FACTA UNIVERSITATIS (NIŠ)
Series: Electronics and Energetics vol. 9, No.1 (1996), 61–73

LAYERED VIDEO CODING TECHNIQUES IN HIGH DEFFINITION TV:

DEVELOPMENTS AND RESEARCH ACTIVITIES

Zoran Bojković

Abstract. This paper seeks to provide some developments, recent effects and research activities in the field of layered video codiug in high definition TV (HDTV). The status of the packet video is briefly reviewed. Benefits and shortcomings of packetized video are pointed out, too. Some types of layered video coding schemes are identified. The approaches in which the layering is performed using subband coding, as well as discrete cosine transform coding with HDTV pictures are demonstrated. Finally, some research directions are recommended.

1. Introduction

Layered techniques offer attractive features for video coding in several applications including high definition TV (HDTV). The coding of HDTV for digital transmission via Broadband Integrated Services Digital Network (BISDN) is of importance since digital fiber optics may represent the most viable means for avoidiug transmission impairements and delivering very high image quality. Some studies have indicated that high quality HDTV coding can be achieved near 135 Mbit/s which is about the BISDN pay load of the SONET channel.

A layered video coding structure for an ATM network is an important concept because the layered structure is useful not only for source flexibility but also for packet loss prediction/recovery. Coding schemes based on discrete cosine transform (DCT) are suitable because spatial frequency

Manuscript received April 21, 1995.

A version of this paper was presented at the second Conference Telecommunications in Modern Satellite and Cables Services, TELSIKS'95, October 1995, Niš, Yugoslavia.

Dr Z. Bojković is with Faculty of Transport and Traffic Engineering, University of Belgrade, Vojvode Stepe 305, 11000 Beograd, Yugoslavia.

components can be classified into layers according to the quality degradation caused by packet loss.

The advantage of a DCT based layered structure is that we can divide the transform coefficients on a block by block basis so as to control the distortion caused by packet loss [11,12]. Subband coding (SBC) is also well suited to an ATM environment [6].

In order to cope with cell loss as a result of congestion, several multilayer coding schemes in conjunction with packet priority have been proposed [5,10). In a multilayer approach, the images are segmented into a main signal and enhancement signals. The main siguals and enhancement signals are packetized into separate cells, and the main signal is labeled with a high priority while the enhancement signals are labeled with lower priorities [17]. These multilayer coding schemes are essentially a type of hierarchical coding since the data to be transmitted are organized in a hierarchical fashion. Namely, the main signal is the top level signal and can be used alone to produce a rough reconstruction of the transmitted picture, while each enhancement signal can be added to produce more picture details [16]. Therefore, if the network allows different loss priorities, the highest priority should definitely be assigned to the main signal and lower priorities to the enhancement signals according to their importance. In [6], subband filtering is used to decompose the images, while in [5,10] DCT is used.

After a short background, Section 2 begins with benefits and shortcomings of packet video, while Section 3 contains HDTV in the Moving Picture coding Experts Group. Compatibility and scalability approach are addresssed in Section 4. Layered coding techniques developments are addressed in Section 5. The corresponding research directions and activities are presented in Section 6, while Section 7 summarizes the paper.

2. Benefits and shortcomings of packet video

In layered coding, the video information is divided into several layers with lower layers containing low resolution information and higher layers the higher resolution information in descending order of importance. Such a model has the potential to enable integration of video telephony, broadcast quality video and high quality video services. In video services where bandwidth is at a premium, lower layers can provide the desired quality and in broadcast applications, a variable number of higher layers can be integrated with the lower ones to provide the quality and the bit rate that is compatible with the service requirement and the receiver. Furthermore, layered approach may also make the introduction of HDTV more feasible by providing compatibility with the current TV.

Potential benefits of packet video are:

a) Integration of video with voice and data in a common packet switched system which offers some cost savings through sharing of switching and transmission resources,

b) Packet internetworking techniques can be applied to provide intercommunication among video users on different types of networks,

c) Significant advantages in channel utilization for video services by allowing convinient accomodation of video terminals with different bit rates and data formats, and

d) Nodal processors use sophisticated error control and packet accounting techniques enabling unsophisticated video terminals to be used in more advanced applications.

