Real-time Requirements of Media Control Applications

19th Euromicro Conference on Real-Time System,
Pisa, Italy
5th June 2007

Francisco Gómez Molinero
R&D Director
VISUAL TOOLS S.A.
fgomez@visual-tools.com

Contents

Contents

Multimedia Information in Embedded Systems

Multimedia information is present in embedded control systems for two purposes:

- To improve the interaction with human beings (operators)
  - By achieving "telepresence": audio/video streaming
  - By enabling "investigation of the past": audio/video storage (foresincs)
- To develop control strategies closer to the behaviour of *"a human performing control tasks"*
  - Feasible by advances in audio and image analysis algorithms
  - Enabling the implementation of more sophisticated control systems

Control Systems vs. Human Control (1/2): **Real-Time Control System Features**

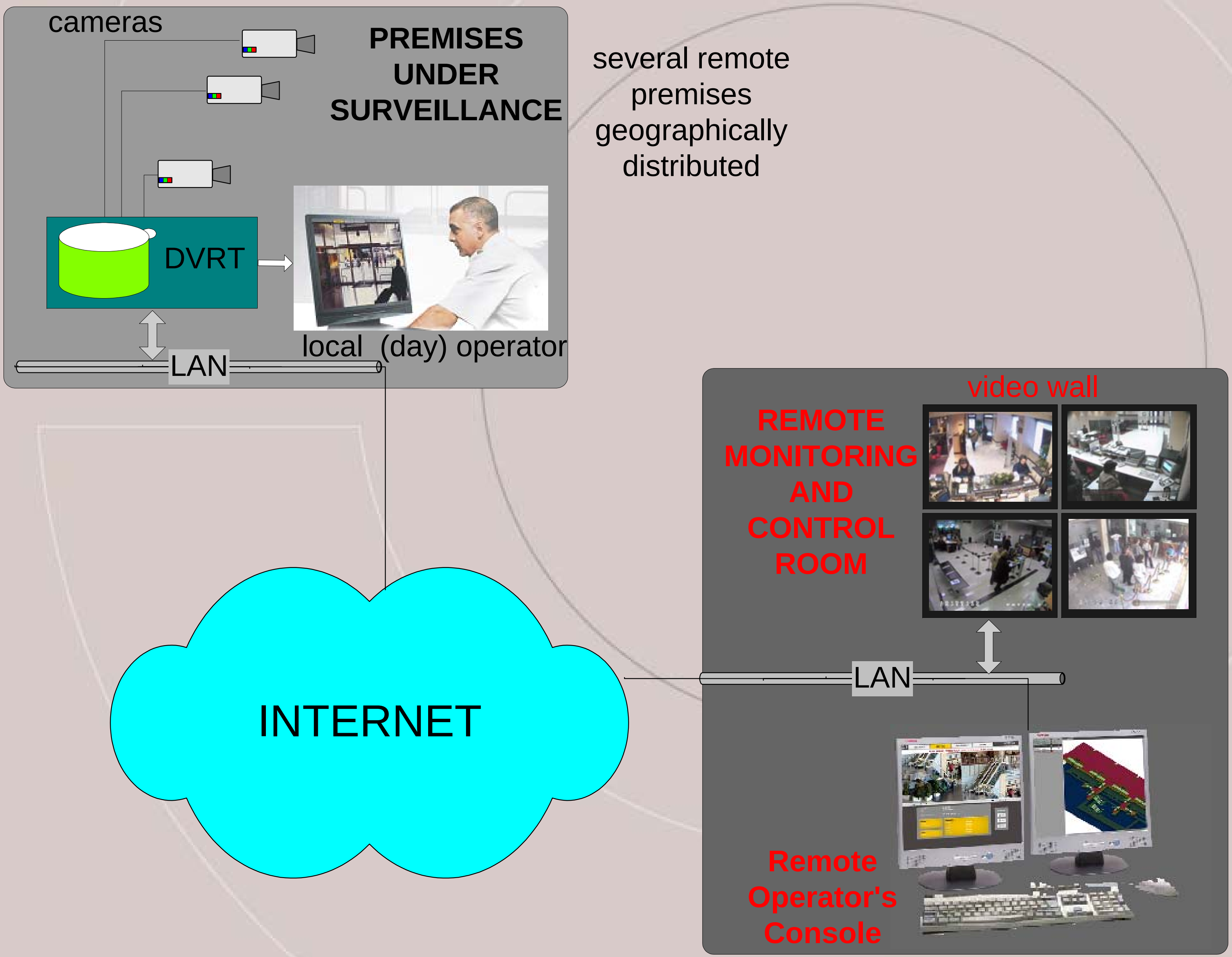
- **Sensors:**
  - limited amount and usually “mono-modal” detection (i.e. Based on the same detection principle)
  - activated by a reduced number of causes
  - relatively simple but robust for the intended detection
  - very little semantics (e.g. detect a contact opened without regard to the ultimate cause that opened it) which is added by the human interpretation based on the environment conditions
- **Embedded Control algorithms:**
  - Fast
  - Limited number of decisions
  - Usually assuming healthy sensors (I.e. limited checking of sensors health, no feedback to sensor devices to improve detection)

