Image and video content analysis

Since its creation in 2001, the Image Processing Department of Multitel has been working in the field of image and video content analysis.

Such analysis is needed to represent objects/assets in a visual scene. The architecture of the vision part is generally divided into three main levels of computation: the image level (acquisition of images and filtering, background evaluation and segmentation), the blob level (tracking level of the mobile regions), and the event level (tracking analysis and understanding).

Multitel’s researchers are specialized in two different approaches they are used to compare and combine.

On the one hand, the traditional approach for segmenting mobile objects/assets, which consists in the evaluation of the scene background and in the extraction of the scene foreground.

The description of regions acts as an interface between the extraction of the foreground and the tracking. The description process consists in the translation of video data into a symbolic representation i.e. descriptors. The objective is to reduce the amount of information to what is necessary and sufficient for the tracking module. The description process calculates the different characteristics observed from the image and the segmentation results.

The part concerning the object tracking is made up of four steps: the prediction of positions, the calculation of the cost matrix, the matching decision and the tracks update.
Moreover, a tracking method makes use of a top-down approach. It consists in the estimation of the position and characteristics of the target (object) in the current image, knowing the position and characteristics of that same object in the previous image. This method is by definition a recursive method. The representative techniques of the top-down approach for asset tracking are the algorithm of Lucas-Kanade, the algorithm of Mean-Shift and the particle filtering. Particle filters (PF) are rather complex techniques for the estimation of patterns. They aim at assessing a sequence of hidden parameters, exclusively based on observations. The asset state is described by the vector $x$, while vector $z$ gathers all observations. Multitel has actually developed an improved version of the particle filter.

In the field of image processing, Multitel publishes about 10 scientific and technical articles each year.

**CANDELA – Content Analysis and Networked Delivery Architectures**

The CANDELA project aims at unifying European knowledge in terms of visual content analysis in a wide range of applications: Video surveillance, road traffic, audiovisual and medical fields. 
www.extra.research.philips.com/euprojects/candela/

**Dynamic scene understanding and behaviour recognition**

The processing of video sequences originating from video surveillance systems allows to analyse scenes in a dynamic way and to understand them. Indeed, you can make a high-level understanding of events in the scene from the detection and tracking of mobile objects in the image. You can also restore persons’ trajectories and actions. Applications in that field can be the fight against insecurity, marketing (study on consumers behaviour), as well as the monitoring of children or ageing people.

Scenarios can be analyzed by means of intelligent agents mechanisms that automatically detect patterns in the tracking results. Those patterns correspond to elementary events that give more complex scenarios when they are put together.
Interactive classification of multimedia data

Nowadays, we have to deal with so large amounts of audiovisual data that tools for the management, structuring, interactive retrieving and browsing of the content have become essential.

Multitel can meet these needs thanks to their researches in terms of interactive classification of images and video scenes.

The following challenges will be described:

How can you retrieve video scenes or specific images from a visual database?

This issue of interactive retrieval is particularly important today. For instance, you have to compare stored images with images the user gives as an example. This occurs without having to use complex metadata formats or annotations. In another addressed issue, the user interacts in the learning process without getting a too heavy workload.

Advanced techniques of artificial intelligence are studied in order to meet such challenges:

- Support Vector Machines,
- Decision Trees,
- Active Learning,
- Multiple Instances Learning,
- Relevance Feedback.

These techniques are based on an interaction between the user and the machine so that the retrieving occurs intuitively and is adapted to each person and his/her resources.

Our scientific approach allows to adapt those techniques to other issues that can appear as quite different, such as indexing of sequences or structuring of interactive video servers.

Multitel owns a video data retrieving system that can also be used for their research work in that field.

Example of project:

IRMA – Interface de Recherche Multimodale en contenu Audio-visuel

The objective of IRMA is the design and development of an innovative modular interface for the retrieving and the multimodal customized browsing in indexed audiovisual databases.

How to adapt a video flow following its semantic content?

The real-time analysis of a video flow allows to detect elements whose relevance has to be taken into account at the moment of the compression, transmission, adaptation of the flow to the user’s material resources and to its storage or indexing.

For instance, the background of a visual scene can be considered as less important and its quality can be reduced in order to guarantee a minimum output when transmitted on a wireless network.

Moreover, in order to protect private life, human faces can be automatically hidden without any human intervention.

The techniques that are used in this field are:

- Asset tracking and segmentation,
- Detection of human faces,
- Interactive classification,
Numerous applications of that concept exist. For instance:

- Private life protection,
- Automatic censure,
- Content adaptation to the restricted resources of a user or a channel,
- Indexing and retrieving of particular regions in sequences,
- Monitoring in marketing, e.g., retrieve logo or inscriptions,
- ...

Example of project:

**WCAM - Wireless Cameras and Audio-Visual Seamless Networking**

*Design and development of a system for the transmission of secured audiovisual content on a wireless network.*

[www.ist-wcam.org](http://www.ist-wcam.org)

**Performance Evaluation of Video Systems**

Vision systems are part of the new and complex technology. For that reason, few people are able to assess them properly. Therefore, we are dealing with that issue. Besides the functional testing of products, there are other reasons for evaluating vision systems: bring our advances in developments to light and compare our research work with others, and of course, ensure the respect of possible legal aspects. Multitel participates in the advance in the evaluation field with the publication of new metrics in international conferences (ex: IEEE International Workshop on Performance Evaluation of Tracking and Surveillance).

In the context of performance evaluation, we intend to know if a vision system is able to find back the tracks of persons appearing on the image. Therefore, we have to precisely know the real tracks. To reach that goal, the making of video sequences by means of image synthesis sometimes turns out to be useful.

**TRICTRAC - Tricks for Tracking**

The goal of this project is the development of algorithms for the tracking of objects in real time in one or more video streams. It will bring innovative solutions to a series of problems for which there is no satisfactory solution at the present time, in particular in sports (remote transmission and coaching) and in remote monitoring.


![Figure IX: Image synthesis for the evaluation of a project for the tracking of football players.](image)