A Motion Compensation Method using Least Squares Motion Estimation Filter in Wavelet Domain

M Y Tjoa; Y Imanishi; N Yamane; Y Morikawa

Department of Communication Network Engineering, Faculty of Engineering, Okayama University

Abstract

In this paper, a new motion compensation (MC) method in discrete wavelet transform (DWT) domain is proposed to realize hierarchical MC. In order to dissolve shift-invariant or phase shift problem caused by down sampling in DWT, a locally adaptive interpolation filter, called MC filter is applied. This filter predicts DWT coefficients in current frame from the coarsely motion compensated DWT coefficients in reference frame, using motion vector estimated from lower resolution image. Simulation results show that the proposed MC filter is effective to compensate the phase shift caused by down sampling and the proposed hierarchical MC has higher coding gain compared to conventional non-hierarchical MC methods.

Keywords–Motion Picture Compression, Discrete Wavelet Transform, Hierarchical Motion Compensation, Least Squares Filter, Phase Compensation

1. Introduction

The demand of multi-resolution image coding applications that allows effective treatment of video data for a wide range of output devices or display, is increasing due to the growth in diversity of imaging technology from a small resolution display like a cell phone to a super-resolution display like a huge billboard. In a multi-resolution coding of images, as the wavelet representation provides good energy localization, both in frequency and time domains, and its coding distortion is not annoying compared with one of DCT, it has been successfully applied to still image compression. Therefore, the requirement for the hierarchical motion picture coding in wavelet domain is more and more increasing. In this paper, a hierarchical motion compensation (MC) method in wavelet domain using a least squares inter-frame prediction filter is proposed.

In the discrete wavelet transform (DWT) coding, an image signal is decomposed into high and low-bands by an orthogonal set of low-pass and high-pass filter in horizontal and vertical direction, and each band signal is down sampled. Then the high-bands (denoted by HL, LH and HH, where HL denote high in horizontal and low in vertical, for example) are encoded, and the low-band (denoted by LL) or the lower-resolution level image is further decomposed and encoded, recursively. So, the hierarchical coding in wavelet domain can be efficiently realized, provided efficient MC for high-bands.

However, in MC for high-band DWT coefficients, the down sampling causes phase shift between coefficients in current and reference frame, called shift-variant property, that yields inefficiency in MC [1]. The well-known conventional DWT based motion picture coding is low-band-shift motion estimation (LBSME)[1] in wavelet-domain but this method is incapable of hierarchical implementation.

In the proposed method, in order to compensate the phase shift due to the down sampling, an inter-frame prediction filter, called MC filter, that predicts DWT coefficients in current frame from the coefficients in reference frame is used. As the support region of MC filter in reference frame should include a signal corresponds to current signal with respect to its motion, the MC filter is applied to preparatory motion compensated reference frame to reduce support size. The motion vector used in the pre-MC is estimated using a block-matching (BM) method from the LL-band image and applied to high-bands (LH, HL, and HH), at each resolution level. In this paper, the MC filter is designed as a regularized least squares prediction filter whose coefficients minimize prediction error variance in local blocks and are coded as side information, block by block.

This paper is organized as follows: Section 2 describes outline of hierarchical MC in DWT domain and detail of the proposed MC filter. Comparative studies between proposed and conventional LBSME methods are presented in Section 3. It is shown that the MC filter is effective to compensate the phase shift caused by down sampling with a little side information and it is also shown that, for images including complex motion, the MC filter is also effective for the phase shift caused by object shape variation. As a result, the proposed method has higher coding gain compared with conventional methods. Finally, conclusions are given in Section 4.
2. Hierarchical Motion Compensation Method in Wavelet domain

The general process of hierarchical implementation in wavelet domain is illustrated in Figure 1 and proceeds as follows:

1. Motion images are transformed into wavelet domain.
2. Coding of frames starts from the top level. The LL-band at this level (the lowest resolution image) in each frame can be encoded by conventional MC using block-matching (BM) method. Then high-bands (HL, LH and HH) at this level are encoded for next resolution frames.
3. Next, the same coding process as the starting process is applied to the next level without the LL-band. This process can be applied hierarchically up to the highest resolution frames.

