Advances in video networking: standards and applications

Christos Grecos, Qi Wang

Audio-Visual Communications and Networks (AVCN) Research Group

School of Computing

University of the West of Scotland

Paisley PA1 2BE, UK

{Christos.Grecos, Qi.Wang}@uws.ac.uk

Abstract

Purpose – The interdisciplinary nature of video networking, coupled with various recent developments in standards, proposals and applications, poses great challenges to the research and industrial communities working in this area. The main purpose of this paper is to provide a tutorial and survey on recent advances in video networking from an integrated perspective of both video signal processing and networking.

Design/methodology/approach – Detailed technical descriptions and insightful analysis are presented for recent and emerging video coding standards, in particular the H.264 family. The applications of selected video coding standards in emerging wireless networks are then introduced with an emphasis on scalable video streaming in multihomed mobile networks. Both research challenges and potential solutions are discussed along the description, and numerical results through simulations or experiments are provided to reveal the performances of selected coding standards and networking algorithms.

Findings – The tutorial helps to clarify the similarities and differences among the considered standards and networking applications. A number of research trends and challenges are identified, and selected promising solutions are discussed. This practice would provoke further thoughts on the development of this area and open up more research and application opportunities.

Research limitations/implications – Not all the concerned video coding standards are complemented with thorough studies of networking application scenarios.

Practical implications – The discussed video coding standards are either playing or going to play indispensable roles in the video industry; the introduced networking scenarios bring together these standards and various emerging wireless networking paradigms towards innovative application scenarios.

Originality/value – The comprehensive overview and critiques on existing standards and application approaches offer a valuable reference for researchers and system developers in related research and industrial communities.

Keywords Video networking, video coding standards, H.264, SVC, mobile networks, wireless sensor networks
Paper type Literature review

1. Introduction

Recent years have witnessed tremendous advances in both visual signal processing and networking technologies, which are paving the way for a broader range of high-quality video applications and continued commercial success in the consumer and industry electronic markets. Due to the interdisciplinary nature of the video networking area, we believe that a unified vision integrating both signal processing and networking perspectives would greatly facilitate further progress and open up more research opportunities.

In the visual signal processing domain, the H.26X family is the mainstream video coding standards. To be examined are the Advanced Video Coding (H.264/AVC), the Scalable Video Coding (H.264/SVC), the Multiview Video Coding (H.264/MVC), and the emerging H.265 standards. H.264/AVC is the next-generation video compression standard, which compared with the MPGE-2 one, offers the same video quality with significantly lower bandwidth requirements for standard-definition and high-definition TV, video conferencing, and many others in wired or wireless networks. H.264/SVC is the scalable extension to H.264/AVC and allows scalable video transmission in temporal, spatial and quality dimensions for graceful downgrading or upgrading of the video experience according to the network conditions and user equipment capabilities. H.264/MVC is a core enabling technology to the emerging three-dimensional (3D) video applications such as 3D TV, free-viewpoint video and immersive teleconferencing. Multiple views are generated from an array of cameras and H.264/MVC was designed to efficiently encode and decode the simultaneous video sequences. In addition, we will discuss the Distributed Video Coding (DVC), which enables efficient encoding with low-complexity encoders in contrast to the conventional approach.

In the network domain, we concentrate on emerging wireless networks since the resource-constrained wireless networks pose serious challenges to satisfactory transmission of the resource-demanding and real-time video applications. The concerned wireless networking paradigms include multihomed mobile networks and wireless sensor networks. A multihomed mobile network is a moving network that consists of networked mobile nodes and is able to access multiple networks simultaneously with the corresponding network interfaces. A wireless sensor network comprises a number of power-limited sensors distributed in a field of interest and at least one sink that collects the sensed data from the sensors. Typically, all the nodes are static yet wirelessly connected. We discuss the technical challenges to deliver video applications in these networks given their particular characteristics and present promising approaches and solutions.

The reminder of the paper will first describe and analyse the video coding standards from Section 2 to Section 6, and then will present a detailed case study of H.264/SVC streaming in multihomed mobile networks in Section 7 and outline distributed video transmissions in wireless sensor networks in Sections 8. Section 9 will conclude the paper.

2. The H.264/AVC standard

In a video codec, a “profile” is a set of coding tools or algorithms that can be used to produce a conformant bit stream. A “level” is a set of constraints on parameters of the bit stream. One profile can have many levels. In H.264/AVC (Wiegand *et al.*, 2003), the simplest profiles (Simple, Main) (4:2:0), have twice as many luminance than chrominance samples. Extended profiles (4:4:4) on the other hand are used in digital TV/cinema. The default colour space is Y:Cb:Cr. Coding can be performed either in frame mode, by combining the top/bottom fields together and coding them as a single frame (frame mode) or by coding them separately (field mode). In field coding, the frame is split into macroblock tiles (16*32 pixel areas) and each tile is encoded as frame or field mode. In the case of field mode, the top macroblock (16*16 area) consists of the even pixel lines of the macroblock tile, while the bottom macroblock of the odd ones. The definition of a slice in AVC is in terms of connected macroblocks. Any non-connected macroblocks have to be restructured to connected formations, to be encoded in the same slices. There are 4 slice types in AVC: I slices (all intra predicted macroblocks), P slices which are a mixture of intra + inter predicted macroblocks with at most 1 prediction signal per macroblock) and B slices which are a mixture of intra + inter predicted macroblocks with possibly 1 or 2 or no prediction signals (direct mode) per macroblock. The fourth slice type, SP/SI slices are used for fast forward, playback or encoder/decoder synchronisation purposes.

The operation of an AVC codec is summarised in Figure 1. As can be seen, it is similar to the older MPEG-2 codecs. However, since AVC can achieve the same quality with MPEG-2 in approximately half the bit rate, an identification of the features of the standard that enable this is required. Such features include variable block size and quarter sample accurate motion compensation (MC), picture boundary extrapolation and associated predicted motion vectors (MV), selection of multiple reference slices for effective MC, decoupling of reference from display order for latency reduction in bi-predictive coding, application of in-loop deblocking filters for better prediction and visual quality, weighting and offsetting of prediction signals for better compression and performance on fades, the use of B pictures as predictors, improved skip and direct mode inference, directional spatial prediction for intra coding, availability of small transform size in order to better represent the signal locally, availability of hierarchical transforms in order to tailor compression efficiency to signal properties, the use of 16 bit word lengths and exact match integer transforms, advanced CABAC and CAVLC coding techniques, advanced data partitioning techniques and use of flexible slice sizes and the use of flexible macroblock and arbitrary slice ordering for increasing error resiliency and reducing network delay.

"Take in Figure 1" – **Figure 1.** AVC architecture (Wiegand *et al.*, 2003)

A major advantage of AVC over MPEG-2 is the ability to choose the optimal block size in the rate distortion sense for temporal and spatial predictions. This is called mode decision. The set of inter modes (i.e. modes determined after ME/MC) contains block sizes from 16*16 to 4*4 pixels. The set of intra modes (i.e. modes determined after spatial predictions) consists of 16*16 and 4*4 mode sets only. The 4*4 set of intra modes uses residual coding of the pixel areas based on neighbouring pixels, thus defining specific edge directions. There are nine such directional modes in the 4*4 set. The 16*16 set of intra modes consists of the horizontal, vertical, DC and plane modes. The 16*16 intra modes are used to code relatively large areas with not many details, as opposed to the 4*4 intra ones. The above inter and intra modes are used in both P and B slices (only intra modes in I slices) with the addition of a special skip mode for P slices. As the name suggests, in the skip mode no errors or MVs are sent to the decoder.

In AVC, 4*4 and 8*8 transforms are applied to all block sizes and for both luminance and chrominance components. The 16 DC coefficients of all 4*4 blocks inside a 16x16 macroblock pass through a second level of transforms using 4x4 Hadamard matrices. This is applicable in 4:4:4 resolution videos, while for 4:2:0 resolutions the 2x2 Hadamard matrices are used for the DC coefficients. The advantage of using a smaller 4*4 transform is that it is essentially as good as an 8*8 transform in compression terms because of the improved ME/MC prediction in the standard. At the same time, for very similar compression to the 8*8 transforms, “mosquito noise” and “ringing artefacts” are reduced. Furthermore, smaller transforms imply less computations (adds and shifts).

The quantization in AVC is performed using perceptual-based quantization scaling matrices. The encoder can specify, for each transform the block size separately for intra and inter prediction.. This allows tuning of the quantization fidelity according to a model of the sensitivity of the human visual system to different types of error. It typically does not improve objective fidelity as measured by mean-squared error or PSNR, but it does improve subjective fidelity, which is really the more important criterion. Default values for the quantization scaling matrices are specified in the standard but the encoder can choose to use customized values instead by sending a representation of those values at the sequence or picture level. The quantization process ends with scaling the coefficients according to appropriate offsets.

