as inferred from the presence of replays or increased audio intensity. Hence, each user gets the opportunity to enrich the content, and the annotation load is distributed within the community of users.

Real-life deployment of this framework is certainly beyond the scope of the project. However, we plan to run preliminary investigations to validate the approach.

Trial scenario:

During the trial, the solutions proposed by APIDIS to summarize the pre-annotated content will be tested both by regular end-users (i.e. the man in the street), and by production professionals (e.g. to search for the content to illustrate daily news for TV or electronic medias).

We will first explain to them how the tool works and what the “prototype aims to do”. The regular end-users and the production professionals will have access to the content, their associated metadata and the “on line” tool that will permit them to recognize the moments of interest.

The users will be able to summarize the content depending on their preferences. The results given by the users will make the annotation more accurate and provide a clear vision of which type of segments of the content is important for the end-users and the professional users.

After this, professionals who generally work in the re-mixing process will have access to the summarized content in order to use the content.

Assessment:

Objective measurements:

Objective assessment will only consider the interactive annotation tool.

[1] As a possible incentive to encourage users to contribute to the annotation, free content access could for example be offered to the users who accept to fill-in the annotation forms.

As explained above, the tool addresses requests to the end-users, so as to locate and collect information about the events-of-interest of the content. In this context, objective assessment can simply rely on counting the number of requests required to collect (part of) the information gathered within a manually pre-defined annotation ground truth. The fewer requests are needed to collect information, the more efficient the annotation tool.

Subjective evaluation:

The core of the trial assessment will be based on subjective evaluation of the functionalities offered by the summarization tool.

Given some pre-annotated content, the users (mainly professionals from MP) and regular end-users will get the opportunity to exploit the tool to search for the content needed to address a number of pre-defined content re-mixing objectives (e.g. locating all actions of a specific type or involving a specific player in one or several games). The users will then give their feedback on how the tool has helped them in reaching each one of the objectives. Furthermore, they will conduct the same task using conventional content remixing tools, thereby allowing comparing the two processes and their respective results.

Assessment will rely on questionnaire-based interviews, and on open discussions with users.

5 Automatic generation of content for Internet portals

Involved partner: UCL (lead), and QMUL.

Context and objective:

The trial considers the automatic and personalized production and summarization of a basket-ball event.

The data sample exploited during the trial has been captured by a set of cameras distributed around the basket-ball field during one of the APIDIS acquisition campaign (April 08 or 09). Salient objects, i.e. the players and the ball, have been located as a function of the time, based on automatic detection, tracking and recognition tools developed in WP4 and WP5. Additional metadata have been generated in a fully automatic manner based on the monitoring of the 24'' clock signal. Each time the clock is stopped or reset, automatic action recognition tools attempt to understand the cause of the clock reset. This information is exploited to personalize the summary according to the actions and players preferred by the user.

Given those inputs, the purpose of the trial is:
- to build an attractive showcase, which displays the achievements of the project,
- to demonstrate the integrated autonomous production and summarization framework envisioned by APIDIS,
- to validate the personalization capabilities of the proposed framework, i.e. to show how the produced summary can be adapted to user requirements.
- to evaluate the robustness of the summarization tool to the potential imperfections of the scene interpretation stages.

Methodology:

The production and summarization stages follow the approaches described in deliverable 6.2 and 7.1. Compared to the functionalities presented in deliverable 7.1, we plan to:
- Extend the summarization framework to the basket-ball context. It implies the implementation of a segmentation strategy, together with the definition of local stories associated to each segment. For this purpose, we plan to rely on game finality (scoring) to define a segment, and to exploit intermediate 24'' resets to construct local stories. The summarization also requires the design of appropriate mathematical expressions to compute the benefit of each local story as a function of the user preferences. Those expressions should be a direct extension of the benefit functions envisioned for football summarization.
- Extend and improve the production tools based on the assessment and user consultations associated to proof-of concept trials. As explained in Deliverable 7.1, close views should be considered in some particular circumstances, e.g.

before a free throw, after a successful action, or even after a foul. This requires both accurate understanding of the actions and recognition of players and referees. Top views can also be added sporadically to illustrate how the ball and players move on one side of the field, before actual scoring attempt.

Trial scenario:

We envision running two distinct trials, during which users will have the opportunity to play with the system, i.e. to run it several times with distinct set of input preferences. For both trials, we assume that all the information describing the position of the objects of interest, as well as the metadata defining the actions of the game have been computed off-line.

The first trial scenario considers the conditions of a real exploitation environment. Hence, it will consider online access to content, through a web interface. In this scenario, the content is composed of a set of video clips that have been produced off-line, based on fully automatic scene analysis and video production procedures. The clips have also been pre-encoded off-line, using any suitable video codec. The local stories associated to a segment of the game are defined exclusively based on the concatenation of those pre-computed video clips. Given the centres of interest (in terms of actions or player) and the narrative preferences of the user, the trial will consists in identifying the clips to include in the summary, so as to stream them to the viewer.

The second trial scenario goes one step beyond the first one, in the sense that it *produces* the clips associated to a segment *in response to* the specific *preferences* expressed by the user. As an example, the production process might focus on a player who is of high interest for the user. The scenario will only be demonstrated in a lab environment since the video production and compression algorithms are not expected to run in real time within the course of the project.

Assessment:

Objective measurements:

From an objective point of view, we mainly plan to provide a report discussing the computational complexity of the algorithms involved in the trial. The report will consider the algorithms needed for production and summarization, but also the ones required to analyze and interpret the scene.

Subjective evaluation:

The core of the trial assessment will be based on subjective evaluation of the summaries provided by the system. Assessment will rely on questionnaire-based interviews, and on open discussions with potential end-users (players, coaches, local TVs, etc.).

6 Conclusions

In this document, we have presented the objectives, and the envisioned scenarios for the end of project trials scheduled between months 30 to 33. A particular attention has been devoted to the definition of relevant and realistic assessment criteria. In addition, the document also explains how APIDIS partners should interact and cooperate to correctly prepare the trials.