Motion Estimation Approach Based on Dual-Tree Complex Wavelets

Changzhu Wu, Qing Wang, Hongxiao Wang
School of Computer Science and Engineering, Northwestern Polytechnical University
Xian 710072, P. R. China
qwang@nwpu.edu.cn

Abstract

The loss of information due to occlusion and other complications has been one of the main bottlenecks in the field of motion estimation. In this paper, we propose a novel motion estimation algorithm based on dual-tree complex wavelet (DT-CWT), which utilizes its approximate shift invariance and directional selectivity. Subbands within different orientations and levels are individually treated to overcome the issue of error propagation induced by frequently used coarse-to-fine searching strategy, as these subbands are approximately independent from each other. Then, the frame next to the current one is introduced to supplement the lost information due to the occlusion and to improve the estimation accuracy within boundary areas. Experimental results show that the proposed method is effective with robustness to intense motion, scene change and occlusion.

1. Introduction

Motion estimation plays an important role in many applications in computer vision, and has been one of the most active research hotspot. However, occlusion and acute motion, frequently encountered in motion analysis, lead to increasing in complexity and thus make it necessary to design a robust method for motion estimation.

In the last decades, a great number of approaches for motion estimation have been proposed in the literatures and we divide them into three groups, which were optical flow based approaches[1], block matching based approaches[2, 3] and wavelet based approaches[4-11]. Although the first two groups are easy to implement, they can hardly deal with occlusion and acute motion. Moreover, block artifacts often deteriorate the visual effect for block matching based methods.

In 1991, Metin Uz[4] first proposed multi-resolution motion estimation in the wavelet domain and since then many approaches with outstanding performance were reported[6-11]. To overcome the shift variance of real discrete wavelet transform (DWT), Kim & Park[8, 9] proposed Low-Band-Shift(LBS) filter, but this induced extra computational cost and memory requirement. On the other hand, Magarey & Kingsbury[5] introduced phase based complex discrete wavelet transform (CDWT), which provides approximate shift invariance and better orientation selectivity. Taking advantage of the relation between local translation and its corresponding phase difference of the subbands coefficients, accurate displacement field can be obtained with CDWT. However, CDWT does not satisfy the perfect reconstruction condition, which makes it impossible to obtain motion compensated image in the wavelet domain directly. Besides, redundant wavelet transform(RWT) was also proposed to overcome the shift variance problem in [11].

To solve to the above-mentioned problem for the CWT, the dual-tree complex wavelet transform (DT-CWT)[12] was proposed and favorable performance with applications of image de-noising[13] and texture analysis[14] was obtained. In conclusion, DT-CWT not only has the advantages of CDWT, but also guarantees the perfect reconstruction (PR) condition. As a result, motion field can be first estimated based upon its approximate shift invariance; and then the motion-compensated image is obtained by the PR condition.

Recently, much attention has been drawn to tackle the loss of information under different circumstances [10, 15, 16]. Occluded regions were directly detected either to prevent oversmoothing of motion [15] or to infer motion vectors with bilateral diffusion [16] within these regions. Moreover, motion estimation of blocks within the boundary area has also been one of the stubborn problems. In [10], LBS was combined with symmetric padding to minimize the effects of boundary discontinuities. However, all the above problems can be addressed properly when the successive frame (called the backward reference frame) is considered.

In the paper, we first estimate the wavelet coefficients of the current frame from the forward- and backward- reference frame with DT-CWT, and then directly reconstruct the motion-compensated image by the inverse transform. The main advantages of the work are twofold: first, it can deal with not only the boundary problem but also the frequently occurred occlusion and acute motion, ascribed to the introduction of the backward reference frame; and second, the motion-compensated image can be directly

978-1-4244-2175-6/08/$25.00 ©2008 IEEE

obtained in the wavelet domain benefiting from the perfect reconstruction condition.

