A NEW APPROACH TO VIDEO CODING BASED ON DISCRETE WAVELET CODING AND MOTION COMPENSATION

Wissal Hassen1, Jérôme Mbainaibye2, Hamid Amiri3 & Christian Olivier4
1, 2, 3 LSTS Laboratory, National Engineering School of Tunis (ENIT)
University El Manar, BP 37 Belvedere Tunis, 1002, Tunisia
4 XLIM Laboratory, SIC Department: Signal, Image and Communications,
University of Poitiers, BP 30179, 86962 Futuroscope Chasseneuil Cedex, France

ABSTRACT

Discrete Wavelet Transform (DWT) is a recent and powerful mathematical tool which has stimulated many developments in several scientific and technical fields, particularly in signal and image processing. Its multi-resolution properties are a major advantage. However, in video coding, motion management using DWT is a major challenge due to the DWT decomposition algorithms which do not preserve translation invariance. This means that, a translation of the original frame does not necessarily imply a translation of the corresponding wavelet coefficients. This is one of the reasons which justify, until now, the absence of the DWT-based video compression standard. However, in Discrete Cosine Transform based video compression standard such as MPEG.2, H.263 and H.264 the motion management in video sequences has reached the maturity. This paper introduces a new framework for color video coding that which can reduce the cost of storage and the bandwidth transmission of video file. The new method is a sub-band coding approach that employs Discrete Wavelet Transform (DWT) and is based on a separate sign coding (SSC) from wavelet coefficients amplitude. Furthermore, it uses motion compensation technique (MC) to overcome the problem caused by the lack of invariance translation DWT which leads to false motion vectors. The video codec that we propose in this work is then called Separate Sign Coding with Motion Compensation (SSCMC). An assessment of the proposed scheme was operated on a set of standard video sequences and the results obtained indicate that SSCMC outperforms H.264-AVC in terms of PSNR quality assessment and in terms of perceptual metric as structural similarity metrics (SSIM). Furthermore, the decoded video sequences visualized presents a competitive visual quality compared to H.264-AVC standard.

Keywords: DWT, motion compensation, arithmetic coding, EZW, video quality assessment

1. INTRODUCTION

The aim of video compression is to remove redundancies in both spatial and temporal domains. Normally, a video compression is a two-stage process: inter-frame coding techniques are used to reduce the temporary redundancies between successive frames of a video sequence and intra-frame coding techniques are used to reduce the spatial redundancies within the difference frame obtained by inter-frame coding.

The intra-frame coding is usually obtained by orthogonal transforms which exploit the intra-image correlation. These orthogonal transforms are for example the Discrete Cosine Transform (DCT) [1] used in image compression standards such as JPEG [2], in some video compression standard such as MPEG.1 [3] and MPEG.2 [4], or the Discrete Wavelet Transform (DWT) [5] such as JPEG2000 [6, 7, 8, 9], which provide an efficient framework of multi-resolution space-frequency representation with promising many applications because of its flexibility in representing non-stationary signals such as images and video sequences. Much research effort has been expended in the area of wavelet based image compression, with the results indicating that wavelet-based approaches outperform DCT-based techniques for still images [10, 11, 12, 13]. However, the motion management in video sequence is a major difficulty and constitutes a challenge in DWT-based video compression. Contrarily to block-based DCT algorithm, DWT algorithm decomposed image or video frame in its full size and the sub-sampling of the filtering output cannot ensure the translation invariance. In consequence, the management of the video sequence motion is a difficult task. In [14], the author has proposed a scheme for image compression based on Embedded Zerotree Wavelet algorithm (EZW) [15] named separated sign coding (SSC). This scheme provides competitive results in terms of objective and subjective qualities compared to JPEG.2000 standard and improves it. Through its performance, we will adopt it in the video coding field. In [16], the author uses the difference between the image in the coder and the reconstructed previous image in the decoder as technique for removing the temporal redundancies. The results obtained on some standard video sequences show a high quality of the decoded video at high bit rates; however, SSC codec is failed at low bit rates and the visualization of the decoded sequence gives a jerk movie. To overcome this drawback, we have integrated in SSC codec a motion compensation technique such as used in MPEG
standard and some well known video coding algorithms [8, 17, 18]. Indeed, until now the most efficient technique for motion estimation is the block matching algorithm (BMA) [19, 20, 21], which was adopted by various video coding standards such as ITU-T H. 261, H.263, MPEG.1, MPEG.2 and MPEG.4. But the best way to encode the resulting motion vector remains to be explored. A motion vector field of high precision is expensive in resources compared to the binary coding of the residual, so it is necessary to optimize the cost of transmission of this vector.

