Technologies for multimedia and video surveillance convergence

D. Nicholson[a] and J. Meessen [b],

[a] Thales Communications, Colombes, France
[b] Multitel asbl, Mons, Belgium

ABSTRACT

In this paper, we present an integrated system for video surveillance developed within the European IST WCAM project, using only standard multimedia and networking tools. The advantages of such a system, while allowing cost reduction and interoperability, is to benefit from the fast technological evolution of the video encoding and distribution tools.

1. INTRODUCTION

In the past decade, we have seen the growth of Internet and the growth of the mobile communications. This success is due to the hardware component evolution allowing the deployment of those technologies at a reasonable cost for the user. The focus has been on physical layer and, thanks to hardware component evolution, modems, WLAN boards or mobile phones are available at always more reasonable costs. In parallel, audio and video coders have leveraged the available CPU power to offer better compression and quality allowing the audio and video digital market success.

In CCTV world exists the same evolution but as it concerns the security, requirements are more strict and so, new technologies are coming slowly. A lot of systems are still full analogue systems, but digital storage and video over IP is taking more and more place. Trend is full digital CCTV system.

After a short overview of video surveillance systems, the work and concepts developed in the European IST project name WCAM (Wireless Cameras and Audio-visual seamless networking) are presented. This project focuses on the technology convergence between video surveillance and multimedia distribution over the Internet.

2. VIDEO SURVEILLANCE SYSTEMS SHORT OVERVIEW

Security of individuals and their property remains a cause of major public concern. The installation of surveillance cameras in public or commercial premises and streets, often carried out in partnership with local community schemes, local authorities, police forces and local business, has assisted in detecting crimes such as personal attacks, theft and drug dealing. They also provide benefits in related areas such as public safety, alarm verification and licence plate recognition.

Surveillance systems are currently employed in a wide range of crime and safety-related applications. These include:

Personnel screening, luggage, freight and vehicle scanning, explosives detection,
Building Security, access control, intruder identification by personal signature,
Public space surveillance for crime detection,
Person detection/ biometrics, visual features, gait, crowd and behavioral analysis,
Vehicle security, traffic analysis, identification and tracking,
Anti-terrorism.

First generation of Electronic surveillance techniques (1960 – 1980) employed basic analogue video cameras and transmission techniques, connecting to a control centre for viewing on an array of monitors. Problems arose because of the attention span of operators and the frequent occurrence of missed events, especially in large systems. From around 1980, improvements and cost reductions in sensor, computing and communications technology led to the emergence of more sophisticated processing for event detection, alarm generation, fault detection, illumination compensation and tracking of objects in scenes.

Such analog video surveillance systems usually include cameras, one or several switching matrix and specific keyboards for controlling the cameras. Video switching in analogue CCTV systems is made using an analogue video matrix, which is centralized video equipment. It means that all video connections must be gathered together at the place where the video matrix device is. It can then involve optic or radio transmission for long distances and, in any case, the deployment of cables around the whole site.

Larger systems need more sophisticated operator interfaces. These man machine interfaces usually use Windows graphics and show the location of the cameras on the site map. In an environment with multiple operating stations, it is worthwhile to install video server software to ensure the linkage among all operator consoles, on the one hand, and all field equipment (cameras, matrixes, digital recorders, video based motion detectors, etc.) on the other. Finally, in the case of very large-scale sites that must manage several matrixes, it also becomes easy to design a "partitioning" into independent surveillance areas. Inter matrix links then permit analogue video flows to move from one area to another, so that any camera can be picked up on any monitor screen on the site.

In the following figure you will find an example of such a system:

