A New Fast Multiresolution Motion Estimation In the Wavelet Detail Level

Kwang-Yong Kim, Kyeong-Hwan Lee, Tae-Ho Lee, and Duk-Gyoo Kim

School of Electronic and Electrical Engineering, Kyungpook National University, 1370 San-Kyuk Dong, Puk Gu, Taegu, 702-701, Korea
Phone: +82-53-950-5525
Fax.: +82-53-950-5505
E-mail: khawk@palgong.knu.ac.kr

Abstract

In this paper, a new hierarchical motion estimation (ME) scheme using the wavelet transformed multi-resolution image layers is proposed. While the coarse-to-fine (CtF) ME, used in previously proposed coding schemes, can provide a better estimate at the coarsest resolution, it is difficult to accurately track motion at finer resolution. On the other hand, in fine-to-coarse (FtC) ME, it can solves this local minima problem by estimating motion track at the finest subband and propagating the motion vector (MV) to coarser subband. But this method causes to higher computational overhead.

This paper proposes a new method for reducing the computational overhead of fine-to-coarse multi-resolution motion estimation (MRME) at the finest resolution level by searching for the region to consider motion vectors of the coarsest resolution subband.

Ⅰ. Introduction

Wavelet transform is a valuable tool in video processing because of its flexibility in representing nonstationary signals. Wavelet-based compression has the advantages of efficient decorrelation of image frames and reduced-complexity multiresolution motion estimation. Several motion estimation techniques have been proposed in the wavelet domain. The coarse-to-fine (CtF) motion estimation techniques[1][2][3][4] generally have a lower complexity at the expense of inaccurate estimation. Several reason of inaccurate estimation are exited. Firstly, it is the potential for inaccurate motion estimation (ME) at the coarsest resolution, due to lack to detail and aliasing effects. These inaccuracies result in suboptimal ME at finer resolutions. Secondly, the small block-size at a coarser resolution level does not provide robust motion estimation. Finally, the amount of aliasing increases with the number of decomposition levels. On the other hand, the fine-to-coarse (FtC) motion estimation techniques[5] provide a superior estimation, but at a higher complexity. In this technique, because accurate motion estimation are formed at the finest resolution and then scaled to coarser resolutions in the encoded process, these motion estimates better track the true motion and exhibit lower entropy than coarse-to-fine estimations, providing higher quality, both visually and quantitatively, but because fine-to-coarse MRME is executed in the full search region at the finest resolution subband, it is caused to increase much computational complexity in relatively coarser energy level.

This paper proposes a new method for reducing the computational overhead of fine-to-coarse MRME at the finest resolution level by searching for the region to consider motion vectors of the coarsest resolution subband. While FtC MRME searches for the full search area in finest resolution level, we process closer estimation in the region of scaling initial motion vectors in the coarsest resolution level, and more sparser estimation in the others. And so we determine the potential motion area and estimate the motion vector at finest resolution level. And then this determined motion vector are scaled to coarser resolutions. Therefore, this method is similar to computational complexity of the CtF MRME technique and very significantly reduces that of the FtC MRME technique.In addition, they provide higher quality than CtF MRME, both visually and quantitatively.

This paper is organized as follows. Section 2 provides several brief MRME techniques; like CtF method, FtC method. In Section 3,the proposed algorithm are detailed. The simulation results are provided in Sec.4, which is followed by conclusions in Sec.5.

Ⅱ. CtF and FtC method

1. Coarse-to-Fine MRME method.

This approach exploits the multiresolution property of the wavelet pyramid in order to reduce the computational complexity of the motion estimation process. We note that Zhang et al.[1] have consided several techniques for motion estimation. Here, we choose the "S_8, W_8 +refine" techniques for motion estimation since it provided superior motion estimation. Fig.1 is shown the typical pyramid construction. In this MRME scheme, the motion vectors at the coarsest level of the wavelet pyramid are first estimated using the conventional block-matching-based motion estimation algorithm. Then the motion vectors at the next level of the wavelet pyramid are predicted from the motion vectors of the preceding level, which are refined at each step. At the same time, it codes motion vector and compensated prediction image at each step.

The process order is following.

Step1.Construct Discrete Wavelet Transformed image with N-level

Step2. Estimate the motion at coarsest level (S_8, W_8^o).

From the MVs, form a motion compensated prediction image $\hat{S}_8$, $\hat{W}_8^o$ and prediction residual,

$\overline{S}_8(=S_8-\hat{S}_8)$, $\overline{W}_8^o(=W_8-\hat{W}_8)$

Code the MVs and the prediction residual information.