This paper describes a new baseline wavelet based video codec that we have called Separate Sign coding using Motion compensation (SSCMC). In a first part, we study the motion vector of each test sequence and we present a methodology for lossless arithmetic coding of motion vector. Then, we have developed tools for simultaneous encoding of vertical and horizontal components of motion vector color component. In a second part, test video sequences were compressed with SSCMC codec using the developed motion vector arithmetic coding. The remaining of the paper is organized as follows: in section two, we present a description of SSC; in the third section we present the proposed color video compression scheme. Section four presents the motion vector coding proposed method; section five presents the results in terms of objective and subjective qualities and discussions. Finally, section six presents conclusion and perspectives of this work.

2. THE SEPARATED SIGN CODING (SSC)

The proposed coding scheme uses two different coding techniques; a spatial coding system called SSC and a motion compensation technique for temporal redundancy. So, to explain our encoding system, we firstly present the details of the SSC codec and motion compensation techniques.

The idea of the SSC codec is to exploit Zero Tree Coding (ZTC) [8] principle similar to Embedded Zerotree Wavelet codec (EZW) [22]. To explain SSC we shall describe EZW algorithm which consists of three main blocks as shown in figure 1. In the first block, the discrete wavelet transform is applied on the input image. This block is characterized by lossless data because the DWT is reversible. The second block is concerned by the quantization of DWT coefficients; this block is characterized by loss of data. The third block is concerned by the entropy coding of the quantized data index and is characterized by lossless of data; this block produces the bit stream for transmission or storage.

![Figure 1. Basic Zerotree Coding principle](image)

EZW proceeds in two passes: Dominant Pass followed by Subordinate Pass. In Dominant Pass, the image is scanned and a symbol is outputted for every coefficient. If the absolute value of DWT coefficient is larger than the predefined threshold T, a POS (positive) symbol is encoded if the coefficient is positive; a NEG (negative) symbol in encoded if the coefficient is negative. A ZTR (Zerotree Root) symbol is used to encode an insignificant coefficient in a wavelet tree. An IZ (isolated zero) symbol is encoded if a coefficient is insignificant but having at least one significant child. The ZTR and IZ symbols are used to inform locations of significant coefficients as efficiently as possible. The symbol Z (Zero) is used for encoding the insignificant coefficients belonging to the three detail sub-bands of the first decomposition level.

The specificity of SSC is to separate the coding of wavelet coefficients amplitude and their signs. Thus, one symbol S is used for coding the significant coefficients that must be greater or equal to a predefined threshold; this symbol replaces POS and NEG used by the EZW. A significant map is progressively generated indicating the presence or the absence of the symbol S; the presence of the symbol S is indicated by the symbol ‘1’ and its absence by the symbol ‘0’.

Since the probability to find a negative coefficient is equal to zero in the approximation sub-band, a sign map is also progressively generated only for the details sub-bands; the presence of positive significant coefficient is indicated by the symbol ‘0’ and the presence of a negative significant coefficient is indicated by the symbol ‘1’ in the detail sub-bands. Figure 2 shows the concept of Zero Tree coding, illustrating the relationship between the roots and descendants of different wavelet coefficients trees (figure 2a and figure 2b) and the scanning order in the coding procedure (figure 2c). During the subordinate pass, one refines the quantization of the significant coefficients. The subordinate list contains the location information of all the coefficients that have been encoded (significant coefficient) in previous passes. Hence, the coefficients corresponding to the values of subordinate list belong to the interval (T, 2T) for the first pass. The subordinate pass gives an output ‘1’, if it is in the upper limit of the interval within (3T/2,2T) and ‘0’, if it is in the lower limit of the interval within (T, 3T/2).
3. THE SEPARATED SIGN CODING WITH MOTION COMPENSATION ENCODER (SSC-MC)

In [16], the Separated Sign Coding is used as a video coding system and the difference between the frame in the coder and the reconstructed previous frame in the decoder is used as technique for removing the temporal redundancies. In video coding, a group of pictures, or GOP structure, specifies the order in which intra- and inter-frames are arranged. The GOP is a group of successive pictures within an encoded video stream. The first frame of each GOP is encoded in intra-mode by SSC algorithm and subsequent frames in the video sequence are encoded by performing the difference between the reconstructed previous frame in the decoder and the current frame in the coder; this difference (residual frame) is then encoded by SSC algorithm. The results show that the system can provides best reconstruction quality as well objectively as subjectively for a higher given bit rate, but in the visualization of the video sequences reconstructed by this method, we see a jerky video. This jerkiness is the consequent of applying the non-redundant DWT on the residual frames. Indeed, since translation invariance is not guaranteed in the non-redundant DWT, this manifests itself in the reconstructed video sequence which is different to the original video sequence. In this work our goal is to overcome this drawback and improve the video sequence qualities by the elimination of the jerkiness.

Our solution in this new video compression scheme based on the SSC codec and motion compensation called SSCMC is to use the motion estimation and compensation technique. The motion compensation (MC) is a technique used in video compression standards such as MPEG.1 [23], MPEG.2 [24], MPEG.4 [25], H.263 [26], H.263+ [27] and H.264 [28]. It allows managing and predicting the motion in video sequences. The motion compensation algorithms operate the processing of each video frame through a set of Macro blocks (MBs).

A compressed file by MC contains, in the one hand the motion vector (MV) between the reference MB and the candidate MB (or actual MB) for the coding process, and in the other hand the error representing the difference between the reference MB and the actual MB. Three types of frame are considered: I frame (or intra-coded frame) which are the key frame; I frame is encoded without any reference to another frame. P frame (or mono-directional predictive frame) which contains only the displaced MBs; P frame is encoded with reference to a previous I or P frames. Finally B frame (or bi-directional predictive frame) which is predicted by two motion compensations: the one from the past I or P frames and the other from the future I or P frames. We recall that B frame must not be encoded with reference to another B frame because the decoder would be unable to decode a B frame without having received the reference frames. In consequence, I frame of the next GOP is sent before the B frames in the current GOP. Figure 3 provides a sample of the relationships between the various types of frames included in a GOP. As shown, I-frames are independent and provide input to support the other frames; this means that an error in the I-frames will have more distortions in the video sequence. The MB is the most common method of the motion estimation and is adopted by most part of video compression standards such as H.261, H.262, H.263, H.264, H.26L, MPEG.1, MPEG.2, MPEG.4 [29, 30, 31]. Estimation and motion compensation can allow a significant coding gain because only the motion vector, as shown in figure.4b, is transmitted instead of transmitting the entire MB (or the difference between two MBs in the case of a differential coding as shown in figure.4a).
Figure 4. Management of motion between the current frame \( I_{n+1} \) and the previous frame \( I_n \) in the video compression: (a) for the differential temporal coding where \( I_n \) and the difference are encoded, (b) for the coder with MC where \( I_n \) and the MV are encoded.

Figure 5 shows the architecture of the proposed video compression scheme where figure 5a shows the encoder and figure 5b shows the decoder. The video sequence is decomposed into GOP (generally a GOP contains 12 frames) and each GOP is passed by a two stages video compression process:

- The first stage is the intra-frame coding technique which uses the SSC algorithm to reduce the spatial redundancies within the difference frame obtained by the proposed inter-frame coding in the I frame. The regulation of the output bit rate is managed by the proposed encoder using an indicator bit rate calculated from the size of the resulting file and the video frequency. Thus, we can allocate more bits for the coding of I frame and less bits for the coding of difference frames;

- The second stage is the inter-frame coding technique or motion compensation which consists to reduce the temporary redundancies between successive frames of a video sequence; the encoder estimates the motion between the current frame (P or B frame) and a previously decoded frame using inverse SSC (SSC\(^{-1}\)) to give a motion vector and a predicted error (residual) where the prediction error is the difference between the current frame and the motion-predicted frame. Both the motion information and the predicted error are required to encode and transmit the bit stream to the decoder (bit stream1 and bit stream2).

The decoder uses also MC to generate the motion compensated frame from previously reconstructed frame which is available at the decoder; the decoded bit stream1 with the motion compensated frame gives the current frame. Bit stream2 is decoded by an inverse arithmetic entropy coding (Entropy coding\(^{-1}\)).

Figure 5. Architecture of the proposed video compression scheme: (a) encoder, (b) Decoder

4. ARITHMETIC MOTION VECTOR CODING

The motion vectors play an important role in the error propagation process between inter-frames. The slightest error in transmission or detection of these vectors affects the encoded video quality. Thus, for a video color, these vectors
are encoded by an entropic coding. The cost of motion vectors depends on various parameters such as image resolution, motion vectors precision (pixel, half pixel, quarter-pixel), the used MB size (which can vary from 4x4 to 16x16) etc. This cost can be very important, which is not a problem at high speed, but can cause problems at low speed. In this part of work we propose an approach for encoding motion vectors. During the motion estimation and compensation, the SSCMC codec decomposes P and B frames into MBs and finds a MV pointing to the best prediction MB in a reference frame or field. The goodness of MB prediction is in general evaluated by minimizing a cost function that may be the absolute error or the mean squared error. The cost function applied to the MB is called block matching algorithm (BMA) [29] and is the most used in video coding. In general, every possible prediction in a given range is evaluated; in this work we have used a full search called also exhaustive search [32]. It must be noticed that the capability to perform good motion estimation is a key point for the quality of a coder. 

MV's size depends on image size and MB. In fact, a color frame which has a resolution M x N, divided into MBs (the size of a MB is B x B) formed by the three color components to match the color space chosen. In particular, for the YCbCr space [33], each color component of motion vector has two coordinates x and y as indicated in equation 1 where $YMV$, $CbMV$ and $CrMV$ are respectively the MV of luminance (Y), the MV of blue chrominance ($Cb$) and the MV of red chrominance ($Cr$):

$$YMV_{frame} = (M \times N)/B^2 \times \begin{pmatrix} \Delta x_y \\ \Delta y_y \end{pmatrix}.$$  

$$CbMV_{frame} = (M \times N)/B^2 \times \begin{pmatrix} \Delta x_{C_b} \\ \Delta y_{C_b} \end{pmatrix}.$$  

$$CrMV_{frame} = (M \times N)/B^2 \times \begin{pmatrix} \Delta x_{C_r} \\ \Delta y_{C_r} \end{pmatrix}.$$  

Where $\Delta x$ and $\Delta y$ are respectively the horizontal displacement of the current MB. The size of global motion vector (GMV) for a sequence of length L which contains V motion is given by equations 2:

$$\text{Size(GMV)} = 2 \times 3 \times (M \times N)/B^2 \times V \times 8 \quad \text{with}$$

$$V = L - \left( \frac{L}{GOP} \right).$$

Each vector value is encoded using 8 bits (1 byte). For example the size of global motion vector necessary to encode 300 frames of Foreman video sequence with the frame resolution 352 x 288, B = 4 and GOP = 12, the size of MVs, applying equations 2, is 83635200 bits (83, 635 200 Mbits). We can notice that the size of MVs is very high and our goal is to reduce this size which consequently can increase the bit budget necessary for the coding process. A study of motion vector symbols and entropy coding is performed on a range of test sequence to achieve our goal. Table 1 gives the characteristics of the video test sequences used in this analysis.

Table 1 The used test sequences; fps means frames per second

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Name</th>
<th>Frames</th>
<th>Resolution</th>
<th>fps</th>
<th>Sequence</th>
<th>Name</th>
<th>Frames</th>
<th>Resolution</th>
<th>fps</th>
</tr>
</thead>
<tbody>
<tr>
<td>Akiyo</td>
<td></td>
<td>300</td>
<td>Qcif (176 x144)</td>
<td>30</td>
<td>Flowers</td>
<td></td>
<td>130</td>
<td>Cif (352x240)</td>
<td>30</td>
</tr>
<tr>
<td>Football</td>
<td></td>
<td>130</td>
<td>Qcif (176 x144)</td>
<td>15</td>
<td>Mobile</td>
<td></td>
<td>300</td>
<td>Cif (352x288)</td>
<td>30</td>
</tr>
<tr>
<td>Foreman</td>
<td></td>
<td>300</td>
<td>Cif (352x288)</td>
<td>30</td>
<td>Mother&amp; Daughter</td>
<td></td>
<td>140</td>
<td>Cif (352x240)</td>
<td>30</td>
</tr>
<tr>
<td>Bus</td>
<td></td>
<td>75</td>
<td>Qcif (176 x144)</td>
<td>30</td>
<td>Grand mother</td>
<td></td>
<td>140</td>
<td>Cif (352x240)</td>
<td>30</td>
</tr>
</tbody>
</table>

180
4.1. Study of symbols
We have studied the frequency distributions of each value in the MV for the video test sequences. We have calculated the occurrence of each value in the MV; the ratio of the occurrence of each value by the sum of all occurrences gives the frequency distribution (or histogram) of each value. Histograms of distribution of all sequence are concentrated around the value "0"; this may be explained by the fact that in general there is a little changing between two successive frames. The consequence is that "0" is the dominant symbol in the motion vectors. We can also find that the motion vectors are formed by positive and negative values belonging to the interval \([-7; 7]\) and we can observe the symmetry of the histograms. Some differences may be observed between luminance (Y) histogram and chrominance histograms (C_r and C_b); these differences may be explained by equation 3 which transforms the RGB space YCbCr space where all coefficients are positive in the calculation of luminance pixels. Figure 6 and figure 7 present the histograms for foreman and Football sequences.

\[
\begin{align*}
Y &= 0.299R + 0.587G + 0.114B \\
C_b &= -0.14713R - 0.28886G + 0.436B \\
C_r &= 0.615R - 0.51498G - 0.10001B
\end{align*}
\]  

(3)

Based on the independence of color component histogram, we propose to encode the global motion vector instead of encoding individual motion vector. Then, in our approach, the luminance motion vector YMV, the blue chrominance motion vector C_bMV and the red chrominance motion vector C_rMV form a single global motion vector noted GMV by using a single arithmetic code. The GMV vector is the concatenation of the three motion vectors given by equation 2:

\[
GMV = YMV \cdot C_bMV \cdot C_rMV
\]  

(4)

Figure 8 presents the probability histogram of each symbol in the GMV for each test sequence; we note that the symbol "0" has a high probability. All symbols will be encoded in our approach by the arithmetic coding.

(a) (b) (c)

Figure 6. Normalized histograms of motion vectors for Foreman sequence: a) Luminance, b) Blue chrominance, c) Red chrominance

(a) (b) (c)

Figure 7. Normalized histograms of motion vectors for Football sequence: a) Luminance, b) Blue chrominance, c) Red chrominance
Figure 8. Normalized histograms of global motion vectors GMV
4.2. Arithmetic coding

Arithmetic coding creates one word-code to be associated with each GMV sequence, contrarily to Huffman coding which assigns variable lengths word-codes for each symbol of the GMV. The code associated with a source is a real number in the interval [0, 1]. We recall in this paragraph the arithmetic coding algorithm [34] and its use in our proposed scheme. Let us consider a source \( S_k \) containing \( N \) symbols with probabilities \( P[S_k] = P_k \). To encode a sequence \( S_M = (S_{a1}, S_{a2}, ..., S_{aM}) \) of \( M \) symbols, we use the following algorithm:

1) The first interval is initialized with two bounds: the lower bound \( L_c = 0 \) and the upper bound \( H_c = 1 \). The size of this interval is thus defined by: \( size = H_c - L_c \).

2) This interval is partitioned into \( N \) sub-intervals \( [L_{S_k}, H_{S_k}] \) according to the probabilities of each symbol \( S_k \) of the source; the \( N \) sub-intervals are the initial partitions. By considering the Length \( (H_{S_k} - L_{S_k}) \) of these sub-intervals which is given by \( L_{S_k} = H_{S_k} = P_k \), as shown in equation (5):

\[
L_{S_k} = L_c + size \sum_{i=1}^{k-1} p_i \quad \text{and} \quad H_{S_k} = L_c + size \sum_{i=1}^{k} p_i.
\]

3) The sub-interval corresponding to the next symbol \( S_\alpha_k \) in the sequence is defined in. Then, the initial interval \( [L_c, H_c] \) is redefined as given by equation (6):

\[
\begin{align*}
L_c &= L_c + size \times L_{S_k} \\
H_c &= L_c + size \times H_{S_k}
\end{align*}
\]

4) This interval is divided again by the same method used in step 2.

5) Steps 2, 3 and 4 are repeated until the word-code representing the complete sequence of symbols sources obtained.

This algorithm is applied on all the video test sequences in table 1. The obtained results are presented in table 3. The bit rate en (Kbps) is also given in table 3. This bit rate will be used in the SSCMC coding and will be added to the residual bit rate coding. Motion vector is coded as a GMV with a single code-word for different MV corporants as orthogonal (y), horizontal (y), luminance (Y), red chrominance (C_r) and blue chrominance (C_b). In the decoder, GMV is decoded and each corporant is determined to be used with the decoded residual to determine the compensated frame.

5. EVALUATION OF SSCMC AND RESULTS

The SSCMC developed is applied on the video test sequences. All frames in each video test sequences are transformed from the RGB space (Red Green and Blue) to the YC_bC_r space. The Daubechies bi-orthogonal wavelets (9, 7) filter bank [35] is used to decompose each frame. The motion compensation is operated with the BMA, the size of MB is 4x4 and the method of block matching is Full search. Decoded frames are compared to H.264-AVC. To evaluate the objective and subjective qualities of the proposed scheme, we used three different metrics. The peak signal to noise ratio or PSNR given by equation (7):

\[
PSNR = 10 \log_{10} \left( \frac{255^2}{MSE} \right)
\]

where MSE is the mean square error calculated by equation (8):

\[
MSE = \frac{1}{HL} \sum_{i=1}^{H} \sum_{j=1}^{L} (X(i, j) - \hat{X}(i, j))^2
\]

with \( H \) and \( L \) are respectively the height and width of each frame of the video sequence, \( X \) and \( \hat{X} \) are respectively the original frame and the reconstructed frame. Note that equations (7) and (8) are applied separately on the components R (red component), G (green component) and B (blue component) of the RGB space. The global PSNR of the components R, G and B is calculated as the average of the PSNR of the 3 components, given by equation 9:

\[
PSNR = \frac{PSNRR + PSNRG + PSNRB}{3}
\]

The second metric SSIM is a structural approach presented by Wang and al. [36]. This method measures the similarity between two images. The similarity computation is operated through a window \( x \) of the luminance component \( Y \) of the original image and the window corresponding to the degraded version \( y \) (the \( Y \) component) by combining three measures: a similarity measure of brightness, a similarity measure of contrast and a similarity measure of structure given by equation 10:

\[
SSIM(x,y) = \frac{(2\mu_x\mu_y+c_1)(2\sigma_{xy}+c_2)}{\mu^2_x+\mu^2_y+c_1(\sigma^2_x+\sigma^2_y+c_2)}
\]
Table 3. Arithmetic coding of MVG vectors for each sequence

<table>
<thead>
<tr>
<th>Symbols</th>
<th>Akiyo</th>
<th>BUS</th>
<th>Foreman</th>
<th>Grand-mother</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Probability</td>
<td>Lower bound</td>
<td>Upper bound</td>
<td>Probability</td>
</tr>
<tr>
<td>-7</td>
<td>0.0015</td>
<td>0</td>
<td>0.0015</td>
<td>0.0258</td>
</tr>
<tr>
<td>-6</td>
<td>0.0017</td>
<td>0.0015</td>
<td>0.003</td>
<td>0.0171</td>
</tr>
<tr>
<td>-5</td>
<td>0.002</td>
<td>0.003</td>
<td>0.005</td>
<td>0.0226</td>
</tr>
<tr>
<td>-4</td>
<td>0.0024</td>
<td>0.005</td>
<td>0.008</td>
<td>0.0294</td>
</tr>
<tr>
<td>-3</td>
<td>0.0035</td>
<td>0.008</td>
<td>0.011</td>
<td>0.0416</td>
</tr>
<tr>
<td>-2</td>
<td>0.0058</td>
<td>0.011</td>
<td>0.017</td>
<td>0.0591</td>
</tr>
<tr>
<td>-1</td>
<td>0.018</td>
<td>0.017</td>
<td>0.035</td>
<td>0.1132</td>
</tr>
<tr>
<td>0</td>
<td>0.933</td>
<td>0.035</td>
<td>0.968</td>
<td>0.3803</td>
</tr>
<tr>
<td>1</td>
<td>0.177</td>
<td>0.968</td>
<td>0.986</td>
<td>0.1106</td>
</tr>
<tr>
<td>2</td>
<td>0.052</td>
<td>0.986</td>
<td>0.991</td>
<td>0.0588</td>
</tr>
<tr>
<td>3</td>
<td>0.003</td>
<td>0.991</td>
<td>0.994</td>
<td>0.0404</td>
</tr>
<tr>
<td>4</td>
<td>0.0022</td>
<td>0.994</td>
<td>0.996</td>
<td>0.0358</td>
</tr>
<tr>
<td>5</td>
<td>0.0015</td>
<td>0.996</td>
<td>0.998</td>
<td>0.0209</td>
</tr>
<tr>
<td>6</td>
<td>0.0013</td>
<td>0.998</td>
<td>0.999</td>
<td>0.0175</td>
</tr>
<tr>
<td>7</td>
<td>0.0012</td>
<td>0.999</td>
<td>1</td>
<td>0.026</td>
</tr>
</tbody>
</table>

Arithmetic code: 0.000387
Bit rate (in Kbps): 0.111788618

Arithmetic code: 0.000198
Bit rate (in Kbps): 0.41666667

Arithmetic code: 0.0000348
Bit rate (in Kbps): 0.11178862

Arithmetic code: 0.0000218
Bit rate (in Kbps): 0.20833333

where: \( \mu_x \) and \( \mu_y \) are respectively the mean of x and y (the luminance indicator); \( \sigma_x^2 \) and \( \sigma_y^2 \) are respectively the variance of x and y (the contrast indicator); \( \text{cov}_{xy} \) is the covariance between x and y; \( c_1=(K_1 L)^2 \) and \( c_2=(K_2 L)^2 \) are two variables which guarantee the stabilization of the division if the denominator value is very low. L is the dynamic of the pixel values (255 for the images encoded with 8 bits) and \( K_1=0 \) and \( K_2=0 \) by default. Thus, more the value of SSIM is close to one better is the coder.