On the other side, it is important to stress that packet networks are not without shortcomings for video communication, some of which are:

- Most packet switched networks are designed for time sharing applications, typically involving slow terminal speeds, asynchronous operation and relatively short messages. However, video applications have a somewhat different set of requirements i.e high data rate, long messages, short end-to-end delay and full duplex real-time mode of operation (videoteleconferencing).
- Large end-to-end delay in packet networks, especially for video and voice applications is a serious problem for real time applications.

In addition to designing and laying out packet switched networks for carrying data, voice and video on the same transfer medium, a number of problems and issues need to be researched so as to realize the benefits of packetized video. Some major problems are: development of packet protocols for call set-up and video transport, strategies for reconstitution of video from packets arriving at nonuniform intervals, statistical analysis of different types of video sequences to understand the characteristics of video data to be transmitted over packet switched networks, development of efficient packet video multiplexing techniques in order to exploit the uncorrelated time variant characteristics of multiple video sources, strategies to handle packet loss, packet jitter and channel errors, strategies that will minimize packet overhead and strategies for effective packet control to enable network links to be heavily loaded without congestion, optimal size of video packets that will minimize both the end-to-end delay and the effect of packet defects and random bit errors on reconstructed image quality, optimal packet video switching strategies to reduce both the packetization delay at each node and the overhead due to header.

Since the issues mentioned are inter–related, it is essential to study the impact of possible solutions to any problem on other parameters and issues. For example, end-to-end delay is affected by several parameters like coding rate, packet protocol, packet size, packet header parameters, channel bandwidth, the number of trunk circuits, routing, error control and protection schemes as well as packet switching speed.

3. HDTV in moving picture coding experts group

The introduction of the Moving Picture coding Experts Group (MPEG) in the International Standard Organization (ISO) can be viewed as a first step towards generic solutions not intended for bounded application areas [3]. The MPEG-1 was dissmissed as not providing sufficient quality for TV broadcasting applications, although it was adopted by consumer electronics and computer companies as a viable compromise between quality and bit rate for many other applications such as compact disc interactive (CDI) [7].

When MPEG began its work towards a second standard encompassing interlaced formats and higher bit rates (MPEG–2), many of the TV broadcasting organizations began to take more interest. ISO and the former CCITT appreciated the desirability of reaching common text for MPEG–2 and H.26x, the latter being the working name for the BISDN video coding standard.

Although MPEG–2 was initially perceived by many as being restricted to resolutions not higher than the former CCIR Recommendation 601, it was evident that the studies for MPEG–2 were also highly applicable to HDTV.

This notion was strengthened by experiments with layered coding methods. These were favoured by those wishing to provide both CCIR 601 and lower resolutions from one encoded bit stream. The concept of being able to extract decoded images at more than one quality level is seen as being important by a growing number of the participating organisations in MPEG.

On May 1993, the digital system proponents announced that they had formed a "Grand Alliance" aimed at developing a US standard for terrestrially broadcast HDTV. In summer 1995, the proposed standard was going to be submitted for final certification to the Federal Communications Commission (FCC). For compression of video signals, the Grand Alliance uses a motion compensated DCT algorithm that was employed by all of the contenders in the first round of testing. DCT exploits spatial redundancy while motion compensation exploits temporal redundancy. DCT was chosen for its good energy compaction properties and the many fast algorithms available afford low cost implementation. In addition, the Grand Alliance system employs source adaptive coding and other techniques in order to improve the coding efficiency. MPEG-2 syntax from the Moving Picture Experts Group will be used. The MPEG-2 tool supports most of the compression algorithms. It will also promote world wide acceptance of the Grand Alliance System which conforms to the MPEG-2 main profile implemented at high level.

4. Compatibility and scalability

MPEG is interested in compatible and scalable coding. There is also an interest in exploiting layered coding to provide compatibility between existing the former CCITT Recommendation H.261 codecs on the narrowband ISDN and potential new higher quality codecs on the BISDN [4].

