

# Anisotropic Double Cross Search Algorithm using Multiresolution-Spatio-Temporal Context for Fast Lossy In-Band Motion Estimation

Yu Liu and King Ngi Ngan

Department of Electronic Engineering,  
The Chinese University of Hong Kong, Shatin, N.T., Hong Kong, China  
Email: {yuliu, knngan}@ee.cuhk.edu.hk

**Abstract.** In this paper, we make use of the anisotropic property of the motion field in wavelet domain to develop an anisotropic double cross search algorithm using multiresolution-spatio-temporal context (MR-STC-ADCS) for fast lossy motion estimation in shift-invariant wavelet domain. Experimental results demonstrate the superiority of the proposed fast lossy MR-STC-ADCS algorithm against other fast lossy algorithms in terms of speed-up and visual quality.

**Index Terms** Fast motion estimation, anisotropic, shift-invariant wavelet domain

## 1 Introduction

Motion estimation and motion compensation (ME/MC) in wavelet domain [1, 2] has received much attention due to its superior performance by comparing to the conventional ME/MC in spatial domain. ME/MC in wavelet domain is basically free from the blocking effects due to the global nature of wavelet transform. However, ME/MC in critically sampled wavelet domain is very inefficient in high bands because of the shift-variant property of wavelet decomposition. To overcome the shift-variant property of wavelet transform, the low-band-shift (LBS) [3] method and the complete-to-overcomplete discrete wavelet transform (CODWT) [4] method are proposed for ME/MC in shift-invariant wavelet domain. These methods avoid the shift-variant property of the wavelet transform and perform ME more precisely and efficiently. However, a major disadvantage of these methods is the computational complexity which mainly comes from full search algorithm.

Full search algorithm can obtain the optimal result by searching exhaustively for the best matching block within a search window, but its high computational cost limits its practical applications. For this reason, several alternative and faster techniques have been developed, based on lossless and

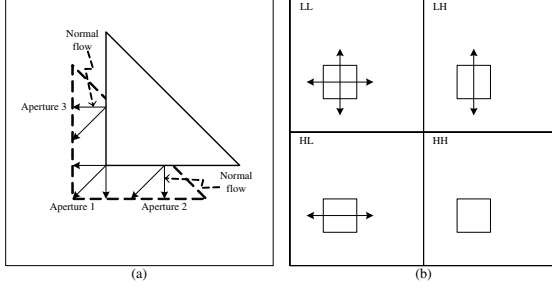
lossy algorithms, that try to reduce computational complexity by reducing the complexity of distortion computation or by pruning the search space. Due to the multi-resolution nature in the wavelet pyramid, many fast lossless and lossy multi-resolution motion estimation algorithms, such as CPME-PDS [5], FMRME [6], have been proposed to reduce the computational complexity. However, these fast lossless and lossy algorithms do not overcome the shift-variant property of the wavelet transform because all these algorithms use the critically sampled subbands of the wavelet reference frame for motion estimation and compensation. Maestroni et al. [7] proposed a fast in-band motion estimation algorithm (FIBME) that works in shift-invariant wavelet domain. However, this approach does not work well in case of large displacement and tends to give a bad estimation when initial coarse MV candidates are not accurate.

In this paper, we make use of the anisotropic property of the motion field in shift-invariant wavelet domain to develop an anisotropic double cross search algorithm using multiresolution-spatio-temporal context (MR-STC-ADCS) for fast lossy motion estimation. Experimental results demonstrate the superiority of the proposed fast lossy MR-STC-ADCS algorithm against other fast lossy algorithms in terms of speed-up and visual quality.

## 2 Background

### 2.1 Motion Estimation in Shift-Invariant Wavelet Domain

In LBS or CODWT, the shift-invariant wavelet coefficient of the  $v$ -pixel-shifted reference frame  $t'$  can be represented by  $S_{t',l,k,v}(i,j) = f_{t',l,k}(dx\%2^l, dy\%2^l, i + \lfloor dx/2^l \rfloor, j + \lfloor dy/2^l \rfloor)$ , where  $l$  denotes the wavelet decomposition level;  $k$  denotes the LL/LH/HL/HH subband type;



**Fig. 1.** (a) Aperture problem in spatial domain, (b) Anisotropic motion model in wavelet domain

$v = \{dx, dy\}$  is the shifting vector or displacement vector;  $\{dx\%2^l, dy\%2^l\}$  indicates the number of shifts in the level  $l$ ; and  $\{i + \lfloor dx/2^l \rfloor, j + \lfloor dy/2^l \rfloