New Accordion based video compression appraoch

Tarek OUNI
National Engineering School of SFAX
Sfax, Tunisia
tarek.ouni@gmail.com

Mohamed ABID
National Engineering School of SFAX
Sfax, Tunisia
Mohamed.abid@enis.rnu.tn

***Abstract*— In this paper, we propose a new approach based on the accordion transform that tends to exploit the both intra and inter frame correlation. The major novelty of the new approach compared to classic one is the dynamic strategy of video frames scan. The direction of the frames scan will depends on a gradient based algorithm which tries to predict the direction where minimal pixels intensities changes are recorded. Both objective (PSNR) and subjective quality evaluation prove the improvement of the compression efficiency, especially in fast and uniform motion video sequences which usually represents the weakness of the classical approach.**

Keywords- Accordion, correlation, gradient, video, compression.

I. INTRODUCTION

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. There are many different approaches which aim to explore these interframe correlations.

A. Predictive methods

Most conventional video coding standard lies on the motion estimation process in order to estimate the motion that occurred between some reference frame and the current frame. The most commonly used ME method is the block-matching motion estimation (BMME) algorithm. In this algorithm, the current frame is first divided into blocks. The motion of each block is then estimated by searching for the best-match block in the reference frame according to some distortion measure. This search is usually restricted to a search window centred on the corresponding block in the reference frame. The motion of the current block is then represented by a motion vector, which is the displacement between the block and its best-match block in the reference frame. Such approach offers a high performance of compression. However, it introduces a high complexity processing which in some cases make it difficult to meet existing technologies [1][2].

B. 3D methods

A new branch of emerging research area in video coding starts to take place in these years where the motion estimation/compensation or prediction step in the temporal domain is omitted. It consists generally in utilizing 3D transforms that consists in extending intra frame image coding methods to inter frame video coding.

A 3-D coding method has the advantage that it does not require the computationally intensive process of motion estimation [3].

Despite their advantages, these methods are not efficient overall and still suffer from several weaknesses. In fact, limiting the transformation to small cubes (8x8x8 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 [4-10].

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.

C. 3D to 2D transform

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 by 2D transform based encoder [11][12].

978-1-61284-860-0/11/$26.00 ©2011 IEEE

Although its efficiency, the Accordion based compression method presents some weakness especially in the fast moving video sequences where the temporal redundancy is no such relevant.

In this paper, we propose a new approach based on the accordion transform. The major novelty of the new approach compared to classic one is the dynamic strategy of video frames scan. The direction of the frames scan will depends on a gradient based algorithm which tries to predict the direction where pixels intensities changes are minimal.

In Section 2 we present an overview on the Accordion based video compression approach. In section 3, we present the basics of the dynamic approach. In Section 4 we discuss the autocorrelation and both objective and subjective compression quality improvement. We conclude with a short summary.

II. CLASSICAL ACCORDION APPROACH

The Accordion 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 in order to exploit the 2D transforms characteristics (DCT, DWT…) in temporal domain.

A. Accordion representation

The input of our encoder is the so called video cube which is made up of a group of frames (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 (c.f. figure 1).

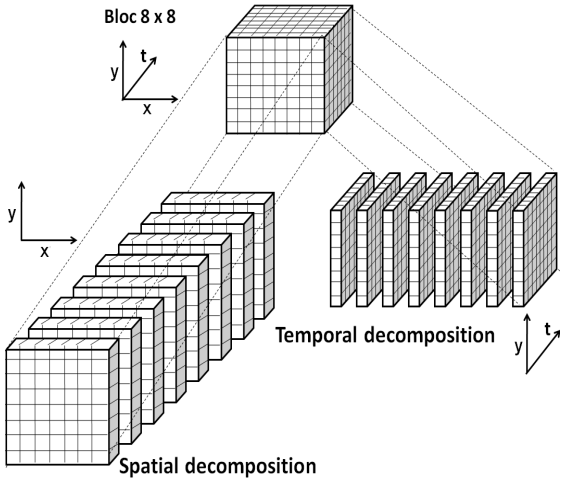

Figure 1. Temporal and spatial decomposition

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 (c.f. figure 2).