Step3. Increase the level.

Using the MVs(at W_8^o), the motion vectors at the next level of the wavelet pyramid are predicted, which are refined at each step

$$V_j^o(x,y)=V_8^o(x,y)\times 2^{N-j}+\Delta(\delta x,\delta y) \quad for \quad j=1,2,,,N \tag{1}$$

$$\Delta(\delta x+\delta y)=\arg\underset{\delta x,\delta y\in\Omega}{Min}\left[\frac{1}{XY}\sum_{p=-X/2}^{X/2}\sum_{q=-Y/2}^{Y/2} I_i(x+p,y+q)-I_{i-1}(x_1+p+x+\delta x,y_1+q+y+\delta y)\right] \tag{2}$$

From this scaling and refining MVs, form a motion compensated prediction image($\hat{W}^o_{2^{N-j}}$), and prediction residual($\overline{W}^o_{2^{N-j}}$),

Code the MVs and the prediction residual information.

Step4. If the proceeding is the finest level, go the next image.

Where V_j^o(x,y) represents the motion vector of the reference block centered at (x,y) for the 0-orientatiom subimage for jth levels of the pyramid. The incremental motion vector $\Delta(\delta x,\delta y)$ is calculated within a reduced search area centered at 2 V_8^o (x,y) and 4 V_8^o (x,y) for level-2 and level-1 pyramids, respectively. The subimages of level-3, level-2.and level-1 pyramids are divided into small blocks of size $n\times n$, $2n\times 2n$, and $4n\times 4n$,respectively.

With this structure, the numbers of blocks in all the subimages are identical. As a result, there is a one-to-one correspondence between the blocks at various levels of wavelet pyramid. The search windows for level-3, level-2and level-1 subimages are p, p/2,and p/4, respectively.

In this method, motion estimation techniques generally have a lower complexity at the expense of inaccurate estimation. For the rest, M.K. Mandal[2] has proposed the method that the motion vectors are dropped adaptively depending on the motion compensation performance, and Karlekar Jayashree, Desai U[3] has proposed the MSAD method to estimate the motion. Although these are all good performance, these methods have the weak point to discard the detail information.

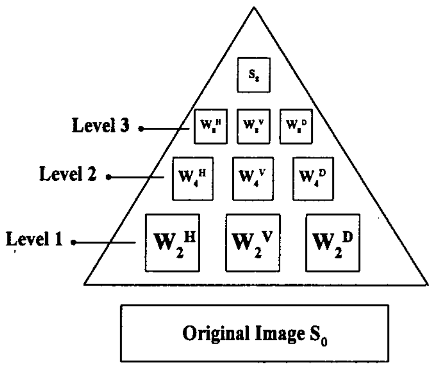


Fig. 1. The Structure of Wavelet Pyramid

2. Fine – to – Coarse MRME method

Conklin et al.[5] have proposed a MRME scheme based on a fine-to-coarse approach. Here, comparing with CtF method, the initial motion estimation is executed in the pixel domain. In the other word, the motion vectors at the finest level of the wavelet pyramid are first estimated using the conventional block-matching-based motion estimation algorithm. Then, scale and refine that at coarser resolutions. Here, for the ordering transmission of the level, FtC method is needed to code the refinements of the MVs and compensated prediction image after the total motion estimation process.

The process order is following.

Step1.Construct Discrete Wavelet Transformed image with N-level

Step2. Estimate the motion at finest level (W_2^o) and as it decrease the level, scale and refine the MVs to coarsest level.

$$V_j^o(x,y)=V_2^o(x,y)\div 2^{N-j}+\Delta(\delta x,y) \quad for \quad j=1,2,,,N \tag{3}$$

Step3. Increase the level.

Using the best set of MVs, form a motion compensated prediction image($\hat{W}^o_{2^{N-j}}$), and prediction residual($\overline{W}^o_{2^{N-j}}$),

Code the MVs and the prediction residual information.

Step4. If the proceeding is the finest level, go the next image.

In this technique, because accurate motion estimation are formed at the finest resolution and then scaled to coarser resolutions in the encoded process, these motion estimates better track the true motion and exhibit lower entropy than coarse-to-fine estimations, providing higher quality, both visually and quantitatively, but because fine-to-coarse MRME is executed in the full search region at the finest resolution subband, it is caused to increase much computational complexity in relatively coarser energy level.

