A complete non predictive video compression scheme based on a 3D to 2D geometric transform

Tarek Ouni*, Walid Ayedi and Mohamed Abid

National Engineering School of Sfax,
Road Sokkra, Km 3 Sfax, Tunisia
E-mail: tarek.ouni@gmail.com
E-mail: ayedi.walid@gmail.com
E-mail: mohamed.abid@enis.rnu.tn
*Corresponding author

Abstract: The proposed method consists on '3D-to-2D' transformation of the temporal frames that allows exploring the temporal redundancy of the video using 2D transforms and avoiding the computationally demanding motion compensation step. This transformation turns video spatial temporal correlation into high spatial correlation. In this paper, we explore the proposed method performances and try to better show what it actually offers to users. The paper presents also the extensions chosen to reduce the perceived artefacts and increase the perceptual as well as objective (PSNR) decoded video quality, which is actually competitive with state-of-the-art video coders, especially when operating complexity is taken into account.

Keywords: non predictive video coding; low complexity; temporal decomposition; correlation; DCT; discrete cosine transformation; DWT.

Reference to this paper should be made as follows: Ouni, T., Ayedi, W. and Abid, M. (2011) 'A complete non predictive video compression scheme based on a 3D to 2D geometric transform', *Int. J. Signal and Imaging Systems Engineering*

Biographical notes: Tarek Ouni received the MS Degree in New technologies of dedicated computer systems from National Engineering School of Sfax, Tunisia in 2006. He is a PhD student at Computer & Embedded Systems Laboratory, affiliated to SFAX University, Tunisia. His research focuses on video processing, and compression techniques, multimedia embedded systems and HW/SW architectures.

Walid Ayedi received the MS Degree from National Engineering School of Sfax, Tunisia in 2008. He is a PhD student at Computer & Embedded Systems Laboratory, Tunisia. He is actually an invited PhD student in University of Technology of Troyes, France, at the Laboratory of Systems Modelling and Dependability. His research interests include image analysis, computer vision and machine learning.

Mohamed Abid received the PhD Degree from the National Institute of Applied Sciences, Toulouse (France) in 1989 and the 'thèse d'état' Degree from the National School of Engineering of Tunis (Tunisia) in 2000 in the area of Computer Engineering & Microelectronics. Actually, he is Head of "Computer Embedded System" Laboratory CES-ENIS, Tunisia, he is working now as a Professor at the Engineering National School of Sfax (ENIS), University of Sfax, Tunisia. His current research interests include: hardware-software co-design, system on chip, reconfigurable system, and embedded system, etc. He has also been investigating the design and implementation issues of FPGA embedded systems.

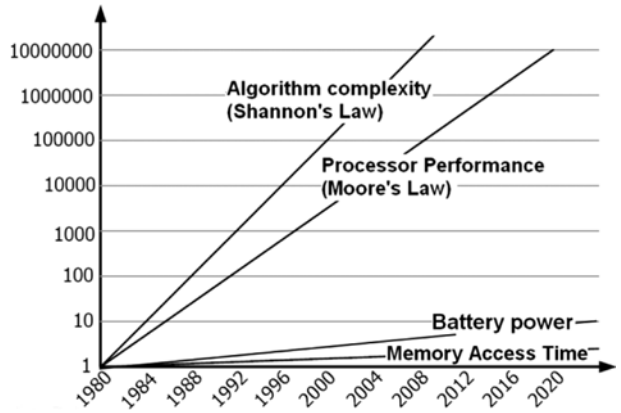
1 Introduction

Video compression is a critical component of many multimedia applications available today. For applications such as DVD, digital television broadcasting, satellite television, internet video streaming, video conferencing, video security, and digital camcorders, limited transmission bandwidth or storage capacity stresses the demand for higher video compression ratios. To address these different scenarios, many video compression standards have been ratified over the past decade. However, such standards are becoming increasingly complex and computationally intensive. Consequently, the current technology has become unable to meet these applications requirements (Molino and Vacca, 2004). Actually, if we refer to the integration density evolution's law (Moore's Law) and algorithmic complexity evolution's law (Shannon's Law) for the evolution of multimedia applications (cf. Figure 1), we can assert that the gap between the two curves keeps on increasing (Kocovic, 2008).

For embedded system industry and portable digital video applications, further particular challenges emerged,

such as energy consumption and real-time constraint. Video compression implementation for such applications becomes an immense trouble especially with the insufficient evolution in chemical energy storage, and memory access time performance (cf. Figure 1). High performance generally entails power sacrifices. The point is to find the ideal equilibrium between the two within a specific design.

Figure 1 Complexity and technology gap



This has traditionally been tackled in embedded systems industry by including a number of design and process strategies for achieving economical performance at system-level, chip-level, and even transistor-level designs, to achieve performance with long battery life (Molino and Vacca, 2004).

In this context, one solution that has been vastly emphasised in the last decade consists in exploring different hardware/software architectures based implementations. Such solutions are still unable to bridge this gap between contemporary technology and the complexity of these systems because of the supplementary costs of the architectures space exploration (time to market).

We only need to point out the H264/MPEG-4 AVC example that has been proposed as a standard in 2003 whereas its implementation is still an indescribable barrier especially for embedded systems.

In addition to this complexity, the MPEG standard is not primarily taking into consideration some new applications requested features such as scalability. To meet these applications requirements, supplementary complex modules have been incorporated into this standard, but it was always at the price of compression performance. In fact, MPEG-4 SVC[1] is less efficient in terms of compression than its predecessor: MPEG-4 AVC (Marpe et al., 2006). Actually, conventional methods are alternatives to a certain extent, because they were developed with different aims and are tailored to specific applications. Overall, the range of available standardised video compression methods covers almost all applications. Some compression algorithms authorise a degree of variation within the standard to meet technology constraints, which decreases considerably the coder performances. As a matter of fact, the ongoing increasing size of the Shannon-Moore gap means that such alternatives alone are not sufficient to close this gap.

One more path that seems to be promising is to sink into other less complex coding schemes, supported by current technology and satisfying the new applications requirements. In this context, researches twisted towards the design of new non predictive coders that intended to reduce the coding complexity by eradicating the motion estimation-compensation module. Most ones looked for the 3D-transforms exploitation in order to utilise temporal redundancy. Coder based on 3D-transform produces video compression ratio which is close to the motion estimation based coding one but with minor processing complexity (Gokturk and Aaron, 2002; Servais, 1997; Burg, 2000; Koivusaari and Takala, 2005).

Nonetheless, 3D-transform based video compression methods, process the 3D video signal temporal and spatial redundancies in the same way. This can decrease these methods efficiency as pixels' values deviations in spatial or temporal directions are not regular and thus, temporal and spatial redundancies have not the same relevance. It is acknowledged that the temporal redundancies are more pertinent than the spatial ones (Gokturk and Aaron, 2002). Hence it is important as possible to achieve more efficient compression by exploiting more and more the redundancies in the temporal dimension; this is the crucial reason of the suggested method.

In this paper, a new non-predictive video coding scheme is presented. It consists in 3D to 2D transformation of the video frames; it will then investigate the video temporal redundancy using 2D-transforms and avoids the computationally demanding motion-compensation step. Above all, the used method projects temporal redundancy of each pictures group and combines it with spatial redundancy into one 2D representation with high spatial correlation. Then, the new representation – called Accordion representation – will be compressed as a still image using 2D-transform based coder. Two video coding schemes were presented respectively in Ouni et al. (2009, 2010). The two coders are derived from the proposed approach, the first is based on 2D-DCT transform, it was called ACC-JPEG, and the second is based on 2D-DWT, it was called ACC-JPEG 2000.

This paper aims at providing a comparison of various features that can be expected from video compression methods derived from the proposed approach. Some features come from the Accordion representation itself, others come from the used image coding algorithms (especially JPEG and JPEG 2000). This synthetic study provides a detail description of the proposed approach's background, it shows the improved correlation given by the Accordion transform compared to traditional spatial decomposition based methods, and it presents performances analysis with detail subjective quality comparison between the proposed coders and the existing standards. To do so, many aspects have been considered including genericity of

the algorithm to code different types of data in lossless and lossy way, and features such as error resiliency, complexity, scalability, etc.

The rest of the paper is organised as follows. In Section 2, we present an overview on existing non predictive video coding techniques. In Section 3, we review the used approach basics. In Section 4, we illustrate the 3D to 2D proposed transform based method. Experimental results are presented in Section 5. Section 6 presents the analysis of the methods characteristics and limitations. Finally, the conclusion is drawn in the last section.

2 Non predictive video coding approaches

The non-predictive video coding is a new branch of an emerging research area in video coding, where the motion estimation/compensation or prediction step in the temporal domain is omitted. One direction was to look for the exploitation of 3D transforms to exploit temporal redundancy. Actually, 3D transform rely on spatial and temporal decomposition of the video source. Spatial decomposition tries to capture most spatial redundancies in neighbouring pixels in a frame. Parenthetically, most conventional video coding scheme is use spatial decomposition. On the other hand, temporal decomposition tries to capture most temporal redundancies in neighbouring frames. Thus, when applying waveform transform along the temporal direction, could benefit from this temporal redundancy. Moreover, it may provide larger decorrelation of data and thus higher energy compaction comparing to ME/MC approach (Molino and Vacca, 2004). Therefore, excluding of ME from the compression process could provide ability of real-time video coding services for mobile devices with restricted computational resources (Molino and Vacca, 2004).

