Reducing Decoding Complexity by Improving Motion Field Using Bicubic and Lanczos Interpolation Techniques in Wyner-Ziv Video Coding

I Made O. Widyantara[1], Wirawan[2] and Gamantyo Hendrantoro[2]
[1] Department of Electrical Engineering, University of Udayana
Denpasar 80361 - Indonesia
[2] Department of Electrical Engineering, Institut Teknologi Sepuluh Nopember (ITS)
Surabaya 60113 - Indonesia
[e-mail: oka.widyantara@unud.ac.id, wirawan@its.ac.id, gamantyo@ee.its.ac.id]
*Corresponding author: I Made O. Widyantara

Received April 5; revised July 16, 2012; accepted August 16, 2012;
published September 26, 2012

Abstract

This paper describes interpolation method of motion field in the Wyner-Ziv video coding (WZVC) based on Expectation-Maximization (EM) algorithm. In the EM algorithm, the estimated motion field distribution is calculated on a block-by-block basis. Each pixel in the block shares similar probability distribution, producing an undesired blocking artefact on the pixel-based motion field. The proposed interpolation techniques are Bicubic and Lanczos which successively use 16 and 32 neighborhood probability distributions of block-based motion field for one pixel in *k*-by-*k* block on pixel-based motion field. EM-based WZVC codec updates the estimated probability distribution on block-based motion field, and interpolates it to pixel resolution. This is required to generate higher-quality soft side information (SI) such that the decoding algorithm is able to make syndrome estimation more quickly. Our experiments showed that the proposed interpolation methods have the capability to reduce EM-based WZVC decoding complexity with small increment of bit rate.

Keywords:, Wyner-Ziv video coding, EM algorithm; Bicubic interpolation; Lanczos interpolation; motion field interpolation

A preliminary version of this paper appeared in IEEE ICICI-BME 2011, November 8-9, Bandung, Indonesia. This version proposes implementation of Bicubic dan Lanczos interpolation in EM-based WZVC and a concrete analysis on GOP size 2, 4 and 8. This work is supported by 2011 Doctoral Dissertation Research Grant and BPPS Scholarship from Indonesian Ministry of National Education. We express our thanks to David Varodayan of Hewlett-Packard Laboratories and David Chen of Information System Lab., Department of Electrical Engineering at Stanford University for sharing the code of WZVC codec.
http://dx.doi.org/10.3837/tiis.2012.09.022

1. Introduction

The traditional method for video compression exploits the correlation between and within video frames at encoder side. This is a computationally intensive process, involving the motion estimation and motion compensation. At the same time, the decoder is far less complex than the encoder. However, with the proliferation of the mobile devices over the last decades, many applications need the dual system, i.e., the low complexity encoders in power-constraint environment and the possibility to move the high complexity decoders in an environment with rich resources.

An emerging approach to the low complexity encoding is distributed video coding (DVC) with side information (SI), where exploiting the unknown statistics source, totally or partially, is performed on the decoder side. The DVC is based on two information theories i.e. Slepian-Wolf (SW) [1] and Wyner-Ziv (WZ) [2]. SW theorem states that optimal rate acquired for lossless encoding and decoding of two joint correlated sources X and Y theoretically can be achieved by performing encoding separately and joint decoding. WZ theorem then generalized this theorem for lossy coding where same rate can be achieved by lossy compression of X when source Y, denoted as SI, is available on decoder side. This lossy coding solution is known as Wyner-Ziv video coding (WZVC).

Currently, WZVC is considered to be a new paradigm for wireless video, such as wireless video sensors, mobile camera phones, and networked camcorders. In these promising applications, in addition to saving bit rate, real time video application will be one important category, in which the too large computation load in decoding will be a high burden. Brites et al. [3] have stated that reducing the decoder complexity and improving the RD performance can be done by optimizing channel coding techniques and improving the quality of the SI. For these applications, design of WZVC codec must consider to balance complexity between the encoder and the decoder. Many WZVC designs adopt the Stanford architecture [4], for example the DISCOVER codec [5]. This codec uses advanced motion-compensated interpolation (MCI) technique to create the SI, based on adjacent decoded frames, one in the past and another in the future. The performance of the SI creation process is limited by the quality of the past and future reference frames as well as the distance and motion behavior between them. This limitation has led to poor rate-distortion (RD) performance of WZVC codec for higher motion sequences and longer group of pictures (GOP) sizes.

To improve the RD performance of WZVC codec, several methods have been proposed such as SI refinement [6][7] and unsupervised motion vector learning [8]. SI refinement method improves decoding quality by updating and improving SI during decoding. The SI generation and refinement mainly depend on the motion-compensated interpolation/extrapolation that is performed either in the temporal domain [6], or the spatial domain [7]. For quick or sudden motion in the video sequence, or for reference frames with large noise, the SI accuracy drops quickly, which inevitably leads to the loss of RD efficiency. In [8], to decode the encoded WZ frame, the decoder has performed an unsupervised motion learning of forward motion vectors during the decoding. The decoder's goal is to recover the reconstruction by using the maximum likelihood estimate of the WZ frame given both a decoded frame previously and the received increment of cumulative syndrome. The motion field that corresponds to the WZ frame and the decoded frame is treated as hidden variables. This condition makes the maximum likelihood must be estimated with incomplete data. To find maximum likelihood estimate of the WZ frame with the incomplete data, an iterative algorithm called the EM algorithm is applied. Fig. 1 shows the EM algorithm on the WZVC [8], which we then write as an EM-based WZVC codec. The E-step updates the motion field distribution with reference to the soft value at the low-density parity-check (LDPC) decoder output. The M-step updates the soft SI with reference to the motion field distribution that is fed to the LDPC decoder. This method is able to produce better RD performance for longer GOP size, which is typically one of the weakest feature in the previous WZVC [5].

Although the EM iteration produces a better RD performance, it takes longer time to converge, leading to high decoding complexity. In addition, the convergence of EM iteration is limited by the accuracy of the motion field. As shown in **Fig. 1**, the motion estimator supplies the LDPC decoder with motion field-compensated (soft) SI (ψ). If soft SI correctly reflects the true state of source (X), then the soft SI will help the EM algorithm to converge towards the correct solution. However, the system makes a simplification in motion estimation where motion field estimation M (or probability distribution) is carried out based on block, instead of frame to simplify computation and to avoid expensive operations from a large number of possible value of M. Next, this motion field is interpolated into its original resolution. This simplification raises a problem that is when one frame is devided into k-by-k block, all pixel within that block will share the same probability distribution of motion field, $P_{app}\{M_{u,v}\}$. This leads to a blocking artefact on the profile of motion field and makes the low accuracy of soft SI.

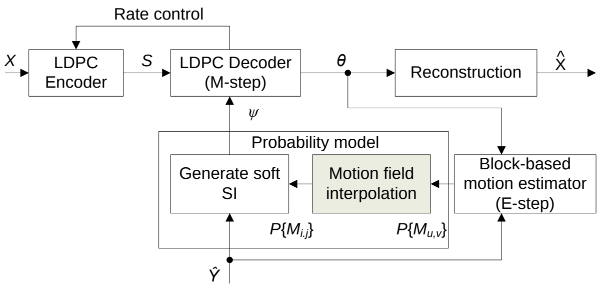


Fig. 1. Motion field interpolation in EM-based unsupervised forward motion vector learning

Several studies have been conducted focusing on improving the interpolation technique to estimate motion field on pixel-by-pixel resolution and reducing decoding complexity on EM-based WZVC codec. A motion field with pixel-by-pixel resolution is created by interpolating a motion field with only block-by-block resolution. Bilinear interpolation technique was proposed by [9] to replace nearest-neighbor (NN) interpolation technique and resulted in significant bit rate saving for WZ codec stereo image. Widyantara, et al. [10] have compared the performance of RD and decoding complexity of EM-based WZVC codec of the two interpolation techniques. The later analysis showed that compared to the NN interpolation technique, Bilinear interpolation technique is able to provide a significant bit rate saving, but on the other hand it also increased the complexity of the decoder up to 15% on low scene video content, and up to 12% in complex scene video content. Furthermore, with Bicubic interpolation techniques, decoding complexity of the existing EM-based WZVC codec that implement Bilinear interpolation technique can be reduced up to 8.29%, while the RD performance can be maintained almost the same [11]. However, performance analysis was still restricted to the size of the group of picture (GOP) 2. Besides improvement of interpolation procedure, a motion search optimization method as proposed by [12], is also applicable in order to reduce decoding complexity of EM-based WZVC.

In this paper we propose Lanczos and Bicubic interpolation techniques to reduce decoding complexity of EM-based WZVC and extend the analysis to the GOP size larger than 2. We chose those interpolation techniques based on the promising results of the Lanczos and Bicubic interpolation technique. The Lanczos interpolation has been implemented for optimizing decomposition based on lossless compression of biomedical images [13] and segmentation of ultrasound breast phantom data [14]. The Bicubic interpolation also has been applied for image zooming [15]. Throughout our knowledge, both interpolation techniques have not been implemented as method of interpolation of motion field in EM-based WZVC. In this research we will compare both interpolation techniques with Bilinear interpolation technique in existing EM-based WZVC codec to obtain the most efficient motion field interpolation method in reducing the decoding complexity.

The rest of this paper is organized as follows. Section 2 reviews about codec with unsupervised forward motion vector learning based on EM algorithm. Next, Section 3 describes Bicubic and Lanczos interpolation techniques for interpolation of motion field from block-based into pixel-based. Evaluation and implementation of this method in the EM-based WZVC codec domain transform will be explained in Section 4. Section 5 shows conclusions of this paper.

2. EM-based Unsupervised Motion Vector Learning

Realistic model of the statistical dependence between source and side information often involve hidden variable. For example, the motion vectors that relate consecutive frames of video are unknown a priori at the decoder of WZVC. EM-based unsupervised motion learning in Fig.1 provides a practical way for the decoder to simultaneously learn the motion field and recover source. The decoder recover the reconstruction by using the maximum likelihood estimate of the source (X) given both the side information ($\hat{Y}$) and the received increments of cumulative syndrome (S), treating the motion field (M) as hidden variables. This can be performed by an iterative algorithm called the EM algorithm [16]. The expectation step fixes the source estimate (θ) and estimates the motion field. The maximization step fixes the motion field estimated and estimates the source using belief propagation iteration of LDPC decoding. Based on [8], EM adaptation for unsupervised motion vector learning scheme in WZVC could be explained as follow.

2.1 Model

Let X and $\hat{Y}$ be consecutive luminance frames of video, with X related to $\hat{Y}$ through a forward motion field M. The residual of X with respect to motion-compensated $\hat{Y}$ is treated as independent Laplacian noise Z. The decoder's a posteriori probability distribution of source X based on parameter θ is modeled as

$$P_{app}\{X\} \equiv P\{X;\theta\} = \prod_{i,j} \theta\left(i,j,X(i,j)\right) \quad (1)$$

where θ(i,j,) = Papp{X(i,j) = } defines a soft estimate of X(i,j) over luminance values $\omega \in \{0,..,2d-1\}$.

2.2 Problem

The decoder aims to calculate the a posteriori probability distribution of the motion field M,

$$P_{app}\{M\} \equiv P\{M \mid \hat{Y}, S; \theta\} \\ \propto P\{M\} P\{\hat{Y}, S \mid M; \theta\}, \tag{2}$$

with the second step by Bayes' Law. The form of this expression suggests an iterative EM solution. The E-step updates the motion field distribution with reference to the source model parameters, while the M-step updates the source model parameters with reference to the motion field distribution. Note that P{M|Ŷ,S;θ} is the probability of observing motion M given that it relates X (as parameterized by θ) to Ŷ, and also given S. The elaboration of P{Ŷ,S|M;θ} is discussed below.

2.3 E-step Algorithm

The E-step updates the estimated distribution on M and before renormalization is written as:

$$P_{app}^{(t)}\{M\} := P_{app}^{(t-1)}\{M\} P\{\hat{Y}, S \mid M; \theta^{(t-1)}\} \tag{3}$$

The large possible values of M make the operation of (3) expensive. Two ways of simplification carried out are: ignore the syndrome S because it is exploited in the M-Step (LDPC decoding), and estimate the motion field M with block-by-block motion vectors Mu, v. When the block is set, every block of θ(t-1) is compared to the collocated block of Ŷ as well as all those in a fixed motion search range around it. For a block θ u,v(t-1) with top left pixel located at (u,v), the distribution on the shift Mu,v is updated as below and normalized:

$$P_{app}^{(t)}\{M_{u,v}\} := P_{app}^{(t-1)}\{M_{u,v}\} P\{\hat{Y}_{(u,v)+M_{u,v}} \mid M_{u,v}; \theta_{u,v}^{(t-1)}\} \tag{4}$$

where $\hat{Y}_{(u,v)+Mu,v}$ is the k-by-k block of $\hat{Y}$ with top left pixel at $((u,v)+M_{u,v})$. Note that $P\{\hat{Y}_{(u,v)+Mu,v}|M_{u,v};\theta_{u,v}^{(t-1)}\}$ is the probability of observing $\hat{Y}_{(u,v)+Mu,v}$ given that it was generated through vector $M_{u,v}$ from $X_{u,v}$ as parameterized by $\theta_{u,v}^{(t-1)}$. This procedure, shown in the left of **Fig. 2**, occurs in the block-based motion estimator.

Probability model iteratively updates the soft SI (ψ) by blending information from pixels in $\hat{Y}$ according to pixel-based motion field distribution. This procedure consists of 2 steps, i.e. motion field interpolation to magnifying the resolution of motion field from block-based into pixel-based and soft SI (ψ) generation. The existing EM-based WZVC codec uses Bilinear interpolation technique to smoothen motion field by using four $P_{app}\{M_{u,v}\}$ for each pixel within k-by-k block. Generally, in soft SI generation, the probability that the blending of SI has value ω in pixel (i, j) is:

$$\psi^{(t)}(i,j,\omega) = \sum_m P_{app}^{(t)}\{M_{i,j} = m\} P\{X(i,j) = \omega \mid M_{i,j} = m, \hat{Y}\} \\ = \sum_m P_{app}^{(t)}\{M_{i,j} = m\} p_Z\{\omega - \hat{Y}_m(i,j)\} \tag{5}$$

where $p_Z(z)$ is the probability mass function of the independent additive noise Z, and $\hat{Y}_m$ is the previously reconstructed frame compensated through motion configuration m.

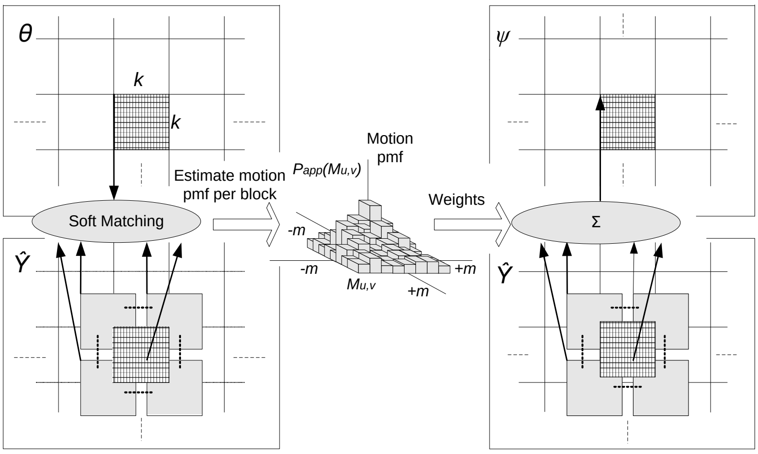


Fig. 2. E-step block based motion estimator (left) and probability model (right)[8]

2.4 M-Step Algorithm

The M-step updates the soft estimate θ, by maximizing the likelihood of Ŷ and syndrome S,

$$\theta^{(t)} := \arg\max_{\Theta} P(\hat{Y}, S, \Theta)$$
$$= \arg\max_{\Theta} \sum_{m} P_{app}^{(t)}\{M = m\} P\left\{\hat{Y}, S \mid M = m, \Theta\right\} \tag{6}$$

where the summation is overall configuration m of the motion field. Due to the true maximization is intractable, an approach that is performed using soft SI values in (5) followed by a joint iteration of LDPC decoding Bitplane [19] to yield θ(t),

$$\theta^{(t)}(i,j,\omega) := \psi^{(t)}(i,j,\omega) \prod_{g=1}^{d} \left(\alpha_g^{(t)}\right)^{1_{[w_g=1]}} \left(1-\alpha_g^{(t)}\right)^{1_{[w_g=0]}} \tag{7}$$

where ω_g denotes the gth bit in Gray mapping of luminance value ω and $1_{[.]}$ denotes the indicator function. M-step also generates a hard estimate of $\hat{X}$ by taking one most probable value for each pixel according to θ.

$$\hat{X}(i,j) = \arg\max_{\omega} \theta(i,j,\omega) \tag{8}$$

By iterating through the M-step and the E-step, the LDPC decoder requests more syndrome bits if the estimates is not convergent. The algorithm terminates when the hard estimate of $\hat{X}$ yields syndrome which is identical to *S*.

Although the numerical convergence of EM algorithm is slow and only guaranteed to converge to a local maximum in the incomplete-data likelihood function, several studies have shown that its implementation to estimate the distribution of disparity in the Wyner-Ziv coding for stereo image compression [9][17][18] and to estimate the distribution of motion field in WZVC [8][10][11][12], results in a good RD performance. The key lies in the proper use of the initial distribution for motion vector/disparity, as a starting point in the EM iterations. Initial distribution of $M_{u,v}$ in [8] is chosen experimentally by setting the largest probability on the colocated block in the reference frame Y (please see (18) in section 4).

As shown in Fig.1, the EM iteration start from the M-step, in which the soft SI (ψ) generated from the blending information from pixels in $\hat{Y}$ according to the initial distribution of $M_{u,v}$ to improve the soft estimate θ. Furthermore, the initial distribution of motion field is updated by the decoder iteratively. The initial distribution and the motion estimator are performed in a blockwise ($M_{u,v}$). All pixels in a block shares the same distribution, that whenever the initial distribution is incorrect the decoder can not converge correctly over time as analysed in [9]. It means that the better of the initial distribution, the shorter the decoding time of EM iteration. In this paper we propose two methods of improvement of the initial distribution of the motion field using Bicubic and Lanczos interpolation. Section 3 describes in detail the proposed method.

3. The Proposed Method

In this section, we describe some methods of motion field interpolation to improve the probability distribution of the motion field into the pixel precision on the EM-based WZVC codec.

On the application of digital image processing, some of linear filtering-based interpolation techniques, such as NN, Bilinear, Bicubic and Lanczos have been widely applied to magnify and minimize image resolution. In this interpolation technique, new pixel values are obtained by assuming that values from image function in new pixel location can be computed as linear combination of the original pixel values closest to the new positions. These interpolation techniques are very efficient in computation, especially for Bicubic interpolation (cubic function uses 16 nearest neighbors) that gives better image visualization [20].

To improve the $\psi^{(t)}$ accuracy, we propose to use Bicubic and Lanczos interpolation techniques as magnifying resolution methods. As shown in **Fig. 1**, after block-based motion estimation is done, motion field interpolation is performed by interpolating motion field from block-based into into pixel-based. **Fig. 3** illustrates the interpolation method.

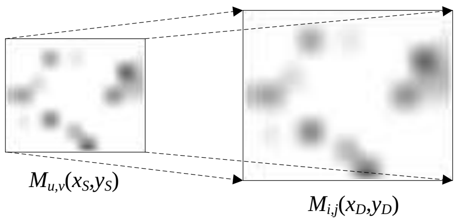


Fig. 3. Motion field interpolation from block-based, $M_{u,v}(x_S,y_S)$ into pixel based, $M_{i,j}(x_D,y_D)$

We adapt Bicubic and Lanczos interpolation techniques from Intel IPP version 6 of documentation library [21]. We integrate this library function into source code, which we explain more detailed in the followings.

3.1 Bicubic Interpolation

To have a better understanding, Let's assume that $M_{u,v}(x_S,y_S)$ denotes *block*-by-*block* motion field and $M_{i,j}(x_D,y_D)$ denotes *pixel*-by-*pixel* motion field. Bicubic interpolation algorithm uses

16 (sixteen) probability distributions $P_{app}\{M_{u,v}(x_S,y_S)\}$ closest to (x_S,y_S) position in block-based motion field, i.e.:

$$\begin{array}{llll} x_{s0} = \text{int}(x_s) - 1; & x_{s1} = x_{s0} + 1; & x_{s2} = x_{s0} + 2; & x_{s3} = x_{s0} + 3; \\ y_{s0} = \text{int}(y_s) - 1; & y_{s1} = y_{s0} + 1; & y_{s2} = y_{s0} + 2 & y_{s3} = y_{s0} + 3; \end{array} \tag{9}$$

where,

- (x_D,y_D) is a coordinate in pixel-based motion field (integer value)
- (x_S,y_S) is a coordinate computed from a position in block-based motion field mapped exactly to (x_D,y_D)
- $P_{app}\{M_{u,v}(x,y)\}$ is a probability distribution on block-based motion field.
- $P_{app}\{M_{i,j}(x,y)\}$ is a probability distribution on pixel-based motion field

Firstly, for each y_{sk}, the algorithm defines 4 (four) cubic polynomials $F_0(x)$, $F_1(x)$, $F_2(x)$, and $F_3(x)$ using below equations:

$$F_k(x) = a_k x^3 + b_k x^2 + c_k x + d_k\,; \quad 0 \le k \le 3 \tag{10}$$

Such that

$$\begin{array}{ll} F_k(x_{s0}) = P_{app}\{M_{u,v}(x_{s0}, y_{sk})\}, & F_k(x_{s1}) = P_{app}\{M_{u,v}(x_{s1}, y_{sk})\}, \\ F_k(x_{s2}) = P_{app}\{M_{u,v}(x_{s2}, y_{sk})\}, & F_k(x_{s3}) = P_{app}\{M_{u,v}(x_{s3}, y_{sk})\} \end{array} \tag{11}$$

