

Wavelet-based Video Coding using Advanced Fractional Motion Estimation Technique

Wissal Hassen and Hamid Amiri

Signal, Image and Information Technologies laboratory
The Electrical Engineering Department of National Engineering School of Tunis TUNISIA
Wissal_hassen@yahoo.fr, hamidlamiri@gmail.com

Abstract—The purpose of this paper is to encode a color video by wavelet transformation. Therefore, we propose a new hybrid approach which combines a fractional motion estimation technique. Several studies were carried out to reduce the spatial and temporal redundancies, hence at the level of spatial video coding, we use a new approach based on sub-bands coding through a discrete wavelet transformation. This technique is based on the principle of the EZW algorithm of Shapiro. It proceeds by separating the encoding of the signs and the magnitudes of wavelet coefficients. Then, at the level of temporal compression, we propose a study of motion estimation with different accuracy based on image interpolation to improve the quality of predicted frame. Next, we present a representation reducing the size of the motion vector field and we compress it by two of entropic coding approaches namely Huffman coding and arithmetic coding.

The proposed video codec was applied on a video sequence with different sizes (CIF and QCIF) and different dynamics. The obtained results, in terms of objective assessment (PSNR, the SSIM and VQM), were satisfactory compared with other video coding standards. We have also proposed a subjective evaluation and the results are compared to those obtained by H.264/AVC standard.

Index Terms—Wavelet transform, H.264/AVC standard, image quality assessment, fractional motion estimation, video coding.

I. INTRODUCTION

Video compression is necessary for storage and transmission in multimedia applications. In a video sequence there are two kinds of redundancy; the spatial redundancy represented by blocks of pixels that repeat in the same image and temporal redundancy that occurs when blocks of pixels are permanent in two or more successive images such as the case of a fixed background. The object of the video compression is to reduce those redundancies.

The first part of this paper is reserved for intra-frame coding. Then, in the second part, we present the chosen inter-frame compression technique and finally, we present the overall video coding algorithm and its results with objective and subjective evaluations. The spatial

compression is based on the image encoding in a frequency domain. Therefore, the image goes through three main stages: frequency transformation, quantification and an encoding step. The standard of image compression JPEG uses the Discrete Cosine Transform (DCT). This transformation is applied on blocks of 8×8 pixels; therefore the decoded image presents artifacts in these blocks. However, JPEG.2000 standard uses Discrete Wavelet Transform (DWT) which is a very powerful tool for image compression and provides both spatial and frequency localization of image energy. Due to the good quality of the DWT-based image coding, several studies have used this technique such as the Embedded Zerotree Wavelet algorithm (EZW) of Shapiro [1] and the Set Partitioning In Hierarchical Trees (SPIHT) coding algorithm, proposed by Said and Pearlman [2]. In our work, we present a new wavelet based codec inspired from the EZW algorithm and based on separate entropy coding of sign and magnitude of wavelet coefficients. Then we show that this algorithm provides a very good quality of compression for video frames by evaluating its performance compared to JPEG.2000 standard. The choices of wavelet decomposition level as well as the proper wavelet filter are discussed in this section.

The main object of the second part of this paper is to reduce the temporal redundancies, then we present the principle of motion estimation by Block Matching and we improve the performance of this technique by introducing a new fractional estimation technique based on image interpolation. Thereafter, we propose a coding technique of motion vector by entropy coding and compare the results found for two coding algorithms: Huffman encoder and arithmetic encoder.

In the third part of this paper, we propose a new video compression scheme which is based on our previous results. The proposed scheme provides very satisfying results which are evaluated relative to other video coding standards.

II. WAVELET-BASED SPATIAL VIDEO CODING

The spatial coding algorithms are based on the decomposition of the image into sub-bands as shown in Fig. 1. In fact the sub-band coding by DCT or DWT is initially used in speech processing [3] and it has been successfully extended to still image compression [4, 5]. A

sub-bands encoder decomposes the image into sub-images. In DCT, the image is split in blocks of 8×8 pixels and the transformation is applied to each block as an independent sub-image. In the case of DWT, the image is decomposed in wavelet domain without block splitting. The DCT is applied by blocks of 8×8 pixels which produce the phenomenon of artifacts that is not the case for the DWT which is applied to the entire image. Thereafter, each sub-band goes through a different encoding depending on desired quality and compression bitrate. Then, a progressive encoding bit-plane by bit-plane is applied. Fig. 2 shows an example of progressive coding and the results for each compression bitrate. The value Bpp (Bits per pixel) indicates the bits required to encode each pixel. The advantage of a progressive encoder is that it allows obtaining a better compression quality for a given bitrate and accomplishing this task in a nested manner. Therefore, the generated bitstream can be cut anywhere of the encoding process without losing the already coded bit plane.

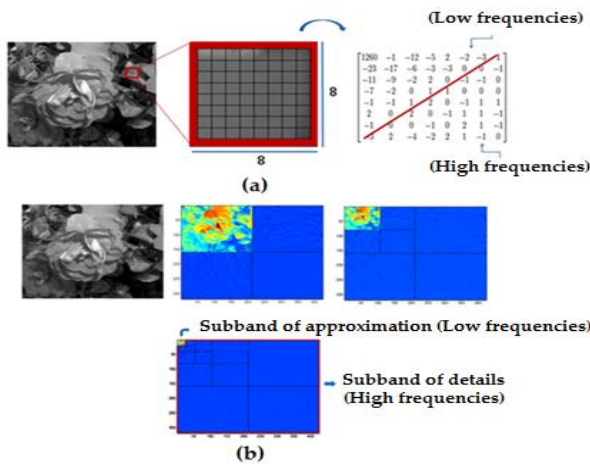


Fig. 1. The DCT is applied by bloc of 8×8 pixels (a) The DWT is applied on the on the entire image (b)

The Embedded Zerotree Wavelet algorithm (EZW) of Shapiro is one of the first progressive encoders, which was improved by the Hierarchical Tree Single Partition algorithm (SPIHT) of Said and Pearlman. The encoder Embedded Block Coding with Optimal Truncation Points (EBCOT) [6] introduced by Taubman is another encoder by bit planes characterized by its performance, but its implementation is more complex. Recent algorithms in the field of bit-plane coding improve the coding principle of EZW. The EZBC encoder [7] for example is also used in video coding [8] under the name MC-EZBC. In this work, we focus on a new encoder called Separate Sign Coding (SSC), which is an inspiration from the EZW algorithm. The SSC algorithm is a new and easy to implement encoder, it offers a quality of compression exceeding that of SPIHT and EZW encoders [9]. In this section we present the principle of SSC codec, and we study its performance using some test video images while changing levels and the kinds of wavelet. The quality of this codec is evaluated relative to the JPEG.2000 standard by objective metrics and human perception.