2.1 3D-DCT based algorithms

A 3D-DCT VC seems to be one of the most promising solutions for limited processing power and energy computation platforms (Servais, 1997; Burg, 2000). Authors have argued that 3D DCT can be effective in compressing video sequences, especially those with little motion (Servais and De Jager, 1997; Lee et al., 1997a; Burg and Keller, 1999, 2000).

Although it stays behind the most advanced MC-DCT appoaches, 3D-DCT VC offers instead significant reduction of computational complexity, improved algorithmic and implementation integrity (Burg, 2000; Song et al., 2000). In addition to this, it is free of error propagation; video coding sides are symmetric and processing delays are reliable (Koivusaari and Takala, 2005). However, it introduces blocking artefacts in reconstructed video. These artefacts become noticeably visible, as discontinuities in the spatial and temporal dimensions of highly compressed video due to the coarse quantisation of transform coefficients. Temporal discontinuities appear as jerky motion of video or as a periodical quality changes within a Group of Frames (GoF) that makes them especially noticeable and annoying. Thus, a postprocessing of 3D-DCT coded video targeting the blocking artefacts reduction is necessary. Unfortunately, there has not been much research on 3D-DCT VC deblocking unlike deblocking methods for two-dimensional DCT block-based compression schemes which have been widely studied (Rusanovskyy and Egiazarian, 2005).

2.2 3D-wavelet based algorithms

3D wavelet transforms have also been investigated as a method for video compression (Sampson et al., 1995; Lee et al., 1997b; Beong and Pearlman, 1997; Vass et al., 1998; Lin and Liu, 1999; Bernabe et al., 2000; Lazar and Averbuch, 2001; Seigneurbieux and Xiong, 2001). Research in 3D Sub-Band Coders (SBC) uses wavelet transform in the temporal axis which significantly reduces computation and design complexity and provides better compression performance than current block-based motion compensated predictive methods. In fact, image and video coding methods using wavelet transform have generated much interest in scientific community as alternative methods to DCT based compression schemes such as MPEG standards. Classical three dimensional wavelet transform algorithm performing into video sequences presents a signal representation very suitable for video compression. These methods perform temporal and spatial decomposition by wavelet transform into original video (Moyano et al., 2001).

Moreover, 3D wavelet decomposition has been combined with EZW (Shapiro, 1993) or SPHIT (Said and Pearlman, 1996) coding to achieve good quality compression. A significant advantage of 3D wavelet transforms or 3D subband coding over other transform methods is that the resulting encoded video is highly scalable, both in the spatial and temporal domains. The scalability and multiresolution properties of wavelet transforms have been utilised in previous work in designing scalable video codecs.

In recent years, research in the use of 3D DWT is continuing. In this context, Akbari and Soraghan (2003) introduces a non-predictive video coding scheme called Adaptive Joint Sub-band Vector Quantisation (AJSVQ) by joining the significant subbands coefficients within a GOF. In Akbari and Soraghan (2003) a group of four frames represent a GOF, and the significant vectors are processed using an Adaptive Vector Quantisation (AVQ) technique as presented in Voukelatos and Soraghan (1997). This eliminates the need for prediction process of the video codec.

In noiseless environment, the non-predictive AJSVQ video codec is reported to perform approximately 2 dBs better than JPEG2000 image compression standard (Lawson and Zhu, 2002) which compresses individual frame from a test video sequence Miss America to emulate the motion non-predictive JPEG2000 video coding scheme. In Akbari and Soraghan (2003), Voukelatos and Soraghan (1997), Lawson and Zhu (2002) AVQ is used in coding the high frequency subbands. In their work the significant coefficients are selected after comparing to a certain

threshold. The quantisation of the significant coefficients uses the LBG algorithm which involves high computation to generate the codebook. The computational load can be significantly reduced by using lattice vector quantisation as in Conway and Sloane (1988). In Salleh and Soraghan (2006) a new non-predictive video coding scheme is presented. The subbands from a GOF are joined and process together to exploit their spatial-temporal redundancies as in Akbari and Soraghan (2003). This work differs from Akbari and Soraghan (2003) in two ways. First, the high frequency subband thresholds are optimised using an adaptive thresholding algorithm for better identification of the significant coefficients. Secondly, the Multi-stage Lattice Vector Quantisation (MLVQ) technique derived from Salleh and Soraghan (2005) is used to quantise the significant coefficients instead of the AVQ. The advantage of having lattice vector quantisation is to reduce the computational load as LBG algorithm demands higher amount of computation to generate codebook. Thus facilitates the new video coding scheme to have multistage processes which reduce quantisation errors and enhance the reconstructed frame quality. The results obtained for the new video coding scheme with adaptive threshold are approximately 4 dBs better than the AJSVQ for Miss America test sequence. It also performs almost 1 dB better as compared to MPEG-2 with I frames only for other test sequences.

2.3 Synthesis

In this section, we focused on 3D transform based non-predictive video compression methods. Despite their advantages, these methods are not efficient overall and still suffer from several weaknesses. In fact, 3D transform based algorithms are still considered as complex processes. They are becoming more and more demanding in computational resources (processing and memory) when many frames have to be considered to be processed at time. Indeed, such algorithms require access to all frames in the GOF to perform the temporal transform which limit their practical implementation especially when the amount of frames in the used GOF is large (Kutil and Uhl, 1999; Khalil et al., 1999). Even when the use of large virtual memory space, the transform process may severely be affected by the requirement of making all information available not found in the main memory. In other hand, if we set a maximum allowable delay, we are putting an upper limit on the number of frames that can be used for temporal decomposition resulting in lower compression efficiency (Khalil et al., 1999; Sikora, 2005). In fact, limiting the transformation to small cubes ($8 \times 8 \times 8$ blocks) also limits the potential for compression since all missed correlations between pixels and frames could be found beyond the 8-pixel surroundings. Therefore, the exploitation of both spatial and temporal redundancy is limited by the number of frames in the GOF and it depends on both available memory resources and the application maximum allowable delay. Moreover, 3D transform consist on one extension of the 2D transform with the temporal axis being the third dimension. It consists on applying the waveform transform on the spatial and the temporal direction. Thus most 3D transform based video compression methods process temporal and spatial redundancies in the 3D video signal in the same way. This can reduce the efficiency of these methods as pixel's values variation in spatial or temporal dimensions is not uniform and so, temporal and spatial redundancies have not the same pertinence.

To resume, we can state that 3D transform based approaches suffer from many limitations which could be an obstacle for its evolution. Obviously a naive application of 2D transforms into the third dimension will not always ensure better compression. Thus more subtle transformations that have to be developed take into account the major characteristics of real world videos which show that the temporal redundancies are more relevant than spatial one (Gokturk and Aaron, 2002). So, the temporal redundancy must be more exploited than the spatial one. It is possible to achieve more efficient compression by more exploiting the redundancies in the temporal domain; in other hand, if the exploitation of temporal redundancy is limited, spatial one is not, so, we foresee that applying some transform with no spatial limit can provide some improvement over traditionally MPEG and 3D transform based coders.

3 Background of the proposed method: exploration of temporal redundancy

Video compression has its own characteristics that make it quite different from still image compression. The major difference lies in the interframe correlation exploitation that exists between successive frames in video sequences, in addition to the intraframe correlation that exists within each frame. The interframe correlation is also referred to as temporal redundancy, while the intraframe correlation is referred to as spatial redundancy. In order to achieve coding efficiency, we need to remove these redundancies for video compression. To do it, we first need to understand these redundancies. Consider a video sequence taken in a videophone service. There, the camera is static most of the time. A typical scene is a head-and-shoulder view of a person imposed on a background. In this type of video sequence the background is usually static. Only the speaker is experiencing motion, which is not severe. Therefore, there is a strong similarity between successive frames, that is, a strong adjacent-frame correlation. In other words, there is a strong interframe correlation. It was reported by Sikora (2005) that when using videophone-like signals with moderate motion in the scene, on average, less than one-tenth of the elements change between frames by an amount which exceeds 1% of the peak signal. Here, a 1% change is regarded as significant. In Kretzmer (1952), experiments on the first 40 frames of the Miss America sequence support this observation. Two successive frames of the sequence, frames 24 and 25, are shown in Figure 2.

Figure 2 Two consecutive frames of the Miss America sequence

Considering a video sequence generated in a television broadcast, it is well known that television signals are generated with a scene scanned in a particular manner in order to maintain a steady picture for a human being to view, regardless of whether there is a scenery change or not. That is, even if there is no change from one frame to the next, the scene is still scanned constantly. Hence there is a great deal of frame-to-frame correlation (Mounts, 1968; Haskell and Limb, 1972; Haskell et al., 1972). In TV broadcasts, the camera is most likely not static, and it may be panned, tilted, and zoomed. Furthermore, more movement is involved in the scene. As long as the TV frames are taken densely enough, then changes between successive frames are due mainly to the apparent motion of the objects in the scene that takes place during the frame intervals. This implies that there is also a high correlation between sequential frames. In other words, there is an interframe redundancy (interpixel redundancy between pixels in successive frames). There is more correlation between television picture elements along the frame-to-frame temporal dimension than there is between adjacent elements in a single frame along the spatial dimension. That is, there is generally more interframe correlation than intraframe correlation. Actually, taking advantage of the interframe correlation leads to video data compression (Netravali and Robbins, 1979).

In the same context, experiments conducted in Gokturk and Aaron (2002) support the hypothesis which says that video stream contains more temporal redundancies than spatial ones. Figure 3 reveals the high correlation in the temporal representation of 3D video signal compared to spatial one (Gokturk and Aaron, 2002).