Control Systems vs. Human Control (2/2): **The Human Being in Control Tasks**

- **Sensors:**
  - Less selective: sensors activated by a large variety of input stimuli as compared to artificial sensors
  - Multimodal: eye (images), ear (sound), touch (pressure, temperature), smell (gas,…), taste
- **Control Algorithms:**
  - More sophisticated
  - Ability to adapt to unexpected changes and learn
  - Sensor fusion information increases robustness
  - Usually structured in two levels:
    - Low processing power background algorithm: “hardwired” unconcious ability to recognize the normal situation
    - Intensive processing triggered by detection of “abnormality”

Contents

Typical System Architecture



Main Functions (1/4): **Video Streaming**

Ability to transmit audio/video information

- Associated processing tasks:
  - A/D conversion
  - Compression
  - Bitrate adaptation
- Advanced features:
  - Ability to efficiently provide multiple streams (at different bitrates)
- Main resources involved:
  - CPU (compression)
  - Network bandwidth
- Measurement of quality:
  - Good image quality
  - Short latency
  - Smooth play

Video Streaming Sample Screen

Main Functions (2/4): **Remote Control of Devices**

Pan-tilt-zoom cameras and other telecontrol devices

- Main functions:
  - Go-to-preset
  - Move continuously (up, down, left, right, zoom-in, zoom-out)
  - Start/stop operation
- Main resources involved:
  - Network reaction time
- Measurement of quality:
  - Short reaction time from command to motion (e.g. through ADSL)
  - Short latency in the video transmitted

Main Functions (3/4):
Video recording and retrieval

- **Associated processing tasks:**
  - Compression
  - Disk access
  - Real-time image analysis for annotation (optional)
- **Main functions:**
  - Event recording activated by external devices (e.g. Door contact) or image analysis (e.g. Video motion detection) or continuous recording at different bitrates
  - Real-time annotation with external device information or with image analysis results
- **Main resources involved:**
  - CPU
  - Bus and disk bandwidth
  - Network bandwidth (in remote video retrieval)
- **Measurement of quality:**
  - Recording 24/7 (i.e. No black-outs) and regular frame rate
  - Smooth replay and efficient filtering based on annotation (for retrieval)

Video retrieval presentation sample screen

Main Functions (4/4): (Local) Video Display

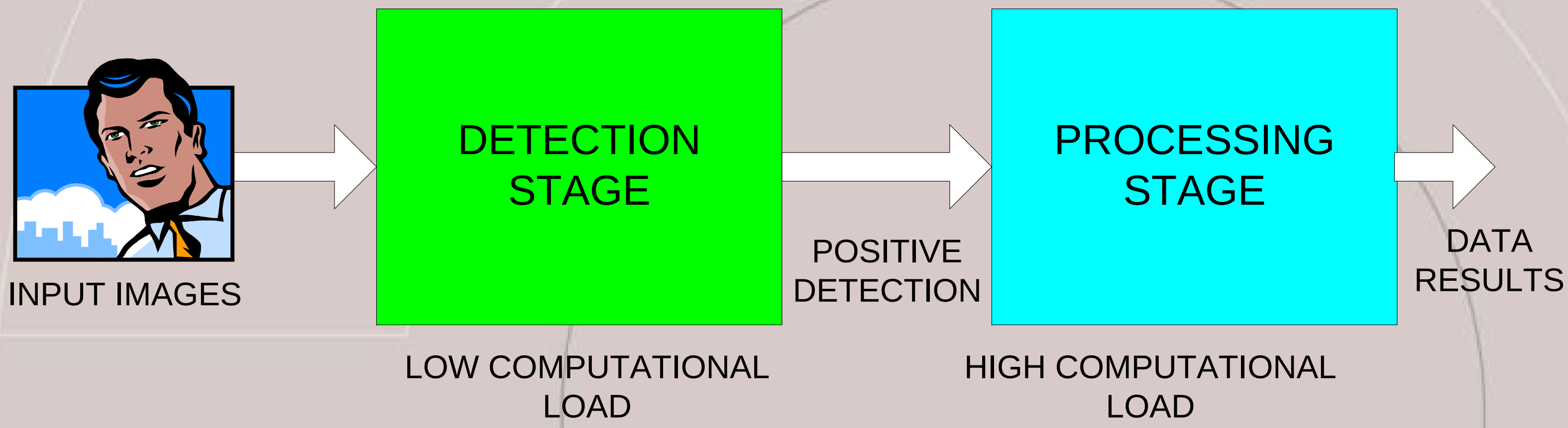
- **Associated processing tasks:**
  - Decompression (for recorded video)
  - Disk access
- **Main functions:**
  - Multiple camera display (Quads)
- **Main resources involved:**
  - CPU (for recorded video)
  - Bus and display memory bandwidth
- **Measurement of quality:**
  - Smooth replay
  - Balanced replay of multiple cameras

