New Non Predictive Wavelet Based Video Coder: Performances Analysis

Tarek Ouni, Walid Ayedi, and Mohamed Abid

National Engineering School of Sfax, Road Sokkra, Km 3 Sfax, 3078 Tunisia
tarek.ouni@gmail.com, walid.ayedi@gmail.com,
Mohamed.abid@enis.rnu.tn

Abstract. A non-predictive video coding is a new branch of 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 temporal decomposition of video frames. The proposed method consists on 3D to 2D transformation of the temporal frames that allows exploring the temporal redundancy of the video using 2D wavelet transforms and avoiding the computationally demanding motion compensation step. Although the many advantages presented by the proposed coder, some annoying artifacts still exist. In this paper, we will explore the performances of the proposed method and try to better show what it actually offers to users. The paper presents also the extensions chosen in order to reduce the perceived artifacts and increase the perceptual as well as objective (PSNR) decoded video quality, which is actually competitive with state-of-the-art video coder algorithms, especially when low computational demands of the proposed approach are taken into account.

Keywords: video coding, temporal decomposition, wavelet, correlation.

1 Introduction

Video compression has generated a lot of discussion and increasing attention from the research in recent years.

Among many proposed methods, motion compensated coding has taken the most attention and taken its place in many standards. These include Mpeg, H.26L, etc. Such encoders exploit inter-frame correlation in order to further improve its compression. However, the main challenge of these methods lies on the motion estimation process which is known to be computationally intensive. Besides, its real time implementation is difficult and costly [1],[2]. Nevertheless, new applications such as sensor networks and portable video devices necessitate a low processing capability for the compression, which makes the encoding complexity a big burden. To deal with this problem, motion-based video coding standard MPEG was primarily developed for stored video applications, where the encoding process is typically carried out off-line on powerful computers. With the explosive growth of video devices ranging from hand-held digital cameras to low-power video sensors, a new class of multimedia devices is required which includes the following architectural requirements: Low

A. Campilho and M. Kamel (Eds.): ICIAR 2010, Part I, LNCS 6111, pp. 344–353, 2010.

power, less- complexity encoding and real time constraint. Therefore, there have been extensive research efforts in video coding in order to give response to the new requirements of video applications different than those targeted by conventional coding schemes in the past years [1]. A non-predictive video coding is a new branch of emerging research area in video coding, where the motion estimation/compensation or prediction step in the temporal domain is omitted.

In [3]-[4], authors exploit 3D transforms in order to exploit temporal redundancy. Coder based on 3D transform produces video compression ratio which is comparable to some motion estimation based coding one but with lower processing complexity [5]. However, 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 variations in spatial or temporal dimensions are not uniform and hence, temporal and spatial redundancies have not the same pertinence. It is known that the temporal redundancies are more relevant than spatial one [2], hence its practical importance. It is more beneficial to utilize the proposed method rather than the 3D based methods, because it is able to achieve higher compression by more exploiting the redundancies in the temporal domain.

This method consists on 3D to 2D transformation of the video frames; it will then explore the temporal redundancy of the video using 2D transforms and avoids the computationally demanding motion compensation step. In particular, the used method projects temporal redundancy of each group of pictures into spatial domain and combines it with spatial redundancy in one representation with high spatial correlation [6]. Then, the new representation will be compressed as still image using wavelet transform based coder (JPEG2000). Actually, the proposed approach presents many advantages. It exploits objectively temporal and spatial redundancy. It omits the temporal prediction step and transforms a 3D processing into 2D one while reducing considerably the complexity processing. Furthermore, it inherits the JPEG 2000 proprieties such as scalability ROI and error resilience.

In this paper, we focus on the analysis of experimental results, solutions and extensions proposed to remove some annoying artifacts of the presented method. Experimental results will show the efficiency of the proposed method at an expense of some annoying artifacts.

The rest of paper is organized as follows. In section 2, we review the basics of the used approach. In section 3, we present some experimental results and explore the method performances and limitations. Section 4 presents the extensions chosen to further improve the compression ratio. Finally, conclusions are drawn in Section 5.

2 Description of the Used Approach

Actually, the used method relies on the following assumption: high frequency data is more difficult to compress compared to low frequency one.