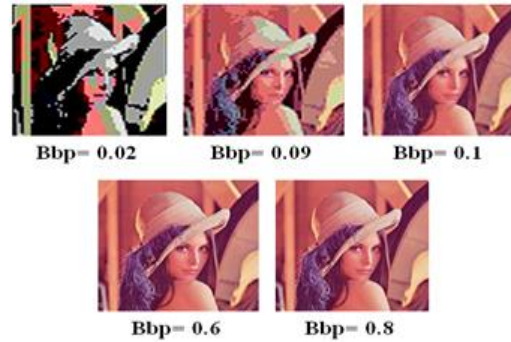


Fig. 2. The An example of a progressive encoder: the image is completely viewable even if you received only 3% of the file. The image is displayed immediately in full, and is refined gradually

The SSC encoder performs two successive passes to determine a list of significant coefficients and refine it. In the first pass, subordinate pass, the image which is decomposed into wavelet is considered as a tree of coefficients constituted from nodes or root and leaves or descendants. These coefficients are arranged from low frequencies to high frequencies. The EZW algorithm codes the coefficients using one of the four symbols: POS, NEG, IZ or ZTR. POS and NEG are used to encode respectively a positive or negative significant coefficient which will be inserted in a List of Significant (LS), IZ is used to encode a non-significant coefficient but has at least one significant descendant and ZTR is used to encode a non-significant coefficient and that his descendants are also not significant (a zerotree). A coefficient is significant if the absolute value of its amplitude is greater than or equal to a threshold T which is initialized to the half the maximum amplitude of the coefficients. After each subordinate pass, coefficients identified as belonging to a previously coded zerotree are replaced by zero. In the refinement pass also called subordinate pass, the algorithm refines the values of the significant coefficients stored in the LS which belong to the interval $[T, 2T]$, then it codes the coefficients of the interval $[T, 3T / 2]$ with the symbol '0', and those of interval $[3T / 2, 2T]$ with the symbol '1'. Once the two passages are performed, the threshold T is halved and the cycle (first pass and second pass) is repeated until obtaining the desired level of compression.

The idea of the SSC algorithm is to exploit the principle of EZW algorithm but the specificity of the first is to encode separately the signs and the amplitudes of the image wavelet coefficients. Therefore, the SSC algorithm replaces POS and NEG with one symbol 'S' which indicates the amplitudes of the significant coefficients, then a map of significant amplitude is progressively generated indicating the presence or absence of the symbol S. The presence of S is encoded by the symbol '1' and its absence with the symbol '0'. Since the probability of a negative coefficient in a sub-band of approximation is zero, a map of signs is also progressively generated only for the detail sub-bands. In this map, the presence of significant positive coefficient is marked by the symbol '0' and the presence of a significant negative coefficient is marked by the symbol '1'. Indeed, the presence of the symbol 'S' allows the decoder to reconstruct the

amplitude of a significant coefficient using the current value of the threshold, and its absence indicates the existence of a zero tree. For the latest of sub-bands of details, the absence of a symbol "S" informs the decoder by the existence of a zero. The amplitude map and the sign map are encoded by entropy coding.

The quality of SSC codec is assessed against the EZW, SPIHT and JPEG.2000 codecs for still images [10]. Here, we assess its quality of SSC encoder for images extracted from video sequences. Therefore, we present the comparison results of this codec relative to JPEG.2000 standard, results are given by histograms presenting the Peak Signal to Noise Ratio (PSNR) relative to the compression ratio in Fig. 3, and subjective results of this comparison are given by Fig. 4. The ratio of compression is calculated as in (1).

$$\text{ratio} = (\text{new_size}/\text{original_size}) \times 100 \quad (1)$$

Where new_size is the new size of the image after compression and original_size is its original size. Test images are of size CIF (Foreman: 352 × 288) and QCIF (Flower and Football: 144 × 176). We see that the quality given by the SSC and codec is better and that the difference of Peak Signal to Noise Ratio (PSNR) may exceed 3 dB for high compression bitrates. At low compression bitrates compression quality degrades for Football and Flower pictures but it remains satisfactory for Foreman image; this can be explained by the size of the Foreman image which is higher than the size of the other images which allows a higher level of wavelet decomposition. We have shown that the SSC codec has the best quality compared to JPEG.2000 standard especially at high compression bitrates. At low bitrates the SSC codec retains better quality for images of high resolution. Thereafter, we study the influence of the level of wavelet decomposition on the SSC compression quality.