The rest of the paper is organized as follows: Section 2 briefly describes the DT-CWT used to compute the motion vectors. In Section 3, some concerns for motion estimation are first discussed and it is then shown that how DT-CWT can be applied to motion estimation, especially for the case of occlusion. Experimental results and discussions are shown in Section 4 and conclusion is drawn in Section 5.

2. Dual-tree complex wavelets

We can achieve the approximate shift invariance with a real DWT by doubling the sampling rate at each level of the tree, where the samples must be evenly spaced. This is equivalent to two parallel fully-decimated trees with the corresponding filters one sample offset from each other, which leads to the birth of DT-CWT [12]. Moreover, the orthogonality of filters guarantees the perfect reconstruction condition of DT CWT.

For 2D signals, DT-CWT can be achieved by separable filtering along columns and then rows, which produces 3 subbands in the first and second quadrants respectively, corresponding to the detailed information along ±15º, ±45º and ±75º directions in the spatial domain. The resultant better orientation selectivity makes it more suitable for directional multi-resolution analysis.

In a word, DT-CWT introduces limited redundancy (2^m:1 for m-dimensional signals comparing to the real DWT) and provides approximate shift invariance and better orientation selectivity while preserving the usual properties of perfect reconstruction and computational efficiency. Its effective applications of image de-noising and texture analysis prove it a powerful tool in signal processing. To the best of our knowledge, however, no related work has been reported on motion estimation with DT-CWT. Experimental results show that improved accuracy can be achieved with the work in this paper in spite of many adverse factors including occlusion and acute motion, etc.

3. Motion estimation based on DT CWT

3.1 Independence of the wavelet subbands

Although the coarse-to-fine search strategy usually adopted by the wavelet-based motion estimation methods can reduce the computational complexity, it can lead to distinct error propagation as the motion vectors at the higher level are obtained by interpolating and refining the ones at the lower level. Furthermore, the frequently used bi-linear interpolation is also quite computationally expensive.

From the frequency domain we can see that all the components of DT-CWT do not overlap with each other, which means the noise within each subband is independent from the rest ones. Therefore, each subband can be processed individually and then integrated to constitute the wavelet coefficients for the compensated frame, and finally the reconstructed image is obtained directly through the inverse transform.

3.2 Inadequateness of a single reference frame

The frequently encountered circumstances, such as scene change, occlusion and entrance of new objects, make the previous frame (called the forward reference frame) insufficient for estimating the current frame, as the boundary and occluded information would disappear, and thus affect seriously the accuracy of motion estimation [16].

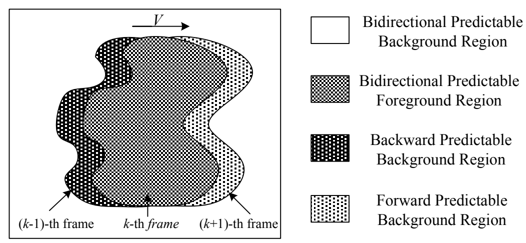


Figure 1 Demonstration of predictable fore-/background regions

Take the simple case of a moving object on the static background for example, as illustrated in Fig. 1, many occluded regions will be produced as the object moves. Assuming the k-th frame is the current frame, then the backward predictable background in Fig. 1 is generally unpredictable from the (k-1)-th frame since it is unobservable in the latter. It is the same case for the forward predictable background for the k-th frame from the (k+1)-th frame. Therefore, the previous and successive frames(called the backward reference frame) are both considered to supplement the lost information.

3.3 Summary of the proposed algorithm

According to the above-mentioned analysis, we can conclude the computation process for the dual-tree complex wavelet-based motion estimation as follows:

Step1 Wavelet decomposition. For the given current, forward- and backward- reference frames, DT-CWT is performed to obtain the scale and wavelet coefficients at the different directions, and abbreviated as W_{cur}, $W_{ref\backslash f}$ and $W_{ref\backslash b}$, respectively.

