New Low Complexity Video Compression/Enhancement Based on DCT

B. DHANANJAYA¹, B. RAMA MURTHY², B. BHASKAR REDDY³
¹Research Scholar, Sri Krishnadevaraya University, Anantapur, AP-INDIA,
E-mail: dhanu.balu@gmail.com.
²Prof, Sri Krishnadevaraya University, Anantapur, AP-INDIA.
³Assoc Prof, Bheema Institute of Technology and Science, Adoni, Kurnool, AP-INDIA.

Abstract: Generally, video signal has high temporal redundancies due to the high correlation between the successive frames. Then, this redundancy have not been exploited enough by current video compression techniques. In this paper, we present a new video compression approach which tends to hard exploit the pertinent temporal redundancy in the video frames to improve compression efficiency with minimum processing complexity. It consists on a 3D to 2D transformation of the video frames that can be allows exploring the temporal redundancy of the video using the 2D transforms and avoiding the computationally and the demanding motion of the compensation step. When this transformation is to returns the spatial-temporal correlation of the video into the high spatial correlation. In, this technique for this transforms for each group of pictures to one picture eventually with high spatial correlation. Thus, the de correlation of this resulting pictures by the DCT makes efficient energy compaction, and therefore produces the high video compression ratio. When the Many experimental tests had been conducted to prove the method of the efficiency is especially in high bit rate and with slow motion video. Then the proposed method seems to be well suitable for the video surveillance applications and for the embedded video compression systems.

Keywords: DCT (Discrete Cosine Transform), Video Compression, Redundancy, Spatial Correlation, Efficiency.

I. INTRODUCTION

Then this the objective of video coding in most video applications is to reduce the amount of video data for storing or transmission purposes without affecting the visual quality when the desired video performances depends on the applications requirements are in terms of quality, disks capacity and the bandwidth. For this portable digital video applications and highly integrated in real-time video compression and decompression solutions are more and more required. Actually, motion estimation is based on the encoders are the most widely used in video compression systems. Such as encoders and exploits inter frame correlation to provide the more efficient compression.

However the motion estimation process is the computationally intensive; it’s a real time implementation and it is difficult and costly [1][2]. This is the why motion-based on the video coding standard MPEG was primarily developed for the stored video applications and the encoding process is typically carried out the off-line on powerful computers. So that it is less appropriate to be implemented as the real-time compression process for a portable recording device (video surveillance camera and fully digital video cameras). In these applications, are the efficient and the low cost/ complexity implementation is the most critical issue. Thus, the researches turned towards the design of new coders more adapted to the new video applications and the requirements. This led is the some researchers to look for the exploitation of 3D transforms in order to exploit temporal redundancy. When the coder based on the 3D transform produces the video compression ratio which is close to the motion estimation based on the coding one with less complex processing.

Then the 3D transform based on the video compression methods treat the redundancies in the 3D video signals in the same way, which can be reduce the efficiency of these methods as pixel's values variation in spatial or temporal dimensions is not uniform and so, the redundancy has not the same pertinence. Which often the temporal redundancies are more relevant than spatial one [3]. When it is possible to achieve the more efficient compression by exploiting more and more the redundancies in the temporal domain; when this is the basic purpose of the proposed method. Then the proposed method consists of projecting temporal redundancy of each group of pictures into spatial domain to be combined with spatial redundancy in one of the representation with high spatial correlation.
II. DCT BASED CODING METHODS

The transform coding developed more than two decades ago, has proven to be a very effective video coding method, especially in spatial domain. Today, it forms the basis of almost all video coding standards. The most common transform based intra-frame video coders use the DCT which is very close to the JPEG. Then the video version is called M-JPEG, wherein the “M” can be thought of as standing for “motion”. The input frame is first segmented into NxN blocks. A unitary space-frequency transform is applied to each block to produce an NxN block of transform (spectral) coefficients that are then suitably quantized and coded. The main goal of the transform is to decorrelate the pixels of the input block. This is achieved by redistributing the energy of the pixels and concentrating most of it in a small set of transform coefficients.

This is known a Energy compaction. Compression comes about from two main mechanisms. First, low-energy coefficients can be discarded with in the minimum impact on the reconstruction quality. Second, the HVS has differing sensitivity to different frequencies. Thus, the retained coefficients can be quantized according to their visual importance. Actually, the DCT, which will be used in our video compression approach, is widely used in most modern image/video compression algorithms in the spatial domain (MJPEG, MPEG). Although its efficiency, it produces some undesirable effects; in fact, when compression factors are pushed to the limit, three types of artifacts start to occur: “graininess” due to coarse quantization and some of the coefficients, “blurring” and due to the truncation of the high-frequency coefficients, and blocking artifacts, which is refer to artificial discontinuities appearing at the borders of neighboring blocks due to the independent processing of each block[7].