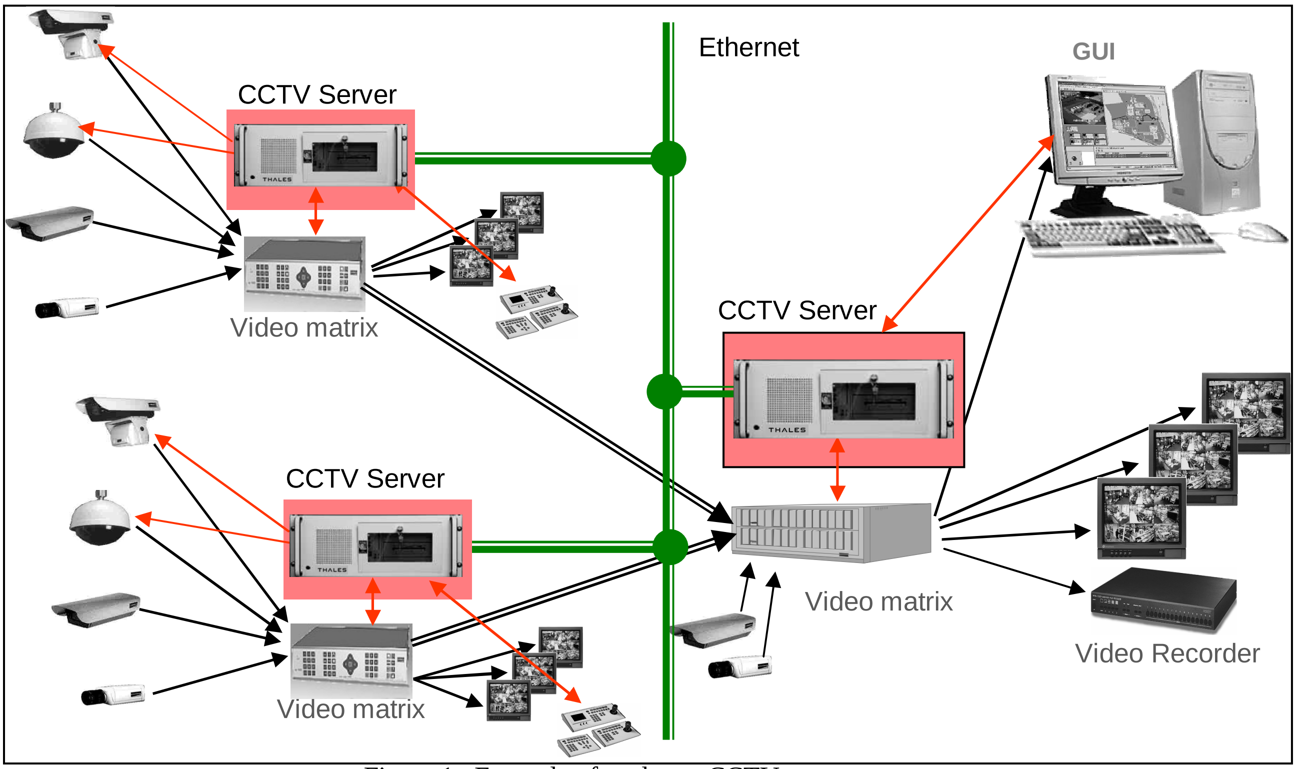

Figure 1 : Example of analogue CCTV system

Digital transmission techniques also emerged during using image and video coding standards such as JPEG and H.261 which facilitated basic levels of digital storage and transmission using ISDN lines and solid state and tape storage. These techniques allowed increased amounts of data to be assimilated due to attention focussing techniques. From around 2000, a third generation of system began to emerge, exploiting new coding standards, broadband communications, open protocols, and offering enhanced processing and intelligent sensorsi. Using internet and wireless access methods, monitoring is now possible from remote sites. It is also possible to "fuse" information from different types of sensor in order to offer improved performance. Commercially available CCTV systems are clearly becoming increasingly sophisticated and methods are progressively migrating from the research laboratory into deployed systems.

First digital devices in CCTV systems have been digital video recorders. Those devices were able to store video sequences on hard disks and to replay them on PC Windows GUIs. Then video encoding and compression have made such progresses that it became possible to distribute the live video through Ethernet network with reasonable bandwidth and good quality. Full digital CCTV systems were then a reality. That is the reason why there are two approaches in terms of products on the market: the first one is based on solutions of recording while the second is based on solutions of transmission. Today, both types of solutions provide similar functionalities (i.e. mainly live video display and recording), but the devices packaging are different. Approach based on recording is rather centralized compared to the one based on transmission. Moreover integrated network cameras over IP become available today.

Video switching in network CCTV systems is made by the network itself. No need here for analogue video switching matrix. That means all composite video signals delivered by analogue cameras are encoded and compressed in order to be distributed over an Ethernet network. The display of live video is made either on PC based GUI applications with software video decoding or on analogue monitors after the video had been decompressed & decoded by a specific decoder.

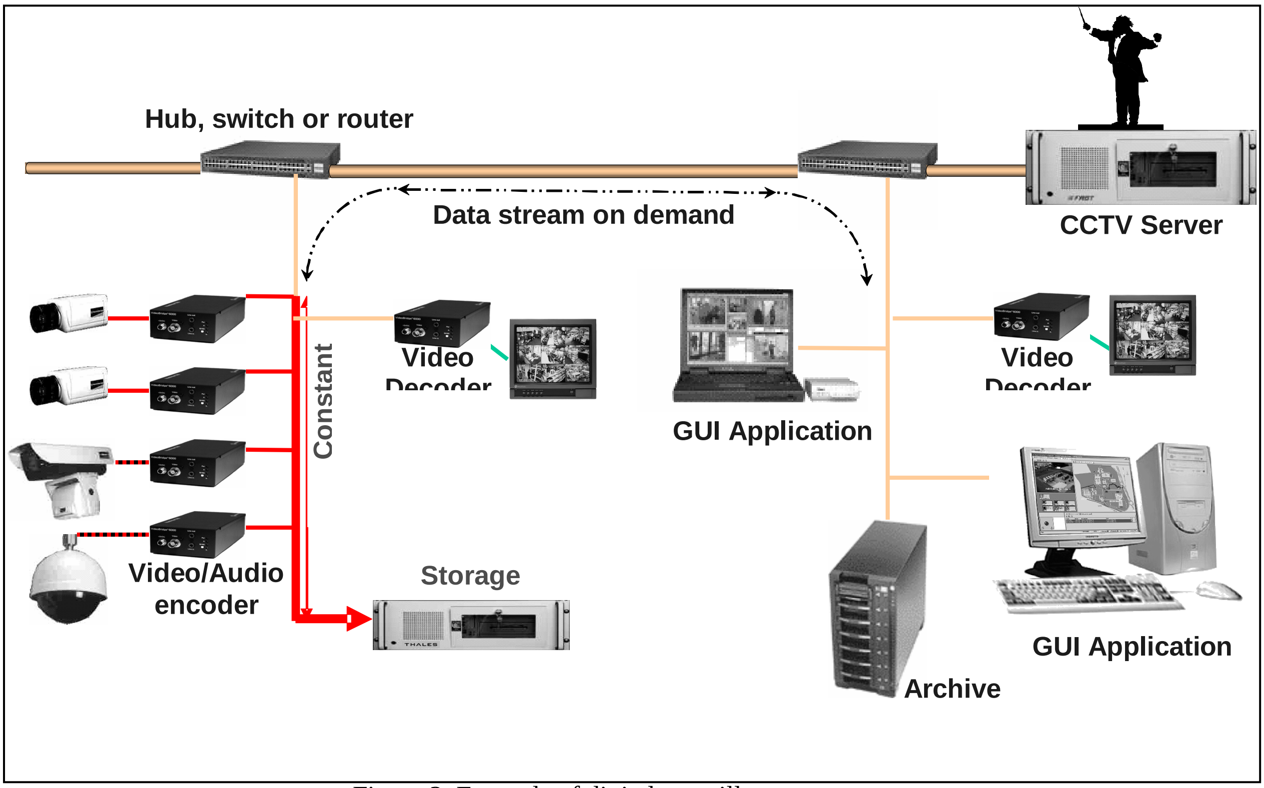

Figure 2: Example of digital surveillance system

Digital connectivity, coding and storage is rapidly overtaking more established analogue systems in the professional and consumer markets but has made slower penetration in CCTV and surveillance markets. The benefits of going to digital are the following:

Digital cameras and data networks are well established in professional and domestic applications. Increased investment and take-up will drive down costs, Digital information can be copied without distortion. Also tape wear is generally less of an issue with digital storage due to the robustness of the encoding method coupled with advanced error protection techniques,

The potential for adaptation to channel variations in terms of bandwidth (compression) and error performance,

Post processing of images and video is possible to facilitate alarm generation and to improve recognition and detection,

Use of standards – the importance of compatibility for multi-sourcing and media interchange, in particular for evidential purposes. On the negative side, the issues that have prevented rapid take-up in the CCTV market have included a lack of standards, access to ready deployed digital communications infrastructure and historical storage and capture costs. In addition, the issue of compatibility between digital storage technologies (e.g. download from integrated hard disk systems) can be a serious issue for police when acquiring evidence.

In digital CCTV, the protection of the data is a very significant point. The main data is the video. It can be live video or stored video. In all cases, the system shall propose a way of protect:

It must be impossible to decode the stream. This item is still more significant regarding wireless technology.

It must be impossible to modify the content of a video stream and the system must ensure that video encoded at the camera side will be the same that the one which will be displayed or stored. Indeed, content in digital form can easily be accessed, manipulated, copied and distributed at negligible cost. In the case of surveillance applications, security of the content will ensure the integrity of the system and the privacy of the users.

Actually, no solution is really available about the content protection. Offers are proprietary, which means that proprietary codec should be used to decode the video stream. But this is not strong protection and has nothing to do with encryption, digital signature or watermarking, only techniques really providing protection.

The challenge of third generation systems is to provide high quality data acquisition, efficient and robust coding, high bandwidth and secure transmission, efficient storage with ease of access, and sophisticated processing to allow flexibility of control, adaptation and enhanced event analysis. In particular these systems should offer:

Reduction of reaction time for alert generation, and information assisted decision

Easy deployment of sensors without large infrastructures, taking advantage of wireless networks technologies, with security features for source authentication and content access protection.

Adaptation to the network conditions, which imply the possibilities of scalability for QoS management
Extraction of metadata from images and video for reporting, indexing and search purposes.

The latest apparition in CCTV systems is the ability to have Video Analysis. The objective of video analysis for surveillance is to extract automatically a first scene interpretation without the need of a human supervision. Today, systems appear enabling real-time events detection like suspect behavior detection, crowd detection, bad way moving, speed detection, object removal, people or car counting, smoke detection etc.

Video analysis also can also provide a rough indexing of the video events or content, which can be stored for further browsing or deeper analysis and indexing.

3. WHY A NEED FOR VIDEO SURVEILLANCE AND MULTIMEDIA CONVERGENCE

The advances in digital video processing and the growing necessity for networked transmission solutions both in the multimedia distribution and surveillance community creates an important market interest for combining these two domains over a common technology. Video surveillance systems can then take advantage of standard solutions, enabling systems interoperability and cost reduction. When existing in the same place, system components can be even shared in a combined application, where the same network infrastructure could be used for conveying video surveillance data and multimedia.

With the recent progresses in wireless network communication, e.g. WLAN, it is possible now to transmit video content from the camera to the user, this one being on a fixed location in a control room or somewhere on the field, accessing video through a Personal Digital Assistant for example. It is then possible to envisage to distribute advertisement and other multimedia data over the surveillance network to be installed in public places and public transports. Moreover, using one common infrastructure for both surveillance applications and multimedia distribution is an optimal way of reducing costs. Furthermore, advertisement or pay services are a good way to fund the video surveillance system installation and maintenance.

It is clear that by using multimedia system tools in a video surveillance application, security aspects have to be covered seriously. Wireless transmission of surveillance data requires strong security for protecting both the content itself and the delivery of the content.

When running both multimedia distribution and video surveillance over a common wireless network, an important traffic management issue appears. Surveillance and multimedia both require Quality Of Service, but their traffic has indeed some differences. Though surveillance can easily be dimensioned when the number of cameras and terminals are determined, this is not the case for multimedia distribution where many people can pay for accessing data at the same time or not. High priority must also be assigned to the video surveillance data in the case of detection of event.

Both video surveillance and multimedia content distribution are concerned with the problem of video analysis for indexing, which opens the door to content–based applications, like event detection for the surveillance, or video retrieval in the case of multimedia.

4. WCAM SYSTEM DESCRIPTION

The WCAM project is studying, developing and validating a wireless seamless and secured end-to-end networked audio-visual system. It takes into account real time aspects as well as security and scalability. The project will improve state of the art technologies in each of the technological components involved in the system and combine them. The WCAM system will be installed and tested with users of both multimedia distribution and video-surveillance communities.

Recent progresses in flexible bit streams representation of video including Region-of-Interest (ROI) have led to new standardization efforts in the frame of JPEG 2000. WCAM will pursue and enforce these works, integrating concerns like wireless, security, and annotations obtained by high-level analysis. WCAM delivery scenario includes both JPEG 2000 and MPEG-4 AVC platforms.

The technologies and systems developed within WCAM will be validated during two trials:

- the first trial will take place in Annecy (France) during the festival of animated movies in June 2005. It will demonstrate the wireless video streaming of MPEG-4 AVC content over a WLAN network.
- the second trial will be done in a video-surveillance site and will demonstrate the ability to operate a video surveillance platform on a wireless network, to secure the transactions an to ensure the authentication of the live video streams and the stored sequences.

The architecture of WCAM system is depicted in the following figure. The WCAM platform will include the following key modules:

- Camera video encoding incorporating scene analysis, object tracking and adaptive and efficient video coding
- Automatic detection of events through scene analysis
- Scalable Video Coding
- Multimedia storage and distribution
- Wireless networking for local connectivity (WLAN)
- Security, both at the network and content levels
- Metadata driven access to content
- Client platforms from PDA to desktop PCs, but also Set Top Boxes for MPEG-4 AVC rendering

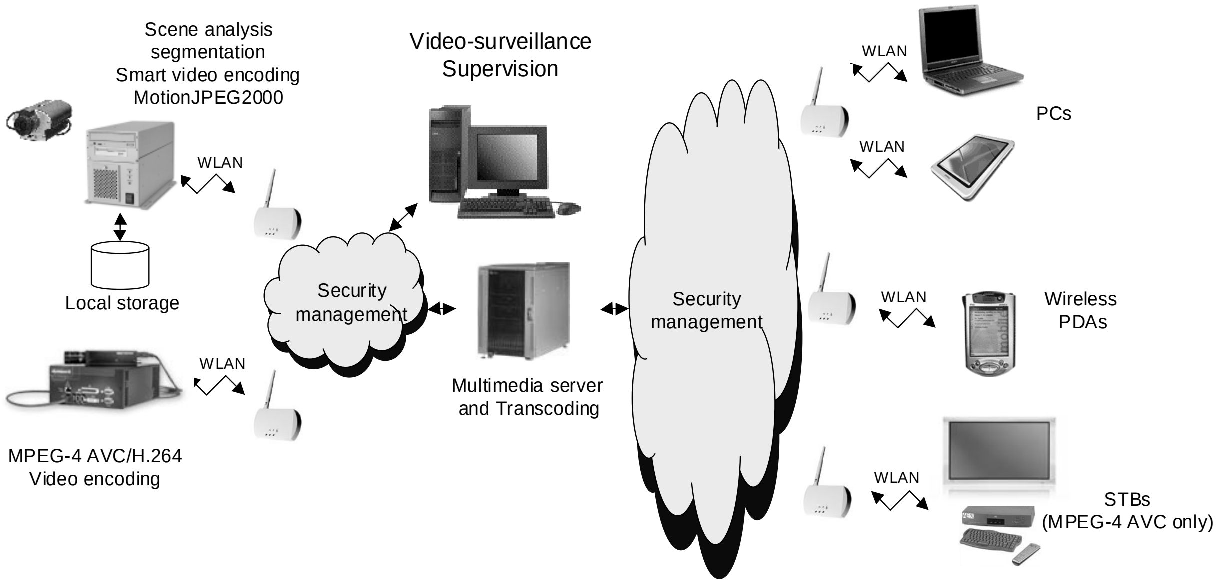

Figure 3: WCAM system architecture

The combination of different technological domains offers new opportunities. As an example, combining video content analysis techniques and video compression offers rich content delivery capacities, as well as the ability of using state of the art image compression tools such as objects and Region Of Interest. WCAM aim is also to provide a convergence between MotionJPEG 2000 and MPEG-4 AVC, the two latest video/image coding standards coming from ISO/IEC/ITU, combining them in the same application framework, wireless connectivity and camera based pre-processing (segmentation and object tracking).

5. VIDEO SCENE ANALYSIS, SEGMENTATION AND INDEXING

The objective of video scene analysis in WCAM is to detect and track Regions Of Interest (ROI) of the video stream in order to generate metadata describing the relevant events for the surveillance application. These metadata are stored using MPEG-7 format [12][13]. These information lead also to easy retrieval of information from the video storage by browsing/searching for events of a certain type.

Scene analysis is also used linked to the video compression module, for selective compression based upon segmented masks. Region Of Interest are coded with a better quality than the background. This allow to guaranty that object of importance appearing in the video scene will have at least a predefined quality. Furthermore, when a scalable video codec is used, like MotionJPEG 2000, and a rate adaptation is required due to reduced bandwidth constraints, quality of the Region Of Interest will remain the same, while only the background will have a reduced quality.

Scene analysis provides also tracking of objects that will be linked to MPEG-4 AVC motion estimation, for reducing motion vectors search range, and fast selection of reference frame. This will help in reducing the algorithmic complexity.

6. VIDEO CODING

In WCAM we are using two types of video codecs, MotionJPEG 2000, MPEG-4 AVC/H.264 that have been recently standardized and we are also experimenting the future MPEG-21 Scalable Video Coding.

JPEG 2000 is the new ISO image compression standard (JPEG 2000 Part 1 Core coding system ISO/IEC 15444-1/ITU-T T.800) [9][10][11]. Part 3 of JPEG 2000, namely MotionJPEG 2000 (MJP2), ISO/IEC 15444-3/ITU-T T.802 [8] is targeting the intra frame coding of video, in direct replacement of MJPEG, while providing every scalability and Region Of Interest features of JPEG 2000.

MJP2 is less efficient than the other video coding schemes like MPEG but nonetheless holds an interest for WCAM:

To replace MJPEG and proprietary frame based video coding schemes used in surveillance applications.