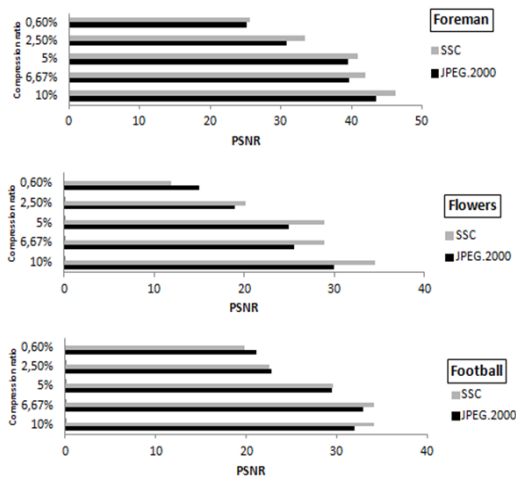


Fig. 3. The The PSNR value relative to the ratio of compression of some video frames (Foreman, Flowers and Football) coded using SSC encoder and JPEG.2000 standard

In Fig. 5 we present the PSNR variation of decoded

image using SSC codec according to the level of wavelet decomposition with different value of bits per pixel (Bpp) used to code each pixel of the decoded image. We used the color “Lena” image which has a resolution of 512 x 512 pixels. We see that as soon as we increase the level of wavelet decomposition, the PSNR value of the reconstructed image is better. In fact, the increase of this parameter (decomposition level) generates the diminution of the energy concentrated around the approximation sub-band and consequently the emergence of more information which gives a good quality of compression. We also see that at a certain level of decomposition - the PSNR value - remains stable. The test performed on the “Foreman” sequence, “Flowers” and “Football” also showed the same results at the exception that the quality does not stabilize a certain level of decomposition because the sizes of the images do not permit to go at higher levels: the level 3 is the maximum for “Football” and “Flowers” sequences and the level 4 is the maximum for “Foreman” sequence.

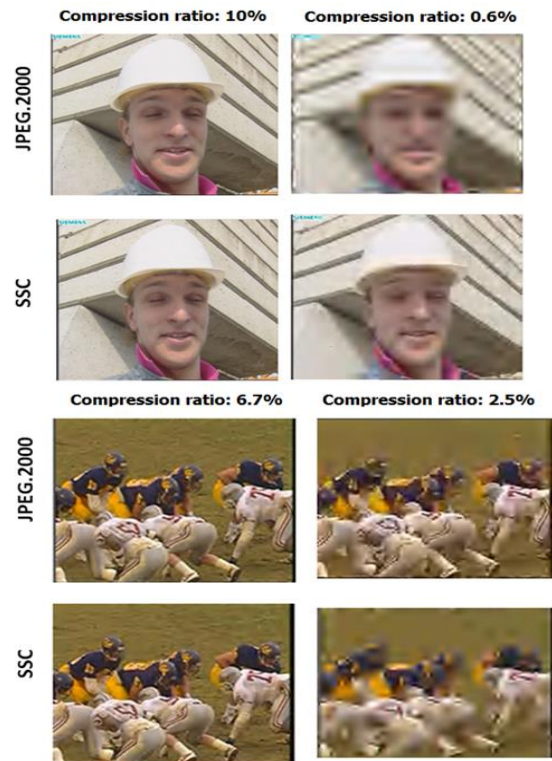


Fig. 4. The Subjective evaluation of SSC encoder and JPEG.2000 standard with variable compression ratio

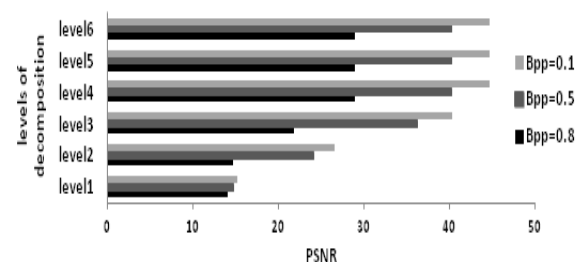


Fig. 5. The PSNR evaluation of SSC codec according to the level of wavelet decomposition with different value of Bpp

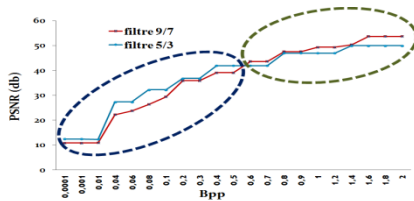


Fig. 6. The The mean value of PSNR relating to CDF 9/7 and Gall 5/3 filters depending on number of Bpp using "Lena", "Flowers" and "Foreman" images

In the context of SSC codec study and the search of optimum parameters of spatial coding, we conducted a comparative study between the two wavelet filters CDF 9/7 and Gall 5/3 which are standardized by the JPEG.2000 codec and are the most adopted in the wavelets based image compression [11].

In Fig. 6, we present the change in the mean value of The PSNR relating to CDF 9/7 and Gall 5/3 filters depending on number of Bpp using "Lena", "Flowers" and "Foreman" images. We find that for very high compression bitrates, the 5/3 wavelet is more efficient than the 9/7 wavelet and better preserves features and for low compression bitrates, the 9/7 wavelet is efficient and gives a good objective quality.

Following the evaluation tests in this part, we can confirm that the SSC codec is more efficient than JPEG.2000 for high resolution images that allow greater levels of decomposition. Later, we will use it in our video compression codec.

III. TEMPORAL VIDEO CODING USING FRACTIONAL MOTION ESTIMATION

The SSC encoder has given very good results for still image compression compared to the JPEG.2000 standard. An attempt to implement SSC encoder for video compression was tested by the use of the difference between images as a technique to reduce temporal redundancies. The first frame of each GOP is encoded by SSC encoder. The rest of the GOP is encoded by calculating the difference between the previous frame (frame reconstructed in the decoder) and the current frame (in the encoder). The difference or the residual image is encoded by SSC. The results show that the system can give a good objective quality of the video sequences which are decoded at high bitrates. However, at low or medium bitrates, the results are not satisfactory.

The comparison of these results with MJPEG.2000 and MPEG.2 standards at using "Flowers" sequence and at 512 kbps shows that the SSC encoder provides a good quality in intra coding mode as shown by the peaks (indicating a very good quality of I-picture compression). However, the quality deteriorates suddenly for the other image of the GOP. Indeed, a video sequence is processed by a Group Of Picture (GOP). Each GOP is composed of 12 pictures: I or P pictures. The first pictures of a GOP must be an I-picture and is coded only with an intra-frame encoder. The rest of the GOP is coded by a differential coding i.e. the difference between I-picture

and every other picture from the rest of GOP, called residual image, is encoded by the SSC codec. The visualization of the decoded video presents saccades at the time of display. These saccades are due to the application of the wavelet transformation on the residual images during compression. This can be explained by the fact that translation is not always ensured by the wavelet transformation, which generates false motions and the reconstructed video is displayed in an irregular manner. To solve this problem, we propose to reduce the temporal redundancies by motion estimation. The principle consists in dividing the sequence in GOP, each GOP is composed of 12 pictures: I or P pictures. The first picture of a GOP is coded by SSC encoder. Predictive pictures (P-pictures) are coded using reference frames which can be a previous I-frame or P-frame.