Moreover, the DCT it can be also used in the temporal domain: In fact, then the simplest way to the extend intra frame image coding methods to interframe video coding is to consider 3-D waveform coding. The 2D-DCT has the potential of easy extension into the third dimension, i.e. 3D-DCT. It includes the time as third dimension into the transformation and energy compaction process [3][4][5][6]. In 3-D transform coding based on the DCT, the video is first divided into the blocks of M N K pixels (M; N; K denote the horizontal and vertical, and then the temporal dimensions). A 3-D DCT is then applied to each block, followed by the quantization and then the symbol encoding, as illustrated has been shown in Figure.1

A 3-D coding method has the advantage that it does not require the computationally intensive process of motion estimation. However, it presents some disadvantages; it requires K frame memories both at the encoder and decoder to buffer the frames. In addition to this storage requirement, the buffering process limits the use of this method in real-time applications because encoding=decoding cannot begin until all of the next K frames are available. Moreover, the 3D DCT based video compression method produce some side effects in low bit rates, for example the effect of transparency produced by the DCT 3D [8]. This artifact is illustrated by figure 2.

The technics of transformed 3Ds was revealed since the 90s, but the research in video compression was oriented towards the coding based on the motion estimation. And then the design tendency of new coding diagrams led some researchers restarting the exploitation of transformed 3Ds in video compression. The coders based on this type of transformation produce high compression ratio with lower complexity compared to motion compensated coding.

Fig.1. 3D DCT video compression

Fig.2. Transparency effect in 3D DCT

3D DCT based video compression methods treat video as a succession of 3D blocks or video cubes, in order to exploit the DCT properties in both spatial and temporal dimensions. The proposed coding method will be based on the same vision. The main difference is how to exploit the temporal and spatial redundancies. That we indeed, then the proposed method was put in to the priority the exploitation of temporal redundancy, which can be more important than the latter assumption will be exploited to make a new representation of original video samples with very high correlation. The new representation should be more appropriate for compression. Detailed approach description will be presented in the next section.
III. PROPOSED APPROACH

The basic idea is to represent video data with high correlated form. Thus, we have to exploit both temporal and spatial redundancies in video signal. The input of our encoder is also called video cube, and which is made up of the number of frames. This cube will be decomposed into the temporal frames which will be gathered into the one frame (2 dimensions). The final step consists of coding the obtained frame. In the following, we detail the method design steps.

A. Hypothesis

Many experiences had proved that the variation of the 3D video signal is much less in the temporal dimension than the spatial one. Thus, pixels, in 3D video signal, are more correlated than the temporal domain than in spatial one [3]; this could be traduced by the following expression: for one reference pixel $I(x, y, t)$ where:

- $I$: pixel intensity value
- $x; y$: space coordinate of the pixel
- $t$: time (video instance)

we could have generally:

$$I(x, y, t) - I(x, y, t + 1) < I(x, y, t) - I(x + 1, y, t) \quad (1)$$

Then 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.

B. “Accordion” based representation

To exploit this succeeding and then the assumption, we start by the carrying out a temporal decomposition of the 3D video signal, the figure 3 shows temporal and spatial decomposition of one 8X8X8 video cube:

“Frames” obtained following the temporal decomposition will be called “temporal frames”. Then these latter are formed by gathering the video cube pixels which have the same column rank. According to the mentioned assumption, these frames have a stronger correlation compared to spatial frames. To increase correlation in Accordion Representation we reverse the direction of event frames. Figure 4 illustrates the principle of this representation.

![Fig.4. ACCORDION Representation](image)

Thus, the “Accordion representation” is obtained as following: first, we start by carrying out a temporal decomposition of the video 3D. When, this the event of temporal frames can be turned over horizontally (Mirror effect). The last step consists of frames successive projecting on a 2D plan further called "I ACC” frame. The “Accordion representation” tends to put in spatial adjacency the pixels having the same coordinates in the different frames of the video cube. This representation transform temporal correlation in the 3D original video source into a high spatial correlation in the 2D representation ("I ACC”). The goal of turning over horizontally the event temporal frames is to more exploit the spacial correlation of the video cube and the frames extremities. It is also minimizes the distances between the pixels correlated in the source. That's could be clearer in figure 5:

![Fig.5. ACCORDION representation example](image)
Figure 6 can show the strong correlation obtained in the “Accordion representation” made of 4 frames which are extracted from “Miss America” sequence.