Figure 3(a) illustrate the spatial correlation into pixels of one frame extracted from the video sequence 'salesman'. Figure 3(b) and (c) respectively show the temporal variations through time of the 80th line and the 100th column.

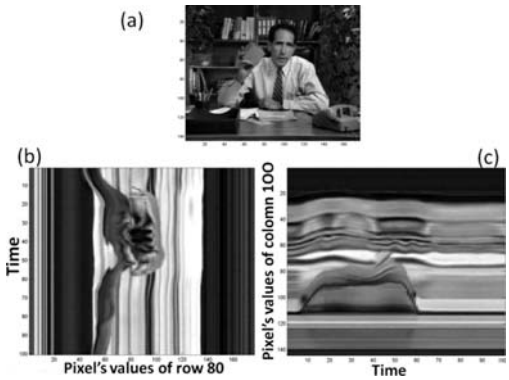
Until now, we are giving a visual demonstration of the relevance of temporal redundancy in video signal. In the next we present some experiments results showing that coding which is based on temporal decomposition is more efficient than this on spatial decomposition.

Conducted experiments consist in coding different $8 \times 8 \times 8$ blocks, with or without movement, with spatial and temporal decompositions, using JPEG technique. Figure 4 illustrates the process described above.

To check our assumption saying that temporal redundancy is more significant than spatial one in a video sequence, we have coding different $8 \times 8 \times 8$ blocks, with or without movement, with spatial and temporal decompositions, using JPEG technique.

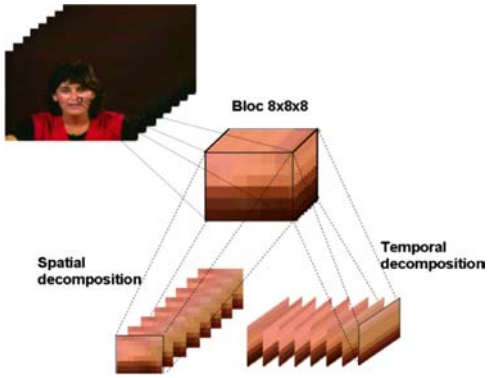
First, we have chosen a sequence of 8 frames from 25th frame to 32nd frame and we have extracted the $8 \times 8 \times 8$ block from space coordinates (80, 80). This block was selected because it contains a part of the lip's movement. The experiment's results show that the size of blocks coded in JPEG following a temporal decomposition (924 bits) is smaller than the size of blocks coded following spatial decomposition (1167 bits). The same experience was done with static blocks (without movements) and it gave more significant results.

Figure 3 Spatial and temporal correlation in salesman sequence



Source: Gokturk and Aaron (2002)

Figure 4 Temporal and spatial decomposition (see online version for colours)



All this can only confirm one hypothesis which seems obvious from the beginning, at least in our eyes; The 3D video signal is more correlated in temporal domain than in spatial one; thus, the 3D video signal variation is much less in the temporal direction than the spatial one.

This could be translated by the following expression: for one pixel $I(x, y, t)$ where:

I: Pixel intensity value

$x; y$: Pixel space coordinate

t: Time (video instance).

We could generally have:

$$I(x; y; t) - I(x; y; t + \mathrm{d}t) < I(x; y; t) - I(x + \mathrm{d}x; y; t). \quad (1)$$

This assumption will be the basis of the proposed method where we will try to put pixels – which have a very high temporal correlation – in spatial adjacency. Thus, video data will be presented with high correlated form which exploits both temporal and spatial redundancies in video signal with appropriate portion that put in priority the temporal redundancy exploitation.

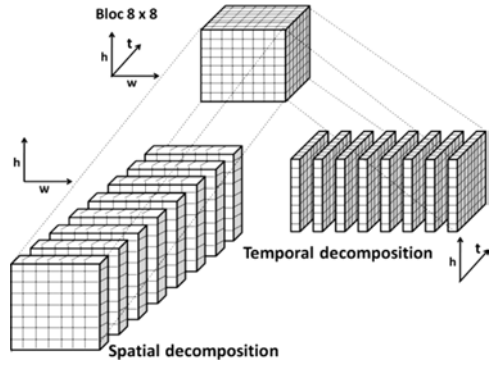
4 Principle of the accordion representation

The key in efficient image and video compression is to explore source correlation so as to find a compact representation of source data. Based on this principle, the proposed method stands on 3D to 2D transformation of the video frames that allows exploring the temporal redundancy of the video in additional of spatial one. This transformation allows representing video data with high correlated form. It consists in projecting temporal redundancy of each group of pictures into spatial domain to be combined with spatial redundancy in one representation with a high spatial correlation (Ouni et al., 2010; Fryza, 2002) in order to exploit the 2D transforms characteristics (DCT, DWT …) in temporal domain.

4.1 Accordion representation

The input of our encoder is the so called video cube which is made up of a GoF. This cube will be decomposed into temporal frames which will be gathered into one 2D representation. Temporal frames are formed by gathering the video cube pixels which have the same column index (cf. Figure 5). These frames will be projected on 2D representation (further called 'IACC' frame) while reversing the direction of odd frames, i.e., the odd temporal frames will be turned over horizontally in order to more exploit the spatial correlation of the video cube frames extremities (cf. Figure 6).

Figure 5 Temporal and spatial decomposition



The Accordion representation tends to put in spatial adjacency the pixels having the same coordinates in the different frames of the video cube. In this way, Accordion representation also minimises the distances between the pixels spatially correlated in the source. Figure 7 illustrates the principle of this representation via an explicative example.

For example, columns 0 and 1 which are adjacent in IACC are also temporally adjacent in the original video sequence, and the columns 2 and 3 that are adjacent in IACC are either in spatial adjacency in the original video sequence.

Figure 6 Accordion representation

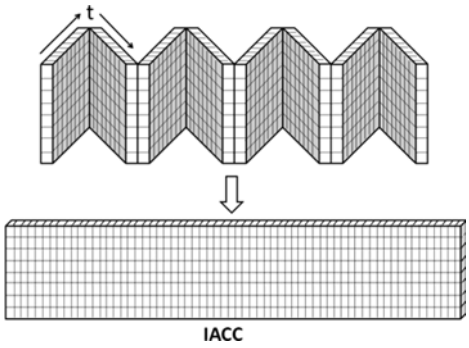


Figure 7 Accordion representation example (see online version for colours)

I1

y\x	0	1	2
0	1	1	1
1	1	1	1
2	1	1	1

I2

y\x	0	1	2
0	2	2	2
1	2	2	2
2	2	2	2

I3

y\x	0	1	2
0	3	3	3
1	3	3	3
2	3	3	3

IACC

y\x	0	1	2	3	4	5	6	7	8
0	1	2	3	3	2	1	1	2	3
1	1	2	3	3	2	1	1	2	3
2	1	2	3	3	2	1	1	2	3

Subsequently, we will show that this representation based on temporal decomposition has more correlation than in the source images.

Let $P(x, y, t)$ any pixel belonging to $I1$, $P'(x + dx, y, t)$ its neighbouring pixel. For the same pixel P, its neighbour in IACC is $P''(x, y, t + dt)$ (or with less probability $P''(x + dx, y, t)$).

According to the so mentioned hypothesis (1), the intensity difference between P and P'' is less than that between P and P' (in the worst condition they are equal). So P and P'' are more correlated than P and P' and then we can state that IACC contains more correlation than original frames.

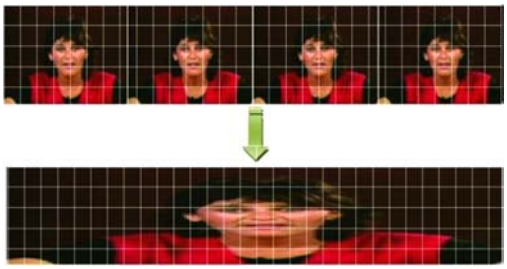
Now, we will check this statement with a real example:

First, Accordion transform is applied to four succeeding frames which are extracted from Miss America sequence. Figure 8 visibly shows the strong correlation in the obtained Accordion representation.

To more prove that the accordion representation provides more correlation than the source images, we will apply the autocorrelation concept on some frames.

By the way, the autocorrelation is a mathematical tool to measure the cross-correlation of a one-dimensional signal itself. The autocorrelation can detect repeated patterns and internal dependencies in a signal. So, a strongly regular and consistent image will have a strong autocorrelation.

Figure 8 Accordion representation made of 4 frames extracted from Miss America sequence (see online version for colours)



In the case of images, we must transform our image into a 1D signal. To do this, we will scan the frame with a curve that goes through all the pixels. The simplest one is to scan image lines after lines or columns after columns.

We will measure the autocorrelation in some frames extracted from the sequence 'MISS AMERICA' in order to compare with the one measured in the Accordion representation as it shown in Figure 9.

Figure 9 Measured autocorrelation comparison between IACC representation and original frame (Miss America) (see online version for colours)

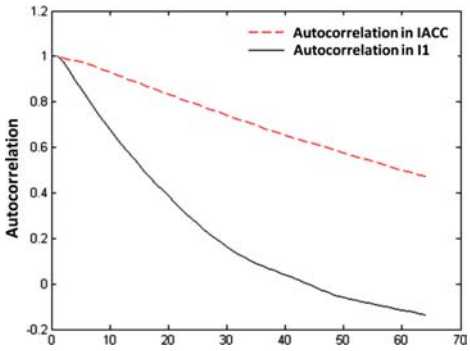


Figure 9 shows a comparison using the local correlation between a source image and IACC representation for MISS AMERICA sequence.