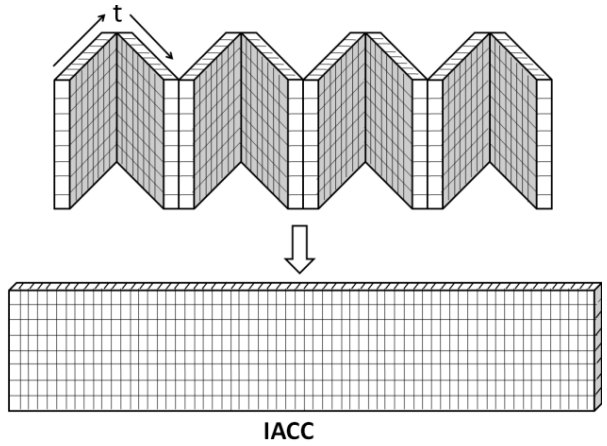

Figure 2. Accordion representation

The encoder block diagram of the Accordion based compression algorithm is in Figure 3.

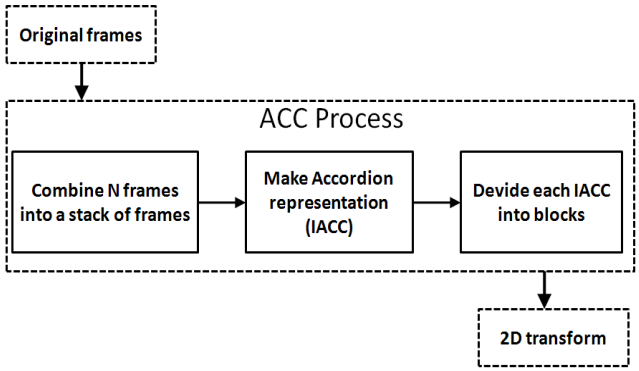

Figure 3. Block diagram of the Accordion representation based encoder

Two Accordion representation based coding schemes have been designed, the first, called ACC_JPEG, uses DCT transform as it is described in [11] , the second, called ACC_JPEG 2000, uses DWT transform [12].

The accordion takes advantage of exploiting the strong temporal correlation in the planar surface (t, y) (cf. figure 1). However, the correlation may be stronger in other directions. In such cases, the accordion based approach compression efficiency is uncertain.

III. DYNAMIC APPROACH

A. Background

The interframe correlation is 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.

Figure 4 (a) illustrate the spatial correlation into pixels of one frame extracted from the video sequence "salesman". Figures 4 (b) and 4 (c) respectively show the temporal variations through time of the 80^{th} line and the 100^{th} column.

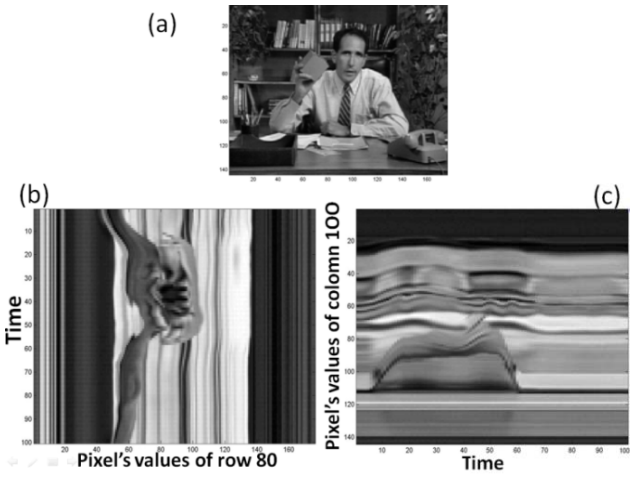

Figure 4. Spatial and temporal correlation in salesman sequence.