MPEG defines a scalable bit stream as one having the property that part or parts of it may be discarded and yet a decoder can still produce useful results [1]. Compatible layered coding provides a degree of scalability, the granularity available depending on the number of layers. Scalable algorithms do not necessarily give compatibility. Great emphasis has been placed on scalable coding for digital terrestrial TV broadcasting. Receivers of varying processing power would be able to decode a particular service from a single encoded bit stream. One such scenario is a HDTV service which embedded extended definition TV (EDTV) and standard definition TV (SDTV). The HDTV and embedded EDTV signal would be decoded by fixed–site receivers.

5. Layered coding techniques developments

Previously investigated work in the literature on layered coding has focused on two areas [9,14]. The first of these is the refinement of coded pixel residuals. That is, actual pixel difference between an input frame and the resulting coded frame is found and these residuals are then coded and sent in the second (refinement) layer. A second approach has been to send additional "frequency" information in the second layer. In the case of the popular DCT, this involves sending a subset of the DCT coefficients in the baseline layer, providing a coarse but lower rate representation of the block frequency content.

5.1 Subband coding as an example of a frequency–domain separation

In frequency domain separation, the pyramid data video structure is formed in terms of their frequency components. The basic idea of the SBC is to divide the frequency band of signal into a number of subband by a bank of bandpass filters. Each subband is then translated to baseband by downsampling and encoded separately. At the receiver, the subband signals are decoded and upsampled back to the original frequency band by interpolation. The signals are then summed to give a close replica of the original signal.

Karlsson and Vetterli [6] have proposed subband coding for compressing video for transmission over packet switched networks. Consider a subband coding scheme shown in Figure 1.

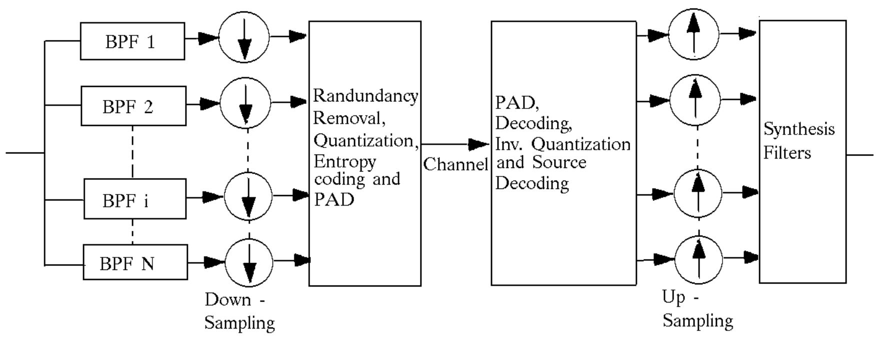

Fig. 1. Subband coding block diagram

The video sequence is passed through a bank of bandpass filters (BPF) and then the signal corresponding to each band is subsampled to its Nyqist rate. Some of the high bands can be thrown away and others are coded using the traditional coding algorithms. In the receiver, the opposite is done. In [6], lowpass and highpass filters are applied in all three dimensions, first in the temporal dimension, and then in the horizontal and vertical dimensions. At each stage of filtering, the signal is subsampled by a factor 2. Thus, 8 bands are generated. The lowest band is again subsampled and bandsplitted in horizontal and vertical dimensions giving rise to 4 more bands. Thus, 11 bands are generated in all with 10 bands displaying a significant reduction in variance. The lowest band, which is highly correlated, is DPCM encoded and the other low variance subbands are PCM encoded with reduced number of quantization levels where the number of bits allocated to pixels in each band depends upon the variance. Finally, the encoded subbands are run-length encoded to exploit the fact that the encoded subbands have large connected areas of zero valued samples in each frame of each subband. The sign PAD means packet assembler/disassembler.

In order to improve the coding efficiency, subband coding can be combined with other coding techniques like DPCM, DCT, vector quantization

(VQ).

In motion compensated interframe SBC approach, the spectrum of each frame of video signal is first decomposed into smaller frequency bands, where each can be coded. To preserve its hierarchical structure, each band is coded independent of higher frequency bands, but can share information with the lower bands.

