INTERNATIONAL ORGANISATION FOR STANDARDISATION
ORGANISATION INTERNATIONALE DE NORMALISATION
ISO/IEC JTC1/SC29/WG11
CODING OF MOVING PICTURES AND AUDIO

ISO/IEC JTC1/SC29/WG11/M13458
July 2006, Klagenfurt, Austria

Title: Distributed Video Coding: Identifying Promising Application Scenarios
Authors R. Leonardi on behalf of DISCOVER consortium
Purpose: Information

DISTRIBUTED VIDEO CODING: IDENTIFYING PROMISING APPLICATION SCENARIOS

TABLE OF CONTENTS

Distributed Video Coding: Identifying Promising Application Scenarios

1. INTRODUCTION

Distributed Source Coding (DSC) is a new coding paradigm based on two major Information Theory results: the Slepian-Wolf [1] and Wyner-Ziv Theorems [2,3]. DSC theory relies on the coding of two or more dependent random sequences in an independent way, i.e. associating an independent encoder to each sequence. A single decoder is used to perform joint decoding of all encoded sequences, exploiting the statistical dependencies between them. Based on the DSC independent encoding-joint decoding configuration, a new video coding paradigm, called Distributed Video Coding (DVC) has been defined. Although the theoretical foundations of distributed source coding have been established in the 1970s, the design of practical video coding schemes based on DSC has been proposed only in recent years [4-13]. A major reason behind these developments is related to the evolution of channel coding, notably the emergence of turbo and LDPC (low-density parity-check) codes, which provide ways to build the channel codes necessary for DVC.

The major objective of this MPEG contribution is to report on the study made by the European project DISCOVER about the application scenarios for which the DVC paradigm may bring major benefits and identify which are these benefits. Note it is not the purpose of this contribution to claim that DVC is the right way to go for any application scenario. Considering the far from mature stage of DVC research, it is too early for final conclusions and claims. The purpose is rather to identify the most promising applications, helping the researchers to focus their work on the most adequate application spots, in order conclusions on the value of DVC for these applications may be taken as soon as possible.

The benefits discussed along this document are valid under the assumption that the major objectives of the DISCOVER project (e.g. flexible allocation of complexity, low encoding complexity, increased error robustness) may be reached, at least within an acceptable degree. Although the literature generally refers that DVC is useful for low complexity and low-power consumption encoders, no detailed application analysis is available on these advantages. It is also believed by the DISCOVER consortium that low complexity is not the single potential DVC advantage, and may not even be the most promising one. This investigation is precisely one of the major research targets of this project.

To achieve the objectives stated above, this contribution identifies and studies which requirements and functionalities are relevant for each application scenario, for example, coding efficiency, error resilience, and encoder-decoder complexity trade-off. This contribution also clusters the application scenarios according to various relevant characteristics, e.g. single/multiple cameras, encoder/decoder complexity, delay constraints, availability of return channel. Finally, a list with the application scenarios for which DVC looks to be more promising will be drawn, following a proposed methodology.

[1] J. Slepian and J. Wolf, "Noiseless Coding of Correlated Information Sources", IEEE Trans. on Information Theory, vol. 19, nº 4, pp. 471 - 480, July 1973.

[2] A. Wyner, "Recent Results in the Shannon Theory", IEEE Trans. on Information Theory, vol. 20, nº 1, pp. 2 - 10, January 1974.

[3] A. Wyner and J. Ziv, "The Rate-Distortion Function for Source Coding with Side Information at the Decoder", IEEE Trans. on Information Theory, vol. 22, nº 1, pp. 1 - 10, January 1976.

[4] R. Puri and K. Ramchandran, "PRISM: A New Robust Video Coding Architecture Based on Distributed Compression Principles", 40th Allerton Conference on Communication, Control and Computing, Allerton, USA, October 2002.

[5] A. Majumdar and K. Ramchandran, "PRISM: an Error-Resilient Video Coding Paradigm for Wireless Networks", First International Conference on Broadband Networks, San Jose, CA, USA October 2004.

2004 Page(s):478 – 485

[6] A. Majumdar, R. Puri, P. Ishwar and K. Ramchandran, "Complexity/Performance Trade-offs for Robust Distributed Video Coding", IEEE Int. Conference on Image Processing, Genova, Italy, September 2005.

[7] A. Aaron, R. Zhang and B. Girod, "Wyner-Ziv Coding of Motion Video", Asilomar Conference on Signals, Systems and Computers, Pacific Grove, CA, USA, November 2002.

[8] A. Aaron, S. Rane, E. Setton and B. Girod, "Transform-Domain Wyner-Ziv Codec for Video", Visual Communications and Image Processing Conference, San Jose, CA, USA, January 2004.

[9] B. Girod, A. Aaron, S. Rane and D. Rebollo Monedero, "Distributed Video Coding", Proceedings of the IEEE, vol. 93, nº 1, pp. 71 - 83, January 2005.

[10] A. Aaron, S. Rane and B. Girod, "Wyner-Ziv Video Coding with Hash-Based Motion Compensation at the Receiver", IEEE Int. Conference on Image Processing, Singapore, October 2004.

[11] Abhik Majumdar, Jiajun Wang, and Kannan Ramchandran, "Drift Reduction in Predictive Video Transmission using a Distributed Source Coded Side-Channel", ACM Multimedia, October 2004.

[12] S. Rane, A. Aaron and B. Girod, "Systematic Lossy Forward Error Protection for Error-Resilient Digital Video Broadcasting - a Wyner-Ziv Coding Approach", IEEE International Conference on Image Processing, Singapore, October 2004.

[13] A. Sehgal and N. Ahuja, "Robust Predictive Coding and the Wyner-Ziv Problem", Data Compression Conference, Snowbird, Utah, USA, October 2003.

2. STUDYING APPLICATION SCENARIOS

Following the proposed methodology, the list of application scenarios selected for analysis in this document is:

1. Wireless Video Cameras
2. Wireless Low-Power Surveillance

3. Mobile Document Scanner
4. Video Conferencing with Mobile Devices
5. Mobile Video Mail
6. Disposable Video Cameras
7. Visual Sensor Networks
8. Networked Camcorders
9. Distributed Video Streaming
10. Multiview Image Acquisition
11. Wireless Capsule Endoscopy

The following subsections will study in detail each of the selected application scenarios.

2.1 WIRELESS VIDEO CAMERAS

2.1.1 GENERAL DESCRIPTION

An important application scenario for DVC is related to the wireless communication of video signals between remote devices. With the new emerging technologies for wireless communication, the possibility of sending video data in a wireless fashion has now become a reality.

From a general point of view, different scenarios where wireless cameras may be used for this purpose can be identified, the first important distinction being between single camera and multi-camera systems. This section mainly addresses the use of single wireless cameras; overlaps with similar applications in other sections are avoided as much as possible, such as networked cameras. Refer to the relevant sections of this document for a more detailed discussion of the latter.

In the present scenario, the situation where a single wireless camera has to send the acquired video data to a central station is the most relevant for the moment. A list with some of important applications where wireless cameras may be of great utility is provided, but it is clear that many other situations could be considered as well.

The first application deals with the possibility of having small portable cameras to be used for home targeted surveillance tasks, see Figure 1. It may be useful for many people to have a device that can be easily placed wherever, in order to monitor if something happens in a room or when someone enters a shop or even how the dog is behaving outside. Also this type of cameras can be used in embedded systems which are information systems integrated into cars, trains, airplanes or any mobile environment. In those situations, the use of a wireless camera is the only viable choice because often it is not possible to use a wired solution, especially if the user wants to have a highly flexible system where the camera can be easily moved from one place to another.

Figure 1 - Ordinary wireless camera [1].

Another situation that may be of great interest and importance is the case of very small wireless cameras for police investigative purposes or for remote sensing of phenomena that are very hard to be physically reached. In both situations in fact one needs to send a video signal from one point to a station while using very small devices and thus with very limited resources; see Figure 2 for examples of such cameras. In those situations, it is important to consider the cases where real time decoding is necessary and when it is not as this has influence on the requirements that will be listed later.

Figure 2 - Tiny wireless cameras [2].

Another example of application of wireless cameras is traffic control, see Figure 3. The advantage deriving from having a non-wired system is that it is possible to rearrange a network as needed without any physical problem. Moreover, wireless cameras show their advantages in the traffic control in problematic locations, where it is difficult to have a wired camera, as for example in suburban areas [3].

Figure 3 - Traffic management center, Arizona, US [3].

Finally, wireless cameras have great value in television production environments, being much used both inside and outside the studio to avoid annoying cables. Starting from 2001, more and more attention has been paid to wireless digital cameras as they “enable more robust transmission of material, more simply, over longer distances” with respect to their analog counterparts [4]. Cameras used for this application are more expensive, larger, and heavier, with respect to the ones for the previous applications, and they are typically intended for high quality video acquisition.

2.1.2 REQUIREMENTS AND FUNCTIONALITIES

For a wireless video camera application scenario, the following requirements and functionalities appear relevant:

- **Low Cost** – For the case of small, personal cameras, the cost must be low in order to encourage the deployment of such systems. It is here important to note that for some applications like TV production, the cost may not be an issue, as expensive high quality devices are generally used.
- **Error Resilience** - A very relevant problem in the case of wireless communications is the presence of a non ideal communication channel with time varying characteristics and the presence of noise. In case the noise is high and possibly time varying, it is important to consider the effects of a particular source coding approach in the overall performance. It is known that the separation principle, which states that source and channel coding can be considered independently for an optimal system, only applies asymptotically, and under particular theoretical hypothesis. This means that in practice the source code choice does affect the performance of the overall system. So, a requirement for this application is the use of coding techniques that provide good error resilience capabilities. In fact even though a reliable estimation of the channel and/or source characteristics can be attempted, this often imposes feedback from the channel, high complexity, and delay which cannot be dealt with in many practical contexts. So, a requirement for this application is the use of coding techniques that provide good error resilience capabilities.
- **Power Consumption** - The importance of power consumption in the case of wireless devices is well understood. In the very general situation, it is important to properly consider the amount of energy used in different tasks, namely, acquisition, processing and transmission. Depending on the application, a reduction in the required power may impact directly on life duration, transmission range or battery size, which are different characteristics all of great interest in the various applications scenarios described above.
- **Small Size** - In the case of wireless micro cameras, the size is of course of fundamental importance. In order to keep the size as small as possible, it is critical to analyze every single component from battery to optical parts, signal processors and transmitting apparatus.
- **Remote Control** - In some applications (typically surveillance) one may need to remotely control the wireless camera in order to zoom on details or to change the optical aperture. In these situations, the wireless camera must incorporate a receiver and this may clearly influence its design. Also, the need for remote control implicitly offers the possibility of having a feedback channel from decoder to encoder, which may be exploited to improve the codec performance.
- **Transmission Range** - An important requirement that may discriminate between different scenarios is the required transmission range (home targeted transmission, mid or long range). Depending on the required range, the device will need different transmission technologies, different error resilience requirements and different power allocation.

2.1.3 FORESEEN DVC BENEFITS

There are two major benefits regarding the usage of DVC in the context of this application scenario:

1. **Error Resilience** - In the video coding community, it is well known that the standard approach based on motion compensation with residual error coding is strongly affected by channel errors (when an error occurs it propagates on successive frames). It has been recently shown in literature that a distributed coding approach is more suitable for this context, as the usage of prediction error is completely eliminated [5, 6]. The predictor used in the standard encoding phase is substituted with the side information at the decoder (unknown and thus not used by the encoder) in the distributed approach; as long as the decoder has good side information, the original signal is recovered regardless of the presence of previous errors, provided enough Wyner-Ziv bits are coming from the encoder.
2. **Low Complexity** - DVC has received a lot of attentions in recent years also because it offers the possibility of shifting computational tasks from encoder to decoder. This has a lot of implications in the design of a video encoding device since a low computational requirement allows the use of simpler devices (which means smaller and cheaper)
3. **Low-Power Consumption** - In addition, the lower complexity may reduce the power consumption, which means longer battery life or reduced size, or more power available for transmission and thus higher transmission range.

2.1.4 CURRENT DVC DRAWBACKS

There are two major drawbacks regarding the usage of DVC in the context of this application scenario:

1. **Decoding Complexity** - As said above, one of the main characteristics of DVC is the shift of complexity from encoder to decoder. In current approaches to DVC, the required decoding complexity seems to be very high and it is difficult to properly evaluate the achievable decoding time performance. In applications requiring real-time decoding, this may be the most serious drawback.
2. **Compression Efficiency** - Another possible drawback of DVC is the limited compression efficiency when compared to the state-of-the-art, namely H.264/AVC. Up to now, it seems to be very difficult to even get close to the compression performance of H.264/AVC by using DVC techniques. Anyway, this drawback may be of secondary importance if a good trade off between compression and encoder complexity is obtainable by DVC.

2.1.5 REFERENCES

[1] http://www.homesecuritystore.com/baby_monitor.html

[2] http://www.pimall.com/nais/e.menu-b.html

[3] http://www.jupiter.com/Documents/CaseStudies/CS-TrafficMangmt.htm

[4] http://www.urbanfox.tv/articles/cameras/c17wireless_cameras.htm

[5] A. Majumdar, K. Ramchandran, "PRISM: an Error-Resilient Video Coding Paradigm for Wireless Networks", BROADNETS 2004, pp. 478-485, San Jose, CA, USA, Oct. 2004,

[6] A. S. Rane, R. Zhang and B. Girod, "Wyner-Ziv Coding for Video: Applications to Compression and Error Resilience", IEEE Data Compression Conference (DCC), Snowbird, UT, USA, March 2003

2.2 WIRELESS LOW-POWER SURVEILLANCE

2.2.1 GENERAL DESCRIPTION

Surveillance is the process of monitoring the behavior of people, objects or processes within systems for conformity to expected or desired norms in trusted systems for security or social control [1]. Surveillance may be covert (without the subject's knowledge) or overt (perhaps with frequent reminders such as "we are watching you"). Surveillance has been an intrinsic part of human history. Sun Tzu's The Art of War, written 2,500 years ago, discusses how spies should be used against a person's enemies [2]. But modern electronic and computer technology have given surveillance a whole new field of operation. Surveillance can be automated using computers, and people leave extensive records that describe their activities.

Surveillance technology grows unstoppably as security is becoming one of the most important values of citizens around the world. These applications are changing the way authorities provide security and contribute to people feeling of safety.

Homeland security is driving most of the new developments in this area but wireless video networks have also important applications that could have significant impacts throughout society. One of its highest priorities is ensuring that no unauthorized person gains entry to a secure facility or sensitive area, such as airports, commercial buildings, and nuclear plants. Private security also demands a wide range of solutions to provide security and safety. Domestic applications such as baby surveillance are also gaining a lot of attraction.

Wireless low-power surveillance network applications are also studied in this section. The key concept of these applications is surveillance, and so security. In order to provide it, different low-power consumption components are interconnected and the communication between them is carried out through wireless communication protocols. The components that provide information to the system are cameras (although other sensors can also be part of these systems) and the images can be captured or displayed by one or multiple devices. The surveillance target can be persons, objects, buildings, areas… so many different applications arise from this application scenario. Figure 4 shows an example of a wireless low-power surveillance network system that would be composed of several wireless video cameras, each packaged with a low-power wireless transceiver that is capable of processing, sending and receiving data, focusing on natural protection in order to prevent natural disasters. Another example is remote sensing applications for land surveying and farming in Earth monitoring applications [3].

Figure 4 - Monitoring wildlife is an emergent application for wireless low-power surveillance networks [4].

As mentioned before, many specific systems and applications can be included into the wireless low-power surveillance scenario. Some of them are enumerated and briefly explained in this section.

Surveillance of Public and Private Spaces

Closed-circuit television (CCTV), as a collection of surveillance cameras (Figure 5) doing video surveillance, regards the use of television cameras for surveillance [5]. Cameras can be wired or wireless depending on the location, availability of energy, transmission networks, cost…. CCTV is often used in areas where there is an increased need for security, such as banks, casinos, and airports. The use of CCTVs in public places has increased, causing debate over security vs. privacy. Closed-circuit cameras are often used to discourage crime. Nowadays systems cover towns, city centers, stations, car-parks, buses, trains...

Sensitive spaces security (nuclear plants, petrochemical facilities…) are also using wireless video networks to prevent unauthorized persons from entering. The most measurable effect of CCTV is not on crime prevention, but on detection and prosecution. Several notable murder cases and robberies have been solved with the use of CCTV evidence. The use of CCTV to track the movements of missing children is commonly used in many countries.

Figure 5 - Video cameras of CCTV monitoring a public space [6].

Among some of most important low-power surveillance applications are:

Surveillance Inside Taxis

In this scenario, cameras are installed inside taxis to reduce violent attacks on drivers. Digital cameras, which are linked to police stations, provide enough evidence to secure convictions in the courts. The compact camera unit deploys a wide angle lens so that images encompass the whole of the back of the cab [7].

Military Reconnaissance

Reconnaissance is the military term for the active gathering of information about an enemy, or other conditions, by physical observation [8]. It is part of combat intelligence. Examples of reconnaissance include patrolling by troops, ships, submarines, or aircraft, or setting up covert observation posts. These actions can be covered by wireless cameras that collect the images about the targets. A camera video transmitter disguised as pens or other small objects with spy or military purposes is another application that falls into this field.

Surveillance Aircraft

Surveillance aircrafts are military aircrafts used for monitoring enemy activity, usually carrying no armament [9]. This term includes military and major civilian aviation activities like ground surveillance for mapping, traffic monitoring, science, and geological survey. In addition, civilian aircrafts are used in many countries for border surveillance, fishery patrols or the prevention of smuggling and illegal migration. Of special interest are airborne systems for civil applications In particular, aerial photography, video productions, surveillance, weather and atmospheric monitoring, disaster relief, emergency communications relay, airborne early warning of storms, and terrorist attacks should be emphasized [10].

Electronic Tagging

Electronic tagging is a form of non-surreptitious surveillance consisting of an electronic device attached to a person or vehicle allowing their whereabouts to be monitored [11]. In general, devices locate themselves using GPS and report their position back to a control centre, e.g. via the cellular phone network. But some cameras can be also added to the person by this way monitoring its environment.

Sousveillance

Sousveillance (pronounced "Sou Veil Lance") refers both to inverse surveillance, as well as to the recording of an activity from the perspective of a participant in the activity (i.e. personal experience capture) [12].

Inverse surveillance, sometimes known by the neologism "hierarchical sousveillance" ("seeing from below" hierarchically), refers to the recording or monitoring of real or apparent authority figures by others, particularly those who are generally the subject of surveillance. Steve Mann, who coined the term [11], describes it as "watchful vigilance from underneath". (The term stems from the contrasting French words sur, meaning "above", and sous, meaning "below", i.e. "surveillance" denotes the "eye-in-the-sky" watching from above, whereas "sousveillance" denotes bring the camera or other means of observation down to human level), see Figure 6.

Inverse surveillance is a type of sousveillance. The more general concept of sousveillance goes beyond just inverse surveillance and the associated 20th century political "us versus them" framework for citizens to photograph police, shoppers to photograph shopkeepers, or passengers to photograph taxicab drivers.

Personal sousveillance is the art, science, and technology of personal experience capture, processing, storage, retrieval, and transmission, such as lifelong audiovisual recording by way of cybernetic prosthetics, such as seeing-aids, visual memory aids, and the like [12]. Even today's personal sousveillance technologies like camera phones and weblogs tend to build a sense of community, in contrast to surveillance that some have said is corrosive to community.

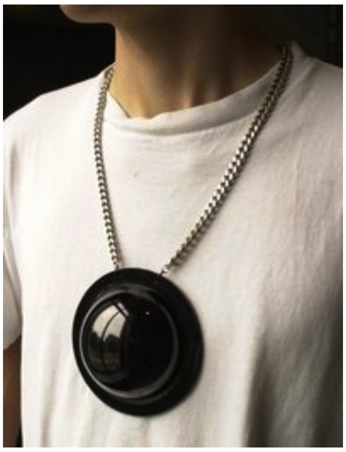

Figure 6 - Wearable wireless webcam imitates surveillance cameras common in casinos and department stores [12].

Monitoring Wildlife

Wireless cameras are placed in strategic points in forests, national parks, nature reserves, etc in order to monitor and alert authorities about natural disasters such as forest fires, and avalanches. The system can be designed as well with transcoders that enable the usage of mobile devices at the reception of the images and alerts. Forest guards will be provided with such devices and can react immediately when the disaster is detected, see Figure 4.

Naval Surveillance

Use of technology could include on-land and at-sea surveillance, video-assisted navigation and ship management and remote monitoring of army training exercises [13].

Estate Surveillance

CCTV patrols are used to monitor estates. This action move follows concerns about anti-social behavior and rising car crime in current society. These camera-equipped vehicles are used to tackle vehicle crime, anti-social behavior and disorders through providing evidences of such actions.

Traffic Monitoring

Many cities and motorway networks have extensive traffic-monitoring systems involving the use of closed-circuit television to detect congestion and notice accidents [5]. In other cities like London, the Congestion Charge (charging a fee for driving private automobiles in its central area) is enforced by cameras positioned at the boundaries of and inside the Congestion Charge Zone, which automatically read the registration plates of cars - if they do not pay the charge that day, they will be fined. Similar systems are also being developed as a means of locating cars reported stolen. Speed cameras are installed in various places, ostensibly to deter speeding.

Facial Recognition

The combination of CCTV with facial recognition technology has been tried as a form of mass surveillance, but has been ineffective because of the low discriminating power of facial recognition technology and the very high number of false positives generated [5]. This type of system has generally been proposed to compare faces at airports and seaports with those of suspected terrorists or other undesirable entrants. Although not directly related to image and video coding, in many occasions face recognition systems use communication systems in order to transport the acquired face to a central station where recognition is applied.

Wireless Home Monitoring and Control System

The wireless devices are mobile (PDAs, mobile phones, webcams…) and can be placed in different locations of the house. The system will capture images constantly and send them to the network base station or transcoder. The images will be decoded and processed there. Once the system detects some event that has been catalogued as an alert, the customer will be informed. The customer will then have the possibility to view directly the images in order to decide if the alert requires some action. This technology could also be used to give homeowners a way to monitor environmental changes in the home, such as water leaks or temperature changes, through the addition of sensors.

2.2.2 REQUIREMENTS AND FUNCTIONALITIES

As the range of applications that are placed into the wireless low-power surveillance network scenario is so wide and varied, the requirements may depend on the application focused. Herein there is a list of the most important requirements.

General Requirements

- **Power Consumption** - The term low-power refers to machines or activities that use less power than other similar machines or activities. Low-power consumption is one of the major requirements of this scenario. This implies that low complexity must be placed in the encoder side. It consumes less electric power than other similar devices. If one computer uses half the power of another computer, then it can run twice as long with the same set of batteries. In electrical engineering, power consumption refers to the electrical energy over time that must be supplied to an electrical device to maintain its operation. One of the major requirements and challenges in deploying wireless low-power surveillance networks is to maintain good video quality while significantly reducing power. In most applications, low-power dissipation is critical because of the limited battery life of the devices.

- **Small Size and Low Weight -** In some applications, like military or espionage, it is crucial to achieve small sizes and low weight in the devices that capture images.
- **Error Resilience -** Depending on the application focused, the quality of the images required will be different. If the application is related with security and must support personal recognition the quality of the images required is higher. Other applications related to entrance controlling, alert systems, object monitoring… may not need such high quality. Anyway the highest possible quality must be provided for a given bitrate and channel conditions. Time variation and unpredictability of the wireless channel makes results in transmission being affected by data errors and losses. It is therefore necessary to use encoders that provide error resilience functionalities in order to avoid possible lost of quality in the image. These errors can be critical for crime resolution purposes or face recognition.
- **Graceful Degradation -** Surveillance systems are usually recording or transmitting during long periods of time. This means that most of the time it is important to maintain the level of quality and disruptions are not allowed (or at least are undesirable). Again depending on the application the level of quality must be maintained constantly and sudden losses of quality and data are not allowed so graceful degradation (this means degradation without sudden big drops) is a requirement for the encoder (sudden losses of quality and data will not be acceptable for some security monitoring systems).
- **Compression Efficiency -** The objective in terms of compression efficiency would be to achieve a coding efficiency similar to the best hybrid video coding schemes (e.g. H.264/AVC). This means that balancing complexity from the encoder to the decoder should ideally not compromise the coding efficiency. In this scenario, cameras normally capture partially overlapped areas and therefore their associated video sequences are correlated. DVC must provide algorithms that enable to explore the correlation between the multiple encoded sequences just at the decoder, providing a low encoding complexity.
- **Delay Constraints -** There are many factors that affect the time performance of this scenario and can introduce delay on it. Image capturing, processing and compression at the encoder, introduces some delays that must be avoided or minimize to provide acceptable level of performance. Depending on the application and the type of surveillance, real time response must be provided and long delays may not be allowed. Other systems record the images and allow later analysis so it might not be a major requirement.
- **Low Cost and Low Complexity Devices -** Although low cost is not a major requirement in many surveillance applications, it can be considered useful or a benefit since some applications use many cameras (with encoders) to cover wide areas. As wireless devices usually have low computational capacity and provide small memory sizes, low complexity becomes a more important requirement in some cases.
- **Confidential Data Treatment -** The first surveillance networks used cameras in public spaces that were crude, conspicuous, in low definition black and white systems without the ability to zoom or pan [5]. Modern cameras use small high definition color cameras that can not only focus to resolve minute detail, but by linking the control of the cameras to a computer, objects can be tracked semi-automatically. The development of surveillance networks in public or private areas, linked to computer databases of people's pictures and identity, presents a serious breach of civil liberties. So the confidential treatment of data is

nowadays being regularized for many governments and any system must be developed according to these laws.

Wireless Related Requirements

- **Wireless Communication Protocol -** The communication between the devices that capture the images and the decoder must be wireless for this application scenario. Several communication protocols can be used and the choice of one of them depends on the specific characteristics of the application system developed. However there are basic features that must be evaluated such as cost, range, power consumption and bandwidth. Table 1 compares some of them (notice that other solutions can be used such as WLAN, wireless bridges…). Table 1 shows that none of the listed solutions can address all the constraints simultaneously. Low-cost and low-consumption protocols fulfill better the wireless low-power surveillance networks requirements. But depending on image sizes, image quality, channel requirements, time constraints, range requirements… other wireless communication protocols can be evaluated and integrated in the application scenario.

Protocol	Cost	Range	Power Consumption	Bandwidth
Wi-Fi	High	High	High	High
Bluetooth	Medium	Short	Low	Short
ZigBee	Low	Very Short	Very Low	Very Short

Table 1 – Comparison between some relevant wireless communication protocols.

- **Wireless Transceiver -** The devices that capture and display images must be provided with wireless transceivers in order to send, process and receive information via wireless communication channels. The transceiver is a device that has a transmitter and receiver which is combined into a one unit. Similar devices include transponders[1], transverters[2], and repeaters[3]. For wireless low-power surveillance network purposes, low-complex and low-cost transceivers must be used.

Image Acquisition/Display Related Requirements

- **Capturing/Displaying Capabilities -** One basic and common requirement for all video surveillance applications is the capability of capturing images. Therefore the devices of this application scenario must be able to capture images and transmit them to the processor component. The image treatment, quality, size, connectivity, speed capturing… are other requirements that drive the choice of the device. Cameras are the most common devices used to this end. Common formats in digital camera images are DCF, DPOF, EXIF, JPEG, RAW. And formats for movies are AVI, MPEG, MOV, WMV, etc. But as capturing images is a necessary requirement for all video surveillance networks, displaying images is only needed for some specific applications. As an example, wireless home monitoring and control system may include this requirement. Once the system detects some event that has been catalogued as an alert, the customer can be informed. The customer will then have the

[1] An automatic device that receives, amplifies, and retransmits a signal on a different frequency.

[2] A radio frequency device that consists of an upconverter and a downconverter in one unit.

[3] An amplifier used to improve the transmission reliability of a usually digital transmission line.

possibility to view directly the images in a portable device in order to decide if the alert requires some action.

- **Single/Multiple Capturing/Displaying Devices -** Depending on the surveillance target one or more cameras are needed. In big surveillance spaces, a high number of cameras are used to cover all the area. In this situation, images taken by the system can be overlapped which can be a requirement in order to used DVC in a profitable way. Another specific requirement for some applications is the multiplicity of the devices capable of displaying the images captured, or the capacity of displaying in the same device all the images captured. For example, in public spaces surveillance, it could be interesting that some police units are able to receive surveillance alerts through their portable devices.

- **Free-View of the Device -** Other requirement that apply to the cameras is the focus mobility. In some use cases, it is needed that the focus of the camera changes. This movement could be random, constant or controlled for the system (automatically or manually).

- **Mobile Cameras -** This requirement refers to the capacity of the cameras to change position. In some applications, cameras are mobile or installed in mobile vehicles (i.e. police unit patrols).

Network Architecture Related Requirements

- **Layered Network Architecture -** Traditional wireless video networks are based on flat, homogeneous architectures in which every camera has the same physical capabilities and can only interact with neighboring cameras in the network. Such networks simply cannot handle the amount of traffic generated by video applications. The amount of processing required on each node in terms of computing and communications and the power required to operate it would make such networks high costly and unfeasible to implement in reality. Therefore a requirement for such systems is the balancing of complexity and signal processing from the coders to the decoders. This way, more cameras can be placed and the cost and complexity of the network decreases. Some solutions have been proposed to this end, such as two-tiered architecture (where low cost video sensing nodes for capturing scenes are combined with more expensive group of control nodes that process the information), architecture based on transcoders, DVC, etc. [4].

- **Scalable Traffic Management -** Layered network architectures will not solve all the scalability issues. There will still be very large amounts of data to be managed. This is compounded by video applications, where each packet within a video flow must meet stringent delay bounds. If a packet of information for a video frame arrives too late, it is useless. Typically, to achieve a guaranteed (or an acceptable) quality of service (QoS), network nodes, routers or transcoders must maintain information on the rate and delay guarantees for each packet in a traffic flow and use that information to schedule the packets.

- **Real Time System** - Depending on the application, real time response must be provided by the system. This requirement implies that delay constraints are critical and the processing time of the images becomes crucial.

- **Transcoder -** In the scenario where both ends of the communication have important memory, computation or power constrains, an intermediate transcoder may be a requirement. The transcoder is located in the fixed part of the network, so the communication is split in two segments; from the emitting terminal to the transcoder, a DVC-based scheme could be used; from the transcoder to the receiving terminal, a traditional scheme could be used in the usual way. Thus, the whole bulk of the computation would be effectively shifted to the fixed part of the network, leaving the terminals free of any heavy calculation.

2.2.3 FORESEEN DVC BENEFITS

According to the description of the application scenario and the requirements and functionalities detailed above, the potential benefits of applying DVC are:

1. **Low Cost** - Lower cost of the cameras as the encoder may be less complex. Also lower cost of maintenance of the system, as most of the algorithms and complexity is placed at the decoder which is unique.
2. **Low-Power Consumption** - Lower consumption at the devices so the life of the network can be enlarged. As the amount of energy can be limited in some scenarios, lower consumption impacts on many aspects, from the amount of information to process at the encoder to the volume of wireless communication that can be carried across large distances. Current deployment of wireless low-power surveillance network solutions is very low in end customer market due to the high cost of deployment. As mobile devices have high penetration in the society, DVC could stimulate this application scenario and bring it near to persons. As devices' life is longer and less energy is required (no power supply is available in some cases), it allows monitoring hard reach areas.
3. **Error Resilience** - High correlation in the video sequences captured allows DVC to better exploit it in the decoder side using some side information. It promises better improvements in terms of error robustness. There is also the higher intrinsic robustness to channel losses due to the absence of the prediction loop at the encoder.
4. **Flexible Allocation of Complexity** - DVC capacity of balancing complexity between encoder and decoder provides flexible solutions to the many different applications that fall in this scenario. Some of them require really low complex encoders without regarding too much the reached quality while others' high priority is graceful degradation.
5. **Size and Weight** - As the complexity of the encoder is supposed to be reduced with DVC approach, the size and the weight of the devices that capture the images will be reduced.

2.2.4 CURRENT DVC DRAWBACKS

Major drawbacks of using DVC in this scenario are:

1. **Compression Efficiency** - Up to now DVC performance in terms of compression efficiency and error robustness are far from the results of state-of-the-art hybrid coding solutions. Also low complexity encoding goal can affect negatively the achievement of efficient coding algorithms.
2. **Transcoder** – In an end-to-end wireless low-power surveillance network scenario, a transcoder inside the network must be used in order to keep both the encoder and the decoder as simple as possible. This might be a bottleneck in future developments. The transcoder has to encode the video with a conventional codec.

2.2.5 REFERENCES

[1] http://en.wikipedia.org/wiki/Surveillance

[2] http://www.omnipresence.mahost.org/surveillance.htm

[3] John R. Jensen "Remote Sensing of the Environment: An Earth Resource Perspective" Prentice Hall, January 2000.

[4] *http://www.ecpe.vt.edu/news/ar03/video.html*

[5] *http://en.wikipedia.org/wiki/Closed-circuit_television*

[6] *http://news.bbc.co.uk/1/hi/england/tees/3529305.stm*

[7] *http://www.mediaeater.com/cameras/news/taxi.html*

[8] *http://en.wikipedia.org/wiki/Reconnaissance*

[9] *http://en.wikipedia.org/wiki/Spy_plane*

[10] *http://www.comets-uavs.org/applications/applic.shtml*

[11] *http://en.wikipedia.org/wiki/Electronic_tagging*

[12] *http://en.wikipedia.org/wiki/Sousveillance*

[13] *http://www.cylab.cmu.edu/default.aspx?id=2000*

2.3 MOBILE DOCUMENT SCANNER

2.3.1 GENERAL DESCRIPTION

The advent of wireless networks and mobile computing has freed businessmen from their offices, allowing them to work on the go. However, some services remain only available at fixed locations. Among them are copy machines, fax machines and image scanners. The large volumes and heavy weights of these machines prevent them from being carried along. This issue needs to be alleviated to allow a truly “anywhere, anytime” working environment. One solution would be to enable mobile phones to be used as portable faxes or scanners that can be used any time, simply by sweeping the phone across the page.

Document scanning on the go with a mobile phone would give wireless carriers the opportunity to provide a host of new services, ranging from the most basic ones like document transmission to email addresses, to printers or to the user’s computer, to more advanced services like Optical Character Recognition (OCR) and instantaneous translation for global travelers, sending back the translated text via instant messaging (see Figure 7). It would also allow queries into remote databases, a service most useful to law-enforcement units trying to collect evidence and identify criminals on the spot.

Scanning an A4 sized page by moving a mobile phone video camera over the document is likely to take about 3 to 5 seconds. Assuming a video frame rate ranging between 5 and 10 frames per second, this is going to produce between 15 and 50 images which a central server must merge together to extract the text and record any images. The application run on the central server must then forward the processed document to the targeted end device, e.g. e-mail, user’ computer, printer, mobile phone.

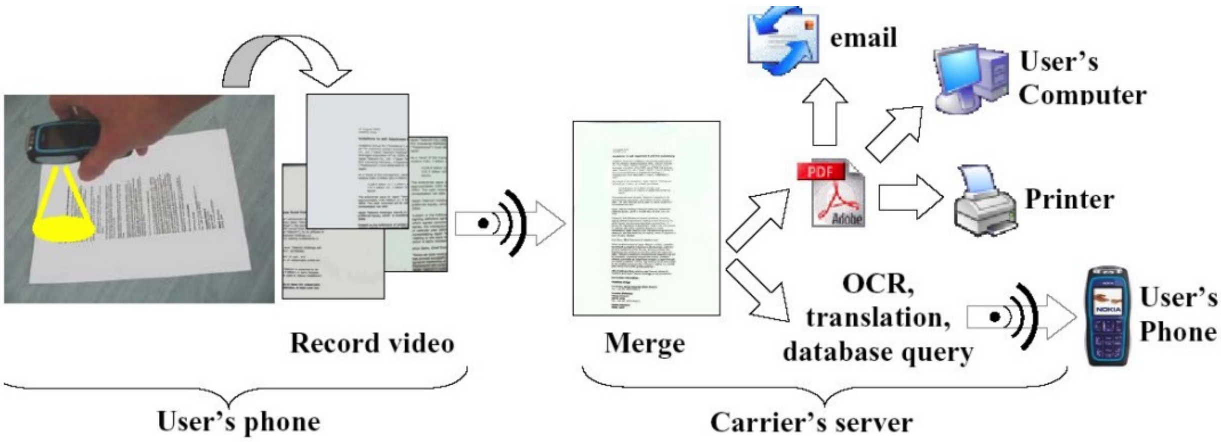

Figure 7 - Document scanning on the go [1].

2.3.2 REQUIREMENTS AND FUNCTIONALITIES

The major requirements regarding this application are:

- **Complexity** – Low-power consumption and low cost for the encoder is a major requirement since the application is foreseen on low cost mobiles devices such as mobile phones. This in turn requires low encoder complexity.
- **Video Post-Processing** – The capability to reconstruct high-quality video via visual processing tasks for mobile devices with low end cameras with small resolutions is requested. Indeed, cost, size, weight and power constraints prevent cell-phone makers from integrating anything but low-end cameras with small video resolutions, with fixed focus, high sensor noise and small field of view into their devices.
- **Central Server Processing** - In the considered "scanner on the move" application, the user moves the camera phone over the whole document, recording it as a video, and merging the video frames into a unique image using super-resolution. This processing cannot be supported today by camera phones because they lack the computational power required by registration, spherical aberration correction, and super-resolution techniques. Thus videos need to be sent over the wireless network to be processed by a central server. A coder/decoder architecture needs thus to be created, with a minimal encoder, most of the work being deferred to the decoder.
- **Image Processing Quality** - On the receiving side, processing of the family of super-resolution processing is required. The decoder has to register key-frames using feature point matching, robust homography estimation and non-linear MSE minimization. It has then to reconstruct the document using super-sampling or super-resolution techniques [2]. Spherical aberrations might need to be corrected [3], depending on the camera characteristics. Matching line features along with point features would make the registration procedure more robust. A homography needs to be applied to the merged image to compensate for the projective distortion and obtain a rectangular document.

2.3.3 FORESEEN DVC BENEFITS

The main DVC benefits for this application scenario can be summarized as follows:

1. **Low Complexity** - With respect to other video coding solutions, the low computational power and limited bandwidth constraints should hopefully be both satisfied. To reduce complexity and bandwidth constraints, one could consider intra coding (using e.g. a still picture encoding solution such as JPEG or JPEG2000) together with a reduced frame rate (that is distant frames). However, if the transmitted frames are too distant, this is likely to have an impact on the quality of the reconstructed document, the performance of the registration, feature point matching and super-resolution techniques being function among other things of the displacement between key frames. DVC would allow increasing the frame rate and sending extra data which will be used to correct registration and projective distortion.
2. **Error Resilience** - The wireless links considered for this type of application are erroneous. The use of traditional UDP/IP based transport of the video will lead to a high rate of packet losses. Traditional video coding techniques in such environments suffer from drift problems leading in turn to loss propagation. The separate encoding of the successive frames as done in DVC limits the problem of drift and error propagation. Using more recent transport protocols, such as UDP-lite, DCCP, and the MAC layer transparent modes of 3GPP and Bluetooth, the data received will be corrupted by bit errors. In this case, DVC will allow to limit temporal error propagation, since the new architecture offers a natural framework for joint source-channel coding.

2.3.4 CURRENT DVC DRAWBACKS

The current DVC drawbacks relevant for this application scenario are:

1. **Compression Efficiency** - The performance of DVC today does not match the one of predictive coding. However, the developments in this area are still in their infancy.
2. **Central Server Scalability** - Possible problems of central server scalability due to increased decoding complexity: In such applications, one can afford to have an increased decoder complexity, since the decoding is performed on a central server, however up to a point related to the scalability of the service, or its capability to support a certain number of users. If the central server has to provide the processing service to a large number of users, the extra burden or complexity for the decoder may also be seen as a drawback. However approaches with a more flexible load balancing between coder and decoder might be very beneficial for such applications. In such scenarios, among the two existing architectures (with return channel or without return channel but extra decoder complexity), DVC architectures with no return channel will be likely preferable to avoid delay and extra frame storage on the mobile sender side.

2.3.5 REFERENCES

[1] M. Maître, "Scanner on the Go", Internal IRISA technical report, 2005.

[2] S.C. Park, M.K. Park and M.G. Kang, "Super-Resolution Image Reconstruction: a Technical Overview", IEEE Signal Processing Magazine, vol. 20, n°3, pp. 21-36, May 2003.

[3] Y. Ma, S. Soatto, J. Kosecka, S. S. Sastry, An Invitation to 3-D Vision: From images to geometric models, Springer Edition, Series "Interdisciplinary Applied Mathematics", 2004.

2.4 VIDEO CONFERENCING WITH MOBILE DEVICES

2.4.1 GENERAL DESCRIPTION

Videoconferencing in its most basic form is the transmission of synchronized image (video) and speech (audio) back and forth between two or more physically separate locations. It allows people at different locations to communicate with the feeling as if the participants were in the same physical conversation. This is accomplished through the use of cameras (to capture and send video from local endpoint), video displays (to display video received from remote endpoints), microphones (to capture and send audio from local endpoint), and speakers (to play audio received from remote endpoints).

Videoconferencing offers a wide variety of benefits to both individual users as well as business organizations. It can speed up business process and procedures in the same way that the fax and the e-mail have revolutionized the way information was exchanged. Used effectively, video conferencing has a dramatic effect on the way people do business and the productivity gains they can achieve.

Tangible benefits are most easily related to actual cost savings. The most obvious quantifiable saving is the cost of travel and the cost of the time spent during travel. Instead of traveling a phone call can be done. However, a phone call is not an alternative to a face-to-face conversation in many situations. The ability to see the remote participants has many benefits over audio conferencing only. The facial expressions and body language of conference participants are both important aspects of communication which are lost with a basic telephone call without video transmission.

Furthermore, communication via a video conferencing system allows a participant to show details of the surrounding field, instead of trying to describe them over the phone. Visually seeing is far more effective and meaningful.

Sometimes it's just not possible or practical to have a face-to-face meeting with two or more people. At other times, a telephone conversation or conference call is adequate. Other times, an email exchange is adequate. Video conferencing adds another possible alternative. Video conferencing should be considered when:

- A live conversation is needed;
- Visual information is an important component of the conversation;
- The parties of the conversation cannot physically come to the same location; or
- The expense or time of travel is an issue.

2.4.2 REQUIREMENTS AND FUNCTIONALITIES

The major requirements regarding this application are:

- **Real-Time** - The coding and decoding process has to be done in real-time. Thus, the latency of the system needs to be low. As a result of many years of telephony research, a benchmark for voice circuit delay was determined and published by the International Telecommunications Union [1]. The accepted delay is up to 150 milliseconds per one-way transmission path, although also longer times, averagely up to 250 ms, may provide end-user acceptable results. Since the video and audio have to be synchronously decoded, the limit of 250 ms applies to video data, too.
- **Low Complexity** - Both the conference participants may use a mobile device for encoding the video and decoding the incoming bit stream. Therefore both, the encoder as well as the

decoder have to be as simple as possible. In order to reach these constraints, the bit stream should be transcoded from a Wyner-Ziv to a standard one (e.g. H.264/AVC) inside the network (see Figure 8), thus enabling a low-complex decoding.

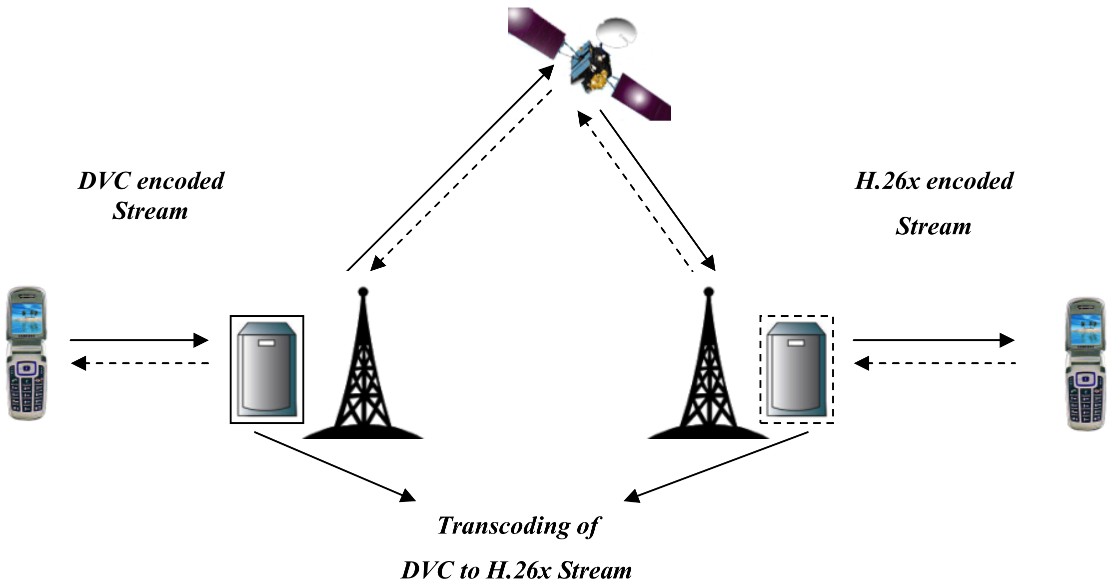

Figure 8 – The transcoding process of Wyner-Ziv bit stream to a conventional H.26x bit stream.

2.4.3 FORESEEN DVC BENEFITS

The main benefits can be summarized as follows:

1. **Low Complexity** - The DVC lower power consumption and lower computational complexity may make smaller devices possible.
2. **Increased Resolution** – Alternatively to lower complexity, the resolution of the captured video may be increased while power consumption or computational complexity are kept.
3. **Drift Robustness** - Another important benefit of the distributed video coding compared to conventional coding techniques is, that it is not affected by a drift between the encoder and the decoder. The drift is an artifact of the predictive coding framework e.g. due to channel losses. Research with coding of distorted video signals has indicated that a Wyner-Ziv coder outperforms a conventional H.263 coder with forward error correction [2].

2.4.4 CURRENT DVC DRAWBACKS

The current DVC drawbacks are:

1. **Compression Efficiency** - Current DVC codecs are not efficient enough to allow competitive solutions for video conferencing.
2. **Transcoder** - In the proposed scenario, a transcoder inside the network is used in order to keep both the encoder and the decoder as simple as possible. This might be the bottleneck in future developments, namely in terms of total end-to-end delay. The transcoder has to encode the video with a conventional codec. One of the challenges is to develop an efficient Wyner-Ziv to e.g. H.264/AVC real-time transcoder.

2.4.5 REFERENCES

[1] ITU-T G.114, "General Recommendations on the Transmission Quality for an Entire International Telephone Connection", Geneva, Switzerland, May 2000

[2] A. Majumdar, K. Ramchandran, "PRISM: An Error-Resilient Video Coding Paradigm for Wireless Networks", pp. 478-485, BROADNETS'04, San Jose, CA, USA, October 2004.

2.5 MOBILE VIDEO MAIL

2.5.1 GENERAL DESCRIPTION

The interest of customers for the new features of mobile devices is growing continuously. Recent statistics show that every two years a mobile phone is replaced by a modern one, allowing new applications, which hadn't been supported before. One of the most popular applications is sending text messages to friends, family or fellow-workers if direct calls are not possible or desired.