Where low delay, high resolution (size and bit-depth), robustness to transmission errors, frame to frame independence, partial access to content and coding features such as Regions of Interest are demanded.

The JPEG 2000 codec may be also used for capturing high resolution still images.

JPEG 2000 provides Region of Interest coding capabilities, as well as fine grain scalability which is very useful for adapting the video encoding to the network bandwidth capability.

New JPEG 2000 parts can provide additional features to Part 1 that can be applied for Motion JPEG 2000. Part 8 (JPEG 2000 Security: JPSEC) provides hooks to security mechanisms such as encryption, Part 9 (JPEG 2000 Interactive Protocol: JPIP) allows interactive access in a client-server architecture to JPEG 2000 images, and part 11 (Wireless JPEG 2000: JPWL) enhances the basic error resilience features of JPEG 2000 part 1.

After finalizing the H.263 standard in 1995, the ITU-T Video Coding Experts Group (VCEG) started working on two further development areas: a “short-term” effort to add extra features to H.263 (resulting in H.263+ and H.263++) and a “long-term” effort to develop a new standard for low bit-rate visual communications. The long-term effort led to the draft “H.26L” standard. In 2001 the Joint Video Team (JVT) was formed, including experts from both MPEG and VCEG in order to develop H.26L into a full international standard (currently the final committee draft is pending approval). The outcome will be two identical standards: ISO MPEG4 Part 10 of MPEG4 and ITU-T H.264. The “official” title of the new standard is Advanced Video Coding (AVC), however it is also known as H.26L and H.264 [1].

The coding scheme defined by H.264 is very similar to that employed in prior video coding standards. It is a hybrid codec that makes use of translational block-based motion compensation followed by block transformation of the displaced frame difference (DFD), scalar quantization of the transform coefficients with an adjustable step size for bit rate control, zigzag scanning and finally run-length VLC coding of the quantized transform coefficients. However, H.264 modifies and enhances almost all of the above operational blocks thus achieving significant performance improvements [2].

Traditionally, the main objective of video coding has been to optimize video quality at a given bit rate. With the widespread usage of very different networks and type of terminals, the interoperability between different systems and networks is becoming more important. Therefore, video servers should provide a seamless interaction between stored content and delivery. More specifically, the video transmission should efficiently adapt to the varying channel bandwidth and terminal capabilities.

Scalable video coding and transcoding are two technologies to achieve this goal. Both address the same problem with two different methods. Basically, transcoding converts the existing data format in order to meet the current transmission constraints. In contrast, scalable video coding defines the compressed bitstream at the encoding stage independently from the transmission environment.

Scalability is a very important feature, especially in heterogeneous environments. Through its syntax and coding representation, a scalable video coding scheme allows for access to the content at multiple resolutions (spatial scalability), frame rates (temporal scalability), qualities (quality or SNR scalability), and image regions. On the one hand, scalability is critical when terminals have differing capabilities in terms of processing power, memory and display resolution. In this case, the terminal will only decode the relevant part of the bitstream according to its capabilities. On the other hand, scalability is also needed when the available network bandwidth is fluctuating. In this case, scalable coding allows for an efficient use of the network bandwidth by adjusting the video bit rate throughput. Scalable video coding can also be useful to transmit video over error-prone networks. More specifically, the layered nature of a scalable video bit stream enables the efficient use of unequal error protection techniques. Namely, it is straightforward to protect more the most important layers, and less the less important layers.

In coding schemes based on motion compensated Discrete Cosine Transform (DCT), such as MPEG-2, MPEG-4 and H.263, scalability performance is unsatisfactory. Indeed, the functionality is rather limited, and induces a significant drop in coding efficiency along with a large increase in complexity.

JPEG2000 has made a significant step forward by developing a wavelet-based embedded coding which supports a flexible and efficient scalability. As a consequence, MotionJPEG 2000, the extension of JPEG2000 to encode video sequences, is efficiently providing the full scalability features. However, Motion JPEG 2000 is simply encoding each video frame independently (i.e. intra-frame coding). Henceforth, it fails to fully exploit the temporal redundancy in the sequence resulting in lower coding efficiency.

Therefore, the problem of scalable video coding remains. For this reason, a new work item for Scalable Video Coding has been launched in MPEG-21 to address this issue. The goal is to provide with very efficient scalability functionality while achieving coding efficiency close to the best available non-scalable video compression schemes (i.e. H.264/AVC).

7. VIDEO STREAMING

For the streaming of video over the network, WCAM is using RTP, RTCP and RTSP protocols. For streaming MPEG-4 AVC, a feature dedicated to network of this standard is used, named the Network Abstraction Layer (NAL).

MPEG-4 AVC/H.264 compressed video can be represented at the Video Coding Layer (VCL), which corresponds to the data representation present at the output of the video encoder, but also at the Network Abstraction Layer (NAL). The NAL defines the interface between the video encoder/decoder and the transport mechanism to be used for conveying this video. The Network Abstraction Layer (NAL) encoder encapsulates the slice output of the VCL encoder into Network Abstraction Layer Units (NAL units), which are suitable for the transmission over packet networks or the use in packet oriented multiplex environments. The exact transport encapsulation of MPEG-4 AVC/H.264 NAL units is specified by the standardization bodies responsible of the definition of the transport protocol used, such as MPEG-2 system and IP/UDP/RTP. Nevertheless, the NAL interface is defined in MPEG-4 AVC/H.264 standard.

A NAL unit consists of a one- byte header and the payload byte string. The header indicates the type of the NAL unit, consisting of a single NAL unit type octet, followed immediately by the NAL payload. NAL units can be transmitted through multiplexing into MPEG-2 Transport streams or an RTP packetization.