Step2 Motion estimation. For each subband at the different scales and directions, we carry out the following sub-steps,

Step2.1 Forward motion estimation. For each wavelet block in the current frame, the corresponding block with minimal MAD (Mean Absolute Difference), as shown in equation (1), in the forward reference frame is found with block matching technology and recorded as *forward_mad*.

$$MAD(dx,dy)=\frac{1}{MN}\sum_{(x,y)\in B}\left|D_{cur}(x,y)-D_{ref}(x+dx,y+dy)\right| \quad (1)$$

where $D_{cur}(x,y)$ and $D_{ref}(x,y)$ are the wavelet blocks of the current and reference frames located at (x,y), respectively; $-P\le dx,dx\le P$ and [-P,P] is the search range; B is set of wavelet blocks with size of $M\times N$.

Step2.2 Selection of areas for the backward motion estimation. With the MADs in *forward_mad* arranged in decreasing order, the first certain number of blocks are selected as the re-estimated areas with the given threshold *thresh* (abbreviated as *ReA*).

Step2.3 Backward motion estimation. For the re-estimated areas labeled in step 2.2, block matching is re-applied with the backward reference frame and the corresponding MADs are recorded as *back_mad*.

Step2.4 Motion compensation. For the re-estimated areas, the blocks with minimal MAD from the forward- and backward- reference frame are selected as the corresponding compensated blocks; and for the rest of the blocks, only the forward reference frame is taken into consideration (see equation (2)).

Step3 Reconstructing motion-compensated image. All the compensated subbands are integrated to reconstruct the motion-compensated image through the inverse transform directly.

4. Experimental results and discussion

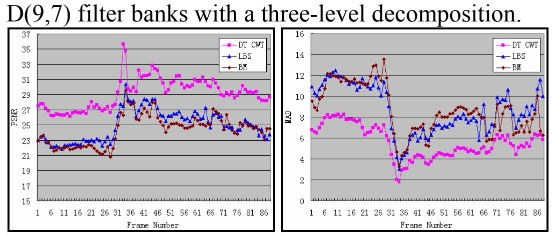
This section shows how the performance of the proposed method compares with two other approaches, block matching based approach with NTSS (BM)[2] and Low Band Shift based approach (LBS)[8], with respect to MAD and PSNR. Several standard video sequences were used for performance comparison, including Football with acute motion (90 frames), Vectra with occlusion (142 frames) and the steady Foreman sequence (300 frames) (all in CIF format, 352×288 pixels).

In experiments, the important parameters were preset as follows: the number of decomposition n=3, where orthogonal near-symmetrical *Fanas* filters for the first level and Q-shift filters for the rest two levels, and the percentage of blocks to be re-estimated *thresh* was set to 0.2. Furthermore, different block sizes and search ranges were adopted as usual and, with the corresponding search strategies, are shown in Table 1.

Table 1. Block size and the corresponding search range and strategy in different levels

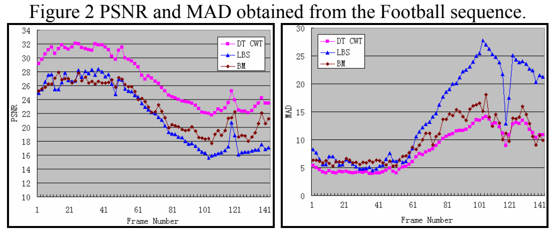
Level	3	2	1
Block size (pixel)	2 × 2	4 × 4	8 × 8
Search range (pixel)	(-2, 2)	(-4, 4)	(-8, 8)
Search strategy	ES	NTSS	NTSS

Comparatively, for BM-based approach, a 16×16-pixel block was taken and the corresponding search range was within [-16,16] with NTSS being the searching strategy. For LBS-based approach, block matching was adopted with parameters the same as BM-based approach, except that the search strategy was ES. Besides, the DWT was performed by the D(9,7) filter banks with a three-level decomposition.