The scanning order of the coefficients is different for frame and field modes and is designed to scan the highest-variance coefficients first and to maximize the number of consecutive zero-valued coefficients appearing in the scan. The Context Adaptive Variable Length Coding mechanism (CAVLC) switches between VLC code tables for improved performance according to conditioned statistics. It is of low implementation complexity (shifts and table look-ups). The Context Adaptive Binary Arithmetic Coding (CABAC) shows improvements over CAVLC, achieving a 5-15% reduction in bit rate for interlaced TV signals. It uses a non-integer number of bits and its very beneficial for symbols with probabilities greater than 0.5. The adaptive coding (both CABAC and CAVLC) permits good adaptation to non- stationary symbol statistics. Previously encoded contexts of symbols are used to estimate conditional probabilities and in turn switch between a variety of probability models.

Blocking artefacts are a trademark of compression methods. In order to remedy this problem, AVC uses quantisation dependent thresholds (QDT) in de-blocking decisions. The basic principle in the de-blocking filter used is that if the distance between samples near block edges is large but bounded from a QDT, then it is likely that an artefact has occurred and smoothing is required. If the distance is greater than QDT though, it probably reflects natural content and no smoothing is performed. Such type of filtering typically achieves 5-10% compression savings for the same objective quality.

The representative application areas for AVC can be categorized according to profiles. In the baseline profile, this standard can be used for conversational services below 1Mbps with low latency, for H320 conversational video services for circuit-switched ISDN-based video conferencing, for 3GPP conversational H324/M services, for H323 conversational services over Internet with best effort IP/RTP, for entertainment video applications between 1-8+ Mbps with moderate latency (0.5-2 sec) and for H222/MPEG-2 systems. In the main profile, this standard can be used for broadcasting via satellite, cable, terrestrial or DSL, for DVD of standard and High Definition video and for VOD on various channels. In the baseline and extended profiles, this standard can be used for streaming services with typically 50-1500 Kbps rates and greater than 2 sec delay and for 3GPP wireless and internet streaming.

3. The H.264/SVC Standard

The general operation of H.264/SVC (Schwarz *et al.*, 2007) can be shown in Figure 2. In this figure, it can be seen that the standard provides three types of scalability, namely temporal (an assortment of fps in Hz), spatial (CIF, SD, HD resolutions etc.) and quality scalability (for low, medium and high quality). The temporal scalability is achieved using hierarchical prediction structures and was a straightforward extension of the AVC picture reference management scheme. The flexibility of this scheme results in improved coding performance and many useful configurations of varying delays, picture buffer capacities etc. A similar approach as in past standards was followed for the spatial scalability dimension. In particular, SVC uses multi-layered coding for each spatial resolution, which can be performed through ME/MC temporally or through inter-layer prediction. The choices for inter-layer prediction include prediction of intra macroblocks through upsampling of intra macroblocks in the reference layer, inferring enhancement layers macroblock information from corresponding marcoblocks in the reference layer and residual prediction through up-sampling the residual from the reference layer. To reduce the decoding latency and picture (slice) buffer requirements, the concept of single loop decoding was also applied in SVC, where the inter layer prediction is constrained and only pictures from the target layer require storing.

“Take in Figure 2” – **Figure 2.** SVC architecture

The methods used in SVC for quality (SNR) scalability can be classified as Coarse Grain Scalability (CGS) and Medium Grain Scalability (MGS). In CGS, a discrete set of rate points is identified for each layer. Conceptually, CGS is very similar to the spatial scalability without up-sampling and inter layer prediction among CGS layers is performed. A texture refinement procedure is also available in CGS, where the residuals of different layers can be requantised by using a finer quantisation step with increasing layer numbers. In this manner, quality is added progressively to the decoded slices. In terms of the rate distortion (RD) performance, there is an optimal number of

CGS layers per slice. If this number of layers is exceeded, RD performance will be degraded. In MGS, a successive refinement of quality is performed within layers through reprocessing of the transformed coefficients. This feature allows graceful degradation of quality within layers. Switching between layers is also possible in any access unit in MGS. Drift may occur due to lost packets when predicting from the base layer in non protected channels In the case of prediction from the enhancements layer(s), the coding efficiency will increase. The concept of key picture is also important in MGS in order to signify if the base or the enhancement layers were used for prediction. Comparing the coding performance of CGS and MGS reveals that MGS outperforms CGS in higher rates. In terms of single versus multiple loop decoding, it is also shown that multiple loop decoding does not provide significant advantages over single loop decoding and this is good news for chip designers (reduces the need for adding more gates). Some coding results for SVC are shown in Figures 3-5. The simulation results for the sequences “City” and “Crew” with spatial scalability from CIF (352x288) to 4CIF (704x576) and a frame rate of 30 Hz are depicted in Figures 3 and 4. For both sequences, results for a GOP size of 16 pictures (providing five temporal layers) are presented while for “Crew,” also a result for IPPP coding (GOP size of one picture) is depicted in Figure 5.

“Take in Figure 3” – **Figure 3.** Simulation results (City with GOP 16) (Schwarz *et al.*, 2007)

“Take in Figure 4” – **Figure 4.** Simulation results (Crew with GOP 16) (Schwarz *et al.*, 2007)

“Take in Figure 5” – **Figure 5.** Simulation results (Crew with IPPP, GOP1) (Schwarz *et al.*, 2007)

For all cases, all inter-layer prediction (ILP) tools, given as intra (I), motion (M), and residual (R) prediction, improve the coding efficiency in comparison to simulcast. However, the effectiveness of a tool or a combination of tools strongly depends on the sequence characteristics and the prediction structure. While the result for the sequence “Crew” and a GOP size of 16 pictures is very close to that form single-layer coding, some losses are visible for “City,” which is the worst performing sequence in our test set. Moreover, as illustrated for “Crew,” the overall performance of SVC compared to single-layer coding reduces when moving from a GOP size of 16 pictures to IPPP coding. It is worth noting that the rate-distortion performance for multi-loop decoding using only inter-layer intra-prediction (“multiple-loop ILP (I)”) is usually worse than that of the “single-loop ILP (I,M,R)” case, where the latter corresponds to the fully featured SVC design while the former is conceptually comparable to the scalable profiles of the MPEG-2 Video, H.263 or MPEG-4 Visual. However, it should be noted that the hierarchical prediction structures, which not only improve the overall coding efficiency but also the effectiveness of the inter-layer prediction mechanisms, are not supported in these prior video coding standards.

4. The H.264/MVC Standard

The H.264/MVC standard (Ho, 2007) is developed for encoding free viewpoint video and 3D video. Its scope is to code efficiently N views, which of course have substantial redundancy among them. Some challenges that the standard has to cope with is misalignment of cameras as well as colour/illumination mismatches. A general layout of the prediction order in MVC is shown in Figure 6.

"Take in Figure 6" – **Figure 6.** Prediction order in MVC (Smolic *et al.*, 2007)

As shown in the Figure 6, prediction structures exploit inter-camera redundancy. Such structures also trade-off memory, delay, computation and coding efficiency. Influencing factors on prediction are frame rates, inter camera distance (baseline) and content complexity (motion intensity, illumination effects etc.). The structure is fully compatible with AVC and due to inter-slice dependencies, a re-organisation of the input slices to a single stream occurs before encoding. The coding schemes used by the standard include predictive coding which applies disparity compensated prediction and subband coding which relies on adaptive subband decomposition (e.g. disparity compensated lifted wavelets). The standardisation efforts have mainly concentrated on the predictive schemes. MVC produces significant rate savings for the same quality as compared to simulcast (IBBP in AVC) and to hierarchical B frames simulcast.

An assortment of techniques contributes to the improved rate distortion and visual quality performance of MVC. Since there is illumination change between neighbouring views, a macroblock (MB) based illumination adaptive ME/MC scheme between views was added to MVC. The illumination compensation for each macroblock is differentially encoded and the Differential Value of Illumination Compensation (DVIC) for a macroblock is calculated based on values of neighbouring macroblocks. The motion compensated prediction data in P and B slices is also scaled in MVC through explicit scaling (weighting factors are encoded in slice header) and implicit scaling (weighting factors are calculated based on the slice positions in the forward and backward reference lists).The optimal MB mode between two views is found by using the Mean Removed SAD (MR_SAD) as the distortion criterion when calculating the cost function in the case of P_16*16 and B_16*16 inter modes and the normal SAD in the case of P_SKIP,B_SKIP and B_Direct modes. The minimal cost function from the examined modes is then used to choose the best mode in the rate distortion sense. As can be seen, the process is very similar to AVC. The way the standard decides that an MB mode is SKIP in inter-view prediction is also worth noticing. In such a case, the motion vectors of the MB to be encoded should be inferred from motion information of MBs in the already encoded neighbouring view. A Global Disparity Vector (GDV) is used in this process and is measured by the MB size of units between the view pair.

The subjective quality is improved in MVC using a deblocking filter in the MB based illumination compensation. The use of this filter reduces the blocking artefacts and causes no PSNR degradation for a given bit rate. MVC also uses adaptive reference filtering to address the focus mismatch problem across different views. Such a filter results in depth dependent blurriness and sharpness changes, enables the accuracy of object segmentation and is easily implementable with a modest number of operations. MVC also uses view synthesis of neighbouring views to create virtual views that are similar to the one to be encoded. To achieve this virtual view generation, a depth map is created and encoded and prediction is used for view synthesis. In terms of depth map generation, the standard provides options for segment based versus block based depth map generation and for explicit disparity to depth conversion versus depth map estimation by using 3D warping. Evidently, accurate depth map estimation improves the performance of the lossless depth map coding which also preserves temporal relations. The view synthesis prediction is performed by encoding the difference image between the input and synthesised images and by adding the synthesised image as an additional reference image as well as by substituting the synthesised image for the current reference image in order to improve the rate distortion performance.

5. Distributed Video Coding (DVC)

In traditional video coding systems (H.26X, MPEG-X etc.), the encoder is significantly more complex than the decoder. Typically, 60-80% of this complexity is due to motion estimation (ME). This suits applications where video is compressed once/decoded many times or in streaming scenarios (VOD etc.) when the encoder has many resources available. Recent DVC systems have been developed (theoretically) where the decoder is significantly more complex than the encoder and ME is performed at the decoder. This suits applications such as wireless sensor based surveillance, transcoding applications involving mobile video phones as encoders, news transmission in conflict hot spots around the world etc. The DVC paradigm comes with two main advantages compared to state-of-the-art coding standards (H.26X, MPEG-X etc.). Firstly, the general idea of shifting the complexity from the encoder to the decoder results in low computational requirements for the encoder. This in turn could lead to low production cost, low power consumption, and very small encoders. Secondly, DVC comes with inherent error robustness since there is no prediction loop in the encoder. This error robustness is further strengthened from the use of error correcting codes.

The theoretical foundations of DVC are in the Slepian-Wolf and the Wyner-Ziv (WZ) theorems. In 1973, Slepian and Wolf proved that two correlated random sequences generated by repeated independent drawings of a pair of discrete random variables X and Y can be coded as efficiently by two independent coders as by a joint encoder, provided that the resulting bit-streams are jointly decoded, and that an arbitrary small residual error probability is allowed. Wyner and Ziv extended this work to lossy compression and showed that the distance between the rates of a WZ codec and a normal predictive codec is greater or equal to zero. The case of this distance being greater than zero occurs when the encoder does not have access to the side information. The equality case holds when we have Gaussian memoryless sources and a mean squared error distortion metric d. Later, these results were extended to more general cases and was shown that the equality also holds for source sequences X that are the sum of arbitrarily distributed side information Y and independent Gaussian noise N (with X = Y +N) and that the rate loss for sources with general statistics and a mean squared error distortion metric d is less than 0.5 bits per sample. A general DVC architecture is shown in Figure 7.

“Take in Figure 7” – **Figure 7.** DVC architecture (Grecos *et al.*, 2010)

The sequence must be partitioned in a certain way, as to obtain the two correlated sources mentioned in the Slepian-Wolf and WZ theorems. For example, the sequence can be partitioned into I frames and W frames, using the temporal direction only. The I frames are coded independently from other frames in the sequence, for example using intra coding techniques available in AVC. Therefore, each I frame can be coded at a rate R(i) close to its entropy H(i), and the decoder can reconstruct the frame without using already decoded frames as references. On the other hand, W frames are coded by exploiting correlation between the frames. However, this correlation is exploited at the decoder only, i.e. the encoder codes each frame (W or I) independently from other frames. To achieve this, the decoder generates side information Y using one or more previously decoded frames (I or W). Hence, according to the WZ theorem, the WZ encoder can now code W at a rate R(W) close to the conditional entropy H(W|Y) so that it can be reliably decoded by the WZ decoder.

The side information Y generated at the decoder can be regarded as a noisy version of the original W. After all, the goal of the side information generation module is to estimate the original W as accurately as possible. Therefore, the correlation between W and Y characterizes a virtual channel: it is as if W has been sent to the decoder over a noisy communication channel, so that instead of W, the corrupted version Y is received at the decoder. Hence, for reliable communication, this channel can be protected using channel codes. Y can be considered as the systematic part of this code, and since Y is already available at the decoder, only the parity bits need to be sent, which require R(W) = H(W|Y) bits. Obviously, the amount of error correcting bits that need to be sent depends on the amount of noise on the virtual channel, i.e., it depends on the correlation between W and Y. If Y is a fairly accurate approximation of W, only few parity bits need to be sent and vice versa. The major problem with the virtual channel is that the virtual noise statistics need to be known by both encoder and decoder: the WZ-encoder needs to know how many parity bits should be sent to the decoder, and the WZ-decoder needs the conditional distribution P(W|Y) for efficient channel decoding (e.g. Viterbi-like decoding). However, the encoder only has access to W while the decoder can only access Y. To solve this problem, two solutions are frequently used in the literature. In the first solution, Y is estimated at the encoder and this information is used to estimate R(W). Information about the conditional distribution P(W|Y) is then sent to the decoder along with the error correcting bits (Puri and Ramchandran, 2002). The disadvantage of this solution is that there is complexity added to the encoder, which is typically not desired in a DVC context. In the second solution (Aaron, 2004), it is the decoder that estimates P(W|Y) and calculates the number of parity bits that are needed to correct Y reliably. Subsequently, a feedback channel is used to request the amount of parity bits from the encoder. However, the use of a feedback channel is impractical in video storage scenarios, and even in streaming scenarios, the use of a feedback channel should be limited to avoid excessive delays.

6. The Emerging H.265 Standard

A recent call for proposals in High Efficiency Video Coding (HEVC) (JCTVC, 2010) resulted in a flurry of submissions. The goal of this proposal was to develop a new video coding standard that would retain the quality at half the bit rate of the AVC standard.

Varieties of tools from these submissions were found to be useful towards this goal (mainly in their cumulative effects). Some of these tools were increased flexibility in partitions (more options, asymmetric rectangular and geometric motion partitions), edge extrapolation capabilities, novel intra prediction modes (plane, angular, arbitrary directional, edge based, combined prediction), explicit removal process for MV predictors, single pass switched Interpolation Filters with Offsets (SIFO), a variety of interpolation filters for luminance components, adaptive scanning orders for coefficients, mode dependent transforms for intra modes, mode dependent directional block transforms for inter modes and application of rotational block transforms.

Some other advanced tools include transforms for larger block sizes (16*16, 32*32, 64*64), new in-loop mode dependent de-blocking filters (luminance, chrominance, planar mode), parallel Variable length bin 2 Variable length encoding/decoding (V2V), partition based illumination compensation, adaptive MV resolutions and finally novel context modelling schemes for variable length coding.

It is expected that the H.265 standard will be finalised in the next couple of years.

7. Scalable video streaming in multihomed mobile networks

After a video is encoded through an appropriate standard, a series of steps are required to achieve the subsequent end-to-end transmission of the coded video stream over a network. In an IP-based wired or wireless network, the major processes include packetisation, packet scheduling, packet delivery, reception and decoding from the perspective of the data (transport) plane.

In this section, we focus on the H.264/SVC coding standard as the major example to demonstrate its application in multihomed mobile networks, and to investigate the challenges and representative potential solutions. Within a multihomed mobile network, a group of mobile nodes move together as a unit and the whole moving network can access multiple external infrastructure networks in parallel. Figure 8 illustrate this emerging networking paradigm, where a video streamer transmits video packets to a mobile network through two independent wireless Internet Service Providers (ISPs), which may provide heterogeneous radio access technologies such as third-generation and beyond cellular systems e.g. the Long Term Evolution (LTE), WiMAX and Wi-Fi.