In this experiment we applied the autocorrelation function while browsing the image line by line.

This experiments show the superiority of the autocorrelation in the accordion representation, that is mean that the presented representation are strongly regular and contain a strong correlation which could increase the eventual applied transform impact.

4.2 Accordion analytic representation

The Accordion representation is obtained following a process having as input the GOFF frames ($I1 \ldots N$) and has as output the resulting IACC representation. The inverse process has as input the IACC representation and as output the coded frames ($I1 \ldots N$).

The first algorithm describes how to make Accordion representation (labelled ACC), The second algorithm represents the inverse process (labelled ACC-1).

These two processes are described by the following algorithms:

Algorithm 1: ACC Algorithm

```
1: for x from 0 to (L * N) - 1 do
2: for y from 0 to (H – 1) do
3: if ((x div N) mod 2) != 0 then
4: n=(N-1) - (x mod N)
5: else
6: n=x mod N
7: end if
8: IACC(x; y)= In (x div N,y)
9: end for
10: end for
```

Algorithm 2: Algorithm of ACC-1

```
1: for n from 0 to N - 1 do
2: for x from 0 to L - 1 do
3: for y from 0 to H - 1 do
4: if (x mod 2) != 0 then
5: XACC= (N -1) - n (x*N)
6: else
7: XACC= n(x _ N)
8: end if
9: In(x,y)=IACC(XACC; y)
10: end for
11: end for
12: end for
```

These two processes analysis leads to the following formulas:

ACC formulas:

$$\mathrm{IACC}(x, y) = \mathrm{In}(x \text{ div } N, y). \quad (2)$$

With $n = (x \bmod 2)\,(N-1) + (1 - 2(x \bmod 2))(x \bmod N)$.

ACC inverse formulas:

$$\mathrm{In}(x, y) = \mathrm{IACC}(\mathrm{XACC}, y). \quad (3)$$

With $\mathrm{XACC} = (x \bmod 2)\,(N-1) + n\,(1 - 2(x \bmod 2)) + x.N$.

Let us note that:

- N is the number of frames of a GOF.
- $\mathrm{IACC}(x, y)$ is the pixel intensity with the coordinates x, y according to accordion representation repair.
- $\mathrm{In}(x, y)$ is the intensity of pixel situated in the Nth frame in the original video source.

4.3 Accordion representation based coding schemes

In this part, we present the coding diagram based on the Accordion representation.

First, the video encoder takes a video sequence and passes it to a frame buffer in order to construct volumetric images by combining N frames into a stack. Then, the obtained stack will be transformed to form the accordion representation (IACC). Here N is the constructed stack depth (N is 8 in our experiments).

Next, each IACC will be divided into $N \times N$ blocks to be processed furthermore by the eventual used 2D transform. The encoder block diagram of the Accordion based compression algorithm is in Figure 10.

Two Accordion representation based coding schemes have been designed, the first, called ACC_JPEG, uses DCT transform as it is described in Shi and Sun (2008),

the second, called ACC_JPEG 2000, uses DWT transform (Ouni et al., 2009).

ACC_JPEG Coding process as follows:

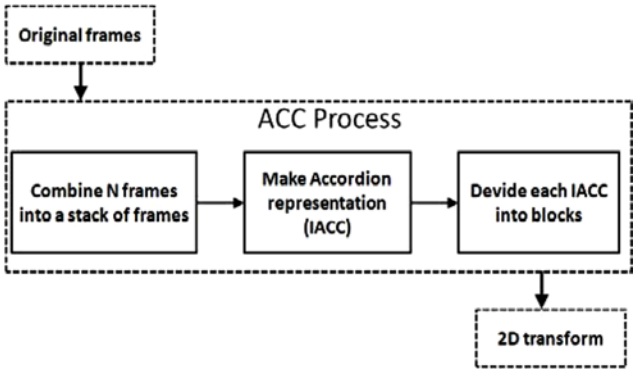
- decomposition of the video in GOF
- GOF Accordion representation
- decomposition of the resulting 'IACC' frame into 8×8 blocks.
- For each 8×8 block:
  - Discrete Cosine Transformation (DCT)
  - quantisation of the obtained coefficients
  - course in zigzag of the quantised coefficients
  - entropic CODING of the coefficients (RLE, Huffman).

The proposed ACC-JPEG2000 coding scheme follows the succeeding steps:

- the decomposition of video sequence into GOF
- GOF Accordion representation
- decomposition of the resulting 'IACC' frame into 64×64 blocks called tiles.
- For each 64×64 tile (block):
  - DWT transform
  - quantisation of obtained coefficients
  - arithmetic coding of obtained coefficients.

It is pointed out that, in our experiments, we use five levels of wavelet decomposition which is mostly sufficient for CIF sequences.

Figure 10 Block diagram of the Accordion representation based encoder



5 Experiments

To show the effectiveness of our approach, we compare the proposed coding methods to the well-known state-of-the-art technologies: M-JPEG, M-JPEG 2000 and MPEG-4.

We choose these technologies for many reasons. In one hand, M-JPEG and M-JPEG 2000 are the most popular video Intra-coding techniques in imaging and video applications. Moreover, M-JPEG and M-JPEG 2000 are respectively comparable to ACC-JPEG and ACC-JPEG 2000 from the complexity point of view. In the other hand, MPEG-4 is among the most popular video inter-coding techniques in contemporary multimedia applications and it is quite judicious to compare the set of features offered by the proposed method to those offered by the other video inter-coding methods.

This choice is taken in order to prove that the proposed method outperforms the comparable existing methods from the complexity point of view in one hand, and offers less complexity compared to the comparable existing methods from the quality point of view in the other. Many experiences have been conducted in order to study the performances of our method; we had chosen different kinds of benchmarks.

We start by 'Hall Monitor' benchmark as the proposed method had been primary tested in video surveillance applications. The two previously proposed coders are evaluated. We start by ACC-JPEG.

5.1 ACC_JPEG

First, we study the performances of ACC-JPEG coder with different N values, it is pointed out that N presents the number of the video cube frames that forms the 'IACC' frame (the GOF).

5.1.1 GOF number impact on coder performances

Experiments show the GOF number augmentation impact on ACC-JPEG compression performance. In slow motion video sequences, the compression rate increases by raising the N value (cf. Figure 11). This compression improvement is due to the temporal redundancies exploitation which become more significant by increasing the GOF's frames number.

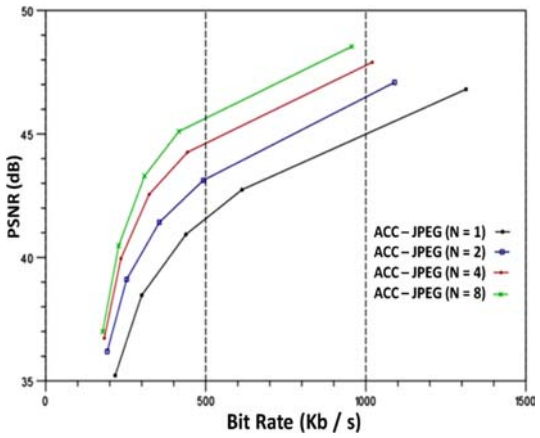
For $N = 1$, it acts as the MJPEG which does not exploit the temporal redundancies. By multiplying the N value by 2, the PSNR increases by about 1–2 dB. The best compression rate is obtained with $N = 8$. Since JPEG process starts with breaking up the image into 8×8 blocks, the Accordion representation does not have considerable interest when it is made up with more than 8 frames.

However, it is necessary to note that for $N = 8$, we lose the additional accordion scan gain. In fact, the Accordion scan allows the exploitation of the spatial correlation in temporal frame ends which is not achieved for $N = 8$ as DCT transform splits the image into 8×8 blocs. This represents one limit to the contribution of ACC-JPEG coder.

This problem does not persist for the second proposed coder (ACC JPEG 2000) where the DCT is replaced with the DWT.

Nevertheless, with fast motion video sequences previous remarks are not still available. In fact, by increasing N value, the video quality worsens, with the transparency artefact apparition. For these sequences, best performances are usually done with $N = 2$.

Figure 11 ACC-JPEG PSNR curve variation according to *N* parameter (Miss America) (see online version for colours)



Furthermore, it should be noted that several other factors must be taken into consideration while selecting the frames number to use in the GOF, particularly, the memory constraints (available memory resources) and the target application requirements (latency constraints, image quality, bandwidth ...).

Thus the frames number in the GOF will be one of our encoder parameters which may be specified according to the user requirements.

5.1.2 Performances evaluation

Both of measured PSNR (cf. Figure 16) and image perceptual quality (cf. Figure 13) prove the ACC-JPEG coder considerable performances.

Actually, as it is shown in Figure 16, the ACC-JPEG outperforms the MJPEG in low and high bit rates as it is shown; it surpasses MJPEG 2000 in high bit rates (from 850 Kbps) and it starts reaching the MPEG 4 performance over 2000 Kbps.

The perceptual quality is acceptable especially with the blocking artefact attenuation as it is shown in Figures 12 and 13 which represent the 114th frames respectively encoded by MPEG-4 and ACC-JPEG 2000.