The main idea of the proposed method is to make some geometric transformation of the 3D data in order to make one representation with very high correlation, and consequently without high frequencies data. We will play on the disposition of pixel’s data in the video cube.

2.1 Hypothesis

The video stream contains more temporal redundancies than spatial ones [2]. 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.

2.2 Accordion Representation

The input of our encoder is the so called video cube (GoF), which is made up of a number of frames. 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.

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. In this way, Accordion representation also minimizes the distances between the pixels spatially correlated in the source. This representation transforms temporal correlation of the 3D original video source into a high spatial correlation in the 2D representation ("IACC") [6]. Figure 1 illustrates the principle of this representation.

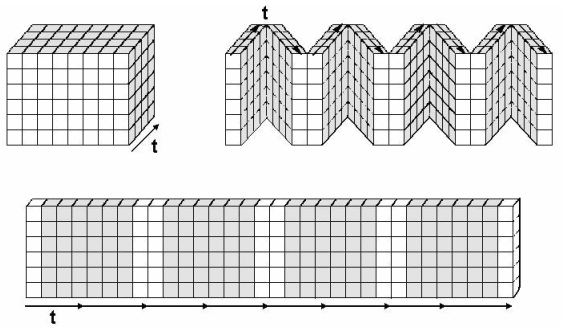


Fig. 1. Accordion representation [6]

In the following, we will present the diagram of coding based on the Accordion representation further called ACC-JPEG2000.

2.3 ACC-JPEG2000 Coding Scheme

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

- The decomposition of video sequence into groups of frames (GOF).
- Accordion representation of the GOF.

- DWT transform.
- Quantization of obtained coefficients (Q).
- Arithmetic coding of obtained coefficients.

Figure 2 presents ACC-JPEG2000 coding scheme.

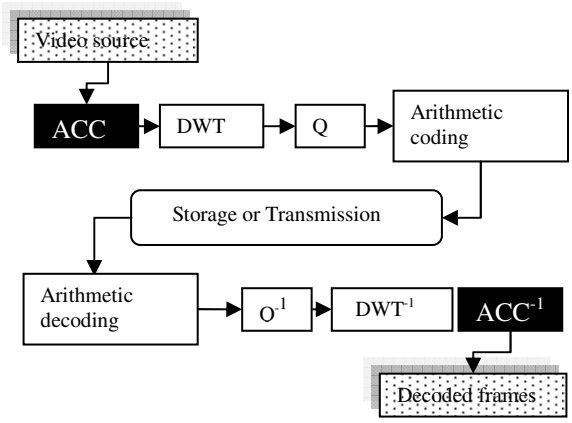


Fig. 2. ACC-JPEG2000 coding scheme

The video encoder takes a video sequence and passes it to a frame buffer. The buffer dispatches a group of frames at a time to Accordion process before being sending to DWT blocks. Each of the DWT blocks performs a 2-D discrete wavelet transform using JPEG2000 wavelet filters coefficients. In this system, we use five levels of wavelet decomposition which is sufficient for CIF sequences.

3 Experiments

In the following, we summarize the experimental results with some analysis and comments.

3.1 PSNR Evaluation

In these experiments, we use the XVID MPEG-4 video coder including P frames. The GOF number of frames relative to ACC-JPEG 2000 is fixed to 8 frames.

The experiments prove the efficiency of ACC-JPEG2000 on slow and uniform motion sequences. In these sequences, temporal redundancy is **relevant**; the spatial representation "IACC" performs a pertinent correlation. In this case, it is expected that the proposed method proves its efficiency. However, the method shows a remarkable sensitivity to very fast motion video sequences.

Among the studied sequences, we have got worst compression performance with “tennis” sequence. Tennis sequence contains very fast motion with fast complex background changes. The generated spatial representation still presents some high frequencies. The ACC-JPEG2000 efficiency decreases with the apparition of transparency effect due to background change, measured PSNR is relatively low with an alternate character. 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. Foreman sequence contains fast non-uniform motion which is caused by the camera as well as the man's face movement. So measured PSNR is relatively low and visual quality suffers from some blur effect, especially in face detail which represents the high resolution data. 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 get better results. 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.