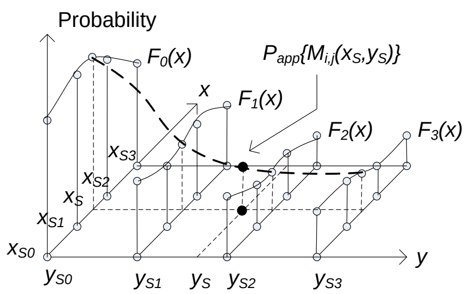


Fig. 4. Bicubic interpolation technique scheme

In **Fig. 4**, these polynomials are shown by solid curves. Next, the algorithm defines a cubic polynomial $F_y(y)$, such that:

$$\begin{array}{l} F_y(y_{s0}) = F_0(x_s), F_y(y_{s1}) = F_1(x_s), \\ F_y(y_{s2}) = F_2(x_s), F_y(y_{s3}) = F_3(x_s) \end{array} \tag{12}$$

Polynomial $F_y(y)$ is represented by dash curve in **Fig. 4**. Finally, the value of probability distribution $P_{app}\{M_{i,j}(x_D,y_D)\}$ is set on $F_y(y_S)$.

In image coding application, Bicubic interpolation technique generates a better performance than a Bilinear interpolation technique because more neighbour samples are incorporated to compute interpolation values [14]

3.2 Lanczos Interpolation

Lanczos interpolation is based on the 3-lobed Lanczos window function as the interpolation function. The interpolation algorithm uses 36 probability distributions $P_{app}\{M_{u,v}(x_S,y_S)\}$ that close to (x_S,y_S) position in block-based motion field, i.e.:

$$\begin{aligned} &x_{s0} = \mathrm{int}(x_s) - 2; \quad x_{s1} = x_{s0} + 1; \quad x_{s2} = x_{s0} + 2; \\ &x_{s3} = x_{s0} + 3 \quad ; \quad x_{s4} = x_{s0} + 4; \quad x_{s5} = x_{s0} + 5; \\ &y_{s0} = \mathrm{int}(y_s) - 2; \quad y_{s1} = y_{s0} + 1; \quad y_{s2} = y_{s0} + 2; \\ &y_{s3} = y_{s0} + 3 \quad ; \quad y_{s4} = y_{s0} + 4; \quad y_{s5} = y_{s0} + 5; \end{aligned} \tag{13}$$

where the definition of variables in the equation above are identical with definition used in Bicubic interpolation.

Firstly, probability distribution P is interpolated along x axis to have 6 intermediate values P_0, P_1, …, P_5, using equation below :

$$P_k = \sum_{i=0}^{5} a_i P_{app}\left\{M_{u,v}\left(x_{si}, y_{sk}\right)\right\}, \qquad 0 \le k \le 5 \tag{14}$$

Then, probability distribution $P_{app}\{M_{i,j}(x_D,y_D)\}$ is computed by interpolating intermediate, P_k, along y axis :

$$P_{app}\left\{M_{i,j}\left(x_D, y_D\right)\right\} = \sum_{k=0}^{5} b_k P_k \tag{15}$$

where a_i and b_k are coefficients that defined as:

$$a_i = L\left(x_s - x_{si}\right); \quad b_k = L\left(y_s - y_{si}\right) \tag{16}$$

where $L(x)$ is *Lanczos windowed sinc* function :

$$L(x) = \sin c(x).Lanczos(x) = \begin{cases} \dfrac{\sin(\pi x)}{\pi x}\dfrac{\sin(\pi x/3)}{\pi x/3}, & 0 \le |x| \le 3 \\ 0 & , \quad 3 \le |x| \end{cases} \tag{17}$$

3.3 Implementation Procedures

The implementation of both Bicubic and Lanczos interpolation techniques is adapted from Intel IPP functions with support regions of interest (ROI). All interpolation process is performed within ROI square defined in block-based motion field (origin) and pixel-based motion field (destination).

As shown in **Fig. 5**, ROI in a motion field is defined by size and offset from motion field origin. The origin of a motion field is stated to be on top left corner, with the value of x increases from left to right and the value y downward. In this paper, ROI size is set the same as block-based motion field resolution size so that offset x and y of ROI is zero. **Table 1** shows detail of parameters used in processing ROI to interpolate block-based motion field into pixel-based motion field.

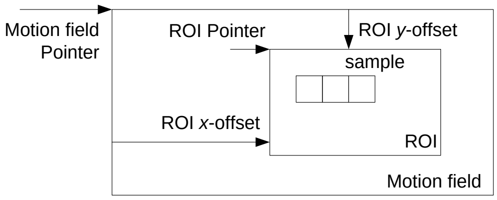


Fig. 5. ROI structure in an motion field

Table 1. Parameters values for motion field interpolation

Parameter	Value
Size of block-based motion field, ($M_{u,v}$)	11 × 9
Magnification factor	8×
ROI size	11 × 9
ROI *x*-offset and *y*-offset	0
Size of pixel-based motion field, ($M_{i,j}$)	88 × 72

4. Experimental Results and Analysis

Motion field interpolation procedures in unsupervised forward motion vector learning based on EM algorithm is shown in **Fig. 1**. We implement motion field interpolation procedures to interpolate block-based motion field onto pixel-based motion field in the existing domain transform EM-based WZVC codec with framework following reference proposed by [8], which implement different interpolation method.

As shown in **Fig. 6**, codec divides video sequence into GOP. First frame in GOP is encoded as a key frame by JPEG encoder. Compressed stream is sent to decoder and its reconstruction is used as a reference frame. The next consecutive frames in GOP, denoted as WZ frames, are encoded according to upper part of **Fig. 6**, and use previously reconstructed frame as SI to decode WZ frames conditionally. By knowing initial value of θ, we start to compute block-based motion field distribution $P_{app}\{M_{u,v}\}$ and interpolate it using Bicubic and Lanczos interpolation techniques to generate probability distribution of pixel-based motion field, $P_{app}\{M_{i,j}\}$. Motion field probability distribution will keep updated iteratively until it approaches the true motion.

To analyse the impact of proposed motion field interpolation approach using Bicubic and Lanczos interpolation on EM-based WZVC codec, several experiments have been performed which are reported in the following. Section 4.1 describes test conditions to justify codec performance. Section 4.2 provides some evaluation of RD performance comparison in different interpolation technique. Section 4.3 describes the evaluation of bit rate and temporal quality comparison, and Section 4.4 provides decoding complexity comparison.

4.1 Test Conditions

All results provided here were generated by following test conditions:

- Test video sequences: we used two video sequences with different characteristics, i.e. Foreman (higher and more complex motion complexity) and Carphone (lower motion complexity). Sample frames of those sequences are shows in **Fig. 7**. The Number of frame in each sequence is 96. This is the same number used to evaluate existing EM-based WZVC codec [8].

- Spatial and temporal: we used QCIF resolution (176 × 144 pixels) and 15 fps for both video sequences. Next, each WZ frame size QCIF was divided into four quadrants, each of size 88 × 77 pixels. Each quadrant was encoded separately using the corresponding quadrant in the previously reconstructed frame as a decoder reference. It is transformed using DCT based on 8 × 8 block to exploit spatial correlation within quadrant.
- GOP size is 2, 4, and 8
- RD points: RD performance was measured in four points that related with four scale factors of JPEG quantization matrix, Q_f, is 0.5, 1, 2, and 4 [22] as shown in **Fig. 8**. The scaling factors were associated with quality index factor, Q_i, of JPEG namely. 75, 50, 25, and 13, so that higher scaling factor has rougher quantization. Quantization indices become LDPC encoder input to reconstruct syndrome (S).