"Take in Figure 8" – **Figure 8.** Video streaming to a multihomed mobile network

7.1 Challenges and solutions

There are various challenges in streaming videos to a multihomed mobile network in an optimal fashion. Firstly, the system needs to keep tracking the locations of the mobile network so that the video packets can be continuously transported to the right places. This demands that either the video steamer is aware of the location updates of the mobile network or the infrastructure deals with this mobility and hides the location changes from the streamer. It is often preferred that the streamer is mobility transparent to facilitate rapid deployment of video services. Mobile IP is

a de-facto mobility management scheme defined by the Internet Engineering Task Force (IETF) and employs a home agent (HA) to redirect incoming packets through tunnelling to the care-of address (CoA) of a mobile node that has registered with the HA; however, Mobile IP is only able to manage single mobile nodes rather mobile networks. To resolve this problem, the IETF has developed a network mobility management protocol called NEMO (Leung, 2009; Devarapalli, 2005), which is an extension to Mobile IP. In NEMO, a mobile router is designated as a mobility management proxy of the mobile nodes within the mobile network. The mobile nodes within the mobile network communicate with external nodes through the mobile router, which can be equipped with multiple network interfaces. The mobile router and the HA together provides mobility transparency to the streamer and the mobile network nodes.
Secondly, multihoming indicates multiple identifiers of the mobile router's interfaces to allow independent and simultaneous access to the corresponding networks. This requirement can be met by introducing multiple IP addresses in IP-based networks. In the context of NEMO, the Multiple Care-of Addresses (MCoAs) scheme (Wakikawa *et al.*, 2009) enables the mobile router to register its interfaces' IP addresses (CoAs) with the HA, and the HA will then be able to establish multiple tunnels between the mobile router and itself for multipath transmissions.
Thirdly, the multiple available access networks should be fully utilised to provide abundant aggregated bandwidth and timely delivery of the video streaming. Generally, flexible policies to distribute traffic flows of various applications can be designed and implemented to achieve dynamic load balancing or "always best connected" service (Wang *et al.*, 2009). To adapt an SVC video streaming instance to the network conditions, an intelligent scheduling algorithm has to be in place and this will be further discussed. In addition, the following subsections will describe other practical issues to be resolved to enable optimised SVC streaming in a multihomed mobile network.

7.2 Packetisation

Since real-time applications such as video transmission are typically delivered over RTP in IP networks, a set of IETF drafts have been proposed to address the RTP level packetisation for various video coding standards such as AVC packetisation (Wang *et al.*, 2010), SVC packetisation (Wenger *et al.*, 2010) and MVC packetisation (Wang and Schierl, 2010). Back compatible with that of the non-scalable AVC, SVC packetisation is rather complicated. Except for a common format of an RTP header, the specific formats of SVC RTP payloads vary significantly depending on different NAL unit types and subtypes, being interleaved or non-interleaved, being single-time or multi-time, being an aggregation packet or not etc. Figure 9 shows an RTP packet containing two non-interleaved multi-time aggregation units (Wenger *et al.*, 2010).

"Take in Figure 9" – **Figure 9.** Example RTP packet in SVC

Subsequently, an RTP packet will be encapsulated with an IP header at the network layer. If the size of the packet exceeds the Maximum Transmission Unit (MTU) of the network, which is common for SVC packets, it will be fragmented into more than one IP packets. In the mobile network scenario, additional tunnelling overhead ("IPv6 in IPv6" encapsulation) will be incurred at the mobility management elements.

7.3 Packet Scheduling over Multiple Paths

Before a video packet is placed on the network, a range of preparation schemes in addition to packetisation are desired to ensure a committed quality of service (QoS) or quality of experience (QoE) better than best effort. We emphasise the path selection beyond standard routing for multipath transmission, the intelligent scheduling that is aware of the video content and network conditions, and QoS control especially QoS signalling and network resource management. It is worth mentioning that signalling based on Session Description Protocol (SDP) etc. is needed prior to packet delivery to identify the used packetisation modes (e.g., non-interleaved mode), the used MST modes (e.g., NI-T mode), and dependencies between RTP sessions before the video stream starts (Schierl *et al.*, 2009).

Path formulation for data traffic is usually down to the routing protocol to identify the best route (and the backup routes in certain routing protocols). To exploit the path diversity in multihomed networks, the routing protocols in use should be able to determine all the available end-to-end routes that are not only independent from each other but can also meet the one-way transmission delay requirement for real-time video. For this purpose, among other proposals, the Two-Phase geographic Greedy Forwarding (TPGF) routing algorithm (Shu *et al.*, 2010) seems a promising solution, which will be further discussed in Section 8. Once the multiple paths are identified (or predefined), a distributed network facility should be in place to consistently measure instantaneous network conditions and report real-time significant changes to the scheduler for the subsequent path selection and packet scheduling decisions.

Intelligent packet scheduling should take into account the decoding deadlines, the measured/predicted end-to-end transmission delays (network condition awareness), the hierarchical structure of the frames (media content/coding awareness), and network diversities such as multiple paths to select from or switch between. A number of scheduling algorithms have been proposed for multipath video transmissions, being aware of the priorities of the packets and assuming real-time knowledge of the predefined paths. The EDPF algorithm (Chebrolu and Rao, 2006) estimates the earliest delivery time for each packet per path and sends each packet on the quickest path, which tends to minimise the packet reordering overhead at the receiver. The TS-EDPF algorithm (Fernandez *et al.*, 2009) is a TDMA adaptation of EDPF: a sender estimates the packet arrival time at the receiver using EDPF and then adjusts the time so that it will fall into the time-slot allocated for the receiver. More complex packet dependencies are considered in the LBA algorithm (Jurca and Frossard, 2007) for improved load-balancing scheduling. A packet's ancestor checking process proactively drops children packets that will not be decodable since their ancestor has been dropped during the scheduling. This operation is a cost saver for the network and the receiver.

Nevertheless, none of the above path selection or scheduling schemes is explicitly designed for SVC applications or multihomed mobile networks. To address this issue, a recent research (Nightingale *et al.*, 2010) has proposed considerable adaptations and optimisations on top of the EDPF and the LBA algorithms for SVC streaming over multihomed mobile networks. In the proposed OPSSA path selection and scheduling algorithm, the tunnelling overhead from the network mobility management and the additional nontrivial delay from the path switching operations have been taken into account, and a trade-off is achieved between bandwidth aggregation and path switching cost. In addition, the ancestor checking functionality proposed in LBA is enhanced by exploring the scalability information available in NAL units. The OPSSA, the EDPF and the LBA algorithms are all implemented in a hardware-based fully functional testbed, and the empirical comparison results indicate a clear-cut performance improvement in terms of PSNR as shown in Figure 10.

"Take in Figure 10" – **Figure 10.** Comparison of video packet scheduling algorithms (Nightingale *et al.*, 2010)

In addition to the awareness of available resources, scheduling schemes may be explicitly combined with QoS management schemes, e.g. to request the network to release and allocate extra resources for prioritised video applications. These QoS management schemes such as resource reservation mechanisms can be specific to the wireless technology or be independent of the underlying medium if operating at the network or higher layers. Examples of higher-layer IETF QoS protocols include DiffServ, RSVP/IntServ and its successor NSIS, the new generation QoS signalling architecture. Additional traffic engineering mechanisms such as MPLS may also be useful. Meanwhile, link-layer-specific radio resource management (RRM) is widely available and usually defined in the standard in nearly every wireless technology although further customisation may still be justified. For example, (Ji *et al.*, 2009) are concerned with SVC streaming over OFDM-based wireless networks such as WiMAX. A gradient-based adaptive scheduling and resource allocation algorithm is proposed to prioritise the transmissions of different users. By exploiting the SVC temporal and quality scalabilities and considering the deadline requirements and transmission history, the algorithm outperforms two benchmark algorithms that are unaware of either the video content or the deadlines in congested OFDM channels. Furthermore, hybrid cross-layer QoS architecture may be considered, e.g. using a link-layer RRM module in a certain wireless system and an upper-layer QoS signalling protocol. For instance, Fernandez *et al*. (2009) propose to couple a TDMA-based bandwidth allocation scheme with a DiffServ-based QoS negotiation procedure, along with the TS-EDPF scheduling.

7.4 Packet Delivery

Once a packet is scheduled out of the video streamer and in transit over a network, it is up to the intermediate network devices such as routers, proxies, base stations, and transcoding/transrating equipment along the end-to-end route to accomplish the actual delivery. Some of these devices are media-aware network elements (MANEs), which are capable of making packet forwarding or discarding decisions based on the scalability information contained in NAL unit headers, RTP and RTP payload headers, and signalling. Among the most desired schemes to be discussed, we highlight selective packet dropping specific to SVC, followed by congestion/corruption control mechanisms.

The spatial, temporal and quality scalabilities in SVC enable flexible prioritised packet delivery in resource-constrained wireless networks. Less important packets will be dropped first as an attempt to cope with network congestion, bandwidth scarcity, or other constraints in network transmission capacity or user equipment capabilities. For example, if a MANE (Media-Aware Network Element) such as a media proxy finds its buffer is going to be overflowed, it will start dropping packets assigned with lower priority, e.g. those belonging to enhancement layer 2 (EL2) whilst attempting to deliver base layer (BL) and EL1 packets, as shown in Figure 11.