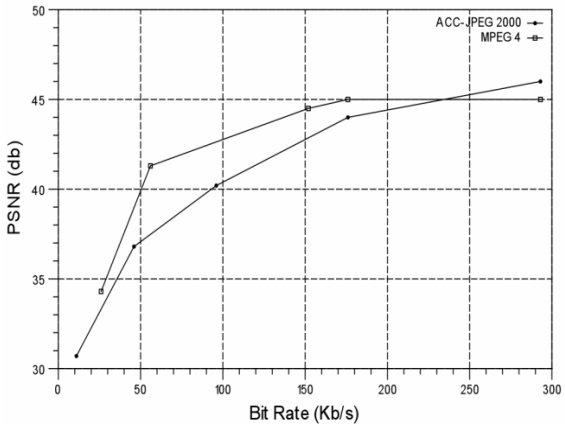


Fig. 3. PSNR evaluation: Miss America (QCIF, 25Hz)

Figure 3 shows results of PSNR based comparative study between ACC-JPEG2000 and MPEG 4 relative to miss america sequence. Up to 230 kb/s, the proposed coder outperforms the mpeg 4, the relative PSNR continue to increase until lossless level. Otherwise, MPEG 4 can not rich less than 22 kbits/s, but ACC-JPEG2000 can go less than 10 kbits/s. we can state that the proposed coder is highly scalable.

Table 1 shows results of PSNR based comparative study between ACC-JPEG2000 and MPEG 4 coder.

Table 1. PSNR EVALUATION

	Bit Rate (Kbits/s)	ACC-JPEG 2000	MPEG 4
Water fall (CIF, 25Hz)	*100*	29	30.1
	1000	35	36.6
Bus (CIF, 25Hz)	*200*	21.3	25.5
	1000	26.6	32
Tennis (CIF, 25Hz)	*100*	25	29
	1000	34	43
Mobile (QCIF, 25Hz)	*100*	22	23
	1000	36	35
Hall monitor (CIF, 25Hz)	*50*	28.3	29.1
	1000	41.2	39.1
Miss America (QCIF, 25Hz)	*25*	33.8	34.2
	300	46.7	44.9

Moreover, the PSNR Curve relative to the ACC-JPEG2000 coding is in continuous alternation from one frame to another unlike MPEG PSNR which is almost stable as it is shown in figure 4.

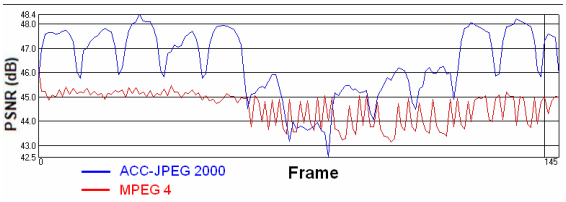


Fig. 4. PSNR evaluation: Miss America (QCIF, 25Hz)

In one hand, ACC-JPEG2000 affects the quality of some frames of a GOP, but on the 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. The example of the “hall monitor” in table 1proves that the video surveillance is one of the best application field to the proposed coder.

Differently to MPEG codec's, ACC-JPEG2000 can reach very low bit rates. In high bit rates it provides a relevant quality (until lossless level).

3.2 Visual Evaluation

Some artifacts existing in DWT based compression methods (MJPEG 2000) and in motion estimation based method (MPEG) such as spatial distortions generated through the massive elimination of the high spatial frequencies (tiling artefact and blocking artefact) as it is presented in figure 6, does not exist in the proposed method as shown in figure 5. It's actually replaced by some less annoying blur artifact.

Blocking artefact in MPEG 4

Blur artefact in ACC-JPEG 2000

Fig. 5. Perceptual evaluation: Tennis (CIF, 25Hz)

In the proposed method, the DWT is exploited in both spatial and temporal domain. Actually, temporal and spatial redundancy is projected on spatial domain forming the IACC representation. The application of the DWT on IACC allows the transformation from the spatial domain to the frequency domain.

After quantification process, we will eliminate the high spatial frequencies of "IACC" frame which actually include the high temporal frequencies of the 3D signal source. As temporal redundancy is more exploited than spatial one, a strong quantification will not seriously affect the quality of image but will rather affect the fluidity of the video. Spatial high frequency is mainly made of fast pixel's values change from one frame to another. 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. As a consequence, the PSNR significantly decreases in very fast motion sequences leading for some annoying blur artifact.