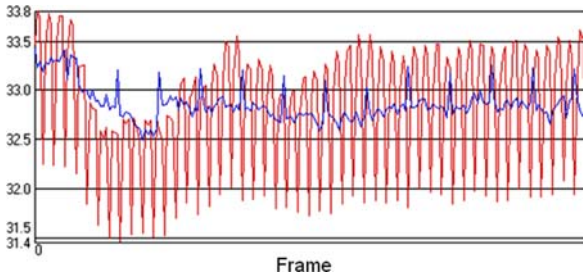
Figure12 Bloking artefact in MPEG4 coder (see online version for colours)

Actually, the ACC-JEPG coder compression efficiency is due to temporal redundancy exploitation. In ACC-JPEG coder, the DCT is mainly applied in temporal domain. Moreover, some artefacts existing in DCT based compression methods such as spatial distortions generated through the massive elimination of the high spatial frequencies (macroblocking) are obviously attenuated in the proposed method as shown in Figure 13. This quality seems suitable for video surveillance application as the proposed method does not seriously affect the spatial details of the image, especially moving object's details.

However, as it is illustrated in Figure 14, the PSNR curve relative to the ACC-JPEG coding is in continuous alternation from one frame to another with a variation between 31.4 dB and 33.8 dB unlike MPEG PSNR which is almost stable.

Figure 13 Attenuation of blocking artefact with ACC-JPEG coder (see online version for colours)

Figure 14 MPEG 4 and ACC-JPEG PSNR variation in Hall Monitor sequence (see online version for colours)



In one hand, ACC-JPEG affects the quality of some frames in GOF, but on the other hand, it provides relevant quality frames in the same GOF, while MPEG produces frames practically with the same quality. In video compression, such feature could be useful for some applications such as video surveillance; generally, we just need some good quality frames in a GOF to identify the objects (i.e., person recognition) rather than medium quality for all the frames.

Previous experiments show the proposed coder superiority compared to both MJPEG in all bit rates and MJPEG 2000 in high bit rates.

However, it is not really competitive especially when compared to MPEG-4 encoder. Due to DCT use, artefacts continue to exist. To overcome these weaknesses,

we thought of applying the DWT instead of DCT. The next proposed coder, namely ACC JPEG 2000, is based on this alteration.

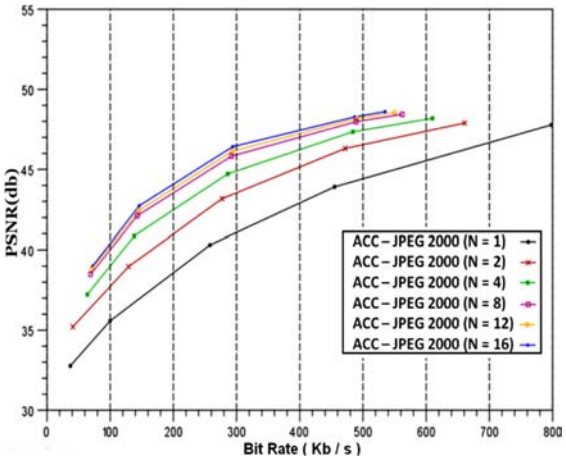
5.2 ACC-JPEG 2000

We start by studying the ACC-JPEG 2000 coder performances with different *N* values.

5.2.1 GOF number impact on coder performances

Figure 15 presents ACC-JPEG 2000 PSNR curves variation according to the *N* parameter used for Miss America sequence (CIF 25 Hz).

Figure 15 ACC-JPEG2000 PSNR curve variation according to *N* parameter (Miss America) (see online version for colours)



With slow motion video sequences, the compression rate increases by raising the *N* value. Up to 8 frames, the *N* variation effect on coder performances decreases. By increasing *N* from 8 to 12, we have got an insignificant improvement (about 0.3 dB for Miss America sequence), which is done at the expense of more memory resources demand and prominent latency. Thus, we suppose that it will be unnecessary to use a GOF size of more than eight images.

For fast motion video sequences, PSNR slightly decreases and some artefacts (blur and transparency artefacts) start to be more and more annoying when *N* increases.

Generally, according to the conducted experiments, 4 and 8 frequently give the best performances.

5.2.2 Performances evaluation

ACC-JPEG 2000 coder compression performances are firstly tested in video surveillance applications. Figure 16 shows PSNR results based on a comparative study between ACC-JPEG2000 and different existing video compression standards relative to Hall Monitor sequence.

N is set to 4 for both ACC-JPEG and ACC-JPEG 2000 coders. Detailed description of the used MPEG-4 coder parameters are given in Table 1.

Figure 16 PSNR evaluation: Hall Monitor (CIF, 25 Hz) (see online version for colours)

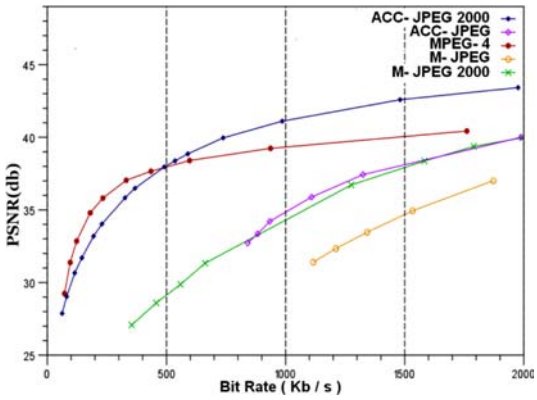


Table 1 MPEG-4 coder's parameters

Implementation	*XVID*
Mode	One pass, average bit rate
Maximum I frame interval	250
Minimum I frame interval	1
Quantisation type	H263
Min I and P frame quantiser	2
Max I and P frame quantiser	31

Experiments conducted with 'Hall Monitor' sequence show that ACC-JPEG2000 can reach very low bit rate (62 Kbps vs. 72 for MPEG 4). Between 100 Kbps and 300 Kbps, MPEG 4 outperforms ACC-JPEG2000 by almost 1 dB. The PSNR relative to ACC-JPEG2000 oscillates from 29.4 dB to 33.9 dB, which decreases the PSNR average value. However, some decoded frames reach the MPEG 4 quality level. Actually, such alternation is not perceptible to the human eye, especially in high frequency (15–30 Hz). In this case, a subjective measure of the ACC-JPEG2000 quality can be more meaningful. Figures 17 and 18 show the 114th frames respectively encoded by ACC-JPEG 2000 and MPEG-4 at 130 Kbps. Although MPEG 4 still outperforms the ACC JPEG 2000 according to measured PSNR, image quality of the latter is somehow close to MPEG-4 even better if we take in consideration the disappearing of the annoying blocking artefact (Figure 17) and the clearness of objects details (Figure 19).

Figure 19(a) and (b), which represent respectively the surrounded areas in Figures 17 and 18, show the object details for respectively ACC-JPEG 2000 and MPEG-4 coder. Actually, ACC-JPEG 2000 provides frames with clearer object's details. This is true even for ACC-JPEG and that was mentioned in Ouni et al. (2009). Such quality is very useful for many video applications (video surveillance, person recognition …).

Up to 500 Kbps, ACC-JPEG 2000 overpass the MPEG 4, it provides pertinent image quality (cf. Figures 20 and 21) with superior average PSNR (39.94 dB vs. 38.40 dB).

Considering the previous results, ACC-JPEG 2000 compression efficiency in video surveillance applications

seems to be unquestionable. First, ACC-JPEG 2000 outperforms MJPEG and MJPEG 2000 universally used in the video surveillance industry. Second, ACC-JPEG 2000 provides useful characteristics and functionalities for video surveillance applications such as its pertinent perceptual image quality with no blocking artefacts, the high quality of some frames in the GOF and Region of Interest (ROI) coding, which are very useful characteristics especially in person recognition.

Figure 17 Perceptual quality given by ACC-JPEG 2000 coding at 130 Kbps (Hall Monitor, 114th frame) (see online version for colours)

Figure 18 Perceptual quality given by MPEG 4 coding at 130 Kbps (Hall Monitor, 114th frame) (see online version for colours)

Figure 19 (a) Perceptual quality in ACC-JPEG 2000 and (b) MPEG-4 at 130 Kbps (Hall Monitor, 114th frame) (see online version for colours)

(a) (b)

It should be noted that ACC-JPEG2000 provides more efficient compression rate compared to ACC-JPEG; In fact, instead of 8×8 size blocs used in DCT, the DWT splits the image into 64×64 sizes blocs based coders which allows more temporal redundancy exploitation.

Figure 20 Perceptual quality given by ACC-JPEG 2000 coding at 600 Kbps (Hall Monitor, 114th frame) (see online version for colours)

Figure 21 Perceptual quality given by MPEG 4 coding at 600 Kbps (Hall Monitor, 114th frame) (see online version for colours)

Previous presented features push to discover the method efficiency in different video categories. This will be detailed in the following.

In most of the studied video sequences, the ACC-JPEG2000 outperforms the MJPEG, MJPEG2000 and ACC-JPEG in low and high bit rates, it often outperforms MPEG-4 in high bit rates.

Figures 22 and 23 can show the global form of the ACC JPEG 2000 PSNR variation in different bit rates compared to MPEG 4 one.

Among the studied coder, only MPEG 4 is competitive to ACC-JPEG 2000. In the next, only MPEG 4 and ACC-JPEG 2000 will be taken into account in our comparative study, all given results are done with $N = 8$.

In the majority of the tested video sequences, we almost find the same forms of the PSNR curves. The MPEG4 outperforms ACC-JPEG 2000 in low bit rates, and the latter outperforms the MPEG 4 in high bit rates. Subsequently, we detail the behaviour of our coder according to the tested videos characteristics.