The MPEG-2 Transport Stream (TS) multiplexing is used for H.264 multiplexing and streamed over the network for multimedia application. The encapsulation of NAL units into MPEG-2 Transport Stream is defined in the Amendment 3 of MPEG-2 system (ISO/IEC 13818-1:2000 Amd3). MPEG-2 TS packets (TS header and payload) can be used directly in IP/UDP packets without the need of additional protocol like RTP, which allows saving bandwidth.

RTP is the short name for real-time transport protocol and provides end-to-end network transport functions suitable for applications transmitting real-time data, such as audio, video or simulation data, over multicast or unicast network services. However it does not guarantee quality-of-service for real-time services. The data transport is augmented by a control protocol (RTCP) to allow monitoring of the data delivery in a manner scalable to large multicast networks, and to provide minimal control and identification functionality. RTP and RTCP have been designed to be independent of the underlying transport and network layers [14].

While the basic RTP header is the same for all transported formats, some specific payload headers can be inserted before the data to take into account the properties of MPEG-4 AVC/H.264. The exact packetization of MPEG-4 AVC/H.264 NAL units into RTP is currently under standardization by IETF [15].

This IETF standard extends the NAL specification, new unit types are defined as well as extensions to the semantics of F and NRI data. According to this RTP payload specification NAL units can be directly encapsulated into RTP packets.

Therefore, the length of the NAL unit directly influences the length of the IP packet, which is a crucial factor in a wireless scenario: The longer the IP packets, the higher the probability that at least one segment is lost, and thus that the whole packet is lost. Hence, shorter IP packets are definitely beneficial in lossy wireless environments. In H.264, this can be achieved by using slice structured coding such that each primary NAL unit generated by the video codec contains a single slice representing a sequence of macro-blocks. Slices within one video frame are independently coded and therefore provide spatially distinct resynchronization points within the video data.

The streaming of MotionJPEG 2000 using RTP is currently under standardization by IETF. The following section details some functionalities already present in the current IETF draft document [16].

To provide a payload format that exploits JPEG 2000 video stream, described in the previous section, the following must be taken into consideration:

> Provisions for packet loss. On the Internet, 5% packet loss is common and this value may become 20% or more. On wireless networks, the loss rate can be even more important. To split JPEG 2000 video streams into RTP packets, efficient packetisation of the code stream is required to minimize problems in decoding due to missing code-blocks. If the main header is lost in transmission, the image cannot be decoded. A system to compensate for the loss of the main header is required.
>
> A packetization scheme that maximizes JPEG 2000 functionality
>
> A packetization scheme so that an image can be progressively transmitted and reconstructed progressively by the receiver using JPEG 2000 functionality would be very powerful. It would allow for maximizing performance over various network conditions and variations in computing resources of clients.

The JPEG 2000 codestream is packetized by packetization units. A packetization unit is defined as either a JPEG 2000 main header, a tile-part header, or a jp2-packet as defined in the ISO JPEG 2000 standard, Part 1.

First, the sender divides the JPEG 2000 codestream into packetization units by parsing the codestream or by getting information from the encoder, and packs the packetization units into RTP packets. The sender puts an arbitrary number of packetization units into an RTP packet, and preserves the codestream order. If a packetization unit with headers is larger than the MTU size, it can be fragmented.

8. WIRELESS NETWORK

The chosen Wireless Local Area Network (WLAN) is the IEEE 802.11b/g solution at 2.4GHz. It implements the IEEE 802.11g [20] functionalities and is backwards compatible with IEEE 802.11b [18]. Note that cards with only IEEE802.11g chipsets are not available and are always backwards compatible with IEEE 802.11b. This standard can support the following high PHY bit rates: 1, 2, 5.5 and 11 Mbits/s (IEEE 802.11b rates) and 6, 9, 12, 18, 24, 36, 48 and 54 Mbits/s (IEEE 802.11g rates). It would therefore allow transmission of several channels (4 in the case of the WCAM project). Each channel should be able to support a total application level throughput of 8-10 Mbits/s, which translates to a minimum of 4 WCAM video channels per carrier.
The IEEE 802.11b standard on its own has too low a bandwidth to support multiple channels (and would struggle to offer a single channel in the presence of other users). The IEEE 802.11a standard [19] at 5GHz would have offered a similar performance to that of 802.11g; however it is not available in Europe. Moreover, operating at 2.4GHz will offer better coverage performance and IEEE 802.11b/g cards are readily available in the marketplace.

9. CONTENT SECURITY ISSUES (DIGITAL RIGHT MANAGEMENT)

Additionally to the basic WLAN and network encryption tools, WCAM is implementing a content security mechanism. A DRM (Digital Rights Management) solution, called OpenSDRM is added to manage all authenticated peers on the network (from end-users to cameras), as well as to manage the rights to access and display conditionally the video data. This whole integrated architecture addresses several security problems such as data encryption, integrity, access control and rights management. Using several protection layers, the level of confidentiality can depend both on content characteristics and user rights, thus also addressing the critical issue of privacy.

10. QUALITY OF SERVICE

Complementary to network Quality Of Service tools, WCAM is looking to the implementation of video content Quality Of Service. Two aspects are covered by the project, the first one being to work on error resilience and concealment tools, these errors being intrinsic to the use of a wireless network. This work is concerning both MotionJPEG 2000, where tools coming from Wireless JPEG 2000 standard will be used and for MPEG-4 AVC where an advanced error concealment method has been already tested by the project.