The first application was a SMS (Short Message Service), a service for transmitting text messages developed for GSM mobile nets and firstly introduced in 1992. Originally offered as a side product free of charge, it developed to be the most profitable service for the providers.

As a successor of SMS, MMS (Multimedia Messaging Service) established on the market. In contrast to SMS, a Multimedia Message may have an arbitrary number of attachments of different types. Thus, short text messages, complex documents, pictures or even videos can be sent. One possible application of MMS is video mail, which can replace SMS in most cases.

The benefits of a video mail over SMS are obvious: instead of typing, which takes a lot of time, only capturing images and freely speaking is needed with different media replacing difficult textual description of emotions or backgrounds, since seeing is believing.

2.5.2 REQUIREMENTS AND FUNCTIONALITIES

The major requirements regarding this application are:

1. **Compression Efficiency -** The objective in terms of compression efficiency would be to achieve a coding efficiency similar to the best hybrid video coding schemes (e.g. H.264/AVC).
2. **Low Complexity -** The encoder must be lightweight to fit into the mobile device.
3. **Low-Power Consumption** – In order to make such kind of application attractive, the encoder must be low-power consuming thus allowing the consumer multiple capturing and encoding per battery charge.
4. **Transcoder -** A transcoder is needed to encode the Wyner-Ziv bitstream into, e.g., a H.264/AVC bitstream, since simpler conventional decoding is used.

2.5.3 FORESEEN DVC BENEFITS

The main DVC benefits for this application scenario can be summarized as follows:

1. **Low Complexity -** DVC may allow low-computational encoding compared to conventional encoders, thus saving battery power.
2. **Increased Resolution -** Alternatively to the previous bullet, the resolution of the captured video can be increased while power consumption is kept.

3. **Error Resilience -** The DVC approach is more robust against channel errors, which are inevitable in mobile communication networks. If the channel capacity is limited and the video stream is corrupted, conventional decoders collapse very fast, while an efficient Wyner-Ziv decoder still could reconstruct the best possible video from an available bitstream.

2.5.4 CURRENT DVC DRAWBACKS

The current DVC drawbacks for this application scenario are:

1. **Compression Efficiency** – DVC solutions still cannot outperform current coding solutions like H.263 and H.264/AVC.
2. **Playback -** Editing or playback of captured video at the encoder side is not possible, since it would require a high-computational decoding processing; therefore only non-professional and rather short video mails come into question.
3. **Transcoder** – Because of high DVC decoder complexity, a transcoder from Wyner-Ziv to e.g. H.264/AVC inside the network is required in order to enable simple decoding at another mobile device.

2.5.5 REFERENCES

-

2.6 DISPOSABLE VIDEO CAMERAS

2.6.1 GENERAL DESCRIPTION

Disposable cameras appeared in the market first for still pictures and only more recently for video. The next sections will provide an overview on disposable video cameras requirements and functionalities as well as on its possible relation with distributed video coding (DVC).

Disposable or single-use photo cameras are basically a simple box camera sold with a roll of film installed, meant to be used only once. Most of these disposable cameras use focus free lenses with by using relatively small apertures increase the depth of field; this simplifies the camera at the cost of image quality (normally this type of lens does not produce very sharp images). Some are equipped with an integrated flash unit, and there are even water proof versions for underwater photography. The whole camera is handed in for processing and some of the cameras are recycled, i.e. refilled with film and resold; this refilling is a central part of the whole business model and thus on the acceptable camera complexity. Disposable digital cameras are a recent innovation (see Figure 9). These types of cameras use digital technology to take pictures, and the cameras are returned for processing in the same fashion as film cameras. In general the one-time-use camera represents a return to the business model pioneered by Kodak for their Brownie camera [1]; they are particularly popular in situations where a reusable camera would be easily stolen or damaged, or simply when a non-disposable camera is not available.

Figure 9 - Disposable video camera by Pure Digital Technologies [2].

The disposable photo camera was invented jointly during the mid-1980s by Kodak and Fujifilm; Kodak's original Fling was originally based on a 110 film but was eventually replaced by the FunSaver line with 35mm film and Advanced Photo System (APS) film, while the Fuji Utsurun-Desu or QuickSnap line used 35mm from the beginning. In Japan, the Fuji disposable cameras were released in 1986, for 1380 yen, and became widely accepted. Because of the immediate appeal, companies like Konica, Canon and Nikon soon produced their own models. To stay competitive, Fuji introduced advanced features to its original model such as panoramic photography, waterproofing and the inclusion of a flash. As of 2006, flash-equipped disposables are the norm. By the early 1990s such cameras were being produced by many companies, and are now a staple of the consumer film camera market [1].

Disposable digital photo cameras have been on the market since 2003, and now cost about €10 for a basic model, €20 if you want more features. Digital one-time-use photo cameras are available in some markets, but are somewhat controversial due to the expense of the process (especially compared to normal digital camera use) and the poor quality of the images compared to a typical digital camera. Buying the digital disposable photo camera itself may not imply saving any money. The price is about the same as that for a midlevel film disposable camera — one with flash, 27 exposures and a film speed of 400-800 for action shots. The savings comes from the photo CDs, which run about €3 less for digital disposables than for film disposables. Still, for those prone to photo mistakes, buying digital disposables instead of film can save in the long run since pictures may be edited, deleted or digitally managed/stored.

Disposable digital cameras are a little different from fancier ones when it comes to processing. Unlike true digital cameras, uploading these photos straight from the camera onto the computer is not possible. They have to be processed at the store chain from which the cameras were purchased. Also the cameras don't perform as well as regular digital cameras. Typically, the images are grainier (due to the small size of the sensor), and they don't always enlarge well (due to the low spatial resolution). The uneven flash can also pose problems.

Disposable photo cameras have been around for years and have carved out a healthy niche in the overall photography market. But nobody had come up with a disposable video camcorder until around June 2005. By that time, a $30 one-time-use camcorder went on sale at CVS stores[4], see Figure 10[3]. Consumers must pay an additional $12.99 to have the camera's 20 minutes of video

[4] CVS Corporation (http://www.cvs.com/CVSApp/cvs/gateway/cvsmain) operates retail drugstores in the United States. It offers prescription drugs, as well as general merchandise, including over-the-counter drugs, beauty products and cosmetics, film and photofinishing services, seasonal merchandise, greeting cards, and convenience foods.

processed and put onto a DVD. The process takes about 30 minutes. The cameras, about the size of an iPod, are not truly disposable in the sense that they are expected to be recycled by the company about five times before their final disposal. Major advantages of these disposable video cameras are:

- Highly portable, the camera fits in a pocket and weighs about 140 grams. The camera doesn't use tape; it stores the video on 128 MB of internal flash memory.
- Can be used for spur-of-the moment events, or when on vacation and the potential user forgot to bring a camcorder. It also can be useful in places where it's not as practical to take a regular camcorder, like on a sandy beach, on a rough roller-coaster ride or on a climb up a steep cliff.
- Easy to share videos on the Internet and via e-mail. When shooting is done, owner turns the camera in and receives a DVD burned with the video clips. Notice that some analysts say that sales of regular camcorders are declining partially because there has been no convenient way to share video online.

However the CVS camera presents the following limitations:

- No zoom capability.
- Can't plug the camera into the TV or computer.
- Can only view the last clip recorded, not the entire 20-minute video.
- At a 640×480 pixel resolution, the video quality is less than for traditional DV (digital video) camcorders.

Figure 10 - CVS disposable video camera [3].

In the CVS camera, currently the only disposable video camera in the market according to the research made, the lens and microphone are in the front of the camera. Images are stored on an internal 128 MB flash memory card. The VGA sensor captures images at 30 frames per second at a resolution of 640×480 pixels, which is better than most comparable, flash-memory-based, non-disposable digital video cameras and combination digital cameras that can also handle short video clips. CVS does not disclose the method by which the video is captured on the one-time-use video camcorder. Footage on the processed disc is MPEG-2. Remember that in order to have access to the footage in a DVD, it is necessary to return the camcorder to a CVS location.

The CVS camera has just four buttons - on/off, playback, record and delete - that give the user the basics on camera control, with no fancy menus to navigate. It also doesn't have a zoom lens, auto focus or image stabilization technology. So the camera works better for close-up shots of people rather than distant objects, and it produces a jittery image when hands shake. The disposable's video quality is good, not great. The company believes it can eventually improve the image quality to rival cameras that cost $300. However, the upside is the ability to keep the camera tucked away until needed for simple, impromptu shots and eliminating the worry of dropping it on the sidewalk or in water. And sharing a video finally becomes as easy as e-mailing a digital photo.

The business model for this type of camera revolves around the fact that the device will be used by multiple customers, allowing CVS to spread the cost of the hardware over a number of purchases - at least, if the camcorder is returned to the store for processing. However, as with the original one-time use digicams from CVS and a number of other retailers, a number of consumers already seem to be working on bypassing the need to return the device for processing. Websites quickly detailed the interface for the digicam variants, and already we've found at least one website describing attempts to access the camcorder variant with modified cables and USB drivers [3].

Disposable video cameras are an emerging type of product which future is still to be seen. It is very likely that more similar products will appear in the market in the next months.

2.6.2 REQUIREMENTS AND FUNCTIONALITIES

Regarding disposable video cameras, there are a number of requirements and functionalities which are essential, notably:

- **Low Cost** – Disposable video cameras must be cheap, from a user point of view, in order to compete with regular video cameras. This does not necessarily means that the device itself has a cost directly related to the price a single user pays for it, if the business model considers the disposable cameras are to be used (and sold) a few times as in the CVS business model described above.
- **Low Complexity** – Whatever the business model adopted, disposable video cameras must be as simple as possible in order to be competitive with regular video cameras; since the video encoder is a rather important module. This typically implies that the video encoder must be as simple as possible. However there is a trade-off between low complexity and quality; it is important to note that the low complexity requirement should not compromise too much the final audiovisual quality.
- **Lightweight Devices** – Disposable video cameras must be lightweight in order they offer an interesting solution for situations where regular video cameras may be cumbersome and too weighty. Again this may imply that the video encoders should be simple and with limited battery consumption.
- **Quality** – Disposable video cameras must use video codecs providing the highest possible quality for a given bitrate; however it is expected that quality may be a bit lower than with regular video cameras due to the simultaneous fulfillment of the cheap, low complexity and lightweight requirements. The efficiency is also closely related with the recording time for a fixed amount of memory. More efficient codecs will allow more recording time or reducing the memory needed for the device (and thus are cheaper); in both cases, the user wins.
- **User Control and Navigation** - Disposable video cameras need a minimum amount of user control, e.g. on/off, playback, record and delete as in the CVS camera, but far from the amount of user control and navigation available in regular video cameras. This requirement relates to the low cost, and low complexity requirements. Notice that this requirement may imply the availability of a decoder at the camera.

- **Easy to Use** – Considering the situations and the users it targets, disposable video cameras must be extremely easy to use. This requirement is in sync with the requirement on user control and navigation which asks for minimum capabilities.
- **A Posteriori Interaction** – Although disposable video cameras have to be simple and cheap, in general, the editing, manipulation, communication, etc. capabilities to be provided after the shooting with the DVD where the footage is stored, and the associated software, must be as powerful as possible in order the competition with regular video cameras is not weak in those grounds.

2.6.3 FORESEEN DVC BENEFITS

Considering the requirements and functionalities listed above for disposable video cameras, the DVC potential advantages in the area of disposable video cameras lie in:

1. **Low Complexity Encoding -** Low complexity encoding even if at the cost of some decrease in terms of compression efficiency/quality, would be a major benefit for this application scenario.
2. **Flexible Allocation of Complexity** – Also the flexible allocation of the codec complexity budget to control how complex is the encoder (at the camera) regarding the decoder (at the retailer or even at the DVD application) would be a useful capability.
3. **Lightweight Devices** - The provision by DVC of low complexity encoders would help to fulfill the requirements on 'low cost', 'low complexity', 'low battery consumption', and 'lightweight devices'. Although some penalty on the video quality may be acceptable compared to regular video cameras, this penalty should not be too high.

DVC advantages associated to error robustness don't seem to be relevant for this application scenario.

2.6.4 CURRENT DVC DRAWBACKS

The major current DVC drawbacks in terms of its application to disposable video cameras regard:

1. **Compression Efficiency -** The low coding efficiency is a drawback, especially if also low complexity encoding is targeted. In fact, current low encoder complexity DVC solutions provide a coding efficiency which is still below conventional coding solutions with low complexity, e.g. hybrid codec without motion estimation. To become a competitive video coding solution for this application domain, DVC must improve its RD performance to the levels stated in the next section for low complexity encoding.
2. **Decoding Complexity –** Although decoding complexity must always lie within reasonable limits, this application may tolerate a rather high decoding complexity to 'buy' a decreased encoding complexity since the decoding/transcoding process can be done off-line.
3. **Lack of Back Channel –** Since many existing DVC solutions require a feedback channel for the decoder to perform rate control, the unavailability of a back channel in this application scenario may be a limitation.

2.6.5 REFERENCES

[1] Wikipedia, The Free Encyclopedia, http://en.wikipedia.org/wiki/Disposable_camera

[2] http://www.washingtonpost.com/wp-dyn/content/article/2005/06/11/AR2005061100169.html

[3] http://www.camcorderinfo.com/content/CVS-One-Time-Use-Video-Camcorder.htm

2.7 VISUAL SENSOR NETWORKS

2.7.1 GENERAL DESCRIPTION

With the proliferation of inexpensive cameras (optical sensors) and non-optical (e.g., electrical, thermal, and biological) sensing devices, and the deployment of high-speed, wired/wireless networks, it has become economically and technically feasible to employ a large number of sensing devices for various applications, including on embedded devices. Embedded networked sensing may reveal previously unobservable phenomena.

Given the scope of DISCOVER, this section is centered on sensor networks using camera sensors. Camera sensor products range from expensive pan-tilt-zoom cameras to high-resolution digital cameras, and from inexpensive webcams and cell phones cameras to even cheaper, tiny cameras such as Cyclops [1]. Due to these advances, the design and deployment of camera sensor networks, or of wireless networks of sensor nodes equipped with cameras, is now feasible and useful in a variety of application scenarios.

There are many sensor networking applications which can significantly benefit from video information. These applications can include both video-only sensor networks or sensor networking applications in which video-based sensors augment traditional scalar sensor networks. Examples of such applications are security surveillance (civil or military), environmental monitoring, health care monitoring and robotics.

In environmental monitoring, a network of wireless camera sensors is used to monitor wild-life habitats, rare species in remote locations. They enable spatially and temporally dense environmental monitoring. Camera sensors can also be used in disaster management scenarios like fire and floods. Since pre-existing infrastructures may be unavailable or destroyed in these settings, a wireless battery powered deployment is necessary.

Surveillance so far has been dealing mostly with single stationary cameras, but the recent trend is indeed towards active multi-camera and sensor systems. In particular, the use of multiple video sensors to view a scene is rapidly increasing in many vision-based defense, security, scientific, and commercial applications. These applications may also combine (fuse) images and data coming from other sensors such as optical and infrared sensors, video, GPS and GIS data, etc.

In a sensor network, multiple sensors will generate signals which need to be sampled, filtered, transmitted, processed, fused, stored, indexed, and summarized as semantic events to allow efficient and effective queries and mining. Video sensor networks provide a formidable challenge to the underlying infrastructure due to the large computational requirements and the size of the captured data. The amount of video generated can consume the same bandwidth as thousands of scalar sensors.

2.7.2 REQUIREMENTS AND FUNCTIONALITIES

Regarding sensor networks, there are a number of requirements and functionalities which are essential, in particular:

- **Power Consumption -** Energy consumption is a primary issue to prolong the sensors network's lifetime. In environments where power is not plentiful, minimizing power can increase the number of sensors that can be economically deployed. It is important to develop techniques to organize nodes into tiers, and for dynamic allocation of tasks to sensors to minimize energy usage while at the same time meeting latency and accuracy requirements. Power management techniques, e.g., turning on a sensor to capture as much video as it can before the battery dies, or turning off some components to save power when the sensor is idle.