5.2. DCT coding

The intrafield DCT coding scheme was developed for efficient HDTV transmission at 135 Mbps [2]. In order to reduce hardware complexity and avoid error propagation from frame to frame, intrafield coding is employed. It is assumed that the input HDTV source is interlaced. The 8×4 DCT is used, since the vertical correlation is smaller than the horizontal correlation in a field. Besides, the transform based on a smaller block size is computationally more efficient. Features of the coding system also include DPCM coding for DC terms, frequency weighting, two–dimensional entropy coding, nonlinear quantization, as well as buffer feedback control. However, the system will suffer very noticeable degradations when cell losses occur.

In the DCT domain, the transform coefficients readily render themselves into a hierarchical structure since an approximate image can be reconstructed based on some low order coefficients and increasingly better approximations can be achieved by adding more coefficients.

In order to partition the DCT coeficients into layers, division of a zig-zag scan of the data is introduced. Here, the point of division in the zig-zag scan is proposed to ensure video quality at a certain level during heavy traffic. It is assumed that the low frequency layer should be assigned to high priority packets and the high frequency layer would be assigned to low priority packets. The division point is determined by evaluating the following distortion along the scan direction starting from the highest frequency coefficient to be coded, i.e

$$D_L = \sum_{k=1}^{L} \left[\frac{q(k)}{1 - a(k)} \right]^2 \tag{1}$$

where $q(k)$ is a quantization step size, while $a(k)$ represents prediction coefficient associated with k-th DCT coefficient in the scan. If D_L exceeds a predetermined threshold D_{th}, then the data scanned to the $(k-1)$-th coefficient would be assigned to a low priority packet. The degradation by packet loss can be evaluated considering the propagation by using equation (1).

The DCT split also suffers from the same difficulty as subband coders over fixed bit rate allocations. This can be overcome by subtracting the quantized values of the lower layer coefficients so that the coding errors in them are passed to the higher layer in addition to the other coefficients [8]. Therefore, higher layers have the opportunity to balance distortion in the low and high resolution components of the decoded composite. When motion compensation is used, drift can arise in decoders which are reconstructing pictures from only part of the total bit stream. The reason is that the reference pictures on which motion compensation is being performed are different at coder and decoder.

A DCT split algorithm was investigated and proposed in [13]. In spite of the mentioned difficulties, this technique has elegance and continuous to be investigated in MPEG. A variant is the data partioning of the coefficients in a single loop encoder. This allows transmission systems to protect parts of the bit stream and permits decoders to ignore parts of it.

5.3 Inter/intra–frame layered DCT coder

The coder uses the inter–frame, also called composite mode or the intra-frame mode, depending upon whether or not there has been significant motion between the previously decoded frame and the present frame, on a frame by frame basis. In this way, we make optimal utilization of either the temporal or the spatial redundancy. To deal with the problem of error propagation from frame to frame while in the inter–frame mode, we propagate a vertical strip of width equal to an integer multiple of the block width. This strip is coded by an intra-frame DCT algorithm, while the rest of the frame is coded by an inter-frame DCT coder. Such a composite arrangement enables complete refreshement of the whole frame in a few frame periods. The inter/intra frame DCT coder is shown in Figure 2.

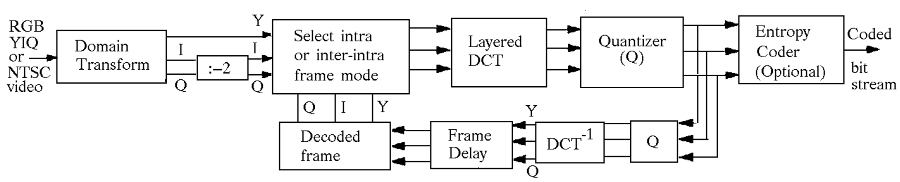

Fig. 2. Inter/intra frame DCT coder

Based on the mean and the variance of a fixed number of pixel differences between the current frame and previously decoded frame at random pixel locations, a statistical parameter is computed and compared against a threshold. If the computed parameter is found to be less than the threshold, an inter–frame mode is selected. Else, an intra-frame mode is selected.