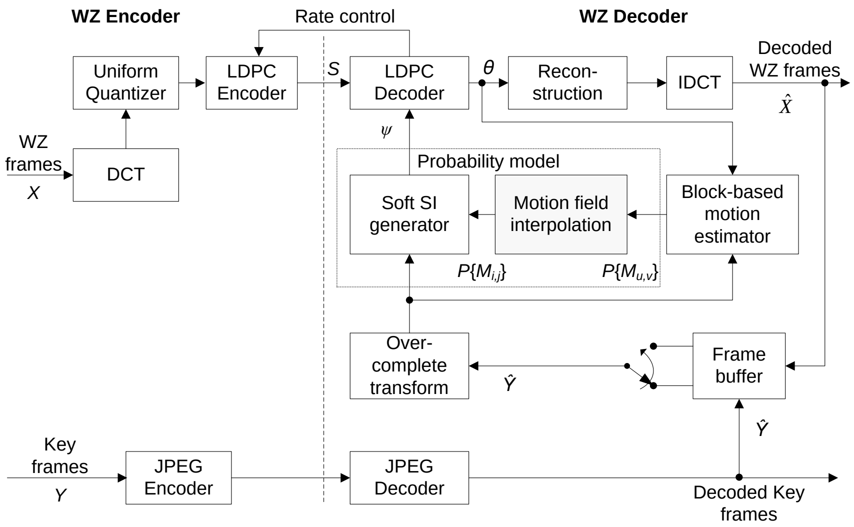
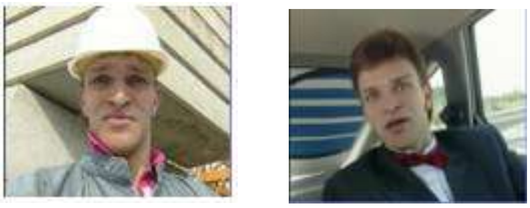


Fig. 6. Architecture of EM-based WZVC codec domain transform



(a) (b)

Fig. 7. Sample of test video sequences, (a) Foreman (frame 61), (b) Carphone (frame 39)

- Block size: 8 × 8 pixels for DCT, block-based motion estimation and probability model. For the latter two blocks, the motion search range is ± 5 pixels vertically and horizontally

- Number of EM iteration: 50. Whenever reconstruction of $\hat{x}$ did not match syndrome conditions after 50 iterations, decoder requested encoder an additional syndrome transmission through feedback channel.

8	6	5	8	12	20	26	31
6	6	7	10	13	29	30	28
7	7	8	12	20	29	35	28
7	9	11	15	26	44	40	31
9	11	19	28	34	55	52	39
12	18	28	32	41	52	57	46
25	32	39	44	52	61	60	51
36	46	48	49	56	50	52	50

$Q_{0.5}$

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

Q_1

32	22	20	32	48	80	102	122
24	24	28	38	52	116	120	110
28	26	32	48	80	114	138	112
28	34	44	58	102	174	160	124
36	44	74	112	136	218	206	154
48	70	110	128	162	208	226	184
98	128	156	174	206	242	240	202
144	184	190	196	224	200	206	198

Q_2

64	44	40	64	96	160	204	244
48	48	56	76	104	232	240	220
55	52	64	96	160	228	276	224
56	68	88	116	204	348	320	248
72	88	148	224	272	436	412	308
96	140	220	256	324	416	452	368
196	256	312	348	412	484	480	404
288	368	380	392	448	400	412	396

Q_4

Fig. 8. Four quantization matrices associated with four different RD performance points

- Rate control: we use rate-adaptive regular degree 3 LDPC accumulate codes with the length of 50688 bits as joint bitplane system platform to rate control.
- Laplacian noise (Z) and motion vector. Decoder was initialized by a good variant value and distribution of block-based motion field like the one used as in existing WZVC codec:

$$P_{app}^{(t)}\{M_{u,v}\} = \begin{cases} \left(\frac{3}{4}\right)^2, & if \quad M_{u,v} = (0,0) \\ \frac{3}{4}\cdot\frac{1}{80} & if \quad M_{u,v} = (0,*),(*,0) \\ \left(\frac{1}{80}\right)^2, & otherwise \end{cases} \tag{18}$$

- Bit rate and PSNR: only luminance component from each frame was used to compute bit rate and PSNR. RD evaluation includes both Key frame and WZ frame.

4.2. RD Performance Analysis

This section presents the proposed EM-based WZVC codec performance and compares it with the existing EM-based WZVC codec [8]. Evaluation is performed in GOP size 2, 4, and 8. **Table 2** and **3** show that EM-based WZVC codec with Bicubic and Lanczos interpolation techniques for motion field magnifying method from block-based into pixel-based generate RD close to the existing EM-based WZVC codec which use Bilinear interpolation method, either for sequence video Foreman or Carphone. In similar rate, both interpolation techniques result the same PSNR gain and constant in every JPEG quantization scale factor Q_f = 0.5, 1, 2 and 4.

In EM-based WZVC codec for motion field learning, block-based motion field consists of probability within interval [0,1]. This means that interpolation technique implementation is identical to mapping probability into probability. In [9], Chen et al. reported that linear convex combination of probability generated by interpolation technique must also continue within interval [0,1]. This identical RD performance to the existing EM-based WZVC codec indicates that both Bicubic and Lanczos interpolation techniques generate linear convex combination of probability within interval [0,1].

Table 2. Average luminance rate and PSNR of WZVC codec using different motion field interpolation method for Foreman video input

GOP	Q_f	Bilinear Interpolation			Bicubic Interpolation			Lanczos Interpolation		
		Rate (Kbps)	PSNR (dB)	Dec time (s)	Rate (Kbps)	PSNR (dB)	Dec time (s)	Rate (Kbps)	PSNR (dB)	Dec time (s)
2	0.5	509.63	35.49	1231	513.66	35.49	1114	512.83	35.49	1115
	1	348.09	32.94	901	351.06	32.94	829	350.29	32.94	827
	2	230.10	30.55	596	231.64	30.55	559	230.99	30.55	552
	4	158.88	28.38	447	159.47	28.38	429	159.23	28.38	425
4	0.5	467.43	35.49	1220	473.37	35.49	1105	471.65	35.49	1056
	1	313.91	32.94	892	318.54	32.94	826	317.05	32.94	777
	2	201.24	30.54	597	203.74	30.54	565	202.78	30.54	525
	4	136.46	28.34	437	137.41	28.34	422	137.41	28.34	397
8	0.5	447.47	35.49	1228	454.29	35.49	1121	452.75	35.49	1061
	1	298.13	32.94	904	303.95	32.94	831	302.35	32.94	793
	2	187.58	30.54	610	190.79	30.54	571	189.42	30.54	534
	4	125.25	28.32	433	126.61	28.32	418	126.73	28.32	394

Table 3. Average luminance rate and PSNR of WZVC codec using different motion field interpolation method for Carphone video input