“Take in Figure 11” – **Figure 11.** Selective dropping of SVC packets

Moreover, different weights can be assigned to different frames considering if a frame belongs to a BL or an EL in SVC and if it is an I, P or B frame. The higher the weight, the more important the packet is. The server can then schedule the packets and the network nodes selectively drop packets e.g. in congestion according to their importance indicated by the weights.

In sophisticated packet-dropping schemes, the full scalabilities in SVC may be explored. For instance, (Monteiro *et al.*, 2008) proposed three different selective packet-discarding strategies with FGS used for quality scalability in error-prone IP networks with random packet loss. In the joint spatial and FGS layers discarding scheme, (0, *, 0) packets are lossless, and the loss probability for (1, *, 0), (2, *, 0) and FGS layers 1 and 2 are P, 2P and 3P respectively, where P is an adjustable probability value. (Character '*' is a wildcard.) In the joint temporal and FGS layers discarding scheme, (*, 0, 0) lossless, (*, 1, 0) and (*, 2, 0) with P, (*, 2, 0) and (*, 3, 0) with 2P, and FGS layers 1 and 2 with 3P. In the FGS layers discarding scheme, the base layer and control packets are lossless whilst FGS layer 1 with P and FGS layer 2 with 3P. Based on experimental results, the authors find that the first two schemes outperform the random discarding scheme in terms of Y-PSNR and mean frame rate values although there is no clear advantage when they are compared with each other. It is the third scheme that shows the best performance of the three. It seems that the experiments are based on artificial loss assignment rather than a real lossy environment and no error-correction or recovery schemes are explicitly mentioned.

Furthermore, recent ongoing IETF standardisation work (Wenger *et al.*, 2010) further has defined a new SVC Payload Content Scalability Information (PACSI) NAL unit, which facilitates a MANE to decide if an aggregation SVC packet should be discarded without deep inspection of the aggregated NAL units in an RTP packet.

To mitigate network congestion and corruption, corresponding control schemes may be introduced. For example, (Nguyen and Ostermann, 2007) proposed a bandwidth estimation based congestion control algorithm that allows the transmission rate of SVC video to be quickly adjusted to the available bandwidth; whilst (Wien *et al.*, 2007) proposed an unequal erasure protection (UEP) scheme to improve the robustness of SVC streaming in error-prone networks.

Finally, at the receiver, out-of-sequence packets will be reordered and late arrival or non-decodable packets will be dropped. As discussed, intelligent path selection and scheduling algorithms such as EDPF, LBA and OPSSA help to reduce such operations thanks to the delay-aware in-order delivery and the packet ancestor checking.

8. Distributed multipath video transmission in wireless sensor networks

8.1 Challenges and solutions

Compared with infrastructure-based wireless networks e.g., the access networks mentioned in Section 7, wireless sensor networks are further constrained by limited network resources such as even lower bandwidths of the links, higher transmission delays due to multiple hops between the sensors and a sink, and by limited node capacities such as short operation lifetime because of battery depletion and low computational capabilities. Nevertheless, conventional video processing in particular encoding requires complex signal processing, and video transmission demands high throughput and bounded delays. In the following, we discuss approaches to mitigate these conflicts between demands and capabilities.

Firstly, to meet the high throughput requirement for video traffic under the condition of low link bandwidths, the multiple paths between a source sensor and the destination sink has to be utilised to obtain sufficient aggregated bandwidth, a similar approach as in multihoming. Therefore, parallel multipath transmission of video packets (which may not be IP based in common cases) are entailed.

Secondly, to meet the delay constraints across multiple hops between the source and the destination, the routing protocol employed has to be able to identify the disjoint paths that will yield the shortest delays from all the available end-to-end paths. It is noted that when combined the multipath transmission requirement aforementioned multiple paths of shortest delays should be determined rather than just the best route. A couple of representative routing algorithms will be further discussed in the next subsection.

Thirdly, to circumvent the challenges of sensors' energy constraints and low complexity for heavy traffic processing and transmission, energy- and complexity-aware schemes such as DVC need to be applied. The application of DVC in sensor networks will be further discussed below. In addition to the application of DVC, energy management has been among the top concerns in the design of a sensor network. Numerous mechanisms have been proposed regarding this aspect from the perspectives of different protocol stack layers such as physical layer, medium access control (MAC) layer, routing layer and so on. Furthermore, the cross-layer methodology has also been investigated e.g. by Cortes *et al*. (2009). Please refer to (Anastasi *et al.*, 2009) for a detailed survey on energy conservation in sensor networks. In addition, (Abid and Qaisar, 2010) presented an analytical method for calculating energy consumption at a sensor in a sensor network employing DVC.

Finally, there are other challenges for multimedia streaming in sensor networks with regard to security, congestion control, reliable transmission, cross-layer optimisation and so on; these topics were comprehensively analysed in (Misra *et al.*, 2008) from each protocol layer's perspective.

8.2 Multipath transmission

In addition to the mentioned improvement of throughput through bandwidth aggregation, multipath transmission (Figure 12) can also contribute to energy saving since the video traffic between the sensor and the destination sink is shared by multiple paths, and thus the battery consumption from handling the traffic is distributed across sensors. However, multipath routing algorithms need to take into account the trade-off between battery drainage due to repeated use of the same set of sensors along a route and extra energy consumption due to suboptimal path selection. Moreover, it is desirable to generate multiple paths based on local decisions to reduce global communication and storage overhead. Motivated by these considerations, a Geographical Power Efficient Routing (GPER) protocol (Wu and Candan, 2007) was proposed and simulations results showed that GPER could reduce almost 50% power consumption compared with traditional geographical routing algorithms. However, it is noted that GPER focuses on energy conservation other than end-to-end delays for video delivery, e.g., by scaling down the transmission ranges, which leads to an increase in the number of hops.

"Take in Figure 12" – **Figure 12.** Multipath transmission in a wireless sensor network

In favour of selecting routes of shortest end-to-end delays, a Two-Phase geographic Greedy Forwarding (TPGF) routing algorithm (Shu *et al.*, 2010) has been recently proposed in a Multi-Path Multi-Priority (MPMP) framework. TPGF includes two phases: Phase 1 explores all the possible routes, and Phase 2 selects routes of the least hop counts with each route's end-to-end transmission delay gathered. The multiple priorities refer to this end-to-end delay and additionally video and audio content "information value", which is calculated based on the volume of the media traffic and the corresponding importance level of the media in a given application context. For instance, in a surveillance context, the visual streams are assigned higher priority than the audio ones whilst this priority allocation is reversed in a deep ocean monitoring application since the visibility in that context is very low. For this purpose, a context-aware multipath selection algorithm (CAMS) is further executed after TPGF to select the paths of highest information values of video/audio stream transmission whilst guaranteeing the end-to-end delay. Simulations were conducted to verify and evaluate the proposed algorithms and results appear promising.

8.3 DVC in wireless sensor networks

As indicated in the DVC section, the conventional approach using complex encoders and lightweight decoders is suitable for broadcasting applications whilst in contrast DVC moves the complexity in encoders to decoders.

Therefore, DVC is well suited for the convergecast fashion in sensor networks, where numerous sensors send data to a sink. It is worth mentioning that a range of networked applications can utilise DVC (Pereira *et al.*, 2008) although we focus on its application in sensor networks.

In (Puri et al., 2006), the authors described the architecture and algorithms of PRISM, which is based on source coding with side information i.e., distributed source coding (DSC). The authors further argued that a transcoding proxy would be desired in sensor networks so that the complexity is located in this intermediate network element whilst both the encoders at the sensors and the decoders can be of low complexity. Experiment results showed that the proposed implementation of PRISM performed nearly as well as H.263+ in terms of PSNR.

The PRISM encoder and the Wyner-Ziv encoder are two typical DVC encoders, and the comparisons of them were carried out in (Ahmad *et al.*, 2009) to evaluate their energy efficiencies through empirical experiments. The results showed that the Wyner-Ziv encoder has consistently higher energy efficiency compared with the PRISM one. Minor modifications were also applied to both encoders to reduce power consumptions. Most interestingly, a counter-intuitive finding was that the main energy consumer was the local video processing rather than the video transmission.

DVC can be further enhanced with other technologies for improved performances or enhanced capabilities in sensor networks. For instance, compressive sensing (CS) is a technology that can capture compressed visual data efficiently and thus (Kang and Lu, 2009) proposed to integrate DVC and CS to simultaneously capture and compress videos. The proposed distributed compressive video sensing (DCVS) framework can shift most computation complexity to the decoder, which can reconstruct the compressed video at faster speed yet higher quality using a modified gradient projection for sparse reconstruction (GPSR) algorithm compared with existing CS reconstruction algorithms.

Moreover, in dense camera sensor networks, the correlation across multiple views can be further exploited at the decoder and this enhancement leads to the emerging multiview DVC systems (Dufaux *et al*., 2009). It is noted that multiview DVC systems are different from H.264/MVC systems since the former do not require communications between the camera sensors whilst the latter do. In addition, the current multiview DVC systems are outperformed by H.264/AVC although it would be interesting to compare the performances of multiview DVC and H.264/MVC in sensor networks in terms of rate-distortion, energy consumption, and communication overhead etc.

Finally, it appears that so far almost all studies on DVC in sensor networks have been either theoretical or based on simulations. Therefore, the practical sensor network testbed presented in (Oldewurtel *et al.*, 2008) is a timely and relevant contribution. The testbed consists of a sink and a number of one-hop sensors using the Telcos platform with 10 KB RAM, 48 KB ROM, and an IEEE 802.15.4-compatible transceiver (CC2420 radio). The testbed actually employs DSC to compress the sensed temperature, humidity and light data other than visual information. Clearly, more advanced testbeds are desired for video applications using DVC over multihop sensor networks.

9. Conclusion

We have reviewed a group of recent video coding standards including the H.264 family (AVC, SVC and MVC), the DVC codec and the emerging H.265 standard. Every standard was designed with targeted application scenarios in mind. We have taken SVC streaming over multipath mobile networks as a detailed example to show that H264 SVC is highly useful as a standard even in such a challenging networking paradigm, thanks to its versatile scalability capabilities. The challenges and solutions for DVC video transmission in wireless sensor networks have also been discussed.

Despite the recent significant advances in video signal processing and ubiquitous networking standards and emerging technologies, numerous challenges have yet to be further dealt with as indicated throughout our paper. The interdisciplinary nature of video networking entails closer collaborations between the visual signal processing community and the networking community towards meeting the ever-increasing expectation from the general public consumers.

References

Aaron, A, Rane, S, Setton, E., and Girod, B. (2004), “Transform-domain Wyner-Ziv codec for video,” in *proceedings of SPIE Visual Communications and Image Processing*, San Jose, CA, USA, January 2004.

Abid, H., and Qaisar, S. (2010), "Distributed video coding for wireless visual sensor networks using low power Huffman coding", in *Proceedings of 44th Conference on Information Sciences & Systems (CISS)*, Princeton, NJ, USA, March 2010, pp. 1-6.

Ahmad, J.J., Khan, H.A., and Khayam, S.A. (2009), "Energy efficient video compression for wireless sensor networks," in *Proceedings of 43rd International Conference on Information Sciences and Systems (CISS)*, Baltimore, MD, USA, March 2009.

Anastasi, G., Conti, M., Di Francesco, M., and Passarella, A. (2009), "Energy conservation in wireless sensor networks: a survey", *Ad Hoc Networks*, Vol. 7, n.3, pp. 537-568.

Chebrolu, K. and Rao, R. (2006), "Bandwidth aggregation for real-time applications in heterogeneous wireless networks", *IEEE Transactions on Mobile Computing*, Vol. 5, No. 4, pp. 388-403.

Cortes, J., Wang, Q., and Dunlop, J. (2009), "Cross-Layer proactive hybrid MAC to prolong lifetime of wireless sensor networks", in *proceedings of the 69th IEEE Vehicular Technology Conference (IEEE VTC2009-Spring)*, Barcelona, Spain, April 2009.

Devarapalli, V., Wakikawa, R., Petrescu, A., and Thubert, P. (2005), "Network mobility (NEMO) basic support protocol", *IETF RFC 3963*.

Dufaux, F., Gao, W., Tubaro, S., and Vetro, A. (2009), "Distributed video coding: trends and perspectives", *EURASIP Journal on Image and Video Processing*, vol. 2009.

Fernandez, J., Taleb, T., Guizani, M., and Kato, N. (2009), "Bandwidth aggregation-aware dynamic QoS negotiation for real-time video streaming in next-generation wireless networks", *IEEE Transactions on Multimedia*, Vol. 11, No. 6, pp. 1082-1092.

Grecos, C., Lambert, P., Slowak, J., Mys, S., Skorupa, J., and Wan De Walle, R. (2010), "Distributed video coding for video communication on mobile devices and sensors", book chapter in *Handheld Computing for Mobile Commerce: Applications, Concepts and Technologies*, IGI Global, ISBN-10: 1615207619, ISBN-13: 978-1615207619.

Ho, Y. (2007), "Recent progress in multi-view video coding", in *proceedings of Pacific Rim Symposium on Image Video and Technology (PSIVT'07),* tutorial, Santiago, Chile, December 2007.

JCTVC site (2010), available at: http://wftp3.itu.int/av-arch/jctvc-site (accessed October 2010).

Ji, X., Huang, J., Chiang, M., Lafruit, G., and Catthoor, F. (2009), "Scheduling and resource allocation for SVC streaming over OFDM downlink systems", *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 19, No. 10, pp. 1549-1555.

Jurca D., and Frossard, P. (2007), “Video packet selection and scheduling for multipath streaming”, *IEEE Transactions on Multimedia*, Vol. 9, No. 3, pp. 629-641.

Kang, L., and Lu, C. (2009), "Distributed compressive video sensing", in *proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Taiwan, China, April 2009, pp.1169-1172.

Leung, K., Dommety, G., Narayanan, V., and Petrescu, A. (2008), "Network mobility (NEMO) extensions for Mobile IPv4", *IETF RFC 5177*.

Misra, S., Reisslein, M., and Xue, G. (2008), “A survey of multimedia streaming in wireless sensor networks”, *IEEE Communications Surveys & Tutorials*, Vol. 10, No. 4, pp. 18-39.

Monteiro, J. M., Calafate, C. T., and Nunes, M. S. (2008), “Evaluation of the H.264 scalable video coding in error prone IP networks”, *IEEE Transactions on Broadcasting*, Vol. 54, No. 3, pp. 652-659.

Nguyen, D. and Ostermann, J. (2007), “Congestion control for scalable video streaming using the scalability extension of H.264/AVC”, *IEEE Journal of Selected Topics in Signal Processing*, Vol. 1, No. 2, pp. 246-253.

Nightingale, J., Wang, Q., and Grecos, C. (2010), “Optimised transmission of H.264 scalable video streams over multiple paths in mobile networks”, *IEEE Transactions on Consumer Electronics*, Vol. 56, No. 4, pp. 2161-2169.

Oldewurtel, F., Foks, M., and Mahonen, P. (2008), “On a practical distributed source coding scheme for wireless sensor networks”, in *proceedings of IEEE Vehicular Technology Conference (VTC-Spring 2008)*, Marina Bay, Singapore, May 2008, pp. 228-232.

Pereira, F., Torres, L., Guillemot, C., Ebrahimi, T., Leonardi, R., and Klomp, S. (2008), “Distributed video coding: selecting the most promising application scenarios,” *Signal Processing: Image Communication*, Vol. 23, No. 5, pp. 339-352.

Puri, R., Majumdar, A., Ishwar, P., and Ramchandran, K. (2006), “Distributed video coding in wireless sensor networks”, *IEEE Signal Processing Magazine*, Vol. 23, No. 4, pp. 94-106.

Puri, R., and Ramchandran, K. (2002), “PRISM: A new robust video coding architecture based on distributed compression principles,” in *proceedings of Allerton Conference on Communication, Control and Computing*, Monticello, IL, USA, October 2002.

Schierl, T., Karsten, G., and Wiegand, T. (2009), “Scalable video coding over RTP and MPEG-2 transport stream in broadcast and IPTV channels”, *IEEE Wireless Communications*, Vol. 16, No. 5, pp. 64-71.

Schwarz, H., Marpe, D., and Wiegand, T. (2007), “Overview of the scalable video coding extension of the H.264/AVC standard”, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 17, No. 9, pp. 1103-1120.

Shu, L., Zhang, Y., Yang, L. T., Wang, Y., Hauswirth, M., and Xiong, N. (2009), “TPGF: geographic routing in wireless multimedia sensor networks”, *Telecommunication Systems*, Vol. 44, No. 1-2, pp. 79-95.

Smolic, A., Mueller, K., Stefanoski, N., Ostermann, J., Gotchev, A., Akar, G.B., Triantafyllidis, G., Koz, A. (2007), “Coding algorithms for 3DTV—a survey,” IEEE Transactions on Circuits and Systems for Video *Technology*, Vol. 17, No. 11, pp. 1606–1620.

Wakikawa, R., Devarapalli, V., Tsirtsis, G., Ernst, T., and Nagami K. (2009), "Multiple care-of addresses registration", *IETF RFC 5648*.