However, some sudden pixels change will be eliminated. This will offer a useful functionality such as the noise removal. Indeed, the very high temporal frequency (sudden change of a pixels value over time) is generally interpreted as a noise.

As it is shown in figure 6, some other artifacts appear in video sequences containing cuts: transparency [7]:

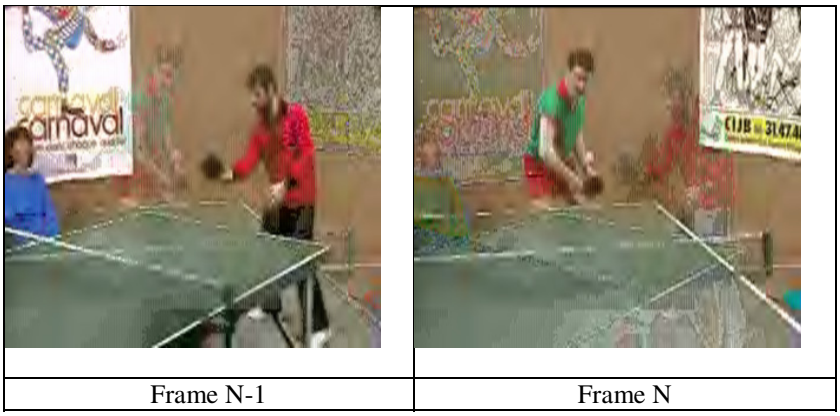


Fig. 6. Transparency artefact: Tennis (CIF, 25Hz)

The input data stream is divided into n frames (in our case n=8) as shown in Fig. 7. These groups of n frames are completely independent to each other. The problem appears when one group contains several types of video sequences. In consequence, particular frames compound images from different video sequences.

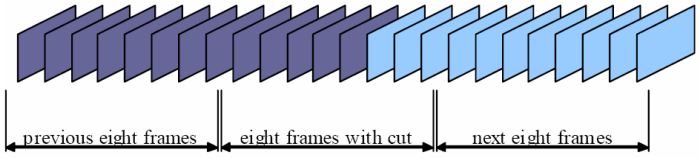


Fig. 7. Video sequence with cut

There are many solutions for this known issue in the prior art [7][8]. However their integration are not well adapted to our coder, and it increase the coder complexity.

4 Performance Improvement

Our current work is directed towards finding solutions to treat certain weaknesses shown by our method. First, the proposed method exhibits significant boundary effects at GOF boundaries. The PSNR drops every N frames, leading to annoying jittering artifacts in video playback. This well known [12] issue can be resolved by extending some temporal filtering indefinitely in time [13][14]. Some spatial filter can be useful when applied on IACC frame. We are currently testing this approach.

Second, the proposed method lose its efficiency in very fast motion sequences especially fast moving objects details are often lost in video playback. Iin this case, we are trying to exploit the ROI propriety of the JPEG 2000. Moreover the extension presented below should clearly decrease this weakness effect.

Another annoying artifact is the transparency; we don't look for some post processing but rather we look for eliminate the cause of this anomaly. Thus, proposed solution is to work with a dynamic strategy in the construction of the GOF, the number of frames will not be previously fixed, but rather will vary according to the semantics of the video in order to avoid cuts in the video inputs GOFs. For this reason an additional inter frame change detection module will be integrated (figure 8).

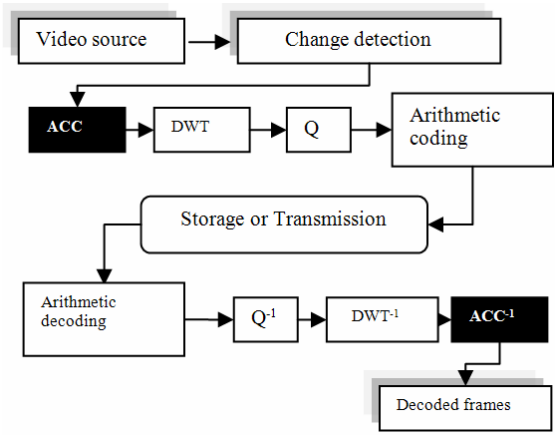
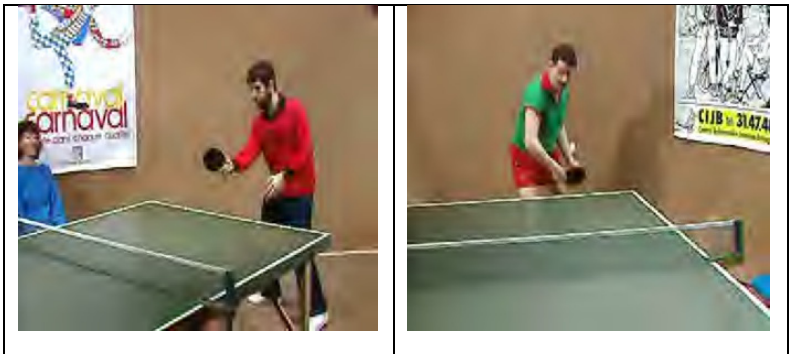


Fig. 8. Integration of the change detection module

There are many existing techniques related to cuts detection [9][10][11], in our case, we don't only look for cuts detection, but also local frames change due to fast moving objects, that's why we proceeded with local comparison with threshold based method. This module is responsible for detecting significant and fast inter-frames changes. This module allows removing transparency artifact 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. The figure 9 shows the disappearance of the transparency artefact after the integration of the change detection module.



Frame N-1	Frame N

Fig. 9. Transparency removal: Tennis (CIF, 25Hz)

5 Conclusion

In this paper, we tended to explore a new non predictive wavelet based video coder; many experiments were conducted in order to prove the method performances and point out its limits. Taking into account its operating simplicity in one hand, and its competitive performances in other hand, we can state that this approach can be useful in large application domains, especially, in embedded systems and video surveillance applications. There are various 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 others possibilities of video representation in order to look for one more correlated one.

References

1. Molino, A., Vacca, F.: Low complexity video codec for mobile video conferencing. In: EUSIPCO, Vienna, Austria, pp. 665–668 (2004)
2. Gokturk, S.B., Aaron, A.M.: Applying 3d methods to video for compression. In: Digital Video Processing (EE392J) Projects Winter Quarter (2002)
3. Servais, M.P.: Video compression using the three dimensional discrete cosine transform. In: Proc. COMSIG, pp. 27–32 (1997)
4. Burg, R.A.: 3d-dct real-time video compression system for low complexity singlechip vlsi implementation. In: Mobile Multimedia Conf, MoMuC (2000)
5. Koivusaari, J.J., Takala, J.H.: Simplified three-dimensional discrete cosine transform based video codec. In: Proc. SPIE-IS&T EI Symposium, San Jose, CA (January 2005)
6. Ouni, T., Ayedi, W.: New low complexity DCT based video compression method. In: International Conference on Telecommunication, Morocco 2009 (2009)
7. Fryza, T.: Compression of Video Signals by 3D-DCT Transform. Diploma Thesis. Institute of Radio Electronics, FEKT Brno University of Technology, Czech Republic (2002)
8. Fryza, T.: Improving Quality of Video Signals Encoded by 3D DCT Transform. In: 48th International Symposium ELMAR 2006 focused on Multimedia Signal Processing and Communications, pp. 89–93 (elmar 2006)
9. Brunelli, R., Mich, O., Modena, C.M.: A survey on the automatic indexing of video data. Journal of Visual Communication and Image Representation 10(2), 78–112 (1999)
10. Idris, F., Panchanathan, S.: Review of image and video indexing techniques. Journal of Visual, Communication and Image Representation 8(2), 146–166 (1997)
11. Ahanger, G., Little, T.D.C.: A survey of technologies for parsing and indexing digital video. Journal of Visual Communication and Image Representation 7(1), 28–43 (1996)
12. Xing, Q., Yan, X.: Tiling artifact reduction for JPEG2000 image at low bit-rate. In: ICME 2004, Taipei, Taiwan (2004)
13. Marusic, B., Skocir, P.: Video post-processing with adaptive 3-D filters for wavelet ringing artifact removal. IEICE Transactions on Information and Systems 88(5), 1031–1040 (2005)
14. Liang, J., Tu, C.: Optimal block boundary pre/post-filtering for wavelet-based image and video compression. In: ICIP 2004, Singapore (2004)