GOP	Q_f	Bilinear Interpolation			Bicubic Interpolation			Lanczos Interpolation		
		Rate (Kbps)	PSNR (dB)	Dec time (s)	Rate (Kbps)	PSNR (dB)	Dec time (s)	Rate (Kbps)	PSNR (dB)	Dec time (s)
2	0.5	403.53	37.55	733	407.10	37.55	679	405.85	37.55	649
	1	275.00	34.67	497	277.86	34.67	475	276.96	34.67	450
	2	184.14	32.10	336	185.75	32.10	325	185.33	32.10	308
	4	128.90	29.57	241	129.49	29.57	235	129.25	29.57	225
4	0.5	367.50	37.55	738	373.38	37.55	661	371.48	37.55	659
	1	243.61	34.67	507	247.53	34.67	461	246.10	34.67	456
	2	157.16	32.10	340	159.29	32.10	314	158.64	32.10	312
	4	104.54	29.53	226	105.84	29.53	215	105.13	29.53	213
8	0.5	347.04	37.55	711	353.64	37.55	657	351.32	37.55	650
	1	225.83	34.67	488	230.41	34.67	461	228.69	34.67	435
	2	142.76	32.10	324	145.19	32.10	315	144.36	32.10	309
	4	91.74	29.51	213	93.22	29.51	212	92.27	29.51	206

4.3. Rate and Quality Analysis

Table 2 and 3 show an average evaluation from the comparison of RD performance between proposed EM-based WZVC codec (using Bicubic and Lanczos interpolation techniques) and the existing EM-based WZVC codec (using Bilinear interpolation technique).

However, those tables can not display the comparison performance in time domain, especially in gain and rate reduction in each frame.

Fig. 9 shows the temporal evaluation comparison for total number bit per frame and PSNR that generated by the proposed and existing EM-based WZVC codec, for Foreman and Carphone video sequences. Those graphics are acquired for GOP 8 and JPEG quantization scaling factor Q_f = 0.5. Bit consumption and PSNR from key frame is also plotted to completely visualize temporal evaluation. Generally, for both video sequences, Bicubic and Lanczos interpolation techniques implementation in EM-based WZVC codec generate identical video reconstruction and number of bit per frame compared to implementation of Bilinear interpolation used in existing EM-based WZVC codec.

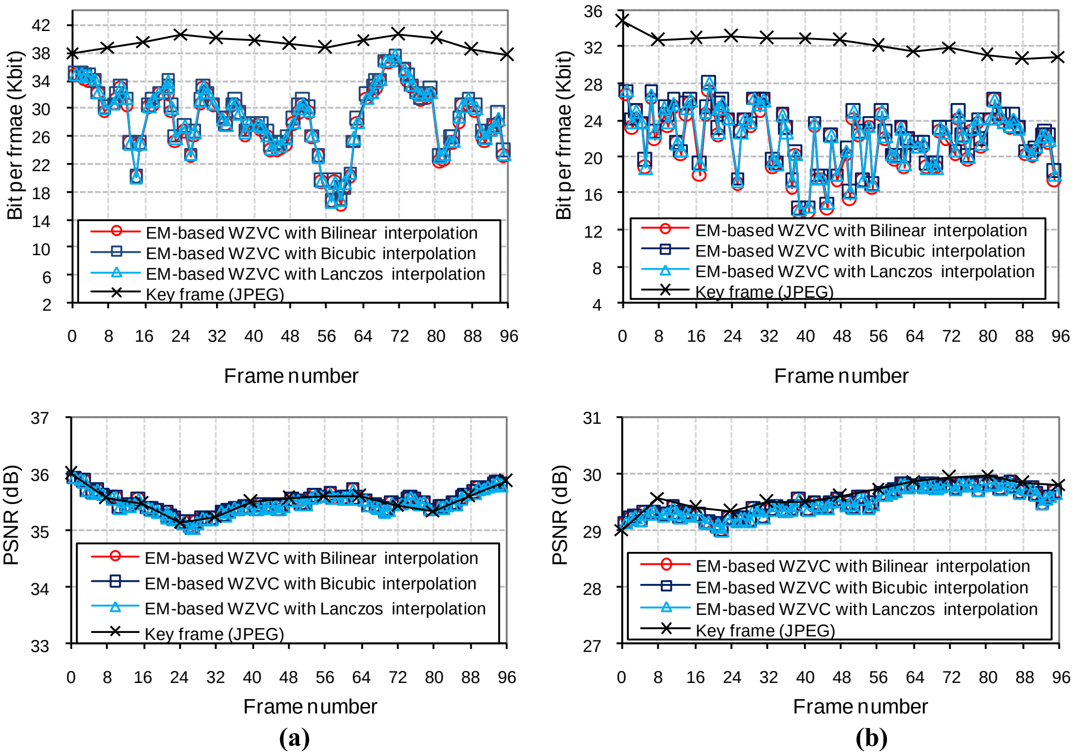


Fig. 9. Number of bit per frame and PSNR on GOP size 8 and Q_f = 0.5 : (a) Foreman, (b) Carphone

4.4. Decoding Complexity Analysis

Decoding complexity was evaluated by measuring average decoding time of EM iteration (in seconds) for every quadrant required by decoder to fulfil syndrome condition. This test was performed using Intel core quad @2.33 GHz processor with 3 GB memory under Windows Vista Operating System for Foreman video sequence. Meanwhile for Carphone video sequence, the test was performed using core 2 quad @2.66 GHz processor, 4 GB memory under Windows XP OS. Source codes were built under C++ programming language with Visual Studio C++ compiler. Bilinear, Bicubic, and Lanczos interpolation functions were implemented on the code using Intel IPP version 6 libraries, i.e. IPPI_INTER_LINEAR, IPPI_INTER_CUBIC, and IPPI_INTER_LANCZOS. This would mean to have a clear comparison mechanism.

Fig. 10 shows the comparison of average decoding time of iteration time per quadrant required to decode 96 frames by proposed EM-based WZVC codec and the existing one. Analysis was conducted on GOP 2, 4, and 8, each on JPEG quantization scaling factor 0.5, 1, 2, and 4. Generally, for both sequences, the implementation of Bicubic and Lanczos interpolation techniques on the proposed EM-based WZVC codec were able to reduce decoding complexity of the existing EM-based WZVC codec. The most decrease in complexity occured in JPEG quantization scaling factor 0.5.

Bicubic interpolation technique in EM-based WZVC codec reduced decoding complexity up to 9.47 % for Foreman video sequence and 7.33% for Carphone video sequence, for GOP 2. For GOP 4, decoding complexity was reduced up to 9.37% for Foreman video sequence, and 10.44% for Carphone video sequence. And for GOP 8, decoding complexity was reduced up to 8.71% for Foreman and 7.64% for Carphone.

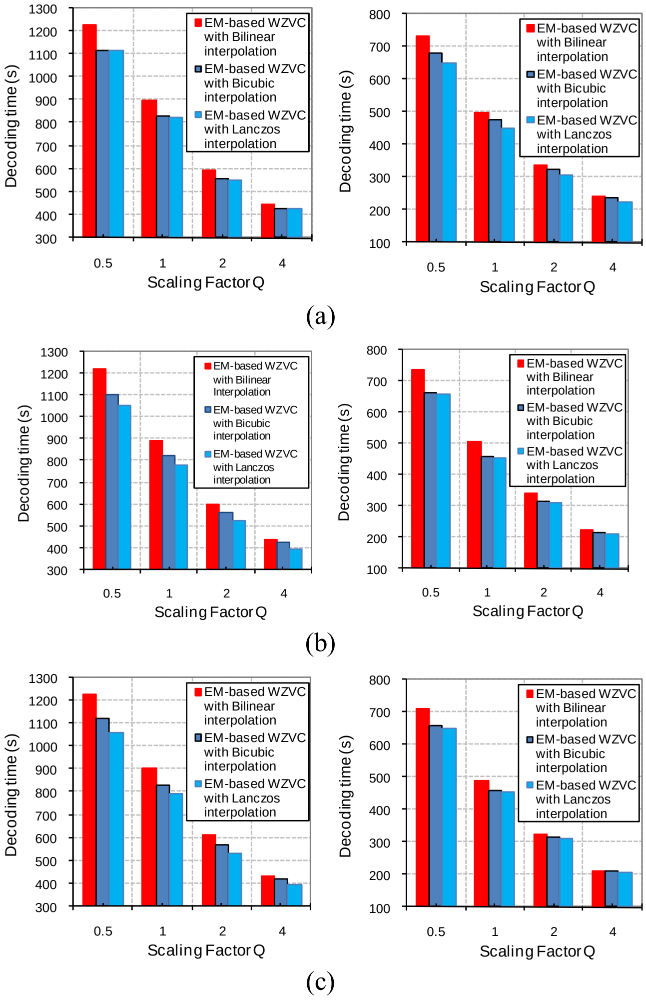


Fig. 10. Comparison of EM-based WZVC decoding complexity for Foreman (left) and Carphone (right) (a) GOP size 2, (b) GOP size 4, and (c) GOP size 8

Lanczos interpolation technique in EM-based WZVC codec has decoding complexity reduction similar to Bicubic interpolation technique in Carphone video sequence. Significant reduction of decoding complexity occurred in Foreman video sequence. In GOP 2, decoding complexity reduction of existing WZVC codec is up to 9.38%. For GOP 4 and 8, the reductions are up to 13.42% and 13.56% respectively. This indicates that Lanczos interpolation technique for upsample method of motion field is suitable for video sequence with high and complex motion.

4.5 Decoding Complexity – RD Tradeoff Analysis

Fig. 11 shows the percentage comparison between the increment of the bit rate and decrement of the decoding complexity required by a decoder for decoding the WZ frame. The percentage is calculated using the average bit rate and decoding time of EM iterations to satisfy the syndrome conditions on EM-based WZVC on implementation of Bicubic and Lanczos interpolation versus Bilinear interpolation. Overall, for Foreman sequence and GOP 8, the implementation of Bicubic and Lanczos interpolation showed a smaller percentage of bit rate consumption increment (1.08% - 1.91% and 0.97% - 1.39% respectively) than the decrement percentage of decoding complexity (3.46% - 8.71% and 8.99% - 13.56% respectively). It can be concluded that a large reduction of average decoding time of the EM iteration results in a small increment as a trade in bit rate.

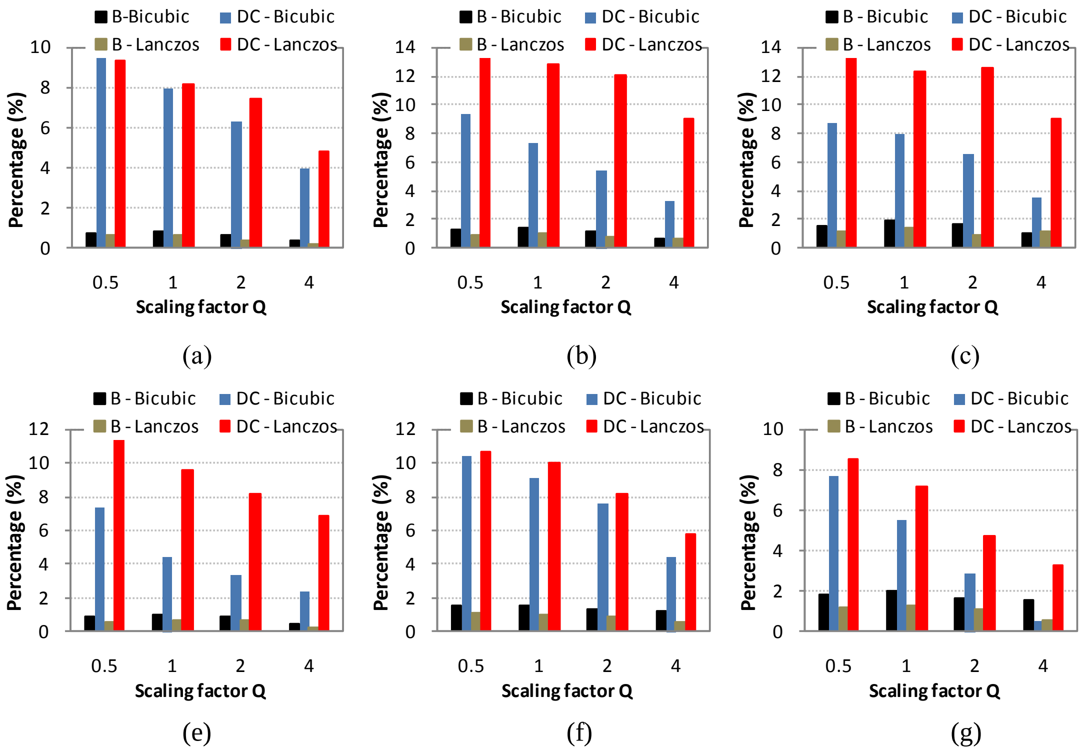


Fig. 11. Decoding complexity-RD tradeoff for Foreman (up) and Carphone (below): (a) and (c) GOP2, (b) and (f) GOP 4, (c) and (g) GOP 8

The best of decoding complexity-RD trade off obtained in JPEG quantization scale of 0.5 in each GOP. Implementation of Bicubic interpolation reduces the decoding complexity (DC-Bicubic) up to 7.33%, 10.44%, and 7.64% with the increment in bit rate (B-Bicubic) is only 0.88%, 1.57%, and 1.86%, in the GOP 2, 4 and 8 respectively for the Carphone sequence. Meanwhile, for the Foreman sequence, the reduction of the DC-Bicubic obtained up to 9.49%, 9.37%, and 8.71%, with the increment in B-Bicubic is only up to 0.79%, 1.25% and 1.50%, in the GOP 2 , 4 and 8 respectively.

The implementation of Lanczos interpolation on the EM-based WZVC has shown that for the input video sequence Carphone, decoding complecity (DC-Lanczos) of the codec can be reduced to 11.43%, 10.73%, and 8.57%, with the increment in bit rate (B-Lanczos) is only 0.57%, 1.07%, and 1.22%, in the GOP 2, 4 and 8 respectively. For the sequence Foreman, the DC-Lanczos reduction of the codec is up to 9.38%, 13.42% and 13.56%, with the increment in the B-Lanczos is only 0.63%, 0.89%, and 1.17% in the GOP 2, 4 and 8 respectively.

Due to the reduction in decoding complexity, EM-based WZVC codec can be optimized as a solution for wireless video applications, especially real time applications, by applying the Lanczos interpolation as a method to improve motion field into pixel precision.

5. Conclusion

In this paper, an improvement of motion field learning based domain transform WZVC codec via EM algorithm is proposed. Within EM algorithm framework, the proposed EM-based WZVC codec updates soft SI accuracy by improving the probability distribution of the motion field to the pixel precision using two interpolation techniques namely Bicubic and Lanczos. The experimental results showed that the improvement motion field decreased the decoding complexity. A large reduction of average decoding time of the EM iteration results in a small increment as a trade in bit rate. The largest decrease decoding complexity is generated by the Lanczos interpolation up to 13.56% in the GOP 8 to encode video sequences with high and complex motion.