Wang, Q., Hof, T., Filali, F., Atkinson, R., Dunlop, J., Robert, E., and Aginako, L. (2007), “QoS-aware network-controlled architecture to distribute application flows over multiple network interfaces”, *Wireless Personal Communications*, Vol. 48, No. 1, pp. 113-140.

Wang, Y., Even, R., Kristensen, T., and Jesup, R. (2010), “RTP payload format for H.264 video”, *IETF draft-ietf-avt-rtp-rfc3984bis-11.txt (work in progress)*, June 2010.

Wang Y., and Schierl, T. (2010), “RTP payload format for MVC video”, *IETF draft-wang-avt-rtp-mvc-05.txt (work in progress)*, April 2010.

Wenger, S., Wang, Y., and Schierl, T. (2010), “RTP payload format for SVC video”, *IETF draft-ietf-avt-rtp-svc-22.txt (work in progress)*, August 2010.

Wiegand, T, Sullivan, G., Bjontegaard, G., and Luthra, A. (2003), “Overview of the H.264/AVC video coding standard”, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 13, No. 7, pp. 560-576.

Wien, M., Cazoulat, R., Graffunder, A., Hutter, A., & Amon, P. (2007), “Real-time system for adaptive video streaming based on SVC”, *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 17, No. 9, pp. 1227-1237.

Wu, S., and Candan, K. S. (2007), "Power-aware single- and multipath geographic routing in sensor network", *Ad Hoc Networks*, Vol. 5, No. 7, pp. 974-997.

About the authors

Full Name: Christos Grecos

Affiliation: School of Computing, University of the West of Scotland, United Kingdom

Contact:

School of Computing

University of the West of Scotland

Paisley PA1 2BE

United Kingdom

Tel: 0044 141 8483301

Fax: 0044 141 8483542

Email: Christos.Grecos@uws.ac.uk

Prof. Christos Grecos is a Professor in Visual Communications Standards, and Head of School of Computing in the University of the West of Scotland (UWS), UK. He leads the Audio-Visual Communications and Networks Research Group (AVCN) within UWS, and his research interests include image/video compression standards, image/video processing and analysis, image/video networking and computer vision. He has published around a hundred research papers in top-tier international publications including a number of IEEE transactions on these topics. He is on the editorial board or served as guest editor for numerous international journals, and he has been invited to give talks in various international conferences. He is the Principal Investigator for several national and international projects funded by UK EPSRC or EU. He received his PhD degree in Image/Video Coding Algorithms from the University of Glamorgan, UK. He is a Senior Member of IEEE and SPIE.

Full Name: Qi Wang

Affiliation: School of Computing, University of the West of Scotland, United Kingdom

Contact:

School of Computing

University of the West of Scotland

Paisley PA1 2BE

United Kingdom

Tel: 0044 141 8483000

Fax: 0044 141 8483542

Email: Qi.Wang@uws.ac.uk

Dr. Qi Wang is a Lecturer in Computer Networking with the University of the West of Scotland (UWS), UK. Previously, he was a Postdoctoral Research Fellow with the University of Strathclyde, UK, and a Telecommunications engineer with the State Grid Corporation of China. He received his PhD in Mobile Networking from the University of Plymouth, UK, and his BEng and MEng degrees from Dalian Maritime University, China. Recently, he has been involved in the European Union FP6 MULTINET project and the UK EPSRC DIAS project. His research interests include Internet Protocol networks and applications, diverse wireless networks, mobility management, multihoming support and intelligent network selection, and cross-layer design. He is a member of IEEE, and on the technical programme committees of a number of international conferences.

Figures

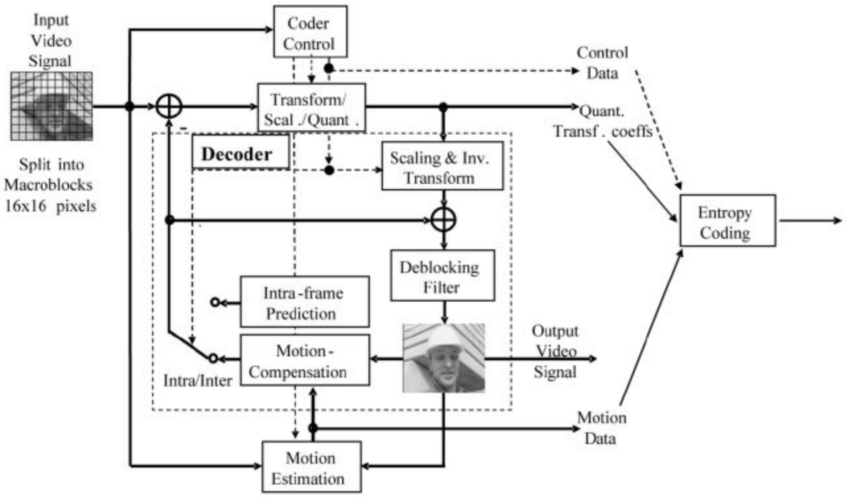

Figure 1. AVC architecture (Wiegand *et al.*, 2003)

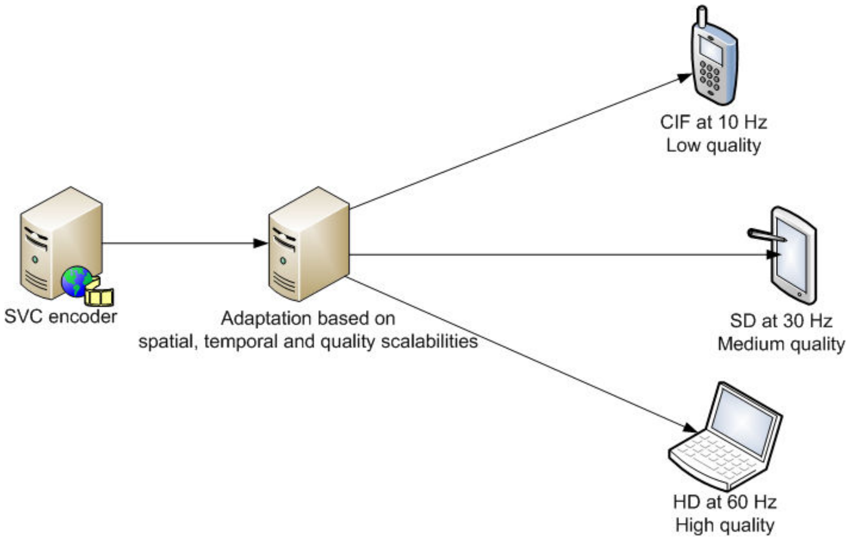

Figure 2. SVC architecture

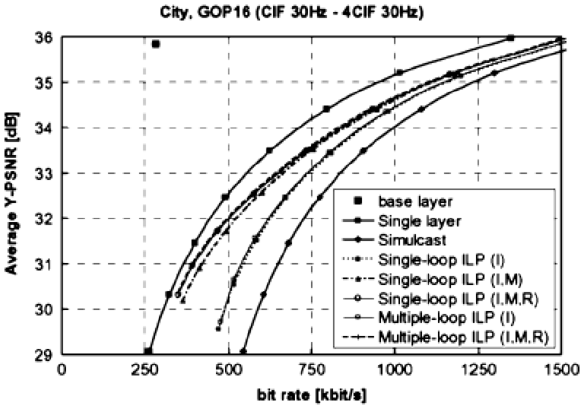

Figure 3. Simulation results (City with GOP 16) (Schwarz *et al.*, 2007)

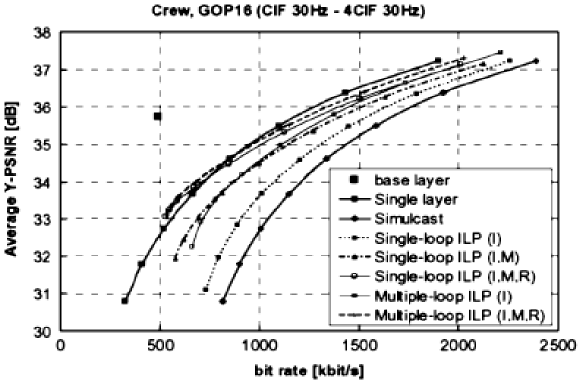

Figure 4. Simulation results (Crew with GOP 16) (Schwarz *et al.*, 2007)