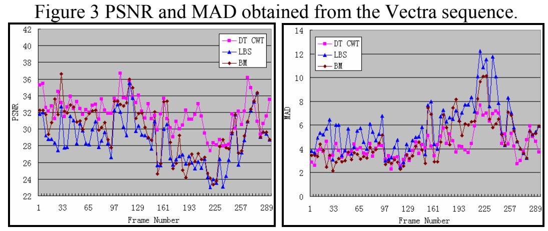
(a) PSNR (b) MAD

Figure 2 PSNR and MAD obtained from the Football sequence.



(a) PSNR (b) MAD

Figure 3 PSNR and MAD obtained from the Vectra sequence.



(a) PSNR (b) MAD

Figure 4 PSNR and MAD obtained from the Foreman sequence.

Figure 2(a) and (b) show the PSNR and MAD from the Football sequence, respectively. DT CWT means Dual-Tree Complex Wavelet Transform, BM means Block Matching and LBS means Low Band Shift, respectively. The sequence can mainly be divided into three stages: several moving objects with acute motion and occlusion (1-30 frames), no moving object with steady background (31-40 frames) and one or two moving objects under more complicated background

$$W_{mc}^{(m,n)}(k)=\begin{cases}W_{ref\backslash f}^{(m,n)}(k), k\notin ReA\cup\left(k\in ReA\cap forward\_mad(k)\le back\_mad(k)\right)\\ W_{ref\backslash b}^{(m,n)}(k), k\in ReA\cap forward\_mad(k)>back\_mad(k)\end{cases} \quad (2)$$

where $W_{mc}^{(m,n)}(k)$, $W_{ref\backslash f}^{(m,n)}(k)$ and $W_{ref\backslash b}^{(m,n)}(k)$ are the wavelet blocks at m-th orientation and n-th scale for the compensated, forward- and backward- reference frame, respectively.

(42-90 frames). It can be seen that, comparing to BM and LBS methods, the proposed method improves remarkably the performance under any of the circumstances, ascribed to the individual processing of each subband for obtaining more detailed information and the introduction of the backward reference frame to supplement the lost information due to occlusion.

In order to illustrate the robustness to occlusion of the proposed method, the PSNR and MAD charts for the Vectra sequence are also shown in Fig 3. The results show that the other two methods are quite sensitive to occlusion, while it can properly be dealt with by the proposed method. This, to a great extent, benefits from the application of the forward- and backward- reference frames, which can supplement the lost information. Fig 4 shows the PSNR and MAD for the Foreman sequence, which proved that DT-CWT based method is also suitable to the less changing circumstances. The mean PSNRs and MADs from the three video sequences are shown in TABLE 2.

Table 2 Mean PSNR and MAD for three sequences

Image Sequence	Method	Average PSNR (dB)	Average MAD (pixel)	Improved PSNR(%)	Improved MAD(%)
Football (352×288)	DT-CWT	29.150	5.5697	—	—
	LBS	24.954	8.6959	16.8	36.0
	BM	23.811	8.4808	22.4	34.3
Vectra (352×288)	DT-CWT	26.958	8.2210	—	—
	LBS	21.931	14.197	22.9	42.1
	BM	22.646	9.8035	19.0	16.1
Foreman (352×288)	DT-CWT	31.949	4.3158	—	—
	LBS	28.472	5.8986	12.2	26.8
	BM	29.591	4.6679	8.0	7.5

5. Conclusion

In this paper, we proposed a new DT-CWT based motion estimation method, taking advantage of its approximate independence of subbands at different levels and directions and better orientation selectivity. Individual processing of each subband enables us to avoid the error propagation and the computational cost of interpolation induced by the traditional coarse-to-fine search strategy; meanwhile, the introduction of the backward reference frame can further improve the detected accuracy of the motion vectors in the case of occlusion and within boundary areas. Experimental results show that, comparing to the LBS and block matching approaches, PSNR can be improved at least 15% (4dB) on average with the proposed method in the cases of acute motion, scene change and occlusion and 2dB for the case of steady sequences.