In inter–frame mode, a difference frame is computed based on the frame replenisment approach and a vertical strip on width equivalent to the block width, or multiple block width, is replaced by intra-frame difference values. A simple predictor is used to derive the intra-frame difference strip which is independent of any data from both the previous frame and the present frame outside the strip. Starting from the left side of the edge of a frame, this strip is advanced in a non-overlapping manner sequentially every time a new frame is coded in a composite mode. For example, in a video sequence whose frame resolution is 512×512 pixels, with strip width of 16 pixels, the whole frame is refreshed in 32 frame periods, provided all the frames are coded in the composite mode. After 32 frame periods, the cycle repeats itself. If the decision is to do intra-frame coding, an error frame can be generated in a number of ways by using different types of predictors. The frame can be divided into vertical strips of equal width, i.e, 64 pixels each and previous pixel predictor can be used in each strip independently. This approach limits the propagation of decoded errors to the strip itself.

The generated difference frame is divided into blocks of 16×16 pixels, which are then coded using a DCT. The DCT coefficients of a block are zig–zag scanned and a set of very high frequency coefficients are discarded. The remaining coefficients are divided into layers - sets of coefficients. The low frequency layer is transmitted first, followed by the higher frequency layers until the last layer corresponding to the desired video quality has been transmitted. If the total energy content in a certain layer is small, that layer and the layers above it, are discarded completely and are not transmitted, then further increasing the compression. These layers are transmitted in such a manner as to enable progressively improving reconstruction of decoded frames at the receiver. A fixed number of bits are assigned to the coefficients of each layer, which may then be entropy coded to further increase the compression. However, employing variable codeword - size entropy coding adversely impacts the resynchronization of video in the event of cell loss unless resynchronization flags are inserted in the cell adaptation header.

5.4. Pel split

The pel split is a hierarchical arrangement. The classical configuration in which the lower layers are produced by filtering and downsampling is shown in Figure 3. The coded versions are upsampled and subtracted from

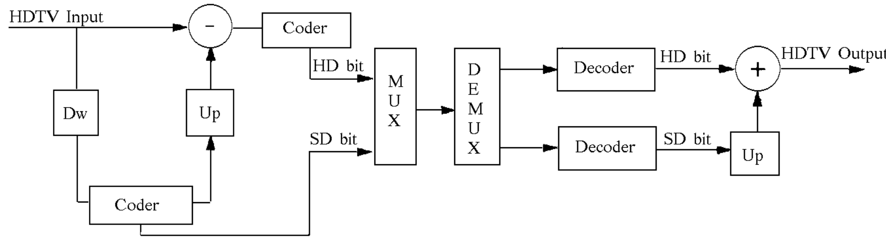

Fig. 3. Classical configuration for the pel split

the originals to yield inputs for higher layers.

By D_w we mean downsampling, while U_p represents upsampling. SDTV is standard definition TV.