A. Fractional motion estimation

In the motion estimation, current frame (the frame to predict) and reference frames are decomposed into rectangular Macro-blocks (Mbs). Then, the displacement of each Mb of the current frame is estimated based in the reference frame. After that, this motion vector is used in the motion compensation stage in order to provide the predicted frame from the reference frame. The error of prediction, named the difference frame or residual, is encoded rather than the current frame itself. Also, the estimated motion information has to be transmitted.

The decoder estimates the current frame from data already decoded: the reference frame, the motion vector and the residual. Each Mb in the current frame passes through a stage of search to determine the 'best' matching Mb or the similar Mb in the reference frame. This search can be carried out by making a comparison between the Mb in the current frame and the possible Mbs in a fixed search area (p) in the reference frame. The search is performed by one of block matching algorithms. A popular matching criterion is the Mean Squared Error (MSE) calculated between the current Mb and the reference Mb and provides a measure of the remaining energy in the difference block. This process of finding the best match is known as motion estimation. The offset between the current Mb and the position of the candidate Mb called motion vector is also transmitted after having encoded by an entropy coding.

Although, the motion estimation is based on the motion vector search in the reference image, the fractional motion estimation increases the size of this image by interpolation to obtain more precision at the searching stage. Therefore, the search in a sub-sampled image requires more computation than that in the original image. In spite of this complexity, sub-pixel motion estimation can significantly outperform integer motion estimation which is due to the fact that object will not necessarily move by an integral number of pixels between successive video frames.

Fig. 7 shows the mean value of PSNR for the predicted first GOP of Football sequence using variable block size (Mb) and search area (p) with different accuracy of motion vector estimation. The 1/2 interpolation pixel is

determined by one interpolation of the reference image, the 1/4 interpolation pixel is determined by two successive interpolations of the reference image and 1/8 interpolation pixel is determined by three successive interpolations of the reference image; according to this study we have shown that the results obtained by sub-pixel estimation are more refined than those obtained by estimation without interpolation.

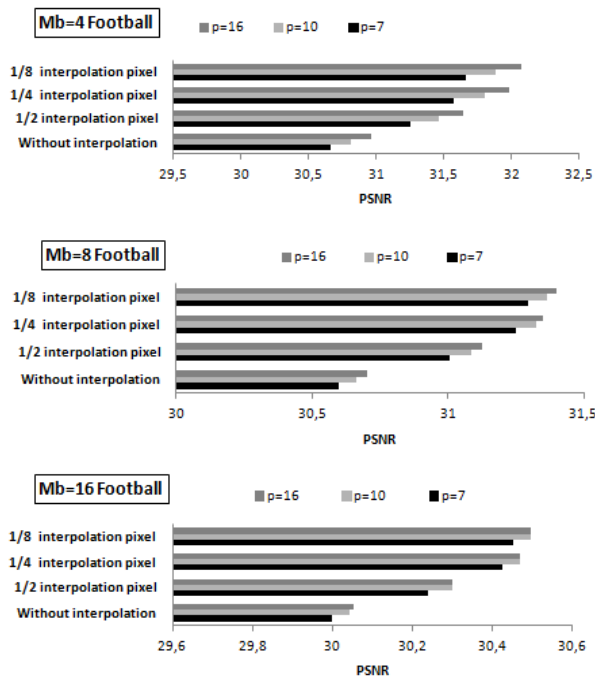


Fig. 7. The mean value of PSNR for the predicted first GOP of Football sequence using variable block size (Mb) and search area (p) with different accuracy of motion vector estimation

B. Motion vector encoding

The reduction of temporal redundancy by motion estimation requires to estimate the motion vector with precision and to encode it. The size of the motion vector depends on the resolution of the sequence and the size of the Mb. In fact, a color image sized, divided into MBs, is formed by three-color components corresponding to the selected color space. In particular for YCbCr space, each color component of the motion vector has two coordinates x and y as shown in Fig. 8 where LMV is the luminance component of the motion vector, C_bMV represents the blue chrominance component of the vector and C_rMV represents the red chrominance component of this vector, Δ_x and Δ_y represent the displacements of coordinates x and y of the current Mb.

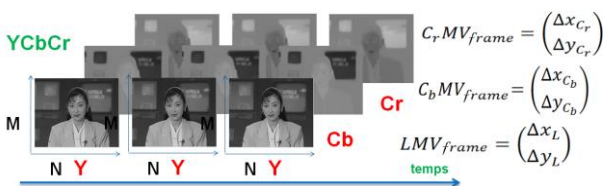


Fig. 8. The motion vector for a color sequence represented in the YCbCr space

Based on the fact that the human eye is more sensitive to luminosity variations than to the color variation, and to reduce the size of motion vector, we introduce a chromatic sub-sampling step before estimating the motion vector. We conducted three tests of Motion estimation. In the first test, we used the video format YCbCr₄:2:0 which maintains for each group of four luminance pixels one chroma pixel. In the second test, we used the format YCbCr₄:1:0 to determine a 1/8 chrominance representation which stores a one chrominance component for every eight luminance components. In the third test, we reserved the chromatic components without modification using the format YCbCr. The result is given by Fig. 9 using two test sequences: Foreman (formed by 250 frames) and Flowers (formed by 130 frames). The PSNR of a sequence is the average value of the PSNR calculated for every predicted frame of the sequence. We see that chromatic sub-sampling step does not much influence on the compressed video quality and that the size of the color motion vector is reduced without great loss of quality. Based on these results we use the YCbCr₄:1:0 format to estimate chrominance motion vectors. This format reduces the size of the motion vector with a rate of 35% without loss of quality.