Contents

Main Features

- Multimedia analysis is a central part of the system
- Main functions:
  - Video acquisition
  - Video image analysis and data extraction
  - Delivery of results
- Very often existing systems provide both types of functions combined together, i.e. the multimedia representation functions described before combined with image analysis

Image Analysis Algorithm Architecture (1/2)



- Very often image analysis is divided in two stages:
  - A first detection stage with low computational load
  - A second processing stage that is only executed if detection is positive
- Types of processing:
  - Processing stage with constant CPU time
  - Processing stage with CPU time proportional to objects detected

Image Analysis Algorithm Architecture (2/2)

- **Constant CPU time image analysis**:
  - Video motion detection
  - People counting
- **CPU time proportional to number of objects detected**:
  - Object tracking
  - Face detection

Typical Image Analysis Techniques (1/7)

- **Video motion detection**:
  - very simple processing giving the motion level of every cell in a rectangular grid of the image
- **Object tracking**:
  - *Detection stage*: moving objects or objects with special features (e.g. colour histogram)
  - *Processing stage*: tracking using different techniques (object overlapping, Kalman filtering, particle filters,...)
  - *Result*: Size, Position, and Speed of the objects appearing on every image of a sequence

Typical Image Analysis Techniques (2/7): **Object Tracking**

Typical Image Analysis Techniques (3/7)

Human Behaviour-related image analysis:

- **Face detection**:
  - *Detection stage*: gross detection of faces present
  - *Processing stage*: finer processing to detect with higher precision and correlate with previous faces
  - *Result*: coordinates of human faces in the scene
- **Face recognition**:
  - *Detection stage*: face detection
  - *Processing stage*: comparison to recognize against suspect's data base
- **Other human behaviour image analysis:**
  - Gait analysis (analysis of human locomotion)
  - Emotion analysis (modelling motion of face muscles)