In the future, we will consider the introduction of some prior information for the images in order to reduce the blur effect in the motion-compensated image, which is one of the challenges for motion estimation in the wavelet domain.

6. Acknowledgement

The work in this paper was supported by Program for New Century Excellent Talents in University (Grant No. NCET-06-0882), National Natural Science Fund (Grant No. 60403008) and Hi-tech 863 Program (Grant No. 2007AA01Z314), P. R. China.

7. References

[1] S. J. Sun, D. Haynor, and Y. M. Kim. Motion Estimation Based on Optical Flow with Adaptive Gradients. In Proc. of ICIP, Canada, Sept. 10-13, 2000:852-855.

[2] R. X. Li , B. Zeng, and L. M. Liou. A Novel Three-Step Search Algorithm for Block Motion Estimation. IEEE Trans. on CSVT, 4(4):438-442, 1994.

[3] V. Christopoulos. A New Predictive Search Area Approach for Fast Block Motion Estimation. IEEE Trans. on Image Processing, 12(6):648-652, 2003.

[4] K. M. Uz, M. Vetterli, and D. LeGall. Interpolative Multiresolution Coding of Advanced Television with Compatible Subchannels. IEEE Trans. on CSVT, 1(1):86-99, 1991.

[5] J. Magarey and N. G. Kingsbury. Motion Estimation Using a Complex-Valued Wavelet Transform. IEEE Trans. on Signal Processing, 46(4):1069-1084, 1998.

[6] X. Li. Scalable Video Compression via Overcomplete Motion Compensated Wavelet Coding. Signal Processing: Image Communication, 19(7):637-651, 2004.

[7] J. W. Zan, M. O. Ahmad and M. N. S. Swamy. Wavelet-based Multiresolution Motion Estimation through Median Filtering. In Proc. ICASSP, Florida, USA, 2002:3273-3276.

[8] H. W. Park and H. S. Kim. Motion Estimation Using Low-Band-Shift Method for Wavelet-Based Moving- Picture Coding. IEEE Trans. Image Processing, 9(4):577-587, 2000.

[9] H. S. Kim and H. W. Park. Wavelet-based Moving-picture Using Shift-invariant Motion Estimation in Wavelet Domain. Sig Pro: Image Com, 16(7):669-679, 2001.

[10] M. F. Fu, O. C. Au and W. C. Chan. Low Band Shift (LBS) Motion Estimation with Symmetric Padding in Wavelet Domain. In Proc. ISCAS, USA, 2002, III: 13-16.

[11] R. Devore, A. Petukhov and R. Sharpley. Motion Estimation with the Redundant Wavelet Transform. Int W. on Digital and Computational video, USA, 2002:53-59.

[12] W. Selesnick, R.G.. Baraniuk, and N. G. Kingsbury. The Dual-Tree Complex Wavelet Transform. IEEE Trans. on Signal Processing, 22(6):123-151, 2005.

[13] N. G. Kingsbury. The Dual-Tree Complex Wavelet Transform-A New Efficient Tool For Image Restoration And Enhancement. In Proc. EUSIPCO 98, Greece, 1998:319-322.

[14] S. Hatipoglu, S. K. Mitra and N. G. Kingsbury. Image Texture Description Using Complex Wavelet Transform. In Proc. ICIP, Canada, September 2000:530-533.

[15] K. P. Lim, A. Das and M. N. Chong. Estimation of Occlusion and Dense Motion Fields in a Bidirectional Bayesian Framework. IEEE PAMI, 24(5):712-718, 2002.

[16] J. Xiao, H. Cheng, H. Sawhney, et al. Bilateral Filtering-Based Optical Flow Estimation with Occlusion Detection. In Proc. of ECCV, Austria, 2006:211-224.