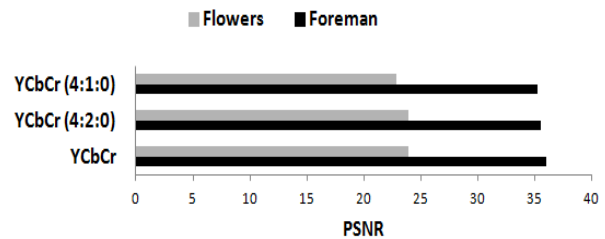


Fig. 9. The mean value of PSNR relating to Flowers and Foreman sequences depending on the used format and the chromatic sub-sampling

The major challenge after the estimation of motion vector is its coding with the minimum loss of information. In this part we study the encoding parameter of this vector. The objective at this stage is to compress and store all the information in the minimum of symbols. This is a conservative compression phase, to encode a motion vector with entropy coding it must be determined by the values of the vector (symbols) and their frequency of occurrence. Later, the encoder maps a symbol or a group of symbols to a certain code. It uses a statistic model calculation with a more or less complex manner, the probability of occurrence of the symbol to be encoded. Thus, a symbol having a high probability of occurrence is coded on very few bits and vice versa. We present in Fig. 10 the distribution of the motion vector symbols characterizing the vertical motion (x-axis) and the horizontal motion (y axis) for each color component of Akyio (300 frames) Flowers (130 frames) and Bus (150 frames) sequences. For each sequence, the luminance motion vector field LMV, the blue and the red chrominance C_bMV C_rMV were presented by two histograms showing the motion along the axes x and y. The motion estimation is performed by the exhaustive

search algorithm, the size of Mb is (4×4) and the search window is (7×7) . Sequences are used with the format $YC_bC_r4:1:0$ and they are treated by GOP of 12 images.

The histograms of "Akyio" sequence are concentrated around the symbol "0" and the probabilities for other symbols are negligible which is obvious seen that this sequence contains small motions. We see also that "Akyio" sequence contains both positive and negative motion, which explains the symmetry of its histograms. In the histograms of "Flowers" sequence we see that the horizontal displacements are more important than the vertical motions. This can be explained by the fact that the sequence has a horizontal scanning of the camera carried on a stationary object (landscape). Also, "Bus" sequence exhibits symmetric horizontal motions caused by the motion of the bus as fast as the sweep of the camera. However, the vertical motions are not negligible given that there was a vertical scan of the camera.

Generally, we find that the luminance, blue chrominance and red chrominance histograms of each sequence are not independent. Indeed, the direction and type of motion in the luminance component are similar in color components but with some difference in the color. This difference derives from the fact that a change of color is seen as motion. Thus, we propose to encode a global motion vector denoted by MVG instead of coding each vector of color component alone. Then, our approach consists in coding to simultaneously LMV , C_bMV and C_rMV as a single global motion vector MVG using one code of entropy. The MVG vector is the concatenation of the horizontal and vertical components of these three vectors as given by (2):

$$MVG = LMV_x \cdot LMV_y \cdot C_bMV_x \cdot C_bMV_y \cdot C_rMV_x \cdot C_rMV_y \quad (2)$$

In Fig. 11 we show the distribution of occurrence probabilities of each value (or symbol in terms of entropy coding) of the global motion vector "MVG" for each test sequence. Our purpose is to code the MVG using entropy coding.

Fig. 12 shows the result of MVG compression by Huffman coding and arithmetic coding. We can deduce that the arithmetic coding is more efficient in compression than Huffman coding. We also note that the arithmetic coding is very efficient for the "Akyio" sequence because it does not present equal probabilities as for other sequences. Following this study, we propose to use arithmetic coding in our video codec. The Gain compression is given by (3).

$$\text{Gain compression} = \left(1 - \frac{\text{new size}}{\text{original size}}\right) \times 100\% \quad (3)$$

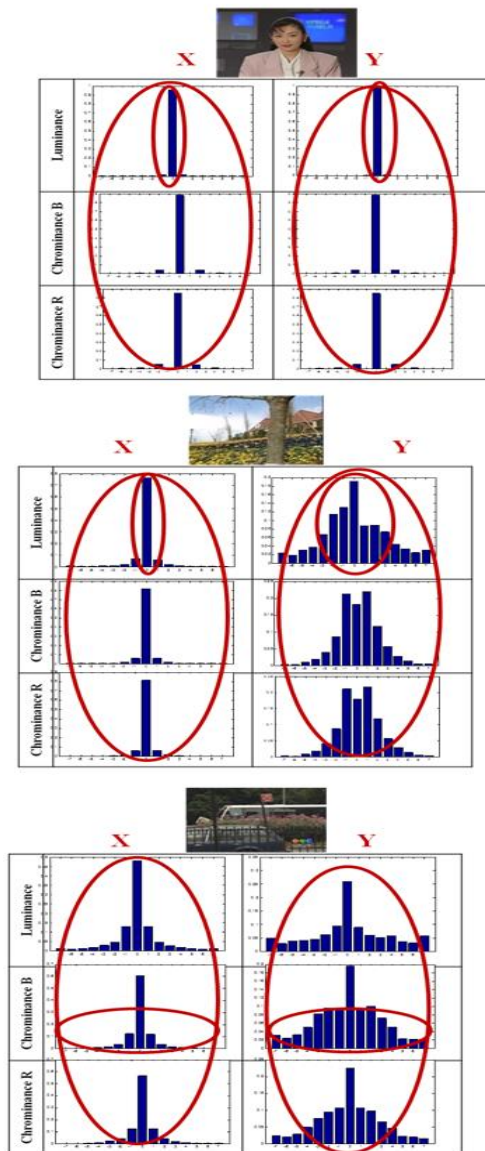


Fig. 10. The The Normalized histograms of the vertical and horizontal motions of Akyio, Flowers and Bus sequences for each color component