Referring to figure 5, we can see that the correlation in a video can be operated with three different linear decompositions:

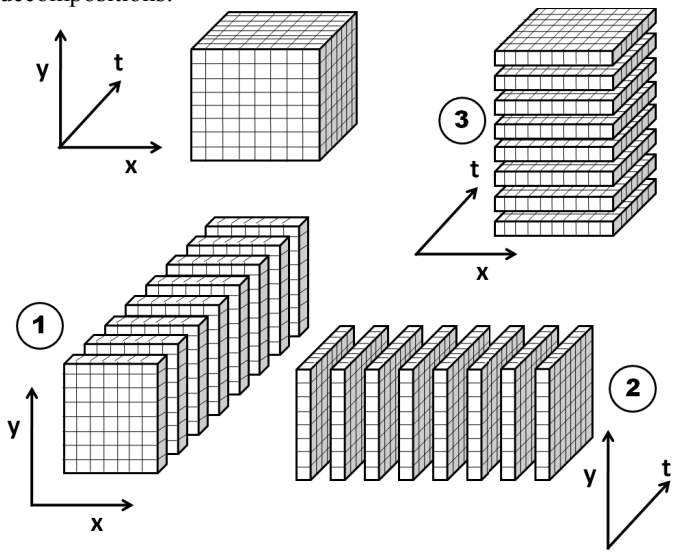

Figure 5. Different 3D signal decomposition

We distinguish the spatial decomposition known, used by most conventional approaches to video compression. This decomposition advantages the correlation in the planar surface (x, y).

We also distinguish two temporal decompositions, the first that supports the correlation in the (y, t) planar surface, the second advantages the correlation in the (x, t) one.

The basic assumption which says that temporal correlation is more relevant than spatial one in video sequences, favors the last two decompositions, whereas the proposed approach only considered the second decomposition, which supports the correlation in the planar surface (y, t). It will be interesting to consider these different decompositions to identify which offers the highest correlation.

The proposed approach tends to more exploit the correlation by looking for directions where pixels intensity change is minimal. This direction can be retrieved once for all the GOP, or calculated for each block (eg block 8x8x8) alone as it is the case in this paper.

B. Direction detection and scan selection

The basic idea is to rank the pixels in such a way that two adjacent pixels have the highest similarity. To overcome this objective we analyze the video pixels activity in order to find the suitable scan that could provides the most correlated representation. The best way is to scan the video frames according to the direction where minimal pixel's change is recorded.

First, for practical purposes, the proposed scan methodology splits each GOP of the original video into isometric 3D blocks. The scan direction is computed for each block individually because each one has its own local auto-correlation characteristic and nearby pixel similarity. The scan direction selection is determinate by the decisional gradient - based algorithm presented below.

The *gradient* of a function of three variables, F(x,y,t), is defined by:

$$\vec{G} = \nabla F = \frac{\partial F}{\partial x}\vec{i} + \frac{\partial F}{\partial y}\vec{j} + \frac{\partial F}{\partial t}\vec{k} \quad (1)$$

and can be thought of as a collection of vectors pointing in the direction of increasing values of F.

The gradient of one three-dimensional bloc F is defined by:

[Gx,Gy,Gt] = gradient(F)

Where Gx (resp. Gy and Gt) corresponds to ∂F/∂x (resp. ∂F/∂y and ∂F/∂t, the differences in x (resp. y and t) direction.

The spacing between two adjacent points in each direction is assumed to be one.

Each point (i,j,t) in the 3D block F is presented by a local gradient:

$$\overrightarrow{g(i,j,t)} = gx(i,j,t)\vec{i} + gy(i,j,t)\vec{j} + gt(i,j,t)\vec{k} \quad (2)$$

The global gradient is:

$$\vec{g} = \sum_{i=0}^{a-1}\sum_{j=0}^{a-1}\sum_{t=0}^{a-1}(gx(i,j,t)\,\vec{i} + gy(i,j,t)\,\vec{j} + gt(i,j,t)\vec{k}) \quad (3)$$

Where a is the dimension of the image block.

Let:

$$\overrightarrow{Gx} = \sum_{i=0}^{a-1}\sum_{j=0}^{a-1}\sum_{t=0}^{a-1} gx(i,j,t)\,\vec{i}$$

$$\overrightarrow{Gy} = \sum_{i=0}^{a-1}\sum_{j=0}^{a-1}\sum_{t=0}^{a-1} gy(i,j,t)\,\vec{j}$$

$$\overrightarrow{Gt} = \sum_{i=0}^{a-1}\sum_{j=0}^{a-1}\sum_{t=0}^{a-1} gt(i,j,t)\,\vec{k}$$

Then, each image block is represented by its own gradient components. Therefore, the direction of the block pixel's change will be approximated as following:

- If ($|Gx| \gg |Gy|\ and\ |Gx| \gg |Gt|$) then

The minimal pixel's change direction may be recorded in the (y, t) plan, consequently the selected accordion will be ACC1 (cf. figure 6).

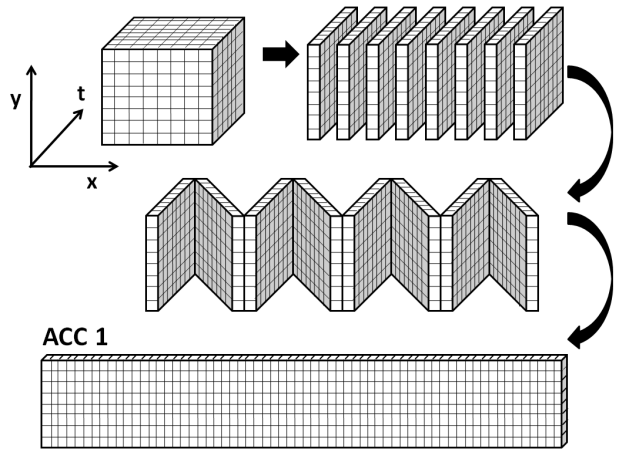

Figure 6. First Accordion representation: ACC1

- If ($|Gy| \gg |Gx| and\ |Gy| \gg |Gt|$) then

The minimal pixel's change direction may be recorded in the (x, t) plan, consequently the selected accordion will be ACC2 (cf. figure 7).

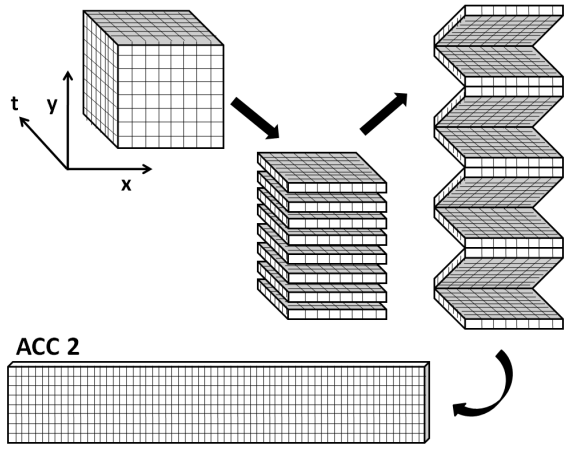

Figure 7. Second Accordion representation: ACC2

- If ($|Gt| \gg |Gx|\ and\ |Gt| \gg |Gy|$) then

The minimal pixel's change direction may be recorded in the (x,y) plan, consequently the selected accordion will be ACC3 (cf. figure 8).

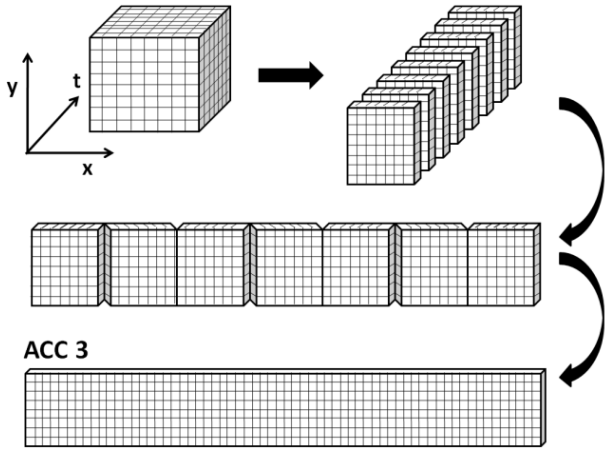

Figure 8. Third Accordion representation: ACC3

C. Dynamic Accordion based video coding scheme

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 is divided into 3D blocks (in this paper the 3D block size is 8x8x8). For each 3D block, the type of the scan is selected according of its global gradient orientation as it is shown in the preceding section. All blocks are projected into the 2D representation (cf. figure 9). This latter will be compressed, as it is the case for the classic accordion based compression approach, using image encoder. In this paper, N, the constructed stack depth, will be fixed to 8 and the used encoder is JPEG.

Figure 9. Dynamic Accordion representation example

Figure 9 shows the new representation, further called IACC+, resulting in the dynamic accordion transformation. This representation is formed by the first four frames of the "Miss America" sequence.

IV. PERFORMANCES EVALUATION

A. Correlation improvment

This figure 9, shows the strong autocorrelation representation "IACC +". Autocorrelation measurement based Comparative between the classic accordion representation the dynamic one confirms this observation as shown in figure 10.

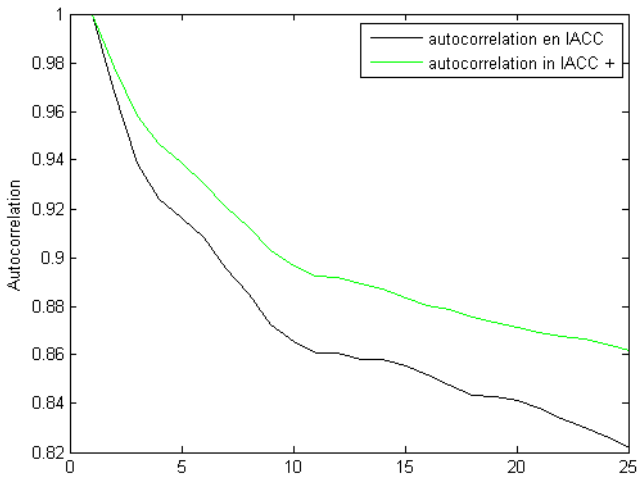

Figure 10. Comparative on autocorrelation

Note also that the autocorrelation in IACC + looks stronger locally than globally. This could be suitable for JPEG coding. Indeed, JPEG splits image into 8x8 blocks.

B. PSNR improvment

For all tested video sequences, the dynamic approach ouperforms the classic one. The improvement is, in the most of time, detected in sequences that contain fast and non uniform motion.

The figure 11 (resp. 12) show the PNSR comparison between ACC-JPEG" and "ACC-JPEG +" for a slow (resp. fast and non uniform) motion video sequence example (Miss Amercia).

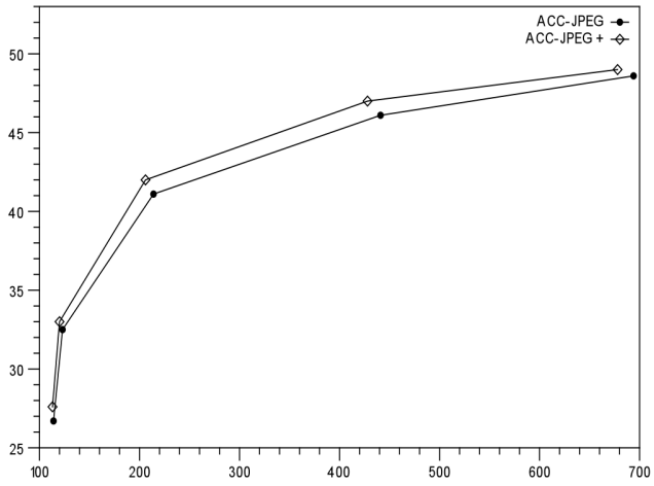

Figure 11. Comparison between "ACC-JPEG" and "ACC-JPEG +" based on PSNR (Miss America)

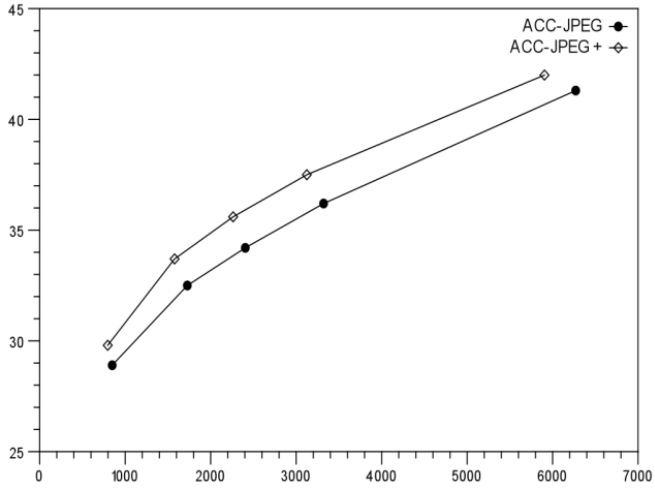

Figure 12. Comparison between "ACC-JPEG" and "ACC-JPEG +" based on PSNR (Foreman CIF, 25Hz)

The improvement in compression performance looks stronger for video containing fast motion, which recognizes the additional complexity of the dynamic approach because these fast and non uniform motion videos have often been the weakness of the classic approach [11][12].

This compression efficiency improvement is confirmed by subjective evaluations.

C. Subjective quality based evaluations

In this part of experimentations, we give a subjective comparative between the new approach and the classical one.

The figures 12 (resp. 14) show the visual image quality of the frame 235 relative to the "foreman sequence" compressed using the classical ACC-JPEG (resp. ACC-JPEG+). These frames which contain a fast and non uniform motion are situated in the moment where the camera suddenly starts to move.

Figure 13. Perceptual quality given by ACC-JPEG (foreman, CIF, 25, 235 th frame)

Figure 14. Perceptual quality given by ACC-JPEG + (foreman, CIF, 25, 235 th frame)

The figures 11 and 12, show clearly that ACC-JPEG + provides better visual quality, especially for image details.
In fact, the dynamic Accordion often provides a "smoother" representation of the video frames while transforming high signal frequencies to low ones.

V. CONCLUSION

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, the Accordion approach comes with a new objective which consists in the subjective way of exploring temporal and spatial redundancy with the minimum complexity processing; although its performances this approach presents some limits in video with fast and non uniform motion. To overcome such limits, a new approach is proposed which consists in a dynamic scan of the video frames. Many experiments were conducted in order to prove the new method's performances at the cost of some additional complexity. With the obvious gains in compression efficiency, we predict that the proposed method could open new horizons in Accordion based video compression exploration.

There are several directions for future investigations. First of all, we will try to combine the new Accordion representation with other image coding techniques (JPEG 2000). Another direction could be to explore other strategies for the scan direction selection.

VI. REFERENCES

[1] A Molino, A., Vacca, F.: Low complexity video codec for mobile video conferencing. In:EUSIPCO, Vienna, Austria, pp. 665–668 (2004)

[2] Y. Q. Shi and Huifang Sun, "Image and Video Compression for Multimedia Engineering Fundamentals, Algorithms, and Standards", CRC Press 2000. Motion Estimation and Compression chapter book, section 3. Publication Date: 2008.

[3] S.B. Gokturk and A.M. Aaron, "Applying 3d methods to video for compression". In: Digital Video Processing (EE392J) Projects Winter Quarter 2002.

[4] M.P. Servais, "Video compression using the three dimensional discrete cosine transform". In: Proc. COMSIG, pp. 27–32 (1997)

[5] R.A. Burg, "3d-dct real-time video compression system for low complexity single chip VLSI implementation". In: Mobile Multimedia Conf, MoMuC 2000.

[6] J.J. Koivusaari, and J.H. Takala, "Simplified three-dimensional discrete cosine transform based video codec". In: Proc. SPIE-IS&T EI Symposium, San Jose, CA, 2005.

[7] M.P. Servais, G. De Jager, "Video Compression using the Three Dimensional Discrete Cosine Transform", Proc. COMSIG 2007, pp. 27-32, 1997.

[8] R.Kutil and andreas Uhl. "hadware and software aspects for 3D wavelet decomposistion on shared Memory MIMD computers", in proceeding of ACPC'99, volume 1557 of lecture notes on computer science, pp 347_356, springer-verlag 1999.

[9] H. Khalil, A. F. Atiya, and S. Shaheen,"Three-Dimensional Video Compression Using Subband/Wavelet Transform with Lower Buffering Requirements", IEEE Transactions on Image Processing, VOL. 8, NO. 6, June 1999.

[10] T. Sikora, "Trends and perspectives in image and video coding," Proceedings of IEEE, vol. 93, no. 1, pp. 6-17, Jan 2005.

[11] T. Ouni, W. Ayedi and M. Abid , "New low complexity DCT based video compression method", ICT 2009, P 202-207, Marrakech, Morocco.

[12] T. Ouni, W. Ayedi and M. Abid , "Non predictive wavelet based video coder", ICIAR 2010, Part I, LNCS 6111, pp. 344–353POVOA DE VARZIM, PORTUGAL.