![Figure 1. Hierarchical implementation starting from lowest resolution level up to highest level.](image)

On the high-bands coding at each level in this procedure, the shift variant or phase shift problem reduces efficiency of MC and prevents hierarchical implementation.

3. Proposed method

In this chapter, a new MC method in wavelet domain using MC filter is proposed to overcome the phase shift problem. In order to compensate the phase shift caused by down sampling and motion between frames, a linear prediction filter that minimizes the prediction error variance estimated by regularized least squares method is used.

3.1 Procedure of Proposed Method

The proposed method is processed as follows:

1. Transform both reference frame and current frame by $S$ -level DWT. Encode LL $(S)$ -band by a method mentioned later and let $s ← S$.
2. Divide all bands LL $(s)$, LH $(s)$, HL $(s)$ and HH $(s)$ of both frames into blocks in size $M(s) × M(s)$.
3. Obtain preparatory motion compensated high-bands, denoted by LH$_{MC}(s)$, HL$_{MC}(s)$ and HH$_{MC}(s)$, from high-bands in reference frame, LH $(s)$, HL $(s)$ and HH $(s)$, respectively, using BM method described in 3.2.
4. Estimate the MC filter coefficient vector $a$ with support size $N × N$ in each block that predict current block using a method described in 3.3, quantize $a$ to $a_0$ using average preserving linear quantizer of step width $Q$ and encode them as side information.
5. Predict high-bands in current frame from pre-motion-compensated high-bands in reference frame using $a_0$, block by block, and encode the prediction error.
6. Let $s ← s − 1$ and repeat the steps 2 to 5 until $s = 0$.

In this procedure, the LL $(S)$ -band is encoded using the same method for high-band, except for the filter coefficients are quantized with step width $Q_0$.

3.2 Preparatory MC

In order to decrease the support size of MC filter, each band signals are coarsely motion compensated, as follows.

1. Estimate the motion vector $v$ of each block from LL-band using BM method.
2. Use $v$ to shift the corresponding block in high-bands in reference frame to obtain preparatory Motion compensated high-bands.
3. Encode $v$ as side information.

<table>
<thead>
<tr>
<th>Image</th>
<th>PSNR (dB)</th>
<th>Level by level motion vector estimation</th>
<th>Top level motion vector estimation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Claire</td>
<td>37.483</td>
<td>37.465</td>
<td></td>
</tr>
<tr>
<td>Cronkite</td>
<td>35.336</td>
<td>35.316</td>
<td></td>
</tr>
<tr>
<td>Mobile</td>
<td>21.287</td>
<td>21.266</td>
<td></td>
</tr>
<tr>
<td>Plant</td>
<td>26.977</td>
<td>26.97</td>
<td></td>
</tr>
<tr>
<td>Train</td>
<td>28.181</td>
<td>28.155</td>
<td></td>
</tr>
</tbody>
</table>

In the case that the number of resolution level $S$ is small, for example 3-levels examined in this paper, the motion vector estimated from the lowest-band can be used for the coarse MC in all levels. Table 1. compares PSNR of prediction image between two methods, one estimates $v$ in
each level and the other estimates $v$ only in the top level $S$ and uses it to all levels, enlarging according to resolution. PSNR is defined as,

$$PSNR = 10 \log_{10} \frac{255^2}{D}$$

where, $D$ denotes mean power prediction error. It is shown that, PSNR of the two methods are almost the same and the proposed method adopts the latter method for simplicity.