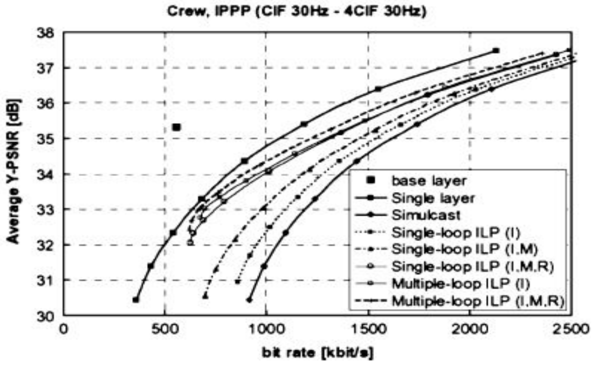

Figure 5. Simulation results (Crew with IPPP, GOP1) (Schwarz *et al.*, 2007)

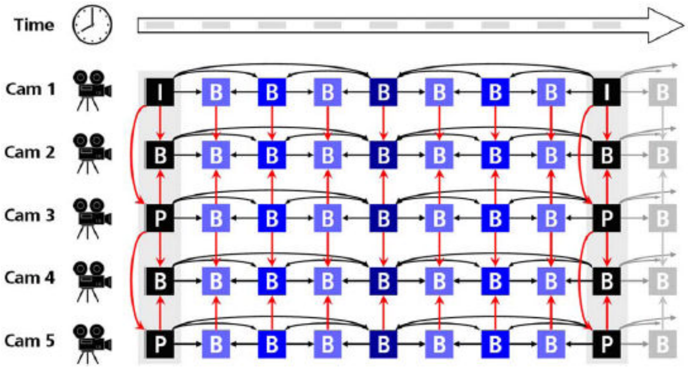

Figure 6. Prediction order in MVC (Smolic, 2007)

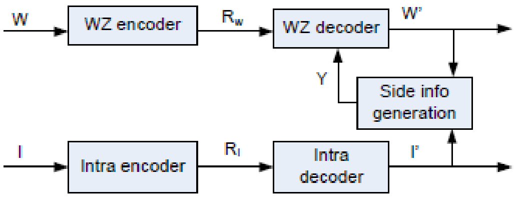

Figure 7. DVC architecture (Grecos, 2010)

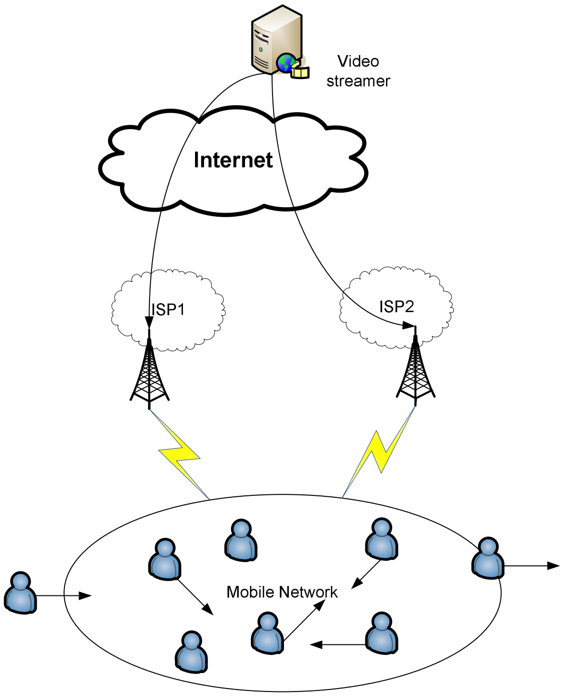

Figure 8. Video streaming to a multihomed mobile network

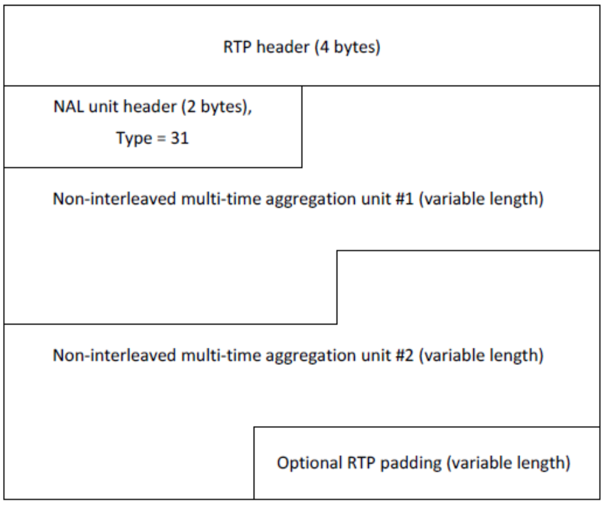
RTP header (4 bytes)
NAL unit header (2 bytes),
Type = 31
Non-interleaved multi-time aggregation unit #1 (variable length)
Non-interleaved multi-time aggregation unit #2 (variable length)
Optional RTP padding (variable length)

Figure 9. Example RTP packet in SVC

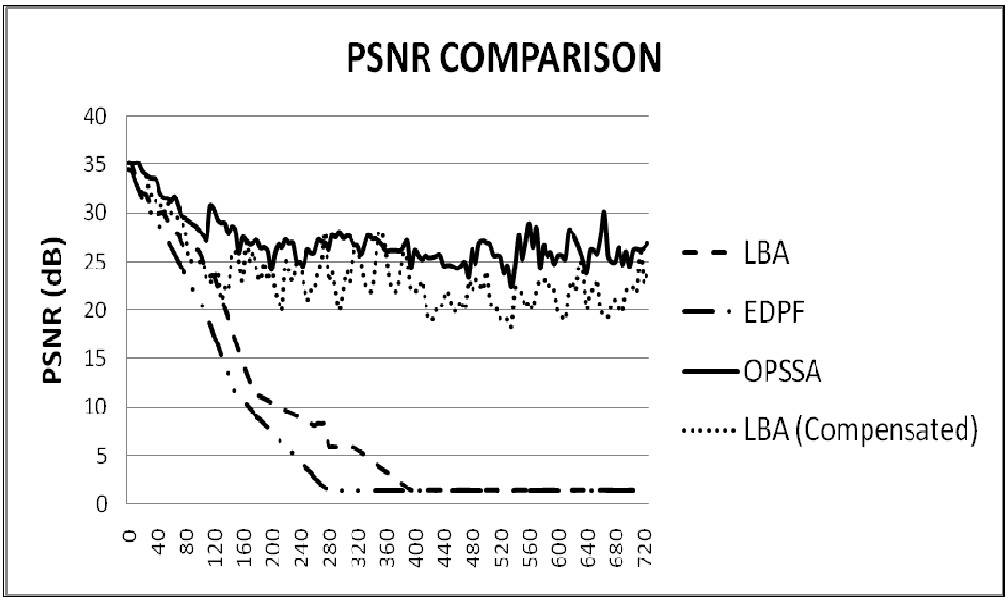

Figure 10. Comparison of video packet scheduling algorithms (Nightingale *et al.*, 2010)

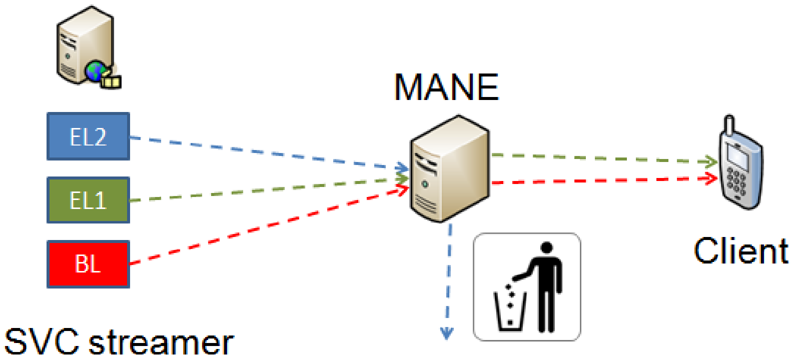

Figure 11. Selective dropping of SVC packets

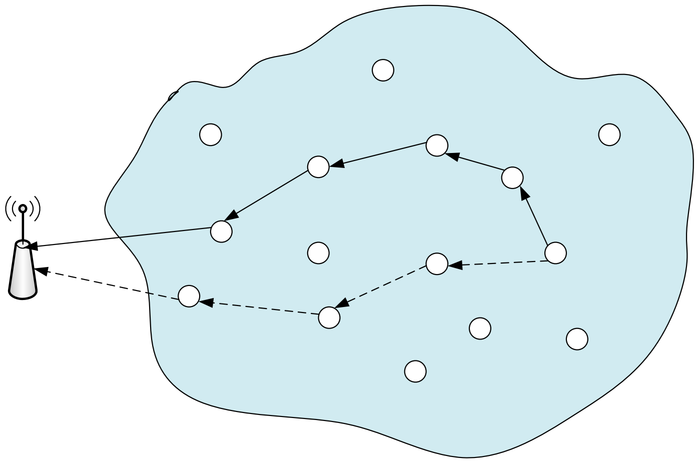

Figure 12. Multipath transmission in a wireless sensor network