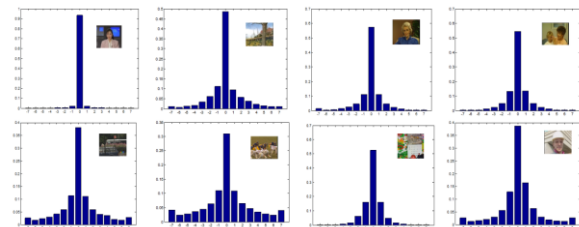


Fig. 11. The The Normalized histograms of the Global Motion Vector MVG of Akyio, Flowers, Grand-mom, Mother-daughter, Bus, Football, Mobile and Foreman sequences

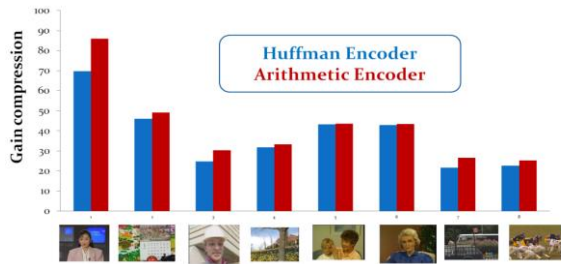


Fig. 12. The Normalized histograms of the Global Motion Vector MVG of Akyio, Flowers, Grand-mom, Mother-daughter, Bus, Football, Mobile and Foreman sequences

IV. THE PROPOSED VIDEO ENCODING AND EVALUATION

Following the study carried out before, we propose in this part a hybrid video coding scheme in Fig. 13 based on the SSC codec and on the Fractional Motion Compensation (SSC-MC) than we have studied these parameters. The video is divided into GOPs 12 images. Then each image is converted to the format $YCbCr4:1:0$ that we adopt in our compression scheme. Each GOP which passes through the encoder is processed in two video compression stages depending on the type of image to be encoded: I picture or P picture. The encoder proceeds by an intra coding pass to encode I-frame using SSC algorithm. Nevertheless, the encoder proceeds by inter coding pass to encode p-frame using the sub-pixel motion estimation. The intra coding pass can encode the I-frame and the prediction error which is determined by the difference between the compensated image and the reference image decoded by SSC. The regulation of compression bitrates is controlled by the encoder. It sets a flag determining the desired compression bitrates for each type of picture (I frame and error frame). Thus, we can allocate more bits to code I-pictures and fewer bits for coding error images. This ensures a good quality for the key frames (I-frames) since the encoding of the rest of the GOP is based on the quality of I-frame. Motion compensation produces a compensated image using only the motion vector and the decoded reference image. Finally, the motion information or motion vector is encoded by entropy coding.

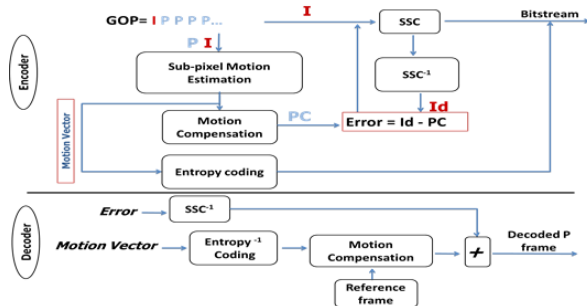


Fig. 13. The proposed wavelet-based video coding schema using fractional (sub-pixel) motion estimation: Id indicates decoded I frame, PC indicates compensated P frame

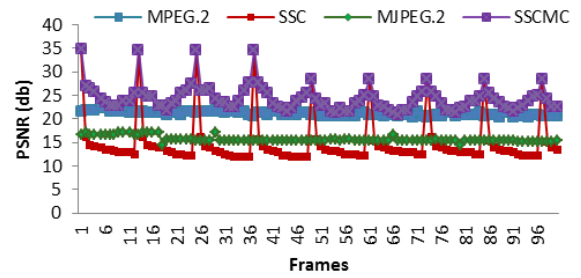


Fig. 14. The Comparison of MPEG.2, MJPEG.2000, SSC and SSCMC encoders using “Flowers” sequence decoded at 512 kbps

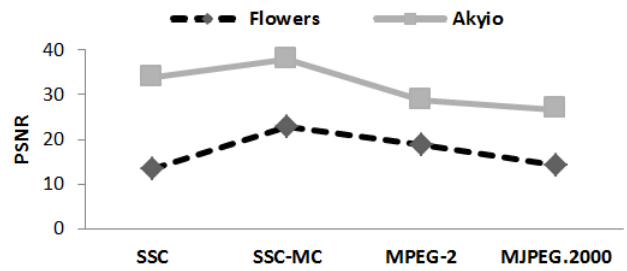


Fig. 15. The Comparison of MPEG.2, MJPEG.2000, SSC and SSCMC encoders using “Flowers” and “Akyio” sequences decoded at 512 kbps

At the decoder, the motion vector passes through an inverse entropy coding. The decoded vector is used in motion compensation to generate the compensated image from a decoded reference image. Then, the current frame is derived by summing the decoded error with the compensated image. The motion vector is coded without loss by arithmetic coding.

Objective and subjective evaluations of our codec are performed on more than one of compression results, coding environment is as follows: the 9/7 wavelet filter is used in the spatial encoding, and the level of wavelet decomposition is 3 or 4 depending on the resolution of the sequence (QCIF and CIF). The technique of motion estimation used is Block-Matching, the research method is the exhaustive search (ES) and the size of Macro Blocks is (4×4) . The motion vector is determined by sub-pixel estimation. The results of this encoder are compared to multiple video standards using several metrics. Indeed, we used the PSNR (Peak Signal to Noise Ratio) [12], The Structural similarity (SSIM) by Wang et al [13], this method is designed to measure the similarity between two images. The SSIM index is a decimal value between 0 and 1. When the SSIM is 0, the correlation between two images is zero; and when the SSIM is 1, the two images are identical. The third metric of evaluation is the Video Quality Evaluation (VQM) [14] which is a metric based on the discrete cosine transform DCT. In this case, if the VQM value is close to 1, the image quality is good.

The first comparisons of the proposed encoder SSC-MC are made with respect to MJPEG.2000 standards and MPEG-2. In fact, we are interested in comparing our algorithm beside MJPEG.2000 since it is the only

standardized encoder using the discrete wavelet transformation.