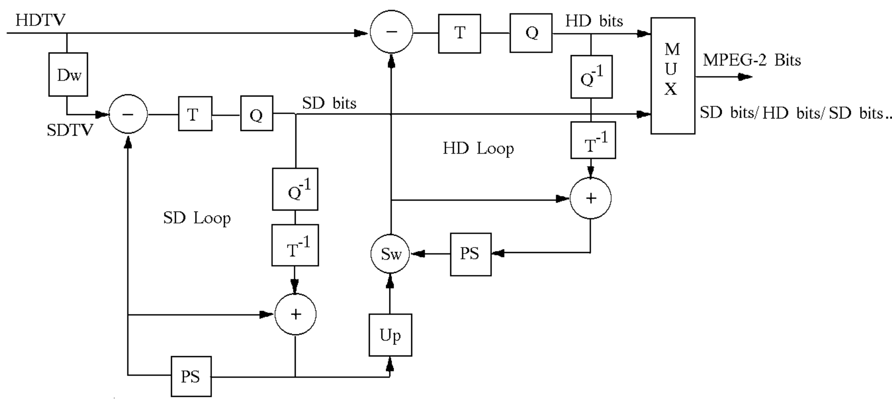

Fig. 4. Two layer version of the multilayer scheme

Figure 4 shows a two–layer version of the multilayer scheme. In fact, there are two loops, each coding a particular picture resolution. The sources for each loop are derived by appropriate downsampling of the HDTV input. Let be: T–transform, Q–quantizer, D_w–downsampling, PS–picture store, T^{-1}–inverse transform, Q^{-1}–inverse quantizer, U_p–upsampling, and S_w–switch. Coding in each loop is independent, but a prediction is available from the SDTV loop to the HDTV loop. Each produces a complete MPEG video bit stream from sequence header down to block level. These, bit streams are multiplexed together at the MPEG system level at suitable intervals to prevent the introduction of undue delay.

The bottom layer is coded independently, the only link with the top layer being the upsampling which is carried out on the reconstructed bottom layer pictures. Therefore, it is possible to code the bottom layer with any algorithm, such as H.261, MPEG-1, or MPEG-2.

5.5. Layred coding of block motion vectors

A layered coding scheme that utilizes motion vector refinements, as well as the pixel difference requantization was proposed in [15]. Blocks in the baseline layer are motion compensated in the usual sense. For the purpose of refinement, the block is then divided into a set of subblocks and the motion of each of these smaller subblocks is also predicted, leading to more accurate prediction of overall motion details. In the case where the baseline layer is coded video, additional motion information will be sent in the refinement layer describing the displacements of the individual blocks within a motion compensated macroblock, leading to more accurate motion prediction. In addition to the macroblock level motion prediction normally carried out in baseline coding process, block level motion prediction is also performed on a parallel operation. For each of the four blocks in a single macroblock, the macroblock's motion vector is subtracted from the individual block motion vectors, the residuals being the block level refinements to the original macroblock motion vector.

At the receiver, the baseline coded video and enhancement bit streams are decoded concurently. The coded video stream, upon decoding, gives the motion compensated location for each macroblock and, thus, the four blocks that comprise the macroblock in question. If the enhancement information for a particular macroblock in a frame indicates additional motion vector refinement is present, then the baseline decoded blocks are replaced by the decoded blocks pointed to by the block level motion vectors, providing a more accurate description of pixel motion.

6. Research directions

Earlier work on packet video was concerned with the effects of packet losses and means for interpolating the signal in the presence of packet losses. This line of research is being reactivated today, because of the increasing importance of the ATM packet technology. An important concept here is that of layered coding. The output of the source encoder is divided into cells of varying significance, typically with two layers, or levels of it. When the packet network is congested, the idea is to drop cells of lower priority. In the case of uncompressed PCM data, the prioritization of encoder bits is straightforward. But in low bit rate coders, the optimization of the communication network will depend on perceptual cues for cell layering and subjective methods for measuring the user acceptance of layered coding.

People first become interested in layered coding as a means of providing resilience against cell loss errors in ATM. The principle of layered coding for error resilience is to separate the coded data into two or more parts which are accorded differing priorities in the network. Data whose loss would be more visible in the decoded pictures are given higher priority. This might include addressing information, the lower spatial frequencies and motion vectors.

Error resilience is also highly desirable in broadcasting environments, such as satellite and terrestrial distribution. Although error correction coding can be employed to reduce transmission errors, there is a point of decreasing returns where it is more economic, in bit rate terms and probably financial terms, to accept that some errors may reach the video decoder and to attempt to conceal them there. Layering is a powerful tool in the overall design of such systems.

An important aspect of terrestrial broadcasting is graceful degradation. As the receivers are placed near the limits of transmitter range, the receivers picture should degrade slowly from high definition resolution to standard definition resolution. The lower layer bit stream is heavily protected through either modulation techniques or error correction. The higher level is less well protected and will be more prone to errors.

More sophisticated techniques can be used through the use of scaled versions of the lower layer motion vectors to provide compensation in the upper layer.