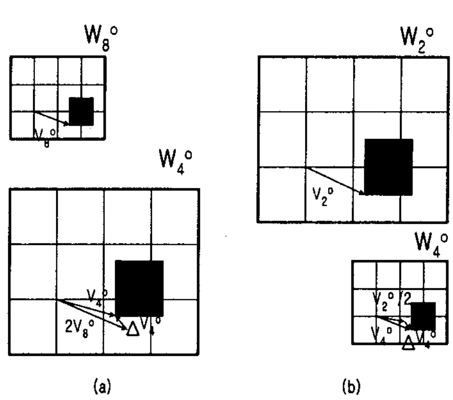


Fig 2. The Procedures of (a) CtF method (b) FtC method

Ⅲ. Proposed Method

This paper's propose is to reduce the computational complexity of the FtC method, similar to that of the CtF method.

Although motion estimation of coarsest level is difficult to track the true motion, due to lack to detail and aliasing effects, it can give the very important clue to estimate motion at finest level.

In upper two method, they independently process motion estimation at the coarsest level about other levels.

So, we choose to use these MVs of coarsest level for reducing the searching area of estimating motion vectors in the finest level subimages.

The technique proposed in this section exploits this characteristic, but to reduce the computational complexity of block-based motion estimation algorithm. We assume that the motion vector at finest level is very likely to exist in the potential area, where scaling the motion vector at coarsest level.

We first estimate the motion vectors at coarsest level in a frame, as indicated by the darkened blocks of Fig.3, by using a block-matching technique including full search algorithm.

Using the MVs, we then limit the searching area by a factor of 2, comparing the original searching area($N \times N$) and in order to increase the searching performance, we shift the limited searching area ($N/2 \times N/2$) to inner point.

As shown fig.4., then we choose the alternatively subsampling point to estimate the motion in a limited searching area, using block-matching algorithm, and choose the vector for which the mean absolute difference(MAD) is the smallest, Then, we again estimate the motion around the chosen vector.

The Process order is following.

Step1.Construct Discrete Wavelet Transformed image with N-level

Step2. Estimate the motion at the coarsest level(S_8).

Step3. Using this method, Estimate adaptively the motion at finest level (W_2^o) and as it decrease the level, scale and refine the MVs to coarsest level.

$$V_2^o(x,y) = V_8(x,y) \times p(x,y) + \Delta'(\delta x + \delta y)$$

$$\bullet\ where\ p(x,y) \begin{pmatrix} x = x \times 2 & if\ \ x = 2 \\ x = x \times 4 & else \\ y = y \times 2 & if\ \ y = 2 \\ y = y \times 4 & else \end{pmatrix} \quad (4)$$

- $\Delta'(\delta x + \delta y)$: alternating subsampling point.

$$V_j^o(x,y) = V_2^o(x,y) \div 2^{N-j} + \Delta(\delta x + \delta y)$$

$$for\ \ j = 1,2,,,N \quad (5)$$

Step4. Increase the level.

Using the best set of MVs, form a motion compensated prediction image($\hat{W}_{2^{N-j}}^o$), and prediction residual($\overline{W}_{2^{N-j}}^o$),

Code the MVs and the prediction residual information.

Step5. If the proceeding is the finest level, go the next image.

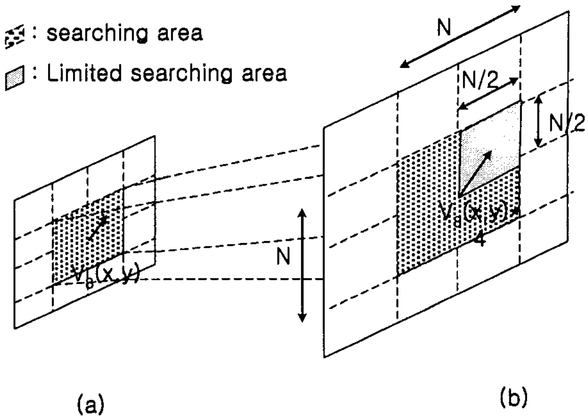


Fig .3. First, motion Estimation starts from (a)the coarsest level, and then this motion vector is scaled to (b)the finest level, and then estimate the motion in the limited searching area.

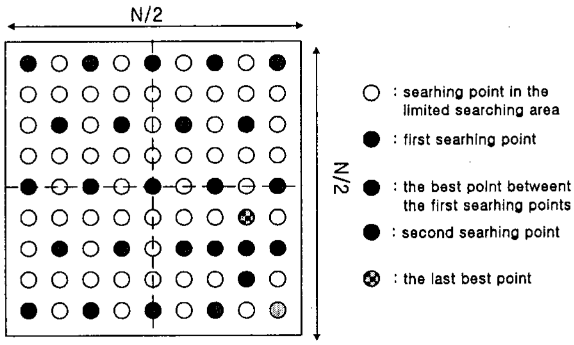


Fig. 4. Searching pattern in the limited searching area

Ⅳ. Simulation and Results

To test the performance of this approach, simulation were run on 720×480 three sequences; 50 frames of Football, Susie, and 40 frames of Table tennis. Football and table tennis sequences offer an interesting combination of still, slow- and fast-moving objects, camera zoom and panning, and objects with relatively difficult sizes. On the other hand, Susie sequences offer slow motion and low spatial detail.