**Algorithm 1 Algorithm of ACC:**

1: for \( x \) from 0 to \((L \times N) - 1\) do 
2: for \( y \) from 0 to \( H - 1\) do 
3: if \(((x \mod N) \mod 2) = 0\) then 
4: \( n = (N \times 1) \times (x \mod N)\) 
5: else 
6: \( n = x \mod N\) 
7: end if 
8: \( \text{IACC}(x,y) = \text{In}(x \div N, y)\) 
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 \((x \mod 2) \neq 0\) then 
5: \( \text{XACC} = (N \times 1) \times (x \mod N)\) 
6: else 
7: \( \text{XACC} = n(x \times N)\) 
8: end if 
9: \( \text{In}(x,y) = \text{IACC}(\text{XACC}, y)\) 
10: end for 
11: end for 
12: end for 

3) \( \text{IACC}(x; y) \) it is the intensity of pixel which is situated in “I ACC” frame with the co-ordinates x, y according to “Accordion representation” repair.

4) \( \text{In}(x; y) \) is the intensity of pixel situated in the Nth frame in original video source.

We can also present the “Accordion Representation” with the following formulas:

**ACC formulas:**

\[
\text{IACC} = \text{In}(x \div N, y)
\]

(2)

with \( n=((x \div N) \mod 2)(N-1) + 1-2((x \div N) \mod 2)(x \mod N)\)

**ACC in verse formulas:**

\[
\text{In}(x, y) = \text{IACC}(\text{XACC}, y)
\]

(3)

with \( \text{XACC} = ((x \div N) \mod 2)(N-1) + n(1-2(x \div N) \mod 2)+x\)

In the following diagram we will present the diagram of coding based on the “Accordion representation”.

**D. Diagram of coding ACC J P EG**

ACC J P EG Coding is proceeded as follows:

1) The decomposition of the video in groups of frames (GOP).

2) The “Accordion Representation” of the GOP.

3) Decomposition of the resulting “I ACC” frame into 8x8 blocks.

4) For each 8x8 block:
New Low Complexity Video Compression/Enhancement Based on DCT

- Discrete cosine Transformation (DCT).
- Quantification of obtained coefficients.
- Course in Zigzag and then the quantized coefficients.
- Entropic Coding of this coefficients (RLE, Huffman)

![Diagram of ACC JPEG diagram coding](image)

**Fig.7. ACC JPEG diagram coding**

**I. EXPERIMENTS**

Many experiences had been conducted in order to study the performances of our method; we had chosen different kinds of benchmarks. In the following, we summarize the experimental results with some analysis and comments.

**A. parameters of the representation**

When we can be start by studying the performances of the proposed method with different NR values, it’s pointed out that NR presents the number of frames of the video cube that forms the “I ACC” frame. When the best compression rate is so obtained with NR=8. Since JPEG process starts with breaking up the image into 8x8 block, the “Accordion representation” does not have any interest with a GOP made up more than 8 frames. Figure 8 presents ACC JPEG PSNR curves variation according to used NR parameter for “Miss America” sequence (CIF 25Hz).

These results reveals the NR influence on ACC JPEG compression performance. By multiplying the NR value by 2, the PSNR increases from 1 to 2 dB. This compression improvement is due to the exploitation of the temporal redundancies which become more significant by increasing the GOP’s frames number. For NR=1, then it is acts as the MJPEG which can be does not exploit the temporal redundancies. By increasing the value of NR, the coder exploits more the temporal redundancies and so offers a better compression performance.

**Fig.8. ACC JPEG PSNR curve variation according to NR parameter (Miss America)**

**B. Compression performance**

In all studied sequences, the ACC JPEG outperforms the MJPEG in low and high bit rates, it outperforms MJPEG 2000 in high bit rates (from 750 kbs) and it starts reaching the MPEG 4 performance in bit rates higher than 2000 Kbh/s. Among the studied sequences, we have got worst compression performance with “Foreman” sequence. The “Foreman” sequence contains more motion than the other studied sequences. In this sequence we have to contains non-uniform and fast motion which are caused by the camera as well as the man’s face movement. The ACC JPEG efficiency decreases, measured PSNR is relatively low with an alternate character, especially in low bit rate. In fact, such results are expected as ACC JPEG eliminate “I ACC” frame’s high frequency data which actually represent the high temporal frequency produced by the fast motion in the Foreman sequence.

**Fig.9. PSNR evaluation (Hall monitor)**

*International Journal of Scientific Engineering and Technology Research*  
*Volume. 02,IssueNo.06, June-2013, Pages:383-390*
Then the “Hall monitor” and then the sequence seems are 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 get better results. Figure 9 shows then the results of PSNR based on the comparative study between ACC JPEG and different existing video compression standards relative to “hall monitor” then the sequence can be best results were given with “Miss America” sequence; “Miss America” it is a low motion sequence. When the motion is confined to the person's have the lips and head. Since the motion is low, temporal redundancy it is high and it is expected that ACC JPEG becomes efficient.