The experiments prove the ACC-JPEG2000 efficiency on slow and uniform motion or translatoric character sequences, and its weakness in fast motion sequences. Among the studied sequences, we have got worst compression performance with Foreman and tennis sequences. Foreman sequence contains fast non-uniform motion which is caused by the camera as well as the man's face movement. Tennis sequence also contains very fast motion with fast background changes.

In such video sequences, the ACC-JPEG2000 efficiency decreases which is proven by measured PSNR (cf. Figure 24) and reconstructed frames present some transparency artefact. In fact, such results are expected as ACC-JPEG2000 eliminates 'IACC' frame's high frequency data which actually contains the high temporal frequency produced by the fast motion.

Hall Monitor sequence seems to involve less motion compared to the Foreman sequence; the motion takes place only in a very concentrated area. Due to the little amount of motion taking place on the overall image, we observed that our method gets better results (cf. Figure 16). For low bit rate, the best results were given with Miss America sequence; Miss America is a low motion sequence. The motion is confined to the person's lips and head. Since motion is low, temporal redundancy is high and it is expected that ACC- JPEG2000 becomes efficient.

Figure 25 shows results of PSNR based comparative study between ACC-JPEG2000 and MPEG 4 relative to Miss America sequence. Up to 230 Kbps, the proposed coder outperforms the MPEG 4, the relative PSNR continues to increase until lossless level contrary to MPEG 4 which reaches saturation at a certain bit rate. Indeed, MPEG 4 cannot reach less than 22 Kbps, but ACC-JPEG2000 can go less than 10 Kbps. We observe this behaviour in most of the studied sequences. Often ACC-JPEG 2000 is capable of achieving very low-bit rates and performance begin to increase for high bit rates to achieve lossless compression, unlike the MPEG-4 which does not support lossless option. We can state that the proposed coder is highly scalable thanks to the DWT concept use. By the way, ACC-JPEG2000 inherits the JPEG-2000 advantages as the scalability and fine-coding ROI which are very useful features for several current applications.

Figure 22 PSNR based comparative study between ACC-JPEG2000 and MPEG 4 (Mobil, QCIF, 25 Hz)

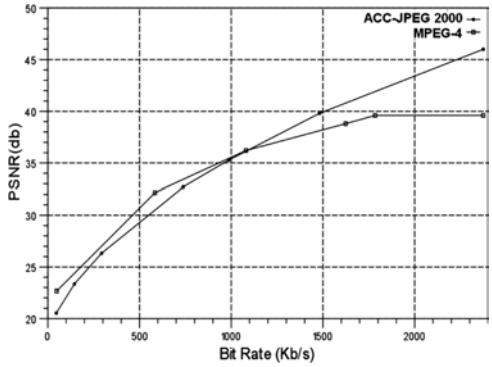


Figure 23 PSNR based comparative study between ACC-JPEG2000 and MPEG 4 (Tempete, CIF, 25 Hz)

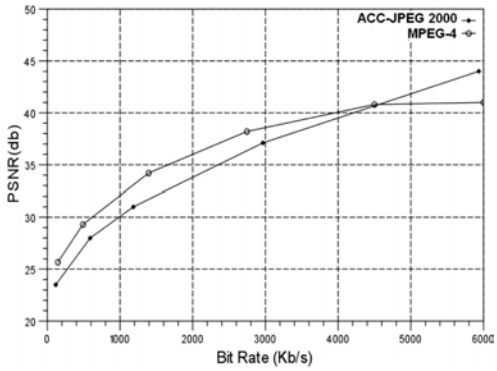


Figure 24 PSNR based comparative study between ACC-JPEG2000 and MPEG 4 (Tennis, CIF, 25 Hz)

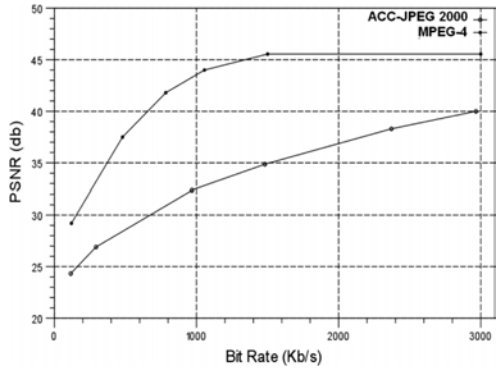
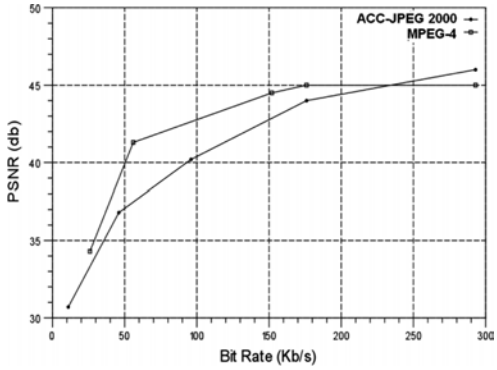


Figure 25 PSNR based comparative study between ACC-JPEG2000 and MPEG 4 (Miss America, QCIF, 25 Hz)



6 ACC-JPEG and ACC-JPEG 2000 features analysis

6.1 ACC-JPEG and ACC-JPEG 2000 artefacts

In ACC-JPEG proposed method the DCT is exploited in both spatial and temporal domain. Actually, temporal and spatial redundancy is projected on spatial domain forming the IACC representation. The DCT application on IACC allows the transformation from the spatial domain to the frequency one. It's pointed out that IACC representation contains more temporal frequency than spatial ones.

After the quantification process, we will eliminate the high spatial frequencies of 'IACC' frame which practically formed by the high temporal frequencies of the original 3D signal source. Actually, the change in the value of a particular pixel from one frame to another can be interpreted as a high frequency in the time domain. The quantisation of DCT block's coefficients does not seriously affect image quality. In fact, when DCT process is applied to ACC representation, resulted blocks contain more temporal information than spatial one, and so, we have more loss in temporal information than in spatial one, this does not affect the image quality but it rather affects the video fluidity, which is sometimes accepted especially in slow motion video sequences. As a result, ACC-JPEG provides better image quality with less blocking artefact but some distortion in video stream.

In ACC-JPEG 2000 proposed coder, the blocking artefact is completely disappeared and replaced by some less annoying 'blur' artefact. In fact, massive elimination of the DWT high frequencies causes some signal distortion. Thus some fast changes over time are somewhat slower. In the time-frequency domain, this means that a foreshadowing of an upcoming event is slightly visible and after the event, the past value will still be visible for a brief moment. This is translated in stable moving objects or slow motion video by blur artefact with unclear moving objects contours (cf. Figure 27), but in video with very fast moving object or video containing cuts this will give some transparency effect (cf. Figure 28).

Figures 26 and 27 present respectively blocking artefacts resulting in MPEG 4 coding and blur artefact resulting in ACC-JPEG2000 coding. As it is shown in Figure 27, the video looks great even under high compression. In fact, the human eye accepts a fuzzy picture more than discrete blocks.

Figure 28 shows the transparency effect resulting in ACC-JPEG 2000 coding when video contains cut.

There are many solutions for this known issue in the prior art (Fryza, 2002, 2006). However their integration is not well adapted to our coder and it increases the coder complexity.

In this work, this problem could be resolved in other way. In fact, the input data stream is divided into n frames as shown in Figure 29. These groups of n frames are completely independent from each other. The problem appears when one group contains several types of video sequences. In consequence, particular frames compound images from different video sequences.

Therefore, the situation of a cut of two different sequences inside encoder input is engaged with help of dynamic method of construction of the GOF. Thus, the proposed solution is to work with a dynamic strategy in the construction of the GOF, the number of frames will not be static, but rather will vary according to the semantics of the video. An inter frame change detection module will be integrated. This module is responsible for detecting significant and fast inter-frames changes. This module allows removing transparency artefact by avoiding cuts in inputs GOFs. It also contributes in the improvement of the video quality by reducing the number of frames in the GOF in fast video sequence.

Our specification was built to a pixel to pixel frame comparison with a threshold. Such algorithm has the advantage of being simple to implement and it can be done while buffering the frames to form the GOF.

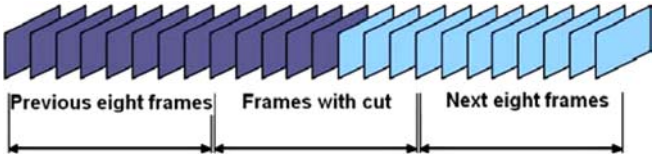
Overall, Accordion based algorithms provide good image quality, and keep the objects details as it is shown in experimental section, it just affect the video fluidity which is often with no damages; in fact, the high temporal frequencies eliminated by the proposed coders are usually a noise, and if it is not the case, it does not make differences as the human eyes cannot detect such high frequency-data.

Figure 26 MPEG 4 blocking artefact, Tennis (CIF, 25 Hz) (see online version for colours)

Figure 27 ACC-JPEG 2000 blur artefact (CIF, 25 Hz) (see online version for colours)

Figure 28 Transparency artefact: Tennis (CIF, 25 Hz) (see online version for colours)

Figure 29 Video sequence with cut (see online version for colours)



6.2 Complexity evaluation

6.2.1 Temporal complexity

First of all, it should be noted that the Accordion transform does not include any real processing; it acts just on the pixel arrangement without any operations on pixels values. Thus the Accordion representation can be integrated in usual 2D transforms such us 2D DCT which results in the further called ACC-DCT. The latter formula is presented as follows:

The definition of the two-dimensional DCT for an input $L \times H$ image I and an output image T is:

$$T_{p,q} = \alpha_p \alpha_q \sum_{x=0}^{H-1} \sum_{y=0}^{L-1} I(x,y) \cos\frac{\pi(2x+1)p}{2H} \cos\frac{\pi(2y+1)q}{2L}$$
$$0 \le p \le H-1, 0 \le q \le L-1. \tag{4}$$

$$\alpha_p = \begin{cases} 1/\sqrt{H} & \text{if } p = 0 \\ \sqrt{2/H} & \text{if } 0 \le p \le H-1 \end{cases}$$

$$\alpha_q = \begin{cases} 1/\sqrt{L} & \text{if } q = 0 \\ \sqrt{2/L} & \text{if } 0 \le q \le L-1 \end{cases}.$$

The values $T_{p,q}$ are called the DCT coefficients of I.