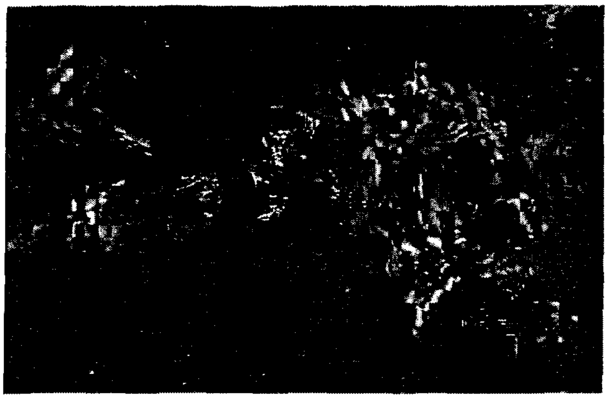
We have used the Daubechies-8 Tab wavelet, which provides good coding performance. The peak signal-to-ratio(PSNR), which defined as

$$PSNR = 10\log_{10}\left(\frac{255^2}{MSE}\right) \quad (6)$$

(where MSE is the mean square error between the signal/images)

has be employed as a measure of the quality of the reconstructed images. For the relative comparison, we have run the CtF method, FtC method, and proposed method. In CtF method, the searching area is $-2 \sim +2$. And, in FtC method, the searching area at finest level is $-8 \sim +8$ and that of other levels is $-2 \sim +2$.

That result is shown the table 1.As shown the table 1, the computational complexity of the proposed method is about 10% than that of FtC method, similar to that of CtF method. And the increase of PSNR is 0.5~0.9dB. Comparing the PSNR's increase of FtC method with that of CtF method, the increase of the proposed method is about 70%.



(a)

(b)

(c)

Fig. 5. the Motion Compensated 1 level- H subband image of the 11th football frame :
(a) CtF method ,(b) FtC method ,(c) Proposed method

Table 1. PSNR and Computational complexity

	PSNR			Matching point/ block (1 level)		
	CtF	FtC	Paper	CtF	FtC	Paper
Fb	23.69	25.03	24.62 (68%)*			
Tt	25.56	26.63	26.26 (66%)*	25	289	27 (9.8%)**
Ss	33.55	34.34	34.03 (65%)*			

()*: the ratio of PSNR of paper-CtF to FtC-CtF
()**: the ratio of matching point of paper to that of FtC

Fig.5 is shown the motion compensated 1 level-H subband image of the football 11th frame. As shown the Fig.5, we can see that our proposed method track the motion at the fine level very well

V . Conclusion

In this paper, we proposed a new method for reducing the computational overhead of fine-to-coarse MRME at the finest resolution level by searching for the region to consider motion vectors of the coarsest resolution subband.

As shown the result, our method of the computational complexity is reduced about 10%, That value is similar to computational complexity of the CtF MRME technique and very significantly reduces that of the FtC MRME technique.

In addition, flicking in textured regions, such as the textured background in Table Tennis or Football, is significantly reduced using our method.

As a result, this method ME provides both higher PSNRs and better visual quality.

References

[1] Y. Q. Zhang and S. Zafar, "Motion-Compensated transform coding for color video compression." IEEE Trans. Circuits Sysst. Video Technol., Vol. 2, No. 3, pp *285-296, 1992.*

[2] M. K. Mandal, E. Chan, X. Wang, S. Panchanathan, "Multiresolution motion estimation techniques for video compression", Optical Engineering, Vol 35, No 1, pp 128-136, Jan, 1996.

[3] K. Jayashree, Desai U., "A New Multiresolution Motion Estimation and Compensation Scheme", Proc. IEEE ISCS'99 , pp459-462, 1999

[4] *K. M. Uz, M. Vetterli, and D. J. Legall,* " Interpolative multiresolution coding of advanced television with compatable subchannels." IEEE Trans, on CSVT, Vol 1, No 1, March, 1991.

[5] G. J. Conklin, S. S. Hemami, "Multiresolution Motion Estimation", Proc. of ICASSP, pp 2873-2876,1997.