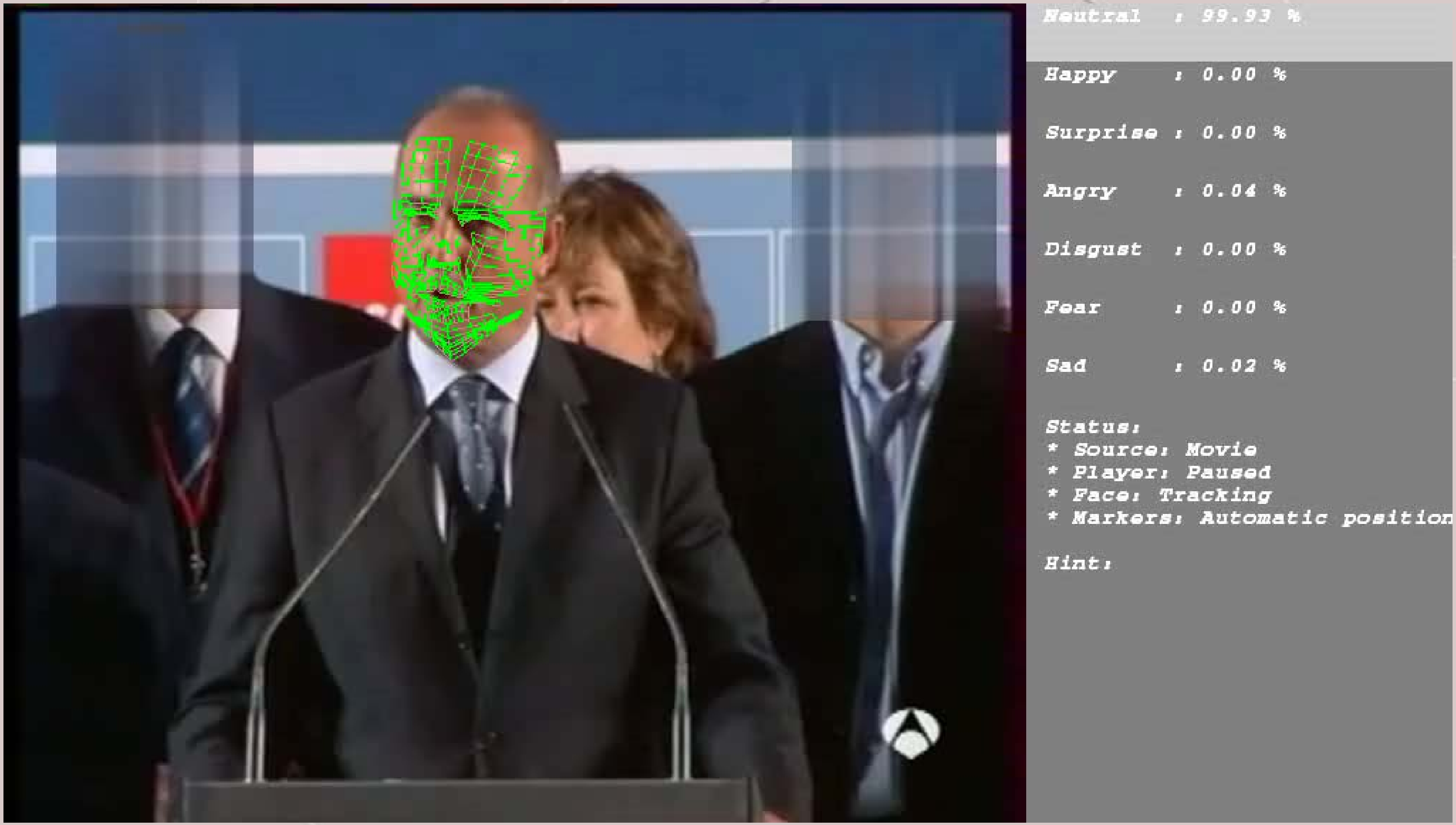
Typical Image Analysis Techniques (4/7):

Face Detection

fgomez@visual-tools.com

Courtesy of Prof. A. Albiol Technical University of Valencia

Typical Image Analysis Techniques (5/7): **Emotion Analysis**



Courtesy of Prof. N. Sebe, University of Amsterdam, work in progress in IST MIAUCE project

Typical Image Analysis Techniques (6/7)

Other image analysis techniques:

- License Plate Recognition
- People Counting (overhead cameras)
- Statistical analysis: generic characterization of “normal situation”

Typical Image Analysis Techniques (7/7): **License Plate Recognition**

Courtesy of Prof. A. Albiol Technical University of Valencia, IST-VISOR BASE Project

Image Analysis Applications (1/5): **Real-Time Requirements**

- The final application of the previous techniques is the crucial factor to specify the real-time requirements, i.e. The same technique applied to different purposes exhibits different timing requirements
- Some very useful products can be developed with image analysis techniques even if not full accuracy of the algorithm is possible (i.e. Some false positives or false negatives exist)
- Typical use of image analysis techniques:
  - *Statistical data extraction*: e.g. people counting
  - *Video annotation*: used to give “some meaning” to an otherwise “flat video sequence” to facilitate further retrieval
  - *Safety critical operations*: Meeting timing constraints is crucial for the correct operation, e.g. using object tracking to detect congestion in escalator

Image Analysis Applications (2/5): **Non Real-Time Applications**

- *Statistical data extraction*: Applying image analysis techniques to extract end-user data that are not required to be provided in real-time (a relatively long delay, e.g. minutes, hours, is acceptable)
- Examples:
  - Counting people to estimate the distribution of visitors in a shop for a full day
  - Recognizing license plates to detect non-authorized access to city areas
- Image analysis can be applied in the background over stored video when processing time is available
- However, in some circumstances the ability to decrease the delay in extracting the data can lead to cheaper products or it can widen the range of applications covered by the system