- **Processing Load Balancing -** A second critical factor in the design of a robust system for such applications is to enable each node to afford a certain level of local processing on the acquired data. This capability ensures better fault tolerance in the network or better robustness to node failure. It also enables a reduction in the amount of data transmission by allowing that only relevant information be extracted and transmitted. Balancing between non centralized and centralized data processing is thus an important issue.
- **Node Synchronization and Collaboration -** The nodes may need to collaborate to accomplish a common task and/or for efficient resource allocation. This requires appropriate sensor synchronization, sensor calibration and appropriate balancing of the different computational requirements and information exchange between the different nodes and between the sensor nodes and the centralized server.
- **Limited Computational Power and Low Bandwidth -** The sensors are limited in power, computing and memory resources. The sensor nodes are in general connected to servers via bandwidth-limited communication links. A representative scenario is one where wireless video sensors are employed in surveillance and the compressed video is transmitted to servers for processing, i.e., in particular for the multiview synthesis and the data fusion task. Note that the sensors may have heterogeneous power and networking requirements.
- **Easy Deployment and Flexible Configuration -** Other critical factors of such sensor networks are the ease of deployment and configuration of the network which may need to be re-configured in real-time (especially when used in hazardous environments). The nodes are prone to failure especially in harsh environments; hence the topology may change frequently. The trade-off between cost and coverage reliability is also an important issue.
- **Compression Efficiency -** A dense deployment is likely to lead to a large amount of data with high redundancy. Compression is thus obviously unavoidable, and the limited computing resources available at the cameras call for transmitting video using asymmetric coding with simple encoders and complex decoders. The video sensors should filter as much of the data captured as possible in order to maximize scalability, minimize the amount of network traffic, and minimize the storage space. Efficient use of bandwidth would also be greatly aided by detecting and identifying events and/or regions of interest in the scene. This could be either supported by the server (preferably in an automatic manner) or by the sensor nodes. In the first scenario, the information would then be fed back to the camera, and/or lead to a selection of the output of the relevant sensors. Objects in motion would typically determine regions of interest, e.g. in surveillance. The use of tracking would be a natural means of identifying such regions in a scene.
- **Data Sampling and Fusion -** All the information captured by the different sensors needs to be aggregated. However in spatially dense sensor networks, one may search for global features. To extract global features, it is necessary to sample the spatially correlated data. Sampling data in uniformly random manner not appropriate. A higher resolution sampling is used where more is going on (i.e. more variance in data values). The "dynamism" of the data is measured with statistical pre-processing. Each node tracks variance of its own data. The nodes locally exchange histograms representing sensor values**.**
- **Visual Processing Tasks -** The applications based on visual sensor networks may require elaborate visual processing such as viewpoint generation from scenes captured with the video cameras. Indeed, a camera sensor network may need to perform several processing tasks such as detection, recognition and tracking, in order to obtain useful information

from the video and images acquired by various camera sensors. This visual information may then be combined and fused with data captured by sensors of different types (infrared or others). Tracking can also be done using multi-source spatio-temporal data fusion, hierarchical description and representation of events and learning-based classification. The system then uses a hierarchical master-slave configuration, where each slave camera station tracks local movements and relays information to the master for fusion, interpretation and global representation.

- **Visual Roaming -** An additional desirable feature for these applications is visual roaming. Visual roaming refers to the possibility of viewing a scene from a chosen user perspective or in function of the data being captured, e.g., data triggering an alarm. Visual roaming requires the capture of a scene with multiple video sensors followed by the hard task of synthesizing new views from user/data-specified vantage points to facilitate visual roaming in a scene. The visual sensor network thus provides a redundant cluster of video streams to observe a remote scene and to supply automatic focus-of-attention with event-driven servoing (motorized control of camera pan-tilt-zoom) to capture desired events at appropriate resolutions and perspectives, e.g., for vehicle tracking and re-identification, for people tracking, face detection-recognition, and activity analysis. Tracking across video sensors is necessary for resolving occlusion and virtual viewpoint generation. Tracking involves the detection and recursive localization of an object or objects of interest based on sequential data measurements.

Continuous surveillance and simultaneous monitoring of multiple events in a scene captured by a sensor network is a formidable task for human observers. Computer vision is likely to play a significant role for simplifying the human tasks in a wide range of surveillance and security applications In the past, considerable amount of effort has been devoted to developing technology for automating the process of video surveillance [1-11].

2.7.3 FORESEEN DVC BENEFITS

The main benefits of DVC in this context are

1. **Coding Efficiency** - A reduction of the transmission rate with respect to the coding solutions used so far (JPEG) with a comparable CPU and power consumption. The expected reduction in the transmission rate with respect to separate encoding and separate decoding of the data captured by the different nodes is critical in wireless sensor networks. This should allow the use of a higher number of sensors, leading to a better coverage of the sensor monitoring or surveillance application.

2. **Low Encoding Complexity** – DVC should allow a reduced encoding complexity which is a key factor for reducing the power consumption and also the cost. Minimizing power and cost can increase the number of sensors that can be economically deployed.

3. **Low-Power Consumption** – A reduced coding complexity leads naturally to reduced power consumption which is critical especially in wireless and mobile sensor networks. In low-power scenarios, the sensor may indeed need to disconnect from time to time. Independent frame encoding as well as data prioritization naturally allowed by DVC coding architectures should facilitate such disconnections and the corresponding decoder re-synchronization.

4. **Error Resilience** - This feature is very critical for wireless sensors and is all the most critical for sensor networks evolving in harsh environments. The lack of prediction loop at the encoder in DVC codecs brings benefits in this area.

2.7.4 CURRENT DVC DRAWBACKS

The main drawbacks of DVC regarding this application scenario today are:

1. **Compression Efficiency** - The performance of DVC today does not match the one of predictive coding or of disparity based coding in a multi-camera set-up. However, the developments in this area are still in their infancy.
2. **Decoding Complexity** - In a dense sensor network, the extra burden or complexity for the decoder may also be seen as a drawback with respect to the scalability of the system or the number of sensors it can support.

2.7.5 REFERENCES

[1] M. Rahimi, R. Baer, J. Warrior, D. Estrin, and M. Srivastava, "Cyclops: In Situ Image Sensing and Interpretation in Wireless Sensor Networks", in Proc. of ACM SenSys, p. 311 (one page demo abstract), Baltimore, 2004

[2] T. E. Boult, R. J. Michieals, X. Gao and M. Eckmann, "Into the Woods: Visual Surveillance of Non-Cooperative and Camouflaged Targets in Complex Outdoor Settings", Proceedings of the IEEE, vol. 89, no. 10, pp. 1456-1477, October 2001

[3] R. T. Collins, A. Lipton, H. Fujiyoshi, and T. Kanade, "Algorithms for Cooperative Multi-Sensor Surveillance", Proceedings of the IEEE, vol. 89, no. 10, pp. 1456-1477, October 2001

[4] R. T. Collins, A. Lipton, H. Fujiyoshi, and T. Kanade, "A System for Video Surveillance and Monitoring", Proc. American Nuclear Society (ANS) Eighth International Topical Meeting on Robotics and Remote Systems, Pittsburgh, PA, USA, April 1999

[5] G. L. Foresti, P. Mahonen, and C.S. Regazzoni, Eds., "Multimedia Video-Based Surveillance Systems: Requirements, Issues and Solutions", Norwell, MA: Kluwer, 2000

[6] G. L. Foresti, V. Murino, C. S. Regazzoni, and G. Vernazza, "A Distributed Approach to 3D Road Scene Recognition", IEEE Trans. Veh. Technol., vol. 43, pp. 389–406, March 1994

[7] J. Heikkila and O. Silven, "A Real-Time Tracker for Visual Surveillance Applications," in Proc. First IEEE Int. Workshop Performance Evaluation of Tracking and Surveillance, PETS'2000, Grenoble, France, pp. 8–15, March 2000

[8] I. Haritaoglu, D. Harwood and L. S. Davis, "W4: Real-Time Surveillance of People and their Activities", in IEEE Trans. on PAMI, vol. 22, no. 8, pp. 809-830, Aug. 2000

[9] A. Iketani, A. Nagai, Y. Kuno, and Y. Shirai, "Real-Time Surveillance Systems Detecting Persons in Complex Scenes", Real-Time Imag. Journal, Special Issue on Video Processing and Communications in Real-Time Video-Based Surveillance, Vol. 7, No. 5, pp. 433-446, Oct. 2001.

[10] L. Marcenaro, F. Oberti, G. L. Foresti and C. S. Regazzoni, "Distributed Architectures and Logical-Task Decomposition in Multimedia Surveillance Systems", Proceedings of the IEEE, vol. 89, no. 10, pp. 1456-1477, October 2001

[11] I. Pavlidis, V. Morellas, P. Tsiamyrtzis and S. Harp, "Urban Surveillance Systems: from the Laboratory to the Commercial World", Proceedings of the IEEE, vol. 89, no. 10, pp. 1456-1477, October 2001.

2.8 NETWORKED CAMCORDERS

2.8.1 GENERAL DESCRIPTION

This section is dedicated to networked cameras understood as networks of cameras. In this context, network cameras will be taken as devices with acquisition, recording and transmission capabilities since this is very common in these days. This type of device is also known as 'camcorder' which is a compounded from 'camera' and 'recorder', the term originated in the early 1980s [1]; in the following, the term 'camcorder' will be mostly used. A camcorder is a portable electronic recording device that is capable of recording live-motion video and audio for later replay through VCRs, TVs, and, in some models, a personal computer. For many camcorders, the transmission capabilities can be quite demanding, e.g. transforming the camcorder in a kind of video server. Another possibility is to allow users to remotely control the camcorder in terms of shooting direction/angle, zooming, etc. Network camcorders are shrinking in size and in price, making them feasible for example for people interested in remote monitoring through a local network or the Internet.

Ordinary consumers use camcorders to film home movies of special events or vacations. Professionals such as professional videographers and filmmakers use camcorders along with other editing and film studio equipment to produce video clips or films for commercial sale. Camcorders come in an assortment of formats, features, and price ranges. When they first arrived, camcorders recorded in one of two analog formats, VHS and Betamax formats, onto video cassettes for replay from the most popular VCRs. These camcorders often produced less than ideal quality images and earlier models were large and cumbersome to use. As technology improved, other formats became available, such as S-VHS, 8mm, Hi-8, and DV (digital video). Many of these formats offered a clearer, sharper picture over the original formats, and, in some cases, allowed more hours of recording on a single tape than previously. These types often required an adapter for playback on a TV or VCR.

A digital camcorder is typically an appliance that has embedded digital video capture capabilities which allow video, audio or other information to be compressed in a variety of coding formats, stored or transmitted over communication networks or digital data link. These camcorders may offer the ability to use intelligent image analysis functions and different types of networking protocol support. This appliance can be used for any type of purpose, either for business or personal use.

The most common application for networks of camcorders is surveillance and monitoring with wired or wireless connections. However these networks of camcorders are also relevant for shooting and recording in other application contexts like entertainment events such as music concerts, sports, etc (see Figure 11, Figure 12). Since there is another section in this document specifically dedicated to surveillance networks, this section will concentrate more on non-surveillance scenarios.

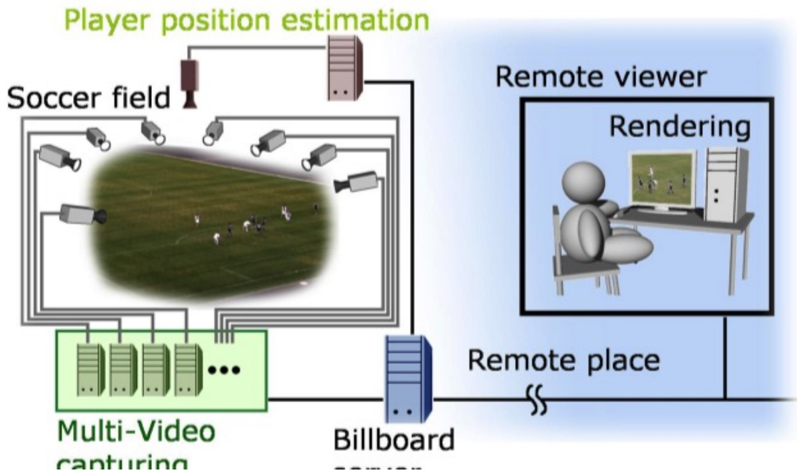

Figure 11 - Network of cameras shooting a sports event [2]; in this case, the cameras may not have recording capabilities.

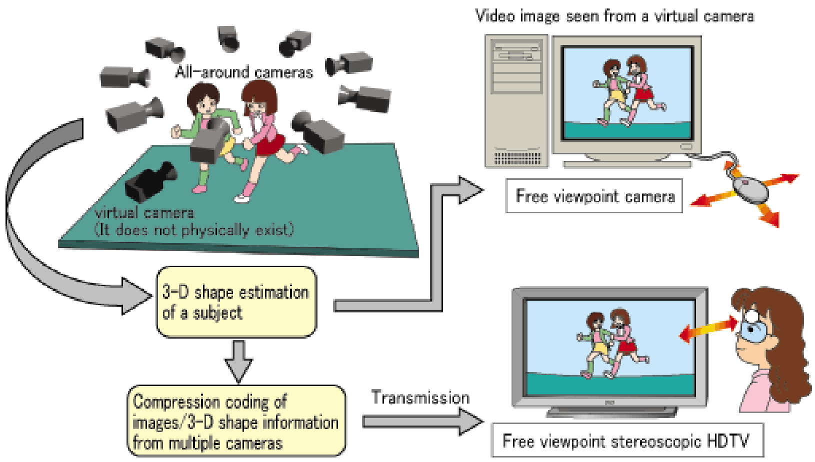

Figure 12 - Network of cameras shooting a scene [3].

This application scenario – networked cameras/camcorders – is mostly characterized by the usage of multiple devices (cameras/camcorders) shooting, recording and streaming the same scene, including the capability of later access on demand via wired or wireless channels to the views corresponding to any of the camcorders. This implies for example that the camcorders don't need to be transmitting continuously and simultaneously (they may be accessed one by one depending on the user needs).

2.8.2 REQUIREMENTS AND FUNCTIONALITIES

Regarding the networked cameras/camcorders scenario, there are a number of requirements and functionalities which are essential, notably:

- **Low Cost** – Although depending on the specific type of application environment in question, network camcorders must typically be cheap since many camcorders have to be used. It is however true that this requirement is not equally critical for a network of camcorders in a broadcasting environments and a surveillance environment.
- **Low Complexity** – Following the same reasoning as for the previous requirement, the camcorders must be as simple as possible, notably for certain application environments. This implies that the video encoder which is a rather important module must be as simple

as possible. Of course, there is a trade-off between low complexity and quality since the low complexity requirement should not compromise too much the final video quality but again this trade-off depends on the specific application environment where the networked camcorders are used.

- **Power Consumption** – Especially for mobile situations, the camcorders must be efficient on the usage of their battery; this implies that also the video encoder should be efficient in this sense, which basically translates to low complexity again.
- **Lightweight Devices** – Depending on the specific application environment, network camcorders may have to be lightweight, notably to be mounted in difficult positions.
- **Quality** – Camcorders must provide the highest possible video quality; again depending on the specific application environment there is a trade-off regarding the 'cheap', 'low complexity' and 'lightweight' requirements. In this application context, high quality shall in principle imply that the network of camcorders exploits the correlation between the multiple views of the same scene taken. To maintain the communications module in each camera simple, in principle this exploitation should happen without requiring the various camcorders to communicate between them.
- **Error Resilience** – Since transmission is important in the networked camcorders scenario, and it is essential to be able to deal with error prone channels (e.g. in wireless links), notably with packet losses, the video coding solution used at the camcorder must provide flexible error resilience capabilities.
- **Random Access** – In order the user may remotely access any part of the recorded views, the video coding solution used must provide random access with enough temporal granularity.
- **User Control and Navigation** – Since camcorders increasingly perform the role of a video server, they must provide a high degree of control and navigation capabilities to the remote user, both wired and wireless. This may imply sending a view (or parts of a view) when other views have already been transmitted, or sending all the views simultaneously, of course always in the most coding efficient way. This may also imply controlling the precise position, angle, or zooming factor to be used for the view captured by the specific camcorder; these parameters may have to be adequately considered by the video coding solution.

2.8.3 FORESEEN DVC BENEFITS

Considering the requirements and functionalities listed above for networked camcorders, the DVC potential benefits are mostly related to:

1. **Low Complexity** - Implementation of low complexity video encoder or at least video codecs with a flexible allocation of complexity between encoders and decoders; this may imply the acceptance of some quality reduction regarding most complex video encoders.
2. **Lightweight Devices and Power Consumption** - Cheap, lightweight devices, and low power usage related to the potential low complexity encoding above mentioned.
3. **Multiview Correlation Exploitation** - Exploitation of the correlation between different video views, either for the simultaneous transmission of all the views or for the differed transmission of one view when others have already been transmitted, targeting high video coding efficiency, without requiring the various camcorders to exchange information among them (however some information like their relative positioning should be known).

4. **Error Resilience** - Increased error resilience capabilities related to the absence of a coding loop at the DVC encoder or increased error resilience capabilities by using a DVC auxiliary channel on top of a standard video coding bitstream.

2.8.4 CURRENT DVC DRAWBACKS

The major DVC drawbacks regarding this application scenario are:

1. **Compression Efficiency -** The major current DVC drawback in terms of its application to networked camcorders regards again the low coding efficiency currently achieved for each single view if also low complexity encoding is targeted in comparison with alternative, conventional solutions, e.g. MPEG-X and H.26X codecs.
2. **Decoding Complexity –** The gains in terms of encoding complexity may happen at the cost of high decoding complexity. It is thus important to be able to control the trade-off between encoding and decoding complexities.
3. **Multiview Coding -** Regarding DVC based multiview video coding, notably exploiting the correlation between cameras, there is very little work reported in the literature and thus an increased research effort must be developed in this area since the DVC coding paradigm fits well the multiview video scenario when significant correlation between views exists.

To become a competitive video coding solution for this application domain, DVC must improve its RD performance to the levels stated in the next section with especial emphasis on low complexity encoding.

2.8.5 REFERENCES

[1] http://searchmobilecomputing.techtarget.com/sDefinition/0,,sid40_gci283966,00.html

[2] http://www.kameda-lab.org/research/project_3dvideo/soccer_3dvideo.jpg

[3] http://www.nhk.or.jp/strl/open2002/en/tenji/id08/e08-1.gif

2.9 DISTRIBUTED VIDEO STREAMING

2.9.1 GENERAL DESCRIPTION

The huge development of the Internet has given the possibility to realize video streaming systems that allow a user to view a video sequence at its own place while downloading it from a remote server or disk. In this setting the user does not want to download first the video sequence in order to see it at a later time, but he wants instead to see the sequence while streaming. With the same idea that led to the development of peer to peer networks used for "distributed" download of files, it is possible to consider the possibility of performing "distributed streaming", in order to give to the receiver the maximum possible data flow (see Figure 13). In this setting, the video stream is sent to the receiver by different senders in a distributed fashion, in order to reduce the bitrate at the sender sides and increase it at the receiver. In this context, it is possible to consider DVC as a new technology that may be used in order to perform a more flexible and reliable video streaming system.

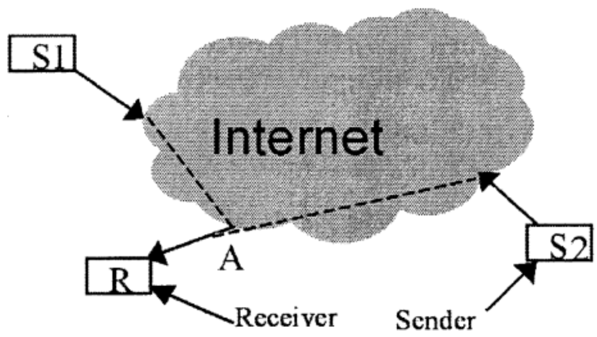

Figure 13 – Streaming through the Internet [1].

2.9.2 REQUIREMENTS AND FUNCTIONALITIES

The following requirements and functionalities should be considered:

- **Compression Efficiency** – The system must have good compression performance, in order to reduce the amount of data flow in the network. However, as there are many senders and only one receiver, it is important to consider that the required compression efficiency may not be very high, as the main target is to reduce the bitrate at the sender side and not at the receiver.
- **Bitrate Allocation** - Depending on the network conditions, different senders should/could be asked for different bitrate in order to find a trade off between receiver requirements and network usage.
- **Flexibility and Error Resilience** - The system must be able to manage the entrance and exit of users from the network, network congestions etc., and it is then required a high level of both flexibility for network user changes and error resilience for congestions and faults.
- **Real-Time Performance** – For the present scenario, it is important that the decoder can operate at the same rate of the incoming flow. This means that the decoding may have some delay, but is not allowed to have a speed regime slower than the incoming rate; otherwise no advantage is obtained by the increased bitrate of the incoming data.
- **Multiple Resolutions Handling** – A nice functionality is the possibility of downloading a video stream at a chosen resolution by taking it from different locations where different resolutions are available. In other words, if the same video is available in two different locations but with different resolution one may be interested in using both sources in order to obtain the required sequence resolution.

2.9.3 FORESEEN DVC BENEFITS

In this application scenario the use of DVC may give important and unexpected results.

1. **Reliability** - Using DVC every sender would provide to the receiver different portions of information without having a precise knowledge of what other senders are doing. This means that in case some of the users disconnect, the system still works as long as sufficient information is globally received from other ones. So, with a DVC approach the distributed streaming could be much more flexible to user changes, and for the same reasons to network faults or rate reallocation.
2. **Multiple Resolutions Handling** - With the use of DVC, it may be possible to construct a video sequence by taking information from two sources that are different, for example in

resolution. This is an interesting point since if the video sequences are not exactly the same, the pure file transfer distributed streaming (like in [1] and as discussed above) cannot be used. By using a DVC approach, one may be able instead to manage this situation allowing a very flexible system. For example, one of the sources can send its own data in a conventionally encoded way while other sources send their data in a WZ fashion. This way the decoder can recover all the incoming data and then use super-resolution techniques in order to construct one single version with a higher quality

2.9.4 CURRENT DVC DRAWBACKS

The main drawbacks of DVC in this application scenario are the following:

1. **Encoding Phase** - The use of DVC in a streaming system implies the necessity of including video encoding into the stream system, which thus changes from a file transfer/decoding to an encoding/transfer/decoding protocol. So, from the transmitter side the streaming changes significantly because encoding of the video sequence must be performed. This problem is anyway related to every distributed streaming system based on encoding techniques (and not only on transfer protocols), like for example the multiple description coding techniques considered above. Thus, considered that the encoding phase in DVC is usually not too complex, this may be a minor drawback.
2. **Decoding Complexity** - The required complexity for a DVC decoder is very high and this implies that the receiver must have a very high computational power.

2.9.5 REFERENCES

[1] T. Nguyen and A. Zakhor, "Multiple Sender Distributed Video Streaming", IEEE Transactions on Multimedia, vol. 6, no. 2, pp. 315-326, April 2004.

2.10 MULTIVIEW IMAGE ACQUISITION

2.10.1 GENERAL DESCRIPTION

Most image and video processing and coding solutions rely on one single camera, referred to as monoview approach. In the last two decades, extensions to two-camera solutions (also referred to as stereo) have been investigated with limited success in both coding and video analysis applications. Although multiview is also used in solutions with two cameras, here the term will only be used for solutions that use more than two cameras. Multiview images of a scene can be used for several applications ranging from free viewpoint television (FTV) to surveillance. In FTV, the user can freely control the viewpoint position of any dynamic real-world scene. This system can cover a limited space. To expand the coverage to a wider area, a distributed sensor network can be used. Such approach can be used for tasks as surveillance, widespread environmental sampling, security, and health monitoring and it is known to make human life safer and easier.

Multiview image and video processing has attracted increasing attention recently and has become one of the potential avenues in future imaging systems, thanks to the reducing cost of cameras. Many tasks can benefit from the availability of multiple views of the same scene, such as interpolation, restoration, segmentation, object recognition, etc. On the other hand, the amount of data captured in multiview imaging is often tremendous. For instance, in the application of image-based rendering, thousands of images are needed to synthesize novel views from arbitrary position. This makes data reduction a key issue in multiview image and video processing. Furthermore, due to the strong correlation between multiple views, multiview data reduction has its own characteristics that differ significantly from traditional image/video compression. As a result, an increasing amount of work on multiview sampling and compression has been proposed in recent years.

Advances in sensor network technology are radically changing the way in which signals of interest are sensed, processed and transported. Sensor networks consist of a large number of sensor nodes that are densely deployed either inside or close to a phenomenon of interest. Each sensor node is an independent, low-power, smart device with sensing, processing and wireless communication capabilities. The range of applications for sensor networks is extraordinary wide and covers a number of different areas such as health, military and home. In this section, the focus is on camera sensor networks that capture overlapped images, that is, each sensor is equipped with a digital camera and transmits the acquired visual information to a common central receiver. The sensors are all observing a certain scene from different viewing positions. The images acquired by different sensors can therefore be highly correlated. If the sensors were allowed to communicate with each other, it would be relatively easy to exploit this correlation in full and transmit only the necessary information to the receiver. However, such collaboration is usually not feasible since it would require a complex inter-sensor communication system that would consume most of the sensors' power, not to mention complex and undesired wiring or wireless network traffic issues. It is therefore necessary to develop separate compression algorithms that would still be able to exploit the correlation between different sensors without requiring any cooperation amongst the sensors themselves.

Distributed compression schemes usually rely on the assumption that the correlation of the source is known *a priori*. It is possible to estimate the correlation structure in the visual information acquired by the cameras by using some simple geometrical constraints and to conceive a coding approach that can exploit this correlation in order to reduce the overall transmission bitrate from the sensors to the common central receiver. The coding schemes proposed in DVC studies allow for a flexible distribution of the bitrates amongst the encoders [1].

Another emerging application field is based on camera arrays, see Figure 14. Large camera arrays can capture multiviewpoint images of a scene, which might be used in numerous novel applications such as movie special effects. For camera arrays built for such applications, one of the challenges is the enormous size of raw data, typically consisting of hundreds of pictures. Hence, compression is needed.

To exploit the coherence among neighboring views, the images are usually encoded jointly. In large camera arrays, however, cameras typically can only communicate with a central node, but not between each other. Since joint coding at the central node requires transmission of all raw images first and excessive memory space to store them temporarily, it is preferable to compress the images directly at each camera, in a distributed fashion. Existing systems either rely on built-in compression capabilities of the capturing devices, thus requiring expensive cameras, or need to add customized circuits to perform some form of standard image compression such as JPEG. With hundreds of cameras involved, the cost of either approach may be prohibitive.

DVC proposes a scheme that assumes no communication between the cameras and requires only a very simple, low-complexity encoder at each camera. The burden of computation is shifted to the centralized decoder, which is assumed to be more sophisticated. In the case of light field compression, the decoder also needs to perform scene geometry estimation, rendering of side information and adaptive rate control [2].

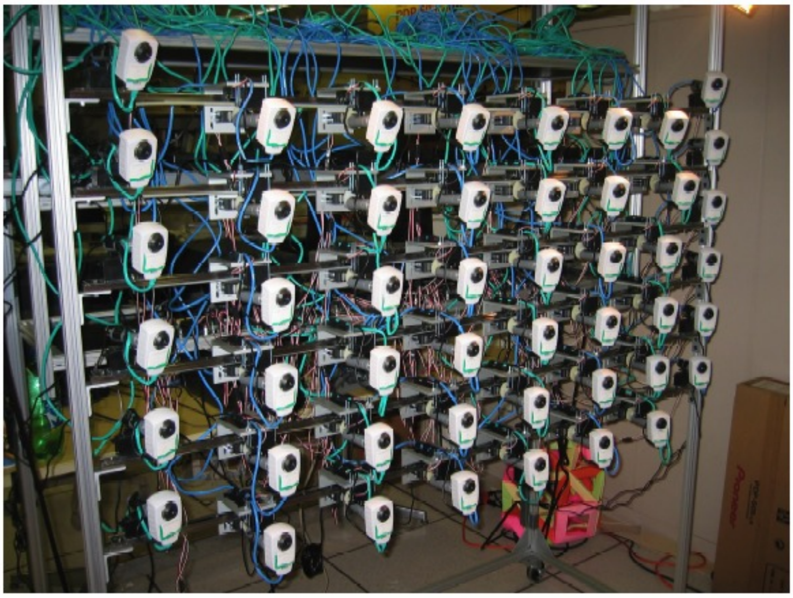

Figure 14 - Camera array system with 48 cameras [3].

It is also important to state that there are several ways to acquire and to arrange multiview images. The images of different view directions are obtained in several ways [4]:

1. The first technique uses a scanner in the image volume of the camera. In this case, multiview images with only horizontal parallax can be obtained. The scanner will effectively divide the input pupil of the camera objective into the same number of the multiview images if the image on the scanner is focused on an image detector.
2. The second technique uses a 2D array of cameras. This is a typical way of taking multiview images with a full parallax. The arrangement of cameras in the array will be either radial or parallel.
3. The third technique uses a translator traveling along the optical axis of the camera. The images obtained are a set of depth-wise sampled images and can be used to generate a volumetric image.
4. The last technique relies on a translator moving a camera in a radial or parallel direction.

A way to assure a "life-like" visual experience is to exploit the binocular aspect of human vision, i.e., the fact that humans perceive depth based on the capture of environment from two slightly different perspectives [5]. The capture of two images by closely-positioned cameras (about 6 cm apart) followed by their separate presentation to viewer's eyes, invokes the perception of depth and is known as "stereoscopic" imaging. When the two views presented to the left/right eyes are allowed to change with viewer head motion, and thus render other perspectives (the so-called "look-around"), terms "multiscopic" or "multiview" imaging are commonly used. In this thrust, there are concerns with problems ranging from data multiplexing for 3D displays, through the generation of virtual views, to 3D image compositing and multiview image sequence compression. Applications of this work can be found in entertainment, medicine, life sciences and tele-operation.

Multiview images are used in various fields and applications. Below some applications envisioned in the literature using multiview imaging are described.

High-Speed Videography

Systems for capturing multi-thousand frame-per-second (fps) video using dense array of cheap image sensors have been lately developed. A benefit of using a dense camera array to capture high-speed video is that higher speeds can be scaled by simply adding more cameras. Even at extremely high frame rates, this array architecture supports continuous streaming to disk from all of the cameras. This allows to record unpredictable events, in which nothing occurs before the event of interest that could be used to trigger the beginning of recording. Synthesizing one high-speed video sequence using images from an array of cameras requires methods to calibrate and correct those cameras' varying radiometric and geometric properties. As semiconductor technology advances, capturing and processing video from many cameras becomes increasingly easy and inexpensive.

Some studies are focused on discovering ways to extend performance by taking advantage of the unique features of multiple camera sensors—parallel compression for very long recordings, and exposure windows that span multiple high-speed frame times for increasing the frame rate or signal-to-noise ratio.

High-speed imaging is used to analyze automotive crash tests, golf swings, explosions, and more. Industrial, research, and military applications have motivated increasingly faster high-speed cameras. Currently, off-the-shelf cameras from companies like Photron and Vision Research can record 800×600 pixels at 4800 fps, or 2.3 Gigasamples per second. These devices use a single image sensor and are typically limited to storing just a few seconds of data because of the huge bandwidths involved in high-speed video [6].

Virtual Environments Generation

The generation of virtual environments for tele-presence systems, interactive viewing, and immersion in remote 3D-scenarios is an expanding and very promising research field. The challenge of this technology is to design systems capable of synthesizing views from any desired perspective using a set of real-scene perspectives. In such a system, two main parts can be distinguished: multiview image-analysis and viewpoint synthesis [7].