Another aspect of Quality Of Service is to take advantage of scalable video for adapting the content, in terms of image size, bit-rate and frame rate to the capabilities of the client terminal and used network. WCAM is currently studying the possibility of using Universal Multimedia Access (UMA) and MPEG-21 Digital Item Adaptation (DIA) for that purpose.

New classes of pervasive computing devices such as personal digital assistants (PDAs), hand-held computers, smart phones, automotive computing devices, and wearable computers allow users more ubiquitous access to information. As users are beginning to rely more heavily on pervasive computing devices, there is a growing need for applications to bring multimedia information to the devices. The basic idea of Universal Multimedia Access (UMA) is to enable client devices with limited communication, processing, storage and display capabilities to access rich multimedia content.

Recently, several solutions focused on adapting the multimedia content to the client devices. UMA can be provided in two basic ways -- the first is by storing, managing, selecting, and delivering different versions of the media objects (images, video, audio, graphics and text) that comprise the multimedia presentations. The second is by manipulating the media objects on-the-fly, such as by using methods for text-to-speech translation, image and video transcoding, media conversion and summarization. This allows the multimedia content delivery to adapt to the wide diversity of client device capabilities in communication, processing, storage, and display.

As the interoperability among numerous contents, service providers and terminals are essential in UMA systems, these descriptions should be described in some standard format. MPEG-21, especially Part 7 Digital Item Adaptation (DIA) provides a rich set of standardized descriptions and tools necessary for adaptation. Some tools in MPEG-7 Part 5 Multimedia Description Scheme (MDS) are also important for Universal Multimedia Access. In this section, we give an overview of MPEG-21 DIA, and MPEG-7 tools for content adaptation.

The goal of MPEG-21 is to define the technology needed to support users to exchange, access, consume, trade, and otherwise manipulate Digital Items (DIs) in an efficient, transparent, and interoperable way [21]. "Digital Item" (DI) is a structured resource (such as video, audio, text, image, etc) with a standard representation, identification, and associated metadata within the MPEG-21 framework. "User" is any entity that interacts within the MPEG-21 environment and/or makes use of DIs. Thus, a User of a system includes all members of the value chain (e.g., creator, rights holders, distributors (service providers) and consumers of Digital Items). It should be noted that at this time the standardization of MPEG-21 is still ongoing.

One goal of MPEG-21 DIA is to provide standardized descriptions and tools that can be used by adaptation engines, which are quite relevant to Universal Multimedia Access [22]. The conceptual architecture of MPEG-21 DIA is illustrated in Figure 4. DIs are subject to a resource adaptation engine, as well as a descriptor adaptation engine, which together produce the adapted DI.

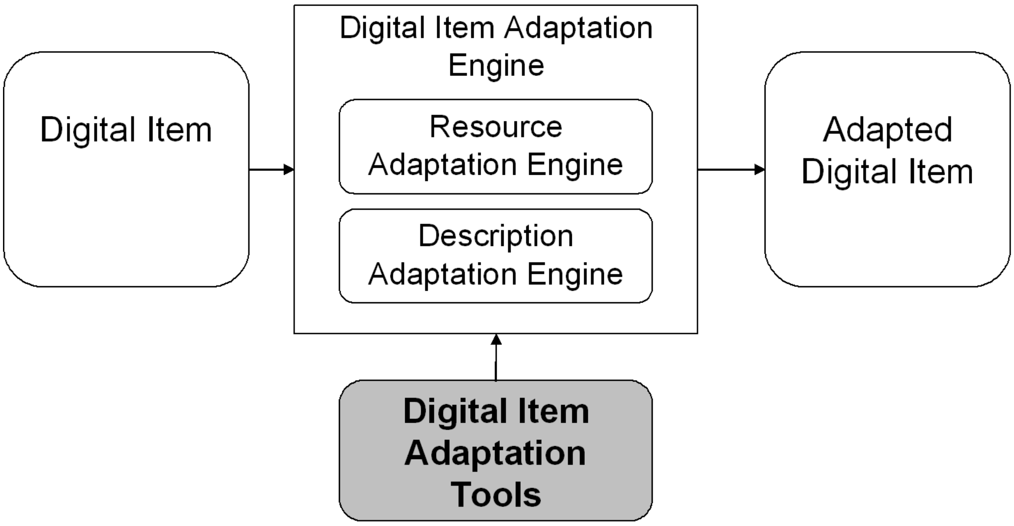

Figure 4 : Illustration of MPEG-21 Digital Item Adaptation.

In WCAM the Digital Item Adaptation is provided by a server that includes transcoding features.

11. TRANSCODING SERVER

For this the video on demand server, WCAM is using a video on demand server coming from the multimedia field, that will allow to distribute MPEG-4 AVC streams included in MPEG-2 Transport Streams, as well as MotionJPEG 2000 using RTP protocol. An important function that will be added to this server is the video transcoding.

Video transcoding is used to convert a coded video signal to another. It is mostly used to reduce a video's bit-rate to send it through a lower capacity channel. It is also used to change the video's spatial and temporal resolution, and to increase error-resilience when the signal is sent on wireless networks. Finally, transcoding is used to connect systems using different technologies, like different video compression standards. A good overview of the transcoding techniques has been published in [5]

The easiest way produce a transcoded signal is to decompress the initial signal, and re-compress it with the new constraints like on the figure below. The cost of this method is of course very high, and more efficient methods have been developed. A compromise has to be found between the output video's quality and the low complexity of the algorithm. Much more efficient algorithms take advantage of the reuse of information in the compressed domain, like motion vectors. This allows avoiding the re-computation of the most costly part of the algorithm.

Transcoding functions play several roles in WCAM:

Bit-rate reduction. It consists in increasing the compression rate usually done by truncation or modification of scale factors.