Where H and L are the row and column size of I, respectively. $I(x, y)$ is the pixel intensity with the coordinates x, y. By combining equations (2) and (4), the definition of the two-dimensional ACCDCT for an input image IACC and output image T will be:

$$TAcc_{p,q} = \alpha_p \alpha_q \sum_{x=0}^{L-1} \sum_{y=0}^{H-1} I_n(x \div N, y) \cos\frac{\pi(2x \div N+1)q}{2L} \cos\frac{\pi(2y+1)q}{2H} \tag{5}$$

$$\text{where } n = \begin{cases} i \bmod N \text{ if } x \bmod 2 = 0 \\ N - x \bmod N + 1 \quad \text{otherwise} \end{cases}$$

where:

- N is the number of frames of a GOF.
- IACC(x, y) is the pixel intensity with the x, y coordinates according to Accordion representation repair.
- In(x, y) is the intensity of pixel situated in the Nth frame in the original video source.

Indeed, the computational complexity of ACC-JPEG is practically the same as M-JPEG one.

For example, if we take in consideration NF $L \times H$-resolution frames, we have:

- For MJPEG algorithm, the number of DCT function calls is: $8 \times 8 \times ((L \text{ div } 8) \times (H \text{ div } 8) \times NF$.
- For ACC-JPEG, the number of ACC-DCT function calls is: $8 \times 8 \times ((L \times NF \text{ div } 8) \times (H \times NF) \text{ div } 8)$ ACC-DCT calls, which is the same number of DCT function calls in M-JPEG.

The same comparison could be done for JPEG-2000 and ACC-JPEG 2000.

Because of its non predictive character, as it avoids the computationally demanding motion compensation step, the proposed scheme is clearly less complex than MPEG/H26x standards. Moreover, ACC-JPEG/JPEG2000 compression process is independent from the video semantic content; whatever the motion complexity, it processes with uniform way on the whole video sequence. In this case, MPEG/H26x process complexity depends on motion complexity. In very fast and non uniform motion sequences, motion estimation processing complexity becomes extremely demanding on processor capabilities.

6.2.2 Spatial complexity

The major drawback of the method is the memory requirements and latency.

Indeed, the proposed technique must await the acquisition of all the N images so it can start coding. For example, for $N = 8$, latency = acquisition time of 8 images + coding time of 8 images.

However, this latency is compensated by an encoding competitive explained by the apparent simplicity of the method which has been shown experimentally in tested applications. In fact, The MJPEG 2000 coder was implanted in a DVR and video surveillance application (Figure 30). The coder can achieve real-time constraints in video recording and transmission using personal computer capabilities (Table 2) for VGA resolution video sequence with 25 frames per second.

Figure 30 DVR and video surveillance application (see online version for colours)

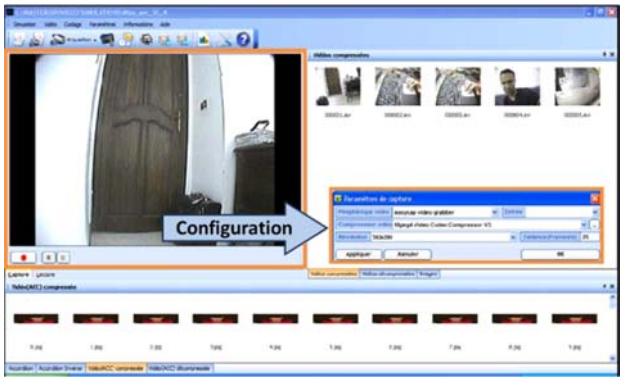


Table 2 Experimental conditions

Processor	Intel Core 2 CPU
Processor frequency	1.6 GHz
RAM	1 GHz
OS	Microsoft Windows XP Professional

6.3 Inherited capabilities

ACC-JPEG 2000 supports a number of functionalities, many of which are inherited from the JPEG-2000 algorithm itself.

First, the generated code-stream is parseable and can be resolution, layer (i.e., SNR), position or component progressive, or any combination thereof. MPEG-4, as ACC-JPEG 2000, is able to produce progressive bitstreams without any noticeable overhead. However, the latter provides more progressive options and produces bitstreams that are parseable and that can be rather easily reorganised by a transcoder on the fly. Along the same lines, ACC-JPEG 2000 also provides *random access* (i.e., involving a minimal decoding) to the block level in each sub-band, thus making possible to decode a region of the image without having to decode it as a whole. These two features could be very advantageous in applications such as digital libraries ...

Another result of the fact that JPEG2000 generates progressive bitstream is the *Region of Interest* functionality which is possible because of the independent coding of the code-blocks and the packetised structure of the codestream. Other advantages can be mentioned such us lossless compression capability and suitability for text …

ACC-JPEG 2000 provides superior rate-distortion performance compared to ACC-JPEG. However, this comes at the price of additional complexity which might be currently perceived as a disadvantage for some applications, as was the case for ACC-JPEG when it was first introduced. Compared to MJPEG 2000, ACC-JPEG 2000 provides superior rate-distortion performance with no perceptible additional complexity.

Compared to the predictive coding schemes (MPEG/h26x), ACC-JPEG2000 provides comparable image quality with significantly less complexity.

Overall, one can say that ACC-JPEG 2000 offers the richest set of features and functionalities when compared to MPEG standards. From this point of view, ACC-JPEG 2000 is a true improvement, providing lossy and lossless compression, progressive and parseable bitstreams, error resilience, ROI, random access and other features in one integrated algorithm. However, while MPEG standards provide higher compression efficiency there is no truly substantial improvement. This is especially true for lossy coding and scalability.

The study shows, as it is presented in Table 3 that the choice of the 'best' standard depends strongly on the application at hand, but that ACC-JPEG 2000 supports the widest set of features among the evaluated standards, while providing considerable rate-distortion performance in most cases.

Table 3 Functionalities based comparison

	M-JPEG	*M-JPEG2000*	*MPEG-4*	*ACC-JPEG*	*ACC-JPEG 2000*
Lossless compression	–	++	–	–	++
Inter-frame correlation	–	–	+++	+	++
Scalability	–	+++	+	–	+++
ROI	–	++	+	–	++
Random access	–	+	–	–	+
Error resilience	+	+++	++	+	+++
Low complexity	++++	++	+	++++	++
Suitability for text	+	+++	+	+	+++

Functionality matrix.

'+' (resp. '–') indicates that the functionality is (resp. is not) supported, the more '+' occurs, the more the associated functionality is supported by the associated method.

6.4 Accordion representation based coder's advantages

The Accordion based approach presents several advantages:

Symmetry: As opposed to motion estimation and compensation based coding schemes of which coding is more complex than decoding, the proposed encoder and decoder are symmetric with almost identical structure and complexity, which makes easier their joint implementation.

Simplicity: The proposed method converts the 3D processing issue to 2D one, which tremendously diminishes the processing complexity. Moreover, the complexity is independent from the compression ratio and semantic contents; whatever the motion complexity is, it processes with uniform way on the whole video sequence. In this case, MPEG process complexity depends on motion complexity.

In very fast and non uniform motion sequences, motion estimation processing complexity becomes gigantic and demanding on processor capabilities.

Moreover, Accordion based coder's complexity is independent on the GOF size unlike 3D transform based method which complexity depends on GOF size.

Subjectivity: Unlike 3D methods that treat temporal and spatial redundancies in the same way, the proposed method is rather discriminatory, it exploits the temporal redundancies more than the space redundancies; what is more objective and more efficient.

Flexibility: The ACC-JPEG parameters offer a flexibility that makes it possible to be adjusted to video applications diversity requirements. The latency time, the compression ratio and the required memory size depend on the N parameter value. Indeed, by increasing the N value,

the compression ratio, the time latency and the reserved memory increase. This parameter makes the compression/ quality compromise optimisation possible, while taking into account memory and latency restrictions. *Implementation advantages*; the proposed method offers the possibility of exploiting existing image compression standards implementations which are broadly available and much optimised for different applications. This lessens their design time to market.

7 Conclusion

Video compression has generated a lot of discussion and increasing attention from the research in recent years. Among many proposed video compression methods, motion compensated coding has taken the most attention because of its performances. However, such methods are computationally intensive and very time consuming. Besides, its real time implementation is difficult and costly and they are less appropriate to be implemented for real-time compression systems and portable recording or communication devices. For this reason, researchers emerge to non predictive less complex methods. In this context, intensive works are investigated in 3D transform based video compression methods. Although their respectable performances and relatively low complexity, such methods suffer from many limitations (memory requirements …) which could be an obstacle for its evolution. Its limited exploitation of spatial and temporal correlation let us state that a naive application of 2D transforms into the third dimension will not always ensure better compression. Thus more subtle transformations that have to be developed take into account the major real world videos characteristics. Actually, the key in efficient image and video compression is to explore source correlation so as to find a compact representation of source data with a reasonable processing complexity. Based on this principle, we tended to explore a new non predictive video coder which subjectively exploits the temporal and spatial redundancy with the minimum complexity processing; many experiments were conducted in order to prove the method's performances and point out its limits. Taking into account its operating simplicity on one hand and its competitive performances on the other hand, we can state that this approach can be practical in outsized application domains, mainly, in embedded systems and video surveillance applications. It gives an interesting compromise between quality and complexity and it provides a response to the latest different applications requirements. With the obvious gains in compression efficiency, and the offered functionalities, we predict that the proposed method could open new horizons in video compression field. There are several directions for future investigations. First of all, we will try to combine the Accordion representation with other image coding techniques. Another direction could be to explore other video representation possibilities in order to look for an extra correlated one.

References

Akbari, A.S. and Soraghan, J.J. (2003) 'Adaptive joint subband vector quantization codec for handheld videophone applications', *lEE Electronic Letters*, Vol. 39, No. 14, pp.1044–1046.

Beong, J.K. and Pearlman, W.A. (1997) 'An embedded wavelet video coder using three-dimensional Set Partitioning in Hierarchical Trees (SPIHT)', *Proc. Data Compression Conference*, Snowbird, UT, USA, pp.251–260.

Bernabe, G., Gonzalez, J., Garcia, J.M. and Duato, J. (2000) 'A new lossy 3-D wavelet transform for high quality compression of medical video', *IEEE EMBS International Conference on Information Technology Applications in Biomedicine*, pp.226–231.

Burg, A. and Keller, R. (1999/2000) *A Real-Time Video Compression System Based on the 3D-DCT*, Diploma-Thesis, Integrated Systems Laboratory ETH-Zurich.

Burg, R.A (2000) '3D-DCT real-time video compression system for low complexity single chip VLSI implementation', *Mobile Multimedia Conf, MoMuC*.

Conway, J.H. and Sloane, N.J.A. (1988) *Sphere Packings, Lattices, and Groups*, Springer-Verlag, New York.

Fryza, T. (2002) *Compression of Video Signals by 3D-DCT Transform*, Diploma Thesis, Institute of Radio Electronics, FEKT Brno University of Technology, Czech Republic.

Fryza, T. (2006) 'Improving quality of video signals encoded by 3D DCT transform', *48th International Symposium ELMAR 2006 Focused on Multimedia Signal Processing and Communications (ELMAR 2006)*, pp.89–93.

Gokturk, S.B and Aaron, A.M (2002) 'Applying 3D methods to video for compression', *Digital Video Processing (EE392J) Projects Winter Quarter*.

Haskell, B.G. and Limb, J.O. (1972) *Predictive Video Encoding Using Measured Subject Velocity*, Patent, US 3,632, 865, January.

Haskell, B.G., Mounts, F.W. and Candy, J.C. (1972) 'Interframe coding of video telephone pictures', *Proc. IEEE*, Vol. 60, No. 7, July, pp.792–800.

Khalil, H., Atiya, A.F. and Shaheen, S. (1999) 'Three-dimensional video compression using subband/wavelet transform with lower buffering requirements', *IEEE Transactions on Image Processing*, Vol. 8, No. 6, June, pp.762–773.

Kocovic, P. (2008) 'Four laws for today and tomorrow', *Journal of Applied Research and Technology*, Vol. 6, pp.133–146.

Koivusaari, J.J. and Takala, J.H. (2005) 'Simplified three-dimensional discrete cosine transform based video codec', *Proc. SPIE-IS&T EI Symposium*, San Jose, CA.

Kretzmer, E.R. (1952) 'Statistics of television signal', *Bell Syst. Tech. J.*, Vol. 31, No. 4, pp.751–763.

Kutil, R. and Uhl, A. (1999) 'Hardware and software aspects for 3D wavelet decomposition on shared Memory MIMD computers', *Proceeding of ACPC'99, Volume 1557 of Lecture Notes on Computer Science*, Springer-Verlag, Las Palmas de Gran Canaria, Spain, pp.347–356.

Lawson, S. and Zhu, J. (2002) 'Image compression using wavelets and JPEG2000', *Electronic & Communication Engineering Journal*, Vol. 14, No. 3, pp.112–121.

Lazar, D. and Averbuch, A. (2001) 'Wavelet-based video coder via bit allocation', *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 11, No. 7, July, pp.815–832.

Lee, G.H., Song, J.H. and Park, R.H. (1997b) 'Three-dimensional DCT/WT compression using motion vector segmentation for low bit-rate video coding', *Proceedings of International Conference on Image Processing*, Vol. 3, pp.456–459.

Lee, M.C., Chan, K.W. and Adjeroh, D.A. (1997a) 'Quantization of 3D-DCT coefficients and scan order for video compression', *Journal of Visual Communication and Image Representation*, Vol. 8. No. 4, pp.405–422.

Lin, G. and Liu, Z. (1999) '3D wavelet video codec and its rate control in ATM network', *Proc. IEEE International Symposium on Circuits and Systems*, Vol. 4, June, Orlando, FL, pp.447–450.

Marpe, D., Wiegand, T. and Sullivan, G.J. (2006) 'The H.264/MPEG4 advanced video coding standard and its applications', *IEEE Communications Magazine*, Toronto, Ontario, Canada, Vol. 44, No. 8, August, pp.134–143.

Molino, A. and Vacca, F. (2004) 'Low complexity video codec for mobile video conferencing', *EUSIPCO*, Vienna, Austria, pp.665–668.

Mounts, F.W. (1968) 'A video encoding system with conditional picture-element replenishment', *Bell Syst. Tech., J.*, Vol. 48, No. 7, pp.2545–2554.

Moyano, E., Quiles, F.J., Garrido, A., Orozco-Barbosa, L. and Duato, J. (2001) 'Efficient 3D wavelet transform decomposition for video compression', *Proceeding of International Workshop on Digital and Computational Video*, IEEE Computer Society Press, February, Tampa, FL, USA.

Netravali, A.N. and Robbins, J.D. (1979) 'Motion-compensated television coding', *Part I, Bell Syst. Tech. J.*, Vol. 58, No. 3, pp.631–670.

Ouni, T., Ayedi, W. and Abid, M. (2009) 'New low complexity DCT based video compression method', *ICT*, Marrakech, Morocco, pp.202–207.

Ouni, T., Ayedi, W. and Abid, M. (2010) 'Non predictive wavelet based video coder', *ICIAR 2010, Part I, LNCS 6111*, Povoa De Varzim, Portugal, pp.344–353.

Rusanovskyy, D. and Egiazarian, K. (2004) *ACIVS 2005*, Antwerp, Belgium.

Said, A. and Pearlman, W. (1006) 'A new, fast, and efficient image codec based on set partitioning in hierarchal trees', *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 63, June, pp.243–250.

Salleh, M.F.M. and Soraghan, J.J. (2005) 'A new Multistage Lattice VQ (MLVQ) technique for image compression', *Proc. European Signal Procession Conference*, September, Antalya, Turkey.

Salleh, M.F.M. and Soraghan, J.J. (2006) 'A new adaptive subband thresholding algorithm for multistage lattice VQ image coding', *Proc. IEEE Int. Conf. Acoustics Speech and Signal Processing (ICASSP 2006)*, Vol. 2, May, pp.457–460.

Sampson, D.G., de Silva, E.A.B. and Ghanbari, M. (1995) 'Low bit-rate video coding using wavelet quantization', *Proc. Inst. Elect. Eng.*, Vol. 142, pp.141–148.

Seigneurbieux, P. and Xiong, Z. (2001) '3-D wavelet video coding with rate-distortion optimization', *Proc. International Conference on Information Technology: Coding and Computing*, April, Las Vegas, NV, USA, pp.322–326.

Servais, M.P. (1997) 'Video compression using the three dimensional discrete cosine transform', *Proc. COMSIG*, South Africa, pp.27–32.

Servais, M.P. and De Jager, G. (1997) 'Video compression using the three dimensional discrete cosine transform', *Proc. COMSIG*, South Africa, pp.27–32.

Shapiro, J.M. (1993) 'Embedded image coding using zerotrees of wavelet coefficients', *IEEE Transactions on Signal Processing*, Vol. 41, No. 12, December, pp.3445–3462.

Shi, Y.Q. and Sun, H. (2008) 'Image and video compression for multimedia engineering fundamentals, algorithms, and standards', *Motion Estimation and Compression Chapter Book*, CRC Press, Section 3, Publication Date: 2008, Boca Raton.

Sikora, T. (2005) 'Trends and perspectives in image and video coding', *Proceedings of IEEE*, Vol. 93, No. 1, January, pp.6–17.

Song, J., Xiong, Z., Liu, X. and Liu, Y. (2000) 'PVH-3DDCT: an algorithm for layered video coding and transmission', *Proc. The Fourth High Performance Computing in the Asia-Pacific Region*, Vol. 2, Beijing, China, pp.700–703.

Vass, J., Chai, B-B. and Zhuang, X. (1998) '3DSLCCA – a highly scalable very low bit rate software-only wavelet video codec', *Proc. IEEE Second Workshop on Multimedia Signal Processing*, December, Redondo Beach, CA, USA, pp.474–479.

Voukelatos, S.P. and Soraghan, J.J. (1997) 'Very low bit rate colour video coding using adaptive subband vector quantization with dynamic bit allocation', *IEEE Transactions. on Circuits and Systems for Video Technology*, Vol. 7, No. 2, pp.424–428.

Note

[1]SVC: Scalable Video Coding.