Cinematographic Special Effects

The requirement for dramatic special effects in films and broadcast material production has led to increased requirements for hybrid synthetic and natural content, object based decomposition methods and 'virtual' view synthesis. These, combined with motion effects based on high frame rate and/or multi-camera capture are beginning to transform film and television content creation.

There are two really good examples of this. Both of which are contained within the Wachowski Brothers' The Matrix Trilogy. The main reason for its use in both of these circumstances was that what was envisioned by the directors could not be done using conventional cameras [8].

Bullet Time

This approach involves a movement that makes a full circle around a subject (in the case of the Matrix movie a person) with high flexibility in terms of changing the speed of this movement.

The rig set up is illustrated in Figure 15. The little black holes are cameras. The green material is called "greenscreen" and is used for chroma-keying segmentation to identify the area to be replaced by the artificial background.

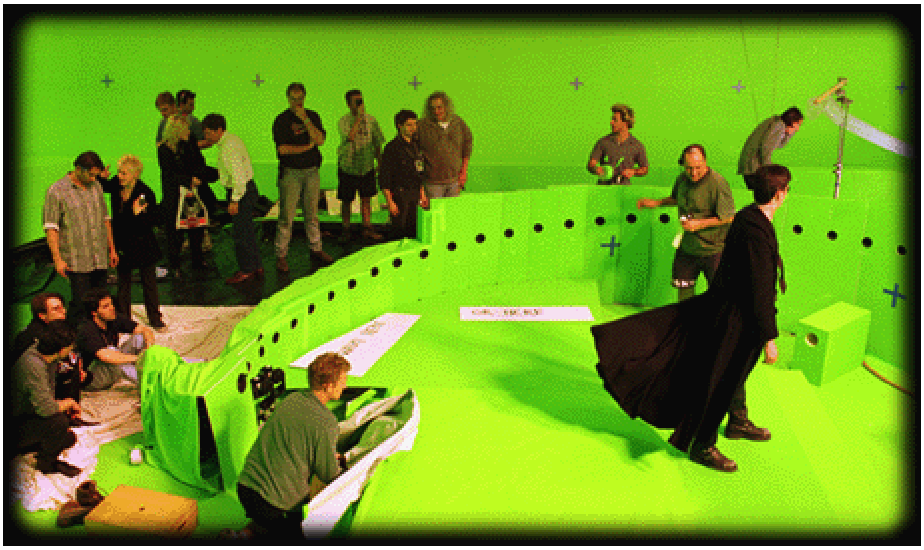

Figure 15 - Rig set up [9].

The cameras all "fire" in sequence as the motion to be captured is made. When the images are sequenced together (often filled in between frames using computer generated frames) it appears that a camera has moved around the object. All that remains then is for the green area and cameras to be removed, and for a virtual area to be super-imposed underneath the image of the object; this results in a completed timeslice sequence, also popularly known under the trademarked name bullet time [9].

Tele-Immersion

When Internet 2-level network performance becomes common place at some point in the future, which applications will make the best use of this giant leap in bandwidth? Virtual laboratories, digital libraries, and distance-independent learning are among some of the advanced applications currently being explored. Jaron Lanier, who helped lead in the development of virtual reality during the 1980s, is now guiding an attempt to validate the Net of tomorrow with a nascent technology known as tele-immersion [10]: long-distance transmission of life-size, three-dimensional synthesized scenes, accurately sampled and rendered in real time using advanced computer graphics and vision techniques. Such replication of visual content in large volumes of everyday reality should lead to more naturalistic teleconferencing work environments (and less business travel), greater fidelity in relaying news and entertainment events (high-definition will seem positively low-resolution), and even Star Trek Holodeck-like tele-presence in remote locales ("beam me up, Jaron") [11].

3D Real World Object Modeling

Obtaining computer models of real world objects is a very active research area with applications in Virtual Reality (VR) and multimedia systems. A common approach to obtain photorealistic object descriptions uses multiple camera views from different positions around the object and attempts to fuse this information into a complete 3D description of the object [12].

2.10.2 REQUIREMENTS AND FUNCTIONALITIES

Major requirements for this application scenario are:

- **Low Cost** - As in the applications depicted above large number of cameras are required, a reduction in the cost of one single device becomes a high amount of money saved.
- **Low Complexity** - Typically, there is high computational cost at the encoder due to the high amount of data size to process. This is a problem in applications where real time is

required and it increases the cost of the encoders (which are in a high number normally). Then, low complexity is considered an improvement and benefit to this scenario.

- **High Number of Cameras** - Most applications based on multiview images required a high amount of cameras that capture different views. Although there are other methods to capture the images that do not required such big number of devices (i.e. camera rotation), typically between 3 and hundreds of sensors are used in image acquisition. These cameras are in most applications from camera arrays.
- **Overlapping** - In such scenario, neighboring cameras of large camera arrays capture overlapped views and therefore views that are (very) correlated.
- **Real Time System** - Although most of the applications of modeling, special effects, rendering, medical diagnosis, etc do not need real time responses, there are others such as surveillance, videoconferencing, etc that do require real time processing. As this scenario manages high amount of images and data, a real-time compression and decompression are required.
- **Camera Parameters and *a Priory* Knowledge of Geometric Information** - Knowledge of the camera parameters and different object boundaries at the encoders is a common necessity in some multiview applications. For example, some systems estimate the correlation structure in the visual information using some geometrical information. This means that complex algorithms must be used (such as shape adaptations techniques) which increases the computational cost at the encoder. These techniques are not directly applicable to encoding of real multiview images.
- **Robustness to Occlusions** - Such systems must be resilient to missing information when capturing images, especially to occlusions of the view that will not let the system to perform a perfect reconstruction and interpolation at the receiver. Depending on the application, this could become a major requirement especially when there is a low number of views which increases the probability of occlusions occurring. As many views and cameras are involved in most of the applications, this requirement is not as important and crucial as it can be in other scenarios. In the camera sensor network scenario, it is clear that the visual occlusions present a challenging problem for any distributed coding technique. In order to reconstruct the position of an object for any virtual camera position, the correct position in at least two different views needs to be known. Nevertheless, using a configuration with more cameras will make it more likely for any object to be visible in at least two views.
- **Fine Control, Compensation and Calibration** - Some applications, such as high speed image capturing, require a combination of fine control over the cameras and compensation for varying geometric and radiometric properties characteristic of image sensors. In other cases, cameras are mobile. When the cameras start to move, their external calibration parameters need to be calculated in real-time. Camera calibration is also performed by the capturing process.
- **Communication vs. Power Consumption** - Camera sensor networks present a significant trade-off between powers consumed by processing versus communication. Communication power costs can vastly exceed today's power efficient processor demands. As a result, in general, developers then strive to process information locally to reduce the data transmitted. The distributed architecture is designed to help the sensor network capitalize on the collective behavior of these complex systems by increasing communication load only when doing so is optimal.

- **Compression Efficiency** - Due to the enormous size of multiview images in most applications, compression is one of the major requirements. Multiview images are usually highly correlated in spatial domain and therefore distributed coding solutions fit perfectly to this scenario. A drastic reduction of the amount of data transmitted will be the major goal and contribution of DVC to this scenario.
- **Graceful Degradation** - Depending on the application focused, the loss of the minimum image quality can be acceptable or not. For 3D object modeling or medical purposes, a minimum loss of data in an image can have drastic effects on reconstruction and diagnosis. For other applications, this is not as crucial. Nevertheless, graceful degradation is a requirement for the encoder.
- **Encoder-Decoder Complexity Trade-Off** - As mentioned previously, one of the major challenges of such applications is the cost of managing enormous data sizes as hundreds of images have to be processed simultaneously. Depending on the application selected and its requirements, the complexity can be balanced to the decoder with the benefits that this implies.
- **Time Constraints** - Depending on the application focus, time constraints become more important. Real time response, synchronization between cameras, speed of capture, …are some of the requirements. The flexibility of balancing complexity from the encoder to the decoder can help to fulfill these restrictions and give better response to current necessities.
- **Distance between Cameras** - If the cameras are packed close together, then the system effectively functions as a single-center-of-projection synthetic camera, which can be configured to provide unprecedented performance along one or more imaging dimensions, such as resolution, signal-to-noise ratio, dynamic range, depth of field, frame rate, or spectral sensitivity. If the cameras are placed farther apart, then the system functions as a multiple-center-of-projection camera, and the data it captures is called a light field. Of particular interest are novel methods for estimating 3D scene geometry from the dense imagery captured by the array, and novel ways to construct multi-perspective panoramas from light fields, whether captured by this array or not. Finally, if the cameras are placed at an intermediate spacing, then the system functions as a single camera with a large synthetic aperture, which allows seeing through partially occluding environments like foliage or crowds. If the array of cameras is augmented with an array of video projectors, a discrete approximation of co-focal microscopy can be implemented, in which objects not lying on a selected plane become both blurry and dark, effectively disappearing.

2.10.3 FORESEEN DVC BENEFITS

According to the requirements and functionalities of this scenario, the potential benefits of applying DVC are enumerated in the following list:

1. **Low Complexity** - The advent of inexpensive digital image sensors has generated great interest in building sensing systems that incorporate large numbers of cameras. In DVC systems, independent encoding of each view can be performed in each camera while a central station has to perform joint decoding, in order to exploit the correlation between views. This will enable to have low complexity encoders and thus to use low-cost cameras, minimizing the total cost of the camera array.
2. **Exploitation of Views Correlation -** Distributed compression schemes take advantage of the redundant information between the views in order to reduce the necessary overall transmission bitrate from the sensors to a central receiver, without requiring any inter-sensor

communication. Flexible distribution of the transmission bitrates amongst the encoders is achieved and this is optimal for many applications.

3. **Compression Efficiency** - Due to the help of side information at the decoder, the Wyner-Ziv coder performs significantly better than the major image coding schemes, such as JPEG and JPEG2000, in the low bitrates range for large camera arrays compression [2].
4. **Image Quality** - The relative quality of the reconstructed images reflects a trend towards higher quality imaging. Thus, new compression schemes with reasonable complexity overhead should be developed and DVC is one alternative to consider for decreasing the complexity especially at the encoder side At low bitrates, JPEG2000 tends to blur out image details and incur ringing effects at object boundaries. The SA-DCT coder preserves the object boundary, but introduces blocking artifacts. The pixel-domain Wyner-Ziv coder, on the other hand, benefits from the rendered side information, and could provide sharper details of the images [2].
5. **Power Consumption** - DVC encoders may be less complex, so consume less power and the processing time is shorter. This will help applications with strong time and power constraints. It is shown that Wyner-Ziv coders provide a significant reduction in complexity and a better time performance when compared to JPEG2000 for large camera arrays compression [2].
6. **Low Maintenance Cost** - Low cost of maintenance of the system, as most of the algorithms and complexity is placed in the decoder which is unique.
7. **Flexible Complexity Allocation** - DVC capacity of balancing complexity between encoder and decoder provides flexible solutions to the many different applications that fall into this scenario. Some adaptive distributed source coding methods can approach to the Slepian-Wolf bound by controlling the quality of encoding [13] which is very useful to adapt solutions to some specific use cases requirements.

2.10.4 CURRENT DVC DRAWBACKS

Major drawbacks of using DVC in this scenario are:

1. **Compression Efficiency** - DVC performance in terms of compression efficiency and error robustness are far from the results of current video coding solutions. As it is shown in [2], at higher bitrates, JPEG2000 and the SA-DCT coder tend to be more efficient, than Wyner-Ziv coding in terms of rate PSNR-performance and reconstructed image quality.
2. **Decoding Complexity** - As said above, one of the main characteristics of DVC is the possible shift of complexity from encoder to decoder. In current approaches to DVC, the required decoding complexity seems to be very high and it is difficult to properly evaluate the achievable decoding time performance. In applications requiring real time decoding, this may be a most serious drawback.
3. **Visual Occlusions** - For the camera sensor network scenario, it is clear that the visual occlusions present a challenging problem for any distributed coding technique.

2.10.5 REFERENCES

[1] N. Gehrig and P. L. Dragotti, "On Distributed Compression in Dense Camera Sensor Networks", Picture Coding Symposium, San Francisco, CA, USA, December 2004

[2] X. Zhu, A. Aaron and B. Girod, "Distributed Compression for Large Camera Arrays", Proc. IEEE Workshop on Statistical Signal Processing, SSP-2003, St Louis, Missouri, USA, Sept. 2003

[3] http://amp.ece.cmu.edu/projects/MobileCamArray/

[4] http://www.ieee.org/organizations/pubs/newsletters/leos/feb04/3ddis.html

[5] http://rstr.bu.edu/iss/Research/MMP/mmp.html

[6] http://graphics.stanford.edu/papers/highspeedarray/highspeed.pdf

[7] E. Izquierdo E, S. Kruse, "Image Analysis for 3D Modeling, Rendering, and Virtual View Generation", Computer Vision and Image Understanding, vol. 71 (2), pp. 231-253, 1998

[8] http://www.bris.ac.uk/eeng/research/spr/

[9] http://en.wikipedia.org/wiki/Virtual_camera

[10] http://www.cs.unc.edu/Research/stc/inthenews/pdf/CGW_2001_jan.pdf

[11] http://cgw.pennnet.com/Articles/Article_Display.cfm?&Section=Articles&SubSection=Display&ARTICLE_ID=89049&PUBLICATION_ID=18&VERSION_NUM=1

[12] P. Eisert, E. Steinbach, B. Girod, "Automatic Reconstruction of Stationary 3-D Objects from Multiple Uncalibrated Camera Views", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 10, nº 2, March 2000

2.11 WIRELESS CAPSULE ENDOSCOPY

2.11.1 GENERAL DESCRIPTION

Many diseases of the human body can only be spotted with images of the ill region. With X-ray, the whole body can be photographed. But these images are not very accurate, and not all diseases can be detected by this technique. An example is to determine the source of gastrointestinal bleeding. Intestinal bleeding occurs when an abnormality on the inner lining begins to bleed. Determining the source of gastrointestinal bleeding that originates in the small bowel[5] is one of the major diagnostic challenges facing gastroenterologists. Many small bowel causes of blood loss go undetected because the small bowel is long, hard to reach and therefore difficult to evaluate. X-ray studies may be unable to pinpoint exact locations of abnormalities. Thus if masses or bleeding lesions are found, their accurate location is difficult to specify to the surgeon for removal. The best way to find most of the causes of small bowel bleeding is to look directly at the small bowel with an endoscope[6]. Most abnormalities that cause small bowel bleeding lie within the reach of either a standard endoscope, or a much longer endoscope called an enteroscope that can reach further into the small bowel. Since the small bowel is more then 5 meters long, which is much longer than any of the instruments currently available, the capsule endoscopy has emerged as an effective way to evaluate the small bowel for bleeding [1].

The endoscopic capsule has the size of a large pill and contains a battery, a strong light source, a camera and a small transmitter, see Figure 16. Once swallowed, the capsule begins to transmit images of the inside of the esophagus, stomach and small bowel to a receiver. The pictures of the capsule passing through the intestine can be analyzed for abnormalities which are possible reasons for bleeding.

---

[5] The area of the intestine between the stomach and the colon.

[6] An endoscope is a tube instrument with a light and camera at one end passed through the mouth.

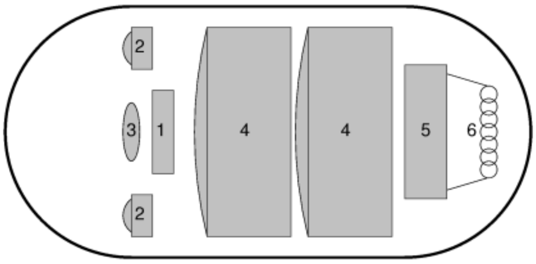

Figure 16 - Wireless capsule endoscope: 1 - CMOS imager; 2 - LEDs; 3 - lens; 4 - batteries; 5 - transmitter; 6 – antenna [2].