In Fig. 14, we present the resolution of problematic presented in the third section, we see the improvement of the quality of SSC decoder by applying the sub-pixel estimation and that the proposed codec SSC-MC presents a quality exceeding given results by SSC, JPEG.2000 and MPEG.2 encoders. Further note that the quality objective has been improved by the proposed codec and that the decoded sequence is no longer jerky when displayed.

The proposed codec has been assessed against MPEG.2, MJPEG.2000 and SSC encoders using "Flowers" and "Akyio" sequences decoded at 512 kbps in Fig. 15 which presents the mean value of PSNR on each frame of the sequence; it is clear that the results given by our approach present a quality that exceeds those given by other codec.

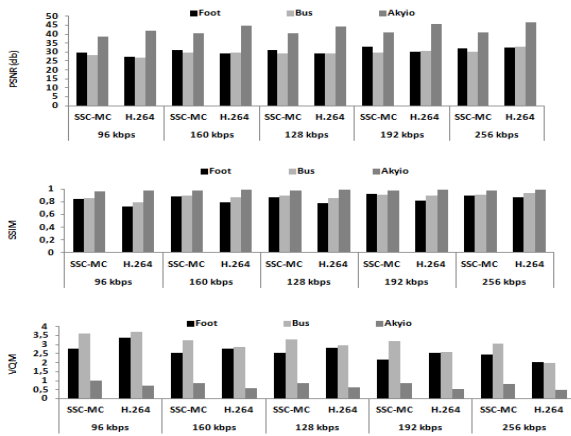


Fig. 16. The Objective assessment with PSNR of SSC-MC at different bitrates using "Football", "Bus" and "Akyio" sequences

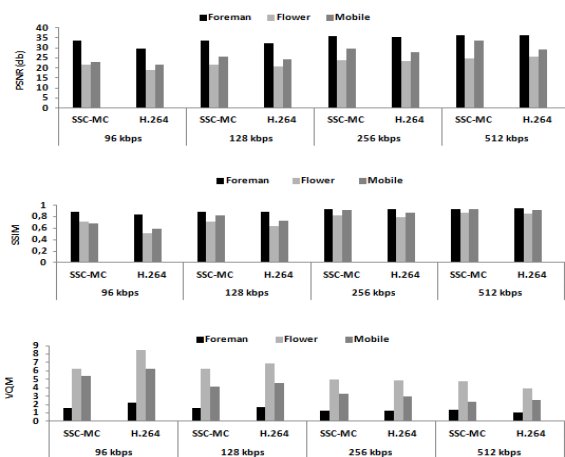


Fig. 17. The Objective assessment with PSNR of SSC-MC at different bitrates using "Foreman", "Flower" and "Mobile" sequences

In Fig. 16 and Fig. 17, we compare the proposed codec to the H.264 / AVC standard [15] with different compression bitrates through the average values of PSNR, SSIM and VQM respectively using QCIF sequences and CIF sequences; we find that the sequences encoded by SSC-MC retain a PSNR value greater than that given by the H.264/AVC at low bitrates. However, for high

bitrates, the results of both encoders are similar and especially for CIF sequences. The evaluation by SSIM and VQM proves these findings. In Fig. 18 and Fig. 19, we present respectively an assessment with PSNR of SSC-MC at low bitrates for "Foreman" and "Flowers" sequences encoded at 128 kbps. We note that the encoding quality of "Foreman" sequence is better than that of the "Flowers" sequence. This is caused by the fact that the last sequence is decomposed until three levels of DWT unlike the "Foreman" sequence which has been decomposed up to four levels of DWT. But, when comparing the two results by H.264/AVC results, we see that the proposed codec has better quality.

In Fig. 20, we present a subjective assessment of the "Football" sequence encoded at 96 kbps through some images of this sequence. We note that even if the selected images have very fast motion, the encoding quality of I-images by our approach exceeds that given by the H.264/AVC standard which is very perceptible by the human eye. We also note that the coding quality of P-frames (frame No. 43 and frame No. 52), which are very sensitive to image coding is acceptable. This quality exceeds that determined by H.264/AVC which uses smoothing filter step to hide artifacts due to coding by DCT. Indeed, this step can produce a loss in fine detail in the sequence. This phenomenon is clearly visible in the images 43 and 52 of "Football" sequence.

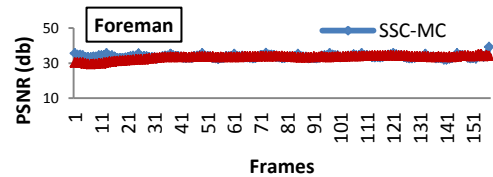


Fig. 18. The Assessment with PSNR of SSC-MC at low bitrate "Foreman" sequence is encoded at 128 kbps.

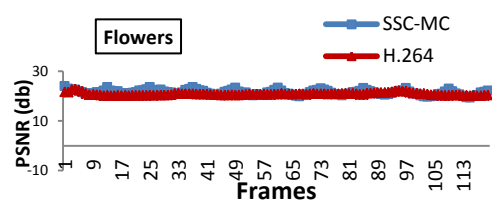


Fig. 19. The Assessment with PSNR of SSC-MC at low bitrate "Flower" sequence is encoded at 128 kbps.

In Fig. 21, we present a subjective assessment of the "Foreman" sequence encoded at 128 kbps through some images of this sequence, we see that the quality of the images coded by the SSC-MC encoder is very similar to that given by the H.264/AVC standard and they are clearer in some cases: for example, the image 155 of "Foreman" is decoded by H.264/AVC and is filtered by smoothing, it is not better compared to the decoded image by SSC-MC. In the image number 155 decoded by SSC-MC, the fingers of the person are clear and away from one another. However, in the results of H.264/AVC encoder, they are close to one another because of the smoothing effect. This effect produces a loss of small details.