7. Conclusions

A layered VBR coding model offers a number of advantages, which are compatible and consistent with the characteristics of the ATM network. A number of experimental packet video coding techniques has been reviewed. It would seem that the frequency domain techniques generally offer a layered structure, immunity to packet loss and packet defects, flexibility to quality and bit rate selection, as well as congestion control.

In order to limit the loss of cells and minimize its effects on video quality degradation, layered coding technique can be recommended. Each layer when upsampled can, but need not, contribute to a prediction of the next layer. This technique permits a very great deal of flexibility. It is also possible to utilize a different balance between coding efficiency and implementation complexity in each layer.

REFERENCES

1. C. Gonzales and E. Viscito: *Flexibly scalable digital video coding.* Signal Processing: Image Communication, vol.5, Nos.1-2, pp. 5-20, February 1993.
2. D. Lee and K. Tzou: *Hierarchical DCT coding of HDTV for ATM networks.* Proceedings ICASSP-90, IEEE, 1990, pp. 2249-2252.
3. D. Morisson: *Standardization by ISO/MPEG of digital video coding for storage applications.* N.Kenyon and C.Nightingale, eds., Audiovisual Telecommunications. Chapman and Hall, London, 1992, pp. 177-193.
4. D. Morisson and D. Beaumont: *Two–layer video coding for ATM networks.* Signal Processing: Image Communication, vol.3, No.2-3, pp. 179-195, June 1991.
5. F. Kishino, K. Munabe, Y. Hayashi and H. Yasuda: *Variable bit rate coding of video signals for ATM networks.* IEEE Journal Selected Areas in Communications, vol.7, No.5, pp. 801-806, June 1989.
6. G. Karlson and M. Veterli: *Subband coding of video signals for packet switched networks.* Proc. SPIE Visual Communication and Image Processing II, Cambridge, MA, USA, vol.845, October 1987, pp. 446–456.
7. J. Van der Meer: *The full motion system for CDI.* IEEE Trans. Consumer Electronics, vol.38, November 1992, No.4, pp 910–920.
8. L. Vandenddorpe: *Hierarchical transform and subband coding of video signals.* Signal Processing: Image Communications, vol.4, Juny 1992, No.3, pp 245–262.
9. M. Ghanbari: *An adapted H.261 two layer video codec for ATM networks.* IEEE Trans. Communications, vol.40, September 1992, No.9, pp 1481–1490.
10. M Ghanbari: *Two-layer coding of video signals for VBR networks.* IEE Journal Selected Areas Communications, vol.7, June 1989, No.5, pp 771–781.
11. M. Nomura, T. Fujii and N. Ohta: *Layered packet loss protection for variablerate video coding using DCT.* II International Workshop on Packet Video, Torino, Italy, September 1988.
12. N. Ohta, M .Nomura and T. Fujii: *Characteristics of variable rate video coding with motion compensated DCT for burst/packetized communications.* IEEE COMSOC International Workshop, Osaka, Japan, paper 4-3, November 1987.
13. R. Ter Host, A. Koster, K. Rijkse. E. Fert, G. Nocture and L. Tranchard: *MUPOS: A multi purpose coding scheme.* Signal Processing: Image Communication, vol.5, February 1993., Nos.1–2, pp. 75–89.
14. S. Blake, S. Rayala, F. Gong and T. Mitchell: *Efficient techniques for two layer coding of video sequences.* Proc.ICIP–94, vol.1, 1994, pp. 253–257.
15. V. Rhee and J. Gibson: *Block–level motion vector refinement in layered video.* 1995 IEEE Workshop on Nonlinear Signal and Image Processing, vol.I, Neos Marmaras, Greece, June 1995, pp 194–197.

16. Z Bojković: *Recent tends in packet video transmission on ATM networks.* Proc. of the Basis of Electronics Workshop – 1994, invited paper, Cluj–Napoca, Romania, October 1994, pp.17–105.

17. Z. Bojković, S. Lazović and M. Bakmaz: *Advance digital HDTV over ATM networks.* Facta Universitatis, Series Electronics and Energetics, vol.7, 1994, pp. 25–37.