The capsule endoscopy is a breakthrough in gastrointestinal diagnostics that overcomes the shortcomings of traditional endoscopy and radiological imaging by offering the following benefits:

- High quality color images of the entire small intestine;
- A non-invasive, patient-friendly procedure that is easily ingested by the patient and does not require sedation;
- No insufflation (forcing of air into the gastrointestinal tract) is required;
- Convenient digital reporting, storage and remote consulting capabilities;
- Cost-effective diagnostic tool for health care providers, because it helps to avoid unnecessary and painful diagnostic tests and can be performed in an outpatient setting environment.

2.11.2 REQUIREMENTS AND FUNCTIONALITIES

The major requirements regarding this application are:

- **Low Complexity** - The complexity of the encoder must be very low to produce this very small device with low power.
- **Small Size** - In case of wireless micro camera and additionally keeping in mind that the patient has to swallow it, the (small) size is of fundamental importance.
- **Coding Efficiency** – Wireless video cameras must use video codecs providing the highest possible quality for a given bitrate and at given constraints like low complexity and lightweight requirements.
- **No Back-Channel** -A back-channel should be avoided, because the needed receiver in the encoder makes the capsule unnecessary complex. A solution without back-channel keeps the capsule smaller thus enabling other useful features. The patient can carry a light-weight receiver, whose sole task is to store the video data. Thus the patient does not have to stay at the hospital while the gastrointestinal examination takes place.

2.11.3 FORESEEN DVC BENEFITS

The main DVC benefits for this application scenario can be summarized as follows:

1. **Low Complexity** - The DVC lower power consumption and lower computational complexity make smaller devices possible.
2. **Small Size** - As the complexity of the encoder is supposed to be reduced with DVC approach, the size and the weight of the wireless cameras will be reduced, thus making them more applicable for this scenario.

3. **Error Resilience** - Current solution uses an analogue representation of the raw video data. There is no way to protect the video against channel errors. Applying conventional digital video codecs would only conditionally solve the problem. Thus, if channel capacity is exceeded, the decoder collapses very fast. With a solution based on Wyner-Ziv coding, the decoder is able to reconstruct also distorted data, while keeping low-complexity encoding processing.

2.11.4 CURRENT DVC DRAWBACKS

The current DVC drawbacks for this application scenario are:

1. **Decoding Complexity –** In this application, the decoding complexity remains the bottleneck. However, this drawback is not critical, since real-time decoding is not required. Furthermore, availability of powerful decoders in hospitals might be assumed.
2. **Compression Efficiency -** Another possible drawback of DVC is the limited compression efficiency when compared to H.264/AVC. But also this drawback may be of secondary importance if a good trade off between compression and encoder simplicity is obtainable with DVC.

2.11.5 REFERENCES

[1] M. Tilson, and J. R. Saltzman; "Small Bowel Bleeding and Capsule Endoscopy", Harvard Medical School, Boston, Massachusetts, http://www.acg.gi.org/patients/gihealth/smallbowel.asp

[2] F. Hagemüller, M. P. Manns, H. G. Musmann, J. F. Riemann, "Medical Imaging in Gastroenterology and Hepatology", Series Falk Symposium, vol.124

3. CLUSTERING OF APPLICATION SCENARIOS

Following the detailed analysis of application scenarios and functionalities performed in the previous section, it is the purpose now to cluster the applications according to relevant criteria and characteristics. Relevant clustering criteria and characteristics to be used are:

Number of Camera Views

Depending on the number of cameras involved in the application scenario, the system can be classified into:

- **Single View System -** One camera captures the single view available.
- **Multiple View System -** Two or more cameras capture the views available.

Real Time System

Depending on the required response time, the system can be classified into:

- **Real-Time Response System -** Systems where the global delay between video capture at the sender and visualization at the receptor is within an acceptable limit to be considered real-time.
- **Off-line Processing System -** Systems where no limit delay exists between video capture and visualization.

It may also happen that for some applications only real-time encoding or real-time decoding is necessary and not both together.

In conventional coding solutions, encoders are typically more complex than decoders which means that real-time performance is more critical at encoders. This is the opposite for DVC solutions where decoders are typically more complex than encoders. This change and the associated flexible allocation of complexity may represent a benefit for DVC codecs.

Availability of a Return Channel

Depending on the availability or not of a return channel, the system can be classified into:

- **Unidirectional System –** No return channel is available.
- **Bidirectional System –** A return channel is available and thus it may be exploited including for compression efficiency purposes.

Compression Efficiency

Depending on the required compression efficiency, the system can be classified into:

- **Compression Efficiency Critical -** Systems where the compression efficiency achieved for the encoding of the video data is critical for the success of the application.
- **Not Compression Efficiency Critical -** Systems where the compression efficiency achieved for the encoding of the video data is important but not critical for the success of the application.

It is obvious that compression efficiency must always be as high as possible. However while some applications are not prepared to trade off compression efficiency with other benefits, e.g. error resilience, lower complexity, other applications are willing to do so.

Error Resilience

Depending on the required error resilience requirements, the system can be classified into:

- **Error Resilience Critical -** Systems where high error robustness is critical for the success of the application.
- **Not Error Resilience Critical -** Systems where the high error robustness is not important or important but not critical for the success of the application.

Power Limitations

Depending on the importance of the low-power constraint, the system can be classified into:

- **Not Limited Power Supply System -** Application scenarios where the cameras are connected to a non limited power supply and thus power limitations are not critical.
- **Limited Power Supply System -** Application scenarios where the cameras are connected to a limited power supply and thus power limitations are critical.

The low-power constraint is one of the most important related to DVC. Therefore the application scenarios where limitations on power supply exist for the capturing, processing and transmission of the images are more relevant for the purposes of the DISCOVER project.

Encoding Complexity

Depending on the criticality of the encoding complexity, the system can be classified into:

- **Encoding Complexity Critical** – Low encoding complexity is a critical requirement for the success of the application.
- **Not Encoding Complexity Critical** – Low encoding complexity is an important requirement but not critical for the success of the application and thus trade offs with other functionalities, e.g. lower compression efficiency, are possible.

Complexity Allocation Flexibility

Depending on the criticality of the complexity allocation flexibility, the system can be classified into:

- **Complexity Allocation Flexibility Critical** – A high flexibility in terms of codec complexity allocation is a critical requirement for the success of the application.
- **Not Complexity Allocation Flexibility Critical** – Flexible allocation of codec complexity is not an issue or at least not an important issue for the success of the application.

Camera Mobility

Depending on the mobility of the cameras, the system can be classified into:

- **Mobile Camera System:** The camera(s) that capture(s) the images are mobile (freely or conditioned, e.g. according to some guides, or inside some vehicle).
- **Fixed Camera System:** The camera(s) that capture(s) the images are fixed although may still be controlled in terms of rotations and zooming.

If the cameras are mobile, it is very likely that the low-power constraint is important as, most likely, they are not connected to any constant energy source (although in vehicles they could be provided with more power through rechargeable batteries). In these cases, the low-power consumption of the DVC encoders could bring major benefits to the application scenario. However some application scenarios with fixed cameras imply the same limitations, i.e. forest surveillance, where cameras are placed in trees.

Size and Weight of Acquisition Devices

- **Size and Weight Critical -** The size and weight of acquisition devices may be critical as for example in endoscopic capsule, or in particular cases of surveillance applications.
- **Not Size and Weight Critical -** The size and weight of acquisition devices may not be critical as for example in multiview systems for image synthesis and rendering applications

Privacy

Depending on how critical is the control of the information to be transmitted in terms of privacy, the system can be classified into:

- **Privacy Critical System -** Application scenarios where the information privacy is critical and thus must be guaranteed by adequate means, e.g. DRM solutions.
- **Non Privacy Critical System -** Application scenarios where the information privacy is not critical and thus no specific means are needed for this purpose.

In principle, privacy issues should not have a major impact on coding solutions.

Degree of Innovation

Depending on the degree of deployment of the application scenario and the number of current solutions available in the market, the system can be classified into:

- **High Innovative System -** Systems with low or no deployment in the market and with no or few solutions currently available.
- **Deployed System -** Systems with many solutions and with high penetration in the market.

While largely deployed application scenarios provide more information for comparison purpose with new DVC based solutions, highly innovative systems may have a higher impact and imply new benefits for the society.

In Table 2, the application scenarios presented in Section 2 are classified in terms of the characteristics listed above. For some cases, it may happen that an application scenario appears on both columns for a certain characteristic if there two relevant 'flavours' of that scenario in terms of that characteristic.

Single-View System Disposable video cameras Wireless video cameras Distributed video streaming Visual sensor networks Mobile document scanner Video conferencing with mobile devices Mobile video mail Wireless capsule endoscopy	**Multiview System** Networked camcorders Distributed video streaming Visual sensor networks Multiview image acquisition Wireless low-power surveillance
Real-Time Systems Disposable video cameras Networked camcorders Wireless video cameras Distributed video streaming Visual sensor networks Multiview image acquisition Wireless low-power surveillance Video conferencing with mobile devices	**Non-Real-Time Systems** Mobile document scanner Wireless video cameras Mobile video mail Wireless capsule endoscopy
Unidirectional Communication Disposable video cameras Wireless video cameras Mobile video mail Wireless capsule endoscopy Visual sensor networks Multiview image acquisition	**Bidirectional Communication** Networked camcorders Distributed video streaming Visual sensor networks Multiview image acquisition Wireless low-power surveillance Mobile document scanner Video conferencing with mobile devices
Compression Efficiency Critical Disposable video cameras	**Non-Compression Efficiency Critical** Multiview image acquisition

Networked camcorders Wireless video cameras Distributed video streaming Visual sensor networks Wireless low-power surveillance Video conferencing with mobile devices	Mobile document scanner Mobile video mail Wireless capsule endoscopy
Error Resilience Critical Networked camcorders Wireless video cameras Distributed video streaming Visual sensor networks Mobile document scanner Wireless low-power surveillance Wireless capsule endoscopy Video conferencing with mobile devices	**Non-Error Resilience Critical** Disposable video cameras Multiview image acquisition Mobile video mail
Low-Power Constraint Disposable video cameras Networked camcorders Wireless video cameras Visual sensor networks Mobile document scanner Wireless low-power surveillance Mobile document scanner Wireless capsule endoscopy Video conferencing with mobile devices	**No Low-Power Constraint** Distributed video streaming Multiview image acquisition
Encoding Complexity Critical Disposable video cameras Networked camcorders Wireless video cameras Distributed video streaming Mobile document scanner Multiview image acquisition Wireless low-power surveillance Mobile document scanner Wireless capsule endoscopy Video conferencing with mobile devices Mobile video mail	**Not Encoding Complexity Critical**
Flexible Allocation of Codec Complexity Critical Disposable video cameras Networked camcorders Wireless video cameras Distributed video streaming Visual sensor networks Mobile document scanner Wireless low-power surveillance	**Not Flexible Allocation of Codec Complexity Critical** Multiview image acquisition

Mobile document scanner Wireless capsule endoscopy Video conferencing with mobile devices	
Size and Weight Critical Disposable video cameras Networked camcorders Wireless low-power surveillance Wireless video cameras Mobile document scanner Video conferencing with mobile devices Wireless capsule endoscopy Mobile video mail Visual sensor networks Multiview image acquisition	**Not Size and Weight Critical** Multiview image acquisition Visual sensor networks Distributed video streaming
Mobile Camera Disposable video cameras Wireless video cameras Visual sensor networks Wireless low-power surveillance Mobile document scanner Wireless capsule endoscopy Video conferencing with mobile devices Mobile video mail	**Fixed Camera** Networked camcorders Wireless video cameras Distributed video streaming Visual sensor networks Multiview image acquisition
Privacy Critical Wireless video cameras Visual sensor networks Wireless low-power surveillance	**Non-Privacy Critical** Disposable video cameras Networked camcorders Wireless video cameras Distributed video streaming Mobile document scanner Multiview image acquisition Mobile document scanner Wireless capsule endoscopy Video conferencing with mobile devices
High Innovation Application Disposable video cameras Networked camcorders Distributed video streaming Visual sensor networks Multiview image acquisition Wireless low-power surveillance Wireless capsule endoscopy	**Low Innovation Application** Wireless video cameras Video conferencing with mobile devices Mobile document scanner Mobile video mail

Table 2 – Characterization of application scenarios.

Based on the characterization of application scenarios presented in Table 2, it is possible to identify some more relevant clusters, notably:

1. Applications using a single view, real-time, bidirectional, compression efficiency and error resilience critical, low-power constraints, encoding complexity critical, flexible allocation of complexity relevant, with mobile or fixed cameras
2. Same as cluster 1 but now with multiple video views
3. Same as cluster 1 but now with unidirectional communication
4. Same as cluster 2 but now with unidirectional communication

4. PROMISING DVC APPLICATION SCENARIOS

Following the detailed description and characterization of application scenarios made in previous sections, this section intends to select the most promising application scenarios for DVC.

4.1 METHODOLOGY FOR SELECTION OF MOST PROMISING DVC APPLICATION SCENARIOS

The methodology proposed here for the selection of the most promising DVC application scenarios is the following:

1. Identification of the application requirements for which DVC may bring significant benefits; these requirements are called in the following the 'DVC golden requirements'.
2. Following the application scenarios characterization made in Table 2, count for each application the number of requirements matching DVC golden requirements.
3. Select as promising the application scenarios with the highest counting in 2. while maximizing at the same time the coverage of the clusters identified in Section 3.

This methodology targets the study of a practical number of significant but also reasonably different application scenarios.

4.2 LIST OF MOST PROMISING DVC APPLICATIONS SCENARIOS

Applying the methodology proposed in Section 4.1, the following outcomes result:

1. According to the literature and the accumulated experience within DISCOVER, the major application requirements where DVC may bring significant benefits, the so-called DVC golden requirements, are:
    a. **Flexible allocation of the global video codec complexity**; this may include low complexity encoding and low battery requirements
    b. **Error robustness** related to the absence of prediction loop
    c. **Exploitation of correlation in multiview** video scenarios
    d. **Codec independent scalability** related to the absence of prediction loop between scalable layers[7]
2. For the application scenarios presented in this document, the counting of the match between application requirements and DVC golden requirements is presented in Table 3.

---

[7] Each layer may use a different, unknown codec since the deterministic knowledge of each layer, spatial or SNR, does need to exist to improve it with a DVC approach.

Application Scenario	Count of Matching between Application and DVC Friendly Requirements
Wireless Video Cameras	2 (a, b)
Wireless Low-Power Surveillance	3 (a, b, c)
Mobile Document Scanner	2 (a, b)
Video Conferencing with Mobile Devices	2 (a, b)
Mobile Video Mail	1 (a)
Disposable Video Cameras	1 (a)
Visual Sensor Networks	3 (a,b,c)
Networked Camcorders	2 (a, b)
Distributed Video Streaming	3 (a, b, d)
Multiview Image Acquisition	1 (c)
Wireless Capsule Endoscopy	1 (a)

Table 3 - Count of Matching between Application and DVC Friendly Requirements

3. Based on the counting included in Table 3, the market relevance and the target to maximize at the same time the coverage of the clusters defined in Section 3, the four application scenarios selected as the most promising ones are:
   a. Video conferencing with mobile devices (monoview scenario)
   b. Wireless low-power surveillance (multiview scenario)
   c. Wireless video cameras (monoview scenario)
   d. Visual sensor networks (multiview scenario)

This selection of application scenarios provides a good balance between monoview and multiview applications as well as between applications with high innovation and applications more established in the market.

5. SUMMARY AND CONCLUSIONS

This document presents in detail application scenarios for which DVC technology may bring major benefits, highlighting the benefits and limitations for each case studied. This document also identifies and studies which functionalities are relevant for each application scenario, notably coding efficiency, error resilience, and encoder-decoder complexity trade-off.

Based on their detailed description, the application scenarios were after classified according to the major features, and clustered in 4 groups of applications.

Finally, this document identifies the 4 most promising application scenarios in terms of exploitation of the DVC technology.