Spatial resolution reduction. It corresponds to the creation of a reduced resolution of the video, e.g. from HDTV to QVGA size

Temporal resolution reduction. It consists in modifying the amount of frames per second, by skipping some of them.

Error-resilience Transcoding. Addition or removal of particular error resilience options by recreating a codestream

Transcoding form one standard to another. Can be done using some compressed domain information or will need a complete decoding/encoding process. WCAM is considering transcoding MPEG-4 AVC from/to MPEG-2 and MPEG-4 AVC from/to MotionJPEG 2000.

Transcoding of encrypted bit streams. This is something implied by the used of secured content fro switching from one Right Management system to another, but also to change access rights.

Scalable video codestream management. When using MotionJPEG 2000, temporal and spatial reductions, as well as bit-rate are managed by removing some part of information in the compressed data. This is ensured by a parser function that will identify the role of each subset of information, and decide to keep them or not depending on the transcoding requirements.

12. CONCLUSION

WCAM project is demonstrating the feasibility of developing a complete video surveillance system (apart from the supervision application) based only on multimedia components. Only standards and not proprietary solutions are used in this system, which is the key issue for interoperability. The project is using standard video compression schemes and metadata descriptions coming from ISO and ITU, streaming coming from IETF, and common wireless network, while adding some important features like error robustness and concealment, automatic selective compression (Region Of Interest) scalability and security. An open implementation of Digital Rights Management is used for the protection of video content complementary to network security.

For the fist time a Digital Right Management system coming from the multimedia domain will be used in the context of a video surveillance system, ensuring protection of the video content and not only the network. This content security system will also allow managing the access rights to the video, enabling different levels of access rights. Scalable video and transcoding are two issues important in multimedia systems that will become more important in the future for video surveillance, allowing the use of different client levels in the same system, up to mobility features.

Not the least important fact is that WCAM system will be tested and asses in a real video surveillance system, with users, demonstrating the validity of the multimedia and video surveillance convergence in a real-life application.

ACKNOWLEDGEMENT

This work has been funded by the EU commission in the scope of the FP6 IST-2003-507204 project WCAM "Wireless Cameras and Audio-Visual Seamless Networking".

REFERENCES

1. I. E. G. Richardson, H.264 / MPEG-4 Part 10 White Papers, www.vcodex.com," 2002.
2. A. Joch, F. Kossentini, H. Schwarz, T. Wiegand, and G. J. Sullivan, "Performance Comparison Of Video Coding Standards Using Lagrangian Coder Control," presented at IEEE International Conference on Image Processing (ICIP), Rochester, USA, 2002.
3. T. Wiegand, G. J. Sullivan, G. Bjontegaard, and A. Luthra, "Overview of the H.264/AVC Video Coding Standard," IEEE Transactions on Circuits and Systems for Video Technology, vol. 13, pp. 560-577, 2003.
4. I. E. G. Richardson, H.264 and MPEG-4 Video Compression: John Wiley & Sons, 2003.
5. Vetro, A. et al., Video Transcoding Architectures and techniques: an overview, IEEE Signal Processing magazine, March 2003, pp 18-29
6. Wee, S. and Apostopoulos, J., Secure scalable streaming enabling transcoding without decryption, Proc IEEE Int. Conf. Image Processing, vol 1, Sept 2001, pp. 437-440
7. Guo, W., et al., MPEG-2 to MPEG-4 transcoding, Proc. of Workshop and Exhibition on MPEG-4, 2001
8. ISO/IEC 15444-3 JPEG 2000 image coding system Part 3: Motion JPEG 2000
9. ISO/IEC 15444-1 JPEG 2000 image coding system Part 1: Core coding system
10. Santa-Cruz, T. Ebrahimi, “An analytical study of JPEG2000 functionalities”, ICIP2000, Vancouver, Sept. 2000.
11. M. Rabbani, R. Joshi, “An overview of the JPEG2000 still image compression standard”, Signal Processing Image Communication, Eurasip, Volume17, No. 1, pp. 3-48, January 2002
12. T. Sikora, “The MPEG-7 Visual Standard for Content Description - An Overview,” *IEEE Trans. on Circuits and Systems for Video Technology*, **11**(6), pp. 696-702, June 2001.
13. P. Salembier and J. Smith, “MPEG-7 multimedia description schemes,” *IEEE Trans. on Circuits and Systems for Video Technology*, **11**(6), pp. 748-759, June 2001.
14. H. Schulzrinne, S. Casner, R. Frederick, V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications", IETF RFC3350, July 2003
15. S. Wenger, T. Stockhammer, M. Hannuksela, M. Westerlund, and D. Singer, "RTP payload format for H.264 video, *Internet Engineering Task Force (IETF)*, “draft-ietf-avt-rtp-h264-11.txt”, Aug. 2004.
16. S. Futemma, E. Itakura, A. Leung, “RTP payload Format for JPEG 2000 Video Streams, *Internet Engineering Task Force (IETF)*, “draft-ietf-avt-rtp-jpeg2000-05.txt”, July 2004
17. IEEE Std 802.11-1999, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications.
18. IEEE Std 802.11b-1999, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Higher-Speed Physical Layer Extension in the2.4 GHz Band.
19. IEEE Std 802.11a-1999, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: High Speed Physical Layer in the 5GHz Band.
20. IEEE Std 802.11g/D1.1-2001, Part 11: Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications: Further Higher-Speed Physical Layer Extension in the 2.4 GHz Band.
21. ISO/IEC 21000-1 FCD: “Multimedia Framework - Part 1: Vision, Technologies and Strategy”, (2003)
22. ISO/IEC 21000-7 DIS: “Multimedia Framework - Part 7: Digital Item Adaptation”, (2003)