Figures 8 shows the results in terms of objective quality metrics (PSNR), figures 10 show the results in terms of subjective quality metrics (SSIM) respectively for Foreman encoded at 156 kbps, Flower encoded at 327 kbps and Mobile encoded at 236 kbps. The size of GOP used for these three video sequences is 12 frames. These results are compared to H.264-AVC. It can be seen that proposed video compression scheme outperforms H.264 standard especially for Flowers and Mobile video test sequences. Table 4 summarizes the average values of PSNR and SSIM obtained by SSCMC compared with those obtained by H.264-AVC codec for each sequence. We can see that SSCMC gives the best result for the three sequences.
In figure 9, SSC-MC curves show slight peaks due to I frames encoded in intra mode with a higher bit rate than the bit rate used to encode the P and B frames because I frames is considered as the reference to predict the rest of the 11 frames of a GOP. Those peaks appear in the beginning of each GOP and appear also for MPEG.2, H.263 + and MPEG.4 encoders. We think that those peaks can be reduced if we adopt two-pass coding technique by making a preliminary analysis of the motion in the video sequence and using a variable size of MBs depending on the complexity of the scene (much or little motion), and a variable GOP structure, such as used in H.264-AVC. However, these peaks are not very clear with other evaluation metrics based on perceptual techniques (SSIM) that are designed to improve the PSNR evaluation method since PSNR is proved to be inconsistent with the human visual system [36].

Figures 12, 13 and 14 present the visual quality of some decoded frames using SSCMC and H.264-AVC for Foreman, Flowers and Mobile video sequences respectively.

(a) Foreman coded at 156 kbps
(b) Mobile coded at 236 kbps
(c) Flowers coded at 327 kbps

Figure 9 Objective assessment with PSNR metric: a) Foreman encoded at 156 kbps, b) Mobile encoded at 236 kbps, c) Flowers encoded at 327 kbps.
Figure 10. Subjective assessment with SSIM metric: a) Foreman encoded at 156 kbps, b) Mobile encoded at 236 kbps, c) Flowers encoded at 327 kbps.

Table 4. Average values of PSNR and SSIM of SSC-MC and H.264-AVC codecs for Foreman, Mobile and Flowers video test sequences encoded at 156 kbps, 236 kbps and 327 kbps respectively.

<table>
<thead>
<tr>
<th></th>
<th>PSNR</th>
<th>SSIM</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>SSC-MC</td>
<td>H.264-AVC</td>
</tr>
<tr>
<td>Foreman (156kbps)</td>
<td>37,686</td>
<td>36,819</td>
</tr>
<tr>
<td>Mobile (236kbps)</td>
<td>28,012</td>
<td>25,147</td>
</tr>
<tr>
<td>Flowers (327kbps)</td>
<td>24,336</td>
<td>21,472</td>
</tr>
</tbody>
</table>
In this paper, we present a new color video compression scheme based on separate sign coding of wavelet coefficients and motion compensation technology. The architecture of the proposed codec is presented and the evaluation is operated on three standard video test sequences. In this work we have detailed our encoder, a study of

6. CONCLUSION AND PERSPECTIVES

Figure 11. Comparative visual qualities of decoded Foreman with SSC-MC (left) and H.264-AVC (right) at 156kbps for the 158th frame in the sequence.

Figure 12. Comparative visual qualities of decoded Flowers with SSC-MC (left) and H.264-AVC (right) at 327kbps for the 64th frame in the sequence.

Figure 13. Comparative visual qualities of decoded Mobile with SSC-MC (left) and H.264-AVC (right) at 236 kbps for the 148th frame in the sequence.
accuracy of motion vector and have also proposed a detailed coding method of motion vector with the three color components by single entropy coding using arithmetic coding.

The coding performance of the new method is compared to H.264-AVC standard in terms of objective metric (PSNR) and subjective metric (SSIM). It is shown that significant PSNR gains can be achieved. In subjective quality using SSIM and VQM, it is shown that the proposed scheme gives a high visual quality competitive with H.264-AVC standard. The comparison is also done in terms of visual quality at low bit rates and it is shown that the decoded frames are visually very similar to those obtained by H.264-AVC.

In the perspectives of this work, we will investigate the influence of noise in transmission channel. Some transmission channel models will be used to evaluate the performance of the proposed scheme in terms of lossy data or transmission errors. We think that it is possible to evaluate the performance of the proposed scheme in terms of execution time compared to H.264-AVC standard because execution time is an important characteristic of the real time coding and transmission systems.

7. REFERENCES


[23]. Standard MPEG-1 : ISO/IEC 11172-2, Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s. 1993
[26]. H.263 : Video coding for low bit rate communication. First recommendation H.263 to the UITT. March 1996.