### 3.3 Motion Compensation Filter

In the proposed method, the motion of each block is estimated using MC filter from pre-compensated reference frame. Let $y$ denote a vector of samples of signal in a block in current frame and $\hat{y}$ denote linear prediction of $y$,

$$\hat{y} = Xa,$$  \(1)\)

where, $X = [x_1, x_2, \ldots, x_{n+1}]$ is a matrix whose column vector $x_i$ is a vector of samples of $i$’th signal in the support region in the pre-compensated block in reference frame and $a$ denotes prediction coefficient vector. In the optimum filtering problem in images, a locally adaptive filter estimated by least squares error criterion $\|y - \hat{y}\|$ subject to appropriate regularization constraint has comparative performance as the global optimum filter in the sense of minimum variance and including non-linear filters[5]. In such filter, the coefficient vector $a$ is derived as following equation (2), in the case that using ridge estimation method as regularization,

$$a = \left[ \frac{1}{M^2} X^T X + \delta^2 I \right]^{-1} \frac{1}{M^2} X^T y,$$  \(2)\)

where $\delta^2$ denotes regularization constant. Note that a preparatory experiment shows that $a$ in (2) can be applied to LL $(S)$-band without unbiasing.

### 4. Simulation

In this section, coding performance of the proposed method is evaluated in coding gain, and compared with conventional method using SIDBA and MPEG images of 256 gray levels. As a conventional method, LBSME is compared, because even though the LBSME is not capable of hierarchical method, it is one of MC realization in DWT domain [1]. Simulation used the first frame of image sequence as reference frame to predict the second frame of each image sequence.

**Table 2. Parameters of proposed method.**

<table>
<thead>
<tr>
<th>Parameter level</th>
<th>level 1</th>
<th>level 2</th>
<th>level 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>$M$</td>
<td>32</td>
<td>16</td>
<td>8</td>
</tr>
<tr>
<td>$N$</td>
<td>3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Q$ ($Q_0$)</td>
<td>0.08 (0.004)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\delta^2$</td>
<td>1.25</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Conditions of the experiment are shown in the followings.

1. Coding gain $G$ is evaluated as followings,

$$G = -\frac{1}{2} \sum_{i=0}^{3} P(i) \log \frac{\sigma_i^2}{\sigma^2} - h,$$  \(3)\)

where $\sigma_i^2$ denotes variance of prediction error in band $i$, $\sigma^2$ denotes variance of current frame and $P(i)$ denotes probability of band $i$. And $h$ denotes entropy of MC filter coefficients and rate for motion vector is ignored.

2. DWT is performed by the D(9,7) filter [3], with $S = 3$ level.

3. Table 2. shows parameters of the proposed method, block size $M$, support size $N$, coefficient quantization step size $Q$ and regularization constant $\delta^2$, determined by a preparatory experiment. MC block size of LBSME is the same as proposed method [1].

4. For the block matching in proposed method, modified logarithmic search method [4] is used with search range [-16, 16] pixels in horizontal and vertical direction, respectively. In LSBME, full search method is used with the same search area.

Coding gains of proposed method and LBSME are compared in Table 3. It is shown that proposed method has higher gain in almost all images except for “Mobile”. Figure 3 shows the coding gain in band-by-band. It is shown that, in the proposed method, coding gain is improved in several bands, especially in high bands. These improvements can be considered as the effect of phase compensation of proposed MC.
Figure 3. Comparison of band-by-band coding gain of prediction error for image sequence (a) Claire, (b) Cronkite, (c) Plant, (d) Football.

Table 3. Comparison of coding gain between proposed method and LBSME method.

<table>
<thead>
<tr>
<th>Image sequence (lines×pixels/line)</th>
<th>Coding Gain $G (\Delta)$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Proposed</td>
</tr>
<tr>
<td>Claire (288×360)</td>
<td>5.478(0.105)</td>
</tr>
<tr>
<td>Cronkite (256×256)</td>
<td>4.819(0.066)</td>
</tr>
<tr>
<td>Mobile (240×352)</td>
<td>1.57(0.099)</td>
</tr>
<tr>
<td>Plant (256×256)</td>
<td>2.113(0.071)</td>
</tr>
<tr>
<td>Slowfg (240×352)</td>
<td>2.68(0.117)</td>
</tr>
<tr>
<td>Football (240×352)</td>
<td>1.900(0.113)</td>
</tr>
</tbody>
</table>

5. Conclusion

In this paper, a hierarchical motion compensation method in DWT domain using inter-frame prediction filter was proposed. Simulation results showed that the proposed method is effective on motion compensation in shift variant and deformed inter-frame signals, and has higher coding gain compared with conventional non-hierarchical methods.

References


