An Improved Complex Spatially Scalable ACC –DCT Based Video Compression Method

Nagabhushana, Aravinda T. V., Krishna Reddy K. R. and Dr. G. Mahadevan

Abstract—In this paper, we propose a low complex Scalable ACC-DCT based video compression approach which tends to hard exploit the pertinent temporal redundancy in the video frames to improve compression efficiency with less processing complexity. Generally, video signal has high temporal redundancies due to the high correlation between successive frames. Actually, this redundancy has not been exposed enough by current video compression techniques. Our model consists on 3D to 2D transformation of the video frames that allows exploring the temporal redundancy of the video using 2D transforms and avoiding the computationally demanding motion compensation step.

This transformation turns the spatial temporal correlation of the video in to high spatial correlation. Indeed, this technique transforms each group of pictures (GOP) to one picture (Accordion Representation) eventually with high spatial correlation. This model is also incorporated with up/down sampling method (SVC) which is based on a combination of the forward and backward type discrete cosine transform (DCT) coefficients. As this kernel has various symmetries for efficient computation, a fast algorithm of DCT-based Scalability concept is also proposed.

For further improvement of the scalable performance, an adaptive filtering method is introduced, which applies different weighting parameters to DCT coefficients. Thus, the decorrelation of the resulting pictures by the DCT makes efficient energy compaction and therefore produces a high video compression ratio. Many experimental tests had been conducted to prove the method efficiency especially in high bit rate and with slow motion video. The proposed method seems to be well suitable for video surveillance applications and for embedded video compression systems.

Keywords—SVC (Scalable Video Coding), Group of Pictures (GOP), ACC-DCT, Spatial and Temporal Correlation.

I. INTRODUCTION

The main 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. The desired video performances depend on applications requirements, in terms of quality, disks capacity and bandwidth. For portable digital video applications, highly-integrated real-time video compression and decompression solutions are more and more required. Actually, motion estimation based encoders are the most widely used in video compression. Such encoder exploits inter frame correlation to provide more efficient compression. However, Motion estimation process is computationally intensive; its real time implementation is difficult and costly [1][2]. This is why motion-based video coding standard MPEG [12] was primarily developed for stored video applications, where the encoding process is typically carried out off-line on powerful computers. So, it is less appropriate to be implemented as a real-time compression process for a portable recording or communication device (video surveillance camera and fully digital video cameras). In these applications, efficient low cost/complexity implementation is the most critical issue.

Thus, researches turned towards the design of new coders more adapted to new video applications requirements. This led some researchers to look for the exploitation of 3D transforms in order to exploit temporal redundancy. Coder based on 3D transform produces video compression ratio which is close to the motion estimation based coding one with less complex processing [3][4][5][6]. The 3D transform based video compression methods treat the redundancies in the 3D video signal in the same way, which can reduce the efficiency of these methods as pixels values variation in spatial or temporal dimensions is not uniform and so, redundancy has not the same pertinence. Often, the temporal redundancies are more relevant than spatial one [3]. It is possible to achieve more efficient compression by exploiting more and more the redundancies in the temporal domain; this is the basic purpose of the proposed method. The proposed method consists on projecting temporal redundancy of each group of pictures into spatial domain to be combined with spatial redundancy in one representation with high spatial correlation. The obtained representation will be compressed as still image with JPEG coder.

The rest of the paper is organized as follows: Section 2 gives an overview of basic definition of three dimensional DCT. Section 3 gives the basics of the proposed method and the modifications made to improve the compression ratio and also reduce the complexity. Experimental results were
discussed in section 4. The section 5 concludes this paper with a short summary.

II. DEFINITION

A. Three Dimensional DCT

The discrete cosine transforms (DCT) [4] [7] has energy packing efficiency close to that of the optimal Karhunen-Loeve transform. In addition, it is signal independent and can be computed efficiently by fast algorithms. For these reasons, the DCT is widely used in image and video compression. Since the common three-dimensional DCT kernel is separable, the 3D DCT is usually obtained by applying the onedimensional DCT along each of the three dimensions. Thus, the N×N×N 3D DCT can be defined as:

\[
X(u,v,w) = \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \sum_{k=0}^{N-1} x(i,j,k)C(i,w)C(j,v)C(k,u)
\]  

(1)

\[
x(i,j,k) = \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} \sum_{w=0}^{N-1} X(u,v,w)C(i,u)C(j,v)C(k,w)
\]  

(2)

where \(C(p,q) = \frac{1}{\sqrt{N}} \cos \left(\frac{\pi p q}{2N}\right)\)

\(, q = 0\)

\(, q \neq 0\)

III. PROPOSED METHOD

The fundamental idea is to represent a video sequence with highly correlated form. Thus we need to expose both spatial and temporal redundancy in video signal. The video cube is the input of our encoder, which is a number of frames. This video cube will decomposed into temporal frames which will be gathered into one 2D frame. The next step consists of coding the obtained frame. Normally, the variation of the 3D video signal is much less in the temporal domain than the spatial domain, the pixels in 3D video signal are more correlated in temporal domain [3].