Image Analysis Applications (3/5): (Soft) Real-Time Applications

- *Video Annotation*: use of image analysis to "give some meaning" to an otherwise very large storage of "flat video sequences"
- Used to improve the video retrieval process, e.g.
  - Video motion detection: retrieve only video sequences with motion in a certain region
  - Face detection: retrieve only images with faces
- Very useful to investigate a past event (e.g. When searching for a suspect, retrieve only the images with faces without having to replay a full day of video)
- This type of application can usually tolerate missing some timing constraints, e.g. If the system skips one second of video motion detection the video annotation is still quite useful, since "the user-in-the-loop" can "compensate" the error.
- The concept of "quality of service" is very appropriate in this case

Image Analysis Applications (4/5): (Hard) Real-Time Applications

- *Safety critical operations*: Meeting the timing constraint is crucial for the correct operation of the system
- Examples:
  - using object tracking to detect congestion in escalator
  - counting people entering a building to avoid passing the legal capacity or to ensure evacuation in case of fire
- Nowadays, this type of applications are mainly limited by the robustness of the image analysis techniques currently available, but…
- … they will increase rapidly when image analysis matures, in particular, with the use of multimodal detection as a means to improve robustness
- Hard real-time scheduling will be clearly useful for these applications

Image Analysis Applications (5/5):
Escalator congestion example (1)

Image Analysis Applications (5/5):
Escalator congestion example (2)

On-going work in National R&D Project Hesperia

Contents

Resources-Related Requirements (1/2)

- **CPU:**
  - Efficient use to optimize cost (volume products)
  - To ensure critical functions are executed first (image capture, critical image analysis, video recording)
- **System bus:**
  - To guarantee image capture rate
  - To control interference between high bandwidth devices (e.g. image capture, hard disk)

Resources-Related Requirements (2/2)

- **Memory:**
  - Optimization of required amount to decrease cost (memory is one of the major contributors to the “bill of materials”)
  - Worst-case memory usage identification
- **Network:**
  - Latency for telecontrol operations, e.g. moving cameras, starting/stopping control devices remotely
  - More critical for Internet communication

Other Timing Requirements

- Designing systems with ability to synchronize video, audio and data coming from different sources opens a wide range of potential applications using the time stamp as a means to relate multimedia information
- Some examples:
  - *Traffic*: synchronized display of video and data traffic to audit traffic installations
  - *Retail*: synchronized display of video and transaction data in the point-of-sales (to detect a wide variety of conflicting situations)
  - *Electrical substations*: synchronized display of video and data acquisition sensors to identify causes of electrical failure (it would need fine-grain synchronization to the ms level)

Contents

Sustained Innovation (Short/Medium-Term)

Step-by-step increase of efficiency and decrease of cost

- *Video streaming*: ability to provide simultaneous streams from the same source at different bitrates
- *Video storage*:
  - Real-time multimedia annotation and indexing
  - More precise (global) time synchronization (using time as global correlation factor)
- *Local video display*: efficient full frame rate display of multiple cameras
- *Arquitectural issues*: progress on standard interfaces adoption or new interfaces defined in certain areas (e.g. to isolate multimedia analysis from further stages)

Disruptive Innovation (Long-Term)

Long-term goal: implementing systems based on media analysis capable to interpret situations as humans

- Promising research lines:
  - Implementing multimodal detection
  - Improving semantics of detection process (using semantic Web technologies)
  - Implementing systems capable to learn the environment, and take their own decisions
  - Developing new computing paradigms to enable the implementation of computing systems more similar to the human brain

Conclusions

The fast cost reduction of multimedia devices and the increase in available processing resources (CPU, memory, bus and network bandwidth) in embedded systems is motivating a fast growing presence of multimedia information

- As a result:
  - Supervised Multimedia Control systems will require less emphasis in “strictly meeting real-time constraints” but the concept of “quality of service” will be more important
  - Multimedia Embedded Systems will evolve to combine critical tasks based on media analysis with less critical functions oriented to the Human-Computer Interaction.
  - Eventually more autonomous control systems (human-like) will be feasible
- Research topics for the real-time community should include new architectures supporting new computing paradigms to implement more human-like computer systems

Thank you!

Paco Gomez
fgomez@visual-tools.com