C. ACC JPEG artifacts

In the proposed method, the DCT is exploited in temporal domain. Some artifacts produced by 3D DCT based compression methods [9][10] persists in ACC JPEG. Actually, the application of the DCT on I ACC allows the transformation from the spatial domain to the frequency domain. After quantification process, we will eliminate the high spatial frequencies of “I ACC” frame which actually present the high temporal frequencies of the 3D signal source. Thus, a strong quantification will not affect the quality of image but will rather affect the fluidity of the video. 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. Once some of the coefficients have been quantized (set to zero) the signal is smoothed out. Thus some fast changes over time is somewhat distorted which explain the alternate character of the ACC JPEG PSNR waveform shown in figure 10.

Indeed, the very high temporal frequency (sudden change of a pixels value over time) is generally interpreted as a noise. Moreover, some artifacts existing in DCT based compression methods such as spatial distortions generated through the massive elimination of the high spatial frequencies (macro blocking) does not exist in the proposed method has been shown in the figure 11. Then the PSNR.

![Fig.10. PSNR waveform comparison between ACC JPEG and MPEG-4 (Miss America)](image)

![Fig.11. MPEG-4 Vs ACC JPEG Curve relative to the ACC JPEG coding is in continuous alternation from one frame to another with in the variation between 31.4 dB and 33.8 dB unlike the MPEG and the PSNR which is almost stable. In one hand, ACC JPEG affects the quality of some frames of the GOP, and other hand, it provides relevant quality frames in the same GOP, while MPEG produces frames practically of the same quality. In video compression, such feature could be useful for video surveillance field; Generally, we just need some good quality frames in a GOP to identify the objects (i.e. person recognition) rather than medium quality for all the frames.

V. ACC JPEG FEATURE ANALYSIS

When the proposed method we can be presents in the several advantages:

**Symmetry:** On this the contrary of coding schemes are based on the motion estimation and compensation whose coding is more complex than decoding, the proposed encoder and decoder are symmetric with almost identical structure and complexity, which facilitates their joint implementation.

**Simplicity:** The proposed method transform the 3D features to 2D ones, which enormously reduce the processing complexity. Moreover, The complexity is independent of the compression ratio and motions.

**Objectivity:** Unlike 3D methods that treat temporal and spatial redundancies in the same way, the proposed method is rather “selective”, it exploits the temporal redundancies more than the space redundancies; what is more objective and more efficient.
Flexibility: The parameters of the ACC JPEG offer a flexibility that makes it possible to be adapted to different requirements of video applications: The latency time, the compression ratio and the size of required memory depend on the value of the NR parameter. Indeed, by NR Value will be increasing and then the compression ratio, the latency time and the reserved memory increase. This parameter allows to optimize the Compression/Quality compromise while taking in consideration memory and latency constraints.

Random Access: 3D transform and motion estimation. Based video compression methods require all the frames of the GOP to allow the random access to different frames. However, the proposed method allows the random frame access. the ACC formula makes it possible to code and/or decode a well defined zone of the GOP (Partial coding). As conclusion, we can state that the ACC JPEG is very efficient for scenes with a translatoric character [9], or with slow motion, especially without change of video plan. However, it loses much of its efficiency in scenes with extremely fast moving objects and very fast change of video plan.

The ACC-JPEG produces images whose details are clearer and without macro-blocking. According to the particular features of ACC JPEG quoted in this section, and others in section 4 (Alternate PSNR) it seems that ACC JPEG can be very suitable to video surveillance applications. In fact, such applications, will be find usually the video with the uniform motion because used cameras are always fixed on specific supports. With such given video, ACC JPEG becomes very efficient, it gives good visual quality with clear image details and identifiable moving objects by exploiting the high quality of some frames in the GOP for further recognizing operations. Furthermore, ACC JPEG seems to be well adapted to embedded or portable video devices such as the IP cameras thanks to its flexibility and its operating simplicity.

VI. CONCLUSION

The video signal has high temporal redundancies between a number of frames and this redundancy has not been exploited enough by current video compression technics. In this research, we suggest a new video compression method which exploits objectively the temporal redundancy. With the apparent gains in compression efficiency we foresee that the proposed method could open new horizons in video compression domain; it strongly exploits temporal redundancy with the minimum of processing complexity which facilitates its implementation of the video embedded systems. And it can be presents some useful functions and features and which can be exploited in some domains as the video surveillance. In this the high bit rate, and it can be gives the best compromise between quality and complexity. It can be provides the better performance than the MJPEG and MJPEG2000 almost in different bit rate values. Over “2000kb/s” bit rate values our compression method performance becomes comparable to the MPEG 4 especially for low motion sequences. There are various directions for future investigations. First of all, we would like to explore others possibilities of the video representation. when the another direction we could be to combine the “Accordion representation” with in the another transformations such as wavelet transformation. Then the latter allows a global processing on the whole of the accordion representation of the contrary of the DCT which is generally acts on the blocks.

VII. REFERENCES