For a single pixel model is denoted as \((x, y, t)\) where \(p\) is pixel value; \(x, y\) are pixel spatial coordinates; \(t\) is video instance at time. The following assumption will be the basis of the proposed model where we will try to put pixels-which have a very high temporal correlation in spatial adjacency.

\[
P(x,y,t)-p(x,y,t+1)-p(x+1,y,t)
\]  

(3)

To exploit the succeeding assumption the temporal decomposition of the 3D video signal will be carried out and the temporal, spatial decomposition of one 8x8 video cube [8] [9] [10] is presented in the Figure.1. Thus the Accordion representation (Spatial Adjacency) is obtained from the basis assumption.

Fig.1 Temporal and spatial decomposition of one 8x8x8 video cube

Accordion representation is formed by collecting the video cube pixels which have the same column rank and these frames have a stronger correlation compare to spatial frames.

To improve correlation in the representation we reverse the direction of event frames. This tends to put in spatial adjacency that the pixels having the same coordinate in the different frames of the video cube. The following example i.e., Figure. 2 clearly projecting the Accordion representation also minimizes the distance between the pixel correlated in the source.

Fig. 2 Accordion Representation Example

Continuation of the Accordion Representation a new concept is originated from scalable video coding (SVC)[16][17][18][19][20] technique; up/down sampling method using the DCT has a large degree of symmetries for efficient computation. Thus, a fast algorithm of the up/down sampling method is also included in our proposed method. For a performance improvement, an adaptive filtering method DCT up/down sampling is applied [13][14][15], which applies different weighting parameters to each DCT coefficient. Then we have to introduce quantization model and Entropy coding (RLE/Huffman) techniques for further performance improvement of the proposed system. An overall constructional detail of the proposed model is explained in Figure.3.
A. Algorithm

1. Decomposition of the video in GOPs (Group of frames/pictures)
2. a) Spatial Adjacency separation (Accordion Representation) of the GOP
   
   For $x=0$ to $(L * N)-1$
   For $y=0$ to $(H-1)$
   If $((x/N)\mod 2)! = 0$ then
   $n=(N-1)-(x \mod N)$
   else
   $n=x \mod N$
   end if
   $I_{ACC}(x,y)=I_{n}(x/N,y)$
   With $n=((x/N) \mod 2)(N-1)+1-2((x/N) \mod 2)(x \mod N)$
   
   (b) For $n=0$ to $N-1$
   For $x=0$ to $L-1$
   For $y=0$ to $H-1$
   If $(x \mod 2)! = 0$ then
   $X_{ACC}=((N-1)-n(x*2N))$
   else
   $X_{ACC}=n(x*2N)$
   end if
   $I_{n}(x,y)=I_{ACC}(X_{ACC},y)$
   end for
   end for
   end for
   
3. Decomposition of the resulting frame into 8x8 blocks.
4. Introduce down sampling filter/ up sampling filter with DCT.
5. Quantization of the obtained coefficients.
6. Zig Zag coding of the obtained coefficients.
7. Entropy (Huffman) coding of the coefficients.

IV. EXPERIMENTAL RESULTS

This section verifies the performance of the proposed low complex scalable ACC-DCT based video compression model. We summarize the experimental results with some analysis and comments. By understanding the performance of the proposed method with different GOP value the best compression rate is obtained with GOP=8. Here Figure.4 GUI model is created to integrate the encoder and decoder sections. Figure.5 shows the progress of frame separation from video sequence. Encoder model is shown in Figure.6 and Figure.7 is the example Accordion representation for one GOP Video cube. Figure.8: GUI for Decoding and Validation Process. Figure.9: GUI for Reconstructed output validation, then history of entire simulation is specified as ans. Finally Figure.10 shows the plot between Frames numbers Vs PSNR (dB), Figure.11 finds the orientation flow estimation of the sample sequence and compared the strength of our proposed model with other leading standards shown in Figure.11.

![Fig.4 GUI Model](image)

![Fig.5 Frame Separation Model](image)

![Fig.6 Encoding Model](image)
The proposed method presents several advantages:

1. The proposed method transforms the 3D features to 2D ones, which enormously reduce the Processing complexity.
2. The proposed encoder and decoder are symmetric with almost identical structure and complexity, which facilitates their joint implementation.

3. It exploits the temporal redundancies more than the space redundancies.

4. Offers 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 GOP parameter.

5. The proposed method allows the random frame access.

V. CONCLUSION

In this paper, we successfully extended and implemented a low complex scalable ACC-DCT based video compression algorithm on MATLAB and provided experimental results to Show that our method is better than the existing methods. We not only improved the coding efficiency in the proposed encoding algorithm but also it reduces complexity. As discussed in the experimental section, proposed method provides benefits of rate-PSNR performance at the good quality of base layer and low quality of enhancement layer. When SVC coding scenario meets these circumstances, proposed method should be useful. With the apparent gain in compression efficiency we fore see 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 it simple mentation in video embedded systems. It presents some useful functions and features which can be exploited in some domains as video surveillance. In high bit rate, it gives the best compromise between quality and complexity. It provides better performance than MJPEG and JPEG2000 almost in different bit rate values. Over 2000kb/s bit rate values; our compression method performance becomes comparable to the MPEG4 standard especially for low motion sequences. Additionally, a further development of this model could be to combine ‘Accordion representation’ with other transformations such as DCT Transformation.

REFERENCES