References

[1] David Slepian. and Jack K. Wolf, “Noiseless coding of correlated information sources,” *IEEE Transaction Information Theory*, vol.19, no.4, pp.471– 480, Jul.1973. Article (CrossRef Link)

[2] Aaron D. Wyner. and Jacob Ziv, ”The Rate-Distortion function for source coding with side information at the decoder,” *IEEE Transaction. Information Theory*, vol.22, no.1, pp.1-10, Jan. 1976. Article (CrossRef Link)

[3] Anne Aaron., Rui Zhang and Bern Girod, ”Wyner–Ziv coding of motion video,” in *Proc. of Asilomar Conference on Signals, Systems and Computers*, vol.1, pp.240-244, Nov.2002Article (CrossRef Link)

[4] Xavier Artegas, Joäo Ascenso, M Dalai, S Klomp, Denis Kubasov and Mourad Ouaret, ”The DISCOVER codec: architecture, techniques and evaluation,” in *Proc. of Picture Coding Symposium*, Nov.2007Article (CrossRef Link)

[5] Catarina Brites, João Ascenso, Jose Quintas Pedro and Fernando Pereira, “Evaluating a feedback channel based transform domain Wyner-Ziv video codec,” *Signal Processing : Image Communication*, vol.23, pp.269-297, 2008. Article (CrossRef Link)

[6] Ricardo Martins, Catarina Brites, João Ascenso and Fernando Pereira, “Refining side information for improved transform domain Wyner-Ziv video coding, ”*IEEE Transactions on Circuits and System for Video Technology*, vol.19, no.9, pp.1327-1341, Sep.2009. Article (CrossRef Link)

[7] Wei Liu, Lina Dong and Wenjun Zeng, “Motion refinement based progressive side-information estimation for Wyner-Ziv video coding,” *IEEE Transactions Circuits System Video Technology*, vol.20, pp.1863–1875, Dec.2010. Article (CrossRef Link)

[8] David Varodayan, David Chen, Markus Flierl and Bernd Girod, “Wyner-Ziv coding of video with unsupervised motion vector learning,”, *EURASIP Signal Processing: Image Communication Journal, Special Issue on Distributed Video Coding*, vol.23, no.5, pp.369-378, Jun.2008. Article (CrossRef Link)

[9] David Chen, David Varodayan, Markus Flierl and Bernd Girod, “Distributed stereo image coding with improved disparity and noise estimation,” in *Proc. of IEEE Int. Conference on Acoustics, Speech, and Signal Processing*, Mar.2008. Article (CrossRef Link)

[10] I Made Oka Widyantara, Wirawan and Gamantyo Hendrantoro, “Efficient motion field interpolation method for Wyner-Ziv video coding’, *Telkomnika*, vol.9, no.1, pp.191-200, Apr.2011. Article (CrossRef Link)

[11] I Made Oka Widyantara, Wirawan and Gamantyo Hendrantoro, “Wyner-Ziv video coding with improve motion field using Bicubic interpolation,” in *Proc of IEEE International Conference of Instrumentation, Communication, Information Technology and Biomedical Engineering*, pp.218-222, Nov.2011. Article (CrossRef Link)

[12] Wang Haifang and Wang Anhong, “An improved unsupervised learning of motion estimation based on diamond searching for distributed video coding,” in *Int. Conf. on Computational Aspects and Social Networks*, pp.642-645, Sep.2010. Article (CrossRef Link)

[13] Jonathan Taquet and Claude Labit,”Optimized decomposition basis using Lanczos filters for lossless compression of biomedical images,” in *Proc. of IEEE Int. Workshop Multimedia Signal Processing*, pp.122-127, Oct.2010. Article (CrossRef Link)

[14] Zhen Ye, Jasjit Suri, Yajie Sun and Roman Janer, “Four image interpolation techniques for ultrasound breast phantom data acquired using fischer’s full field digital mammography and ultrasound system (FFDMUS) : a comparative approach,” in *Proc. of IEEE Interantional Conference of Image Procesing*, pp.1238-1241, Sep.2005. Article (CrossRef Link)

[15] Mehdi Hajizadeh, Mohammad Sadegh Helfroush and Askhan Tashk,”Improvement of image zooming using least directional differences based on linear and cubic interpolation,” in *Proc. of IEEE Int. Conf. on Computer, Control and Communication*, pp.1-6, Feb.2009. Article (CrossRef Link)

[16] Arthur Dempster, Nan Laird and Donald Rubin, “Maximum likelihood from incomplete data via the EM algorithm,” *Journal of the Royal Statistical. Society, Series B*, vol.39, no.1, pp.1–38, 1977Article (CrossRef Link)

[17] David Varodayan, Aditya Mavlankar, Markus Flierl and Bernd Girod, “Distributed coding of random dot stereogram with unsupervised learning of disparity,” in *Proc. of IEEE International Workshop Multimedia Signal Processing*, pp.5-8, Oct.2006. Article (CrossRef Link)

[18] David Varodayan, Yao-Chung Lin, Aditya Mavlankar, Markus Flierl and Bernd Girod, “Wyner-Ziv coding of stereo image with unsupervised learning of disparity,” in *Proc. of Picture Coding Symposium*, Nov.2007. Article (CrossRef Link)

[19] David Varodayan, Anne Aaron, and Bernd Girod,”Rate-adaptive codes for distributed source coding,” *EURASIP Signal Processing Journal*, vol.86, no.11, pp.498-519, Nov. 2006 Article (CrossRef Link)

[20] Nicola Asuni and Andrea Giachetti,”Accuracy improvements and artifacts removal in edge based image interpolation” in *Proc. of Inrernational Conference on Computer Vision Theory and Applications*, pp.58-65, Jan.2008. Article (CrossRef Link)

[21] *Intel Integrated Performance Primitives for Intel® Architecture*, Reference manual, vol. 2: Image and video processing, 2009.

[22] *Information Technology - Digital Compression and Coding of Continuous-tone Still Image-Requirements and Guidelines*, ISO/IEC 10918-1, ITU-T Recommendation T.81, 1993

I Made Oka Widyantara received the M.Eng degree in telecommunication information system from Institut Teknologi Bandung, Bandung, Indonesia, 2001. Currently he is pursuing his Dr degree at the Department of Electrical Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia. His research interests include distributed video coding and video processing. He is a student Member of IEEE.

Wirawan received the DEA degree in informatic and signal processing from Ecole Superieure en Sciences Informatiques Sophia-Antipolis, France, 1996 and Dr degree from Telecom Paris Tech, Paris, France, 2003, in the same field. He is currently a lecturer and head of multimedia communication lab in the Department of Electrical Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia. His research interests include wireless multimedia sensor networks and multimedia signal processing. He is a Member of IEEE.

Gamantyo Hendrantoro received the M.Eng and Ph.D degree in Electrical Engineering from Carleton University, Canada, 1997 and 2001 respectively. He is a Professor in the Department of Electrical Engineering, Institut Teknologi Sepuluh Nopember, Surabaya, Indonesia. He was one of the recipients of the Post-Graduate Research Award from CITR (Canadian Institute for Telecom Reseach) in 2002 and of the Young Scientist Award form URSI (International Union of Radio Science) in 2005. Also in 2005 he was named lecturer with Best Achievement in the Institut Teknologi Sepuluh Nopember. His research interests include wireless digital communication with emphsis on radio channel characterization and modeling. He is a Member of IEEE.