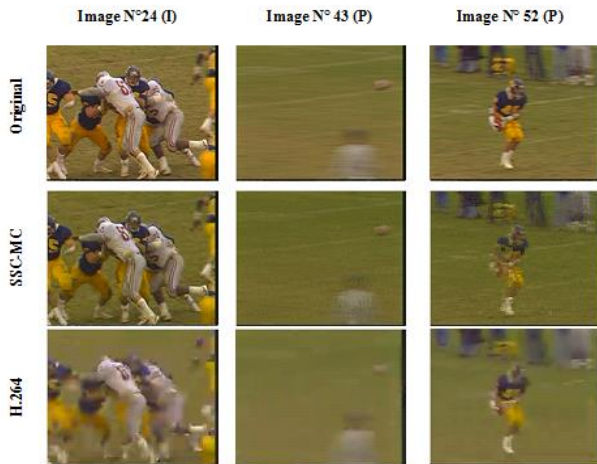


Fig. 20. The Subjective assessment of the sequence "Football" encoded at 96 kbps



Fig. 21. The Subjective assessment of the sequence "Foreman" encoded at 128 kbps

V. CONCLUSION

In this paper, we propose a progressive encoding scheme with an estimation and motion compensation technique by interpolation. We present a Fractional estimation of the vector and we propose an optimal representation of the motion vector to encode it with an entropic encoder at level of temporal compression. However, at the level of spatial compression, we propose to use the SSC algorithm and we determine its optimal compression parameters. The last part of this paper presents the overall scheme of the proposed video coding with an objective evaluation through PSNR, SSIM, and VQM added to a visual assessment. We have shown that the proposed codec is very efficient especially at low bitrates and high resolution video coding.

REFERENCES

[1] ZHANG YU-JIN. Image project(media), image analysis. Shapiro JM. Embedded image coding using zerotrees of wavelet coefficients, *Signal Processing, IEEE Transactions on* 1993; 41, 3445-3462.

- [2] Said A, Perlman W. A new, fast, and efficient image codec based on set partitioning in hierarchical trees, *Circuits and Systems for Video Technology, IEEE Transactions on* 1996; 6, 243-250.
- [3] Galand C. Codage en sous-bandes: th orie et application   la compression num rique du signal de parole. Diss, Universit de Nice, 1983.
- [4] Vetterli M. Multi-dimensional sub-band coding: some theory and algorithms. *Signal processing* 1984; 6, 97-112.
- [5] Woods JW, O'neil SD. Subband coding of images. *Acoustics, Speech and Signal Processing, IEEE Transactions on* 1986; 34, 1278-1288.
- [6] Taubman D. High performance scalable image compression with EBCOT. *Image Processing, IEEE transactions on* 2000, 9, 1158-1170.
- [7] Acharjee, Suvojit, and Sheli Sinha Chaudhuri. Fuzzy Logic Based Four Step Search Algorithm for Motion Vector Estimation. *International Journal of Image, Graphics and Signal Processing (IJIGSP)*, 2012, vol. 4, no 4, p. 49.
- [8] Devarinti, Krishna Kaveri, T. Sai Lokesh, and Gangadhar Vukkesala. Bit Serial Architecture for Variable Block Size Motion Estimation. *International Journal of Image, Graphics and Signal Processing (IJIGSP)*, 2013, vol. 5, no 8, p. 63.
- [9] Mbainabeye J, Ellouze N, Olivier C. Wavelet Based Color Image Compression and Mathematical Analysis of Sign Entropy Coding. *International Journal of Computer Science* 2011; 8 (6).
- [10] Mbainabeye J, Ellouze N. Optimal Image Compression Based on Sign and Magnitude Coding of Wavelet Coefficients. *International Journal of Signal Processing* 2006; 3, 243-251.
- [11] Hassen W, Amiri H. The 5/3 and 9/7 wavelet filters study in a sub-bands image coding. In: 39th IEEE IECON Annual Conference of the IEEE Industrial Electronics Society; 10-13 November 2013; Vienna, Austria: IEEE. 136-139.
- [12] Watson AB, Kreslake L. Measurement of visual impairment scales for digital video. In *Photonics West 2001-Electronic Imaging*. International Society for Optics and Photonics 2001; 79-89.
- [13] Wang Z., Bovik AC, Sheikh HR, Simoncelli EP. Image quality assessment: from error visibility to structural similarity". *Image Processing, IEEE Transactions on* 2004; 13: 600-612.
- [14] Xiao F. DCT-based video quality evaluation. Final Project for EE392J, 769, 2000.
- [15] Wiegand T, Sullivan GJ, Bjontegaard G, Luthra A. Overview of the H. 264/AVC video coding standard. *Circuits and Systems for Video Technology, IEEE Transactions on* 2003; 7: 560-576.

Authors' profiles



Dr. Wissal Hassen, received a Ph.D. degree in 2012 from the National Engineering High School of Tunisia, a master degree in Automatic and Signal Processing in 2006 from the same High School and a Bachelor's degree in Computer Science from Faculty of Sciences of Tunis in 2004. She is a member of Signal, Image and Information Technologies laboratory in the National Engineering High School of Tunisia. Her research activities include Signal Processing, Image

Processing, Image and Video coding, Wavelet Transform and its applications.

Email: wissal_hassen@yahoo.fr



Pr. Hamid Amiri, received the Diploma of Electrotechnics, Information Technique in 1978 and the PHD degree in 1983 at the TU Braunschweig, Germany. He obtained the Doctorates Sciences in 1993. He was a Professor at the National Engineering High School of Tunisia, from 1987 to 2001. From 2001 to 2009 he was at the Riyadh College of Telecom and Information. Currently, he

is again at the National Engineering High School of Tunisia. He is now Director of Signal Processing Research Laboratory LSTS, and is in charge of Control and Signal Processing Master degree at this National Engineering High School. His research is focused on: image processing, speech processing, document processing and natural language processing.

* Signal, Image and Information Technologies laboratory in the National Engineering High School of Tunisia

How to cite this paper: Wissal Hassen, Hamid Amiri, "Wavelet-based Video Coding using Advanced Fractional Motion Estimation Technique", *IJIGSP*, vol.7, no.8, pp.66-75, 2015. DOI: 10.5815/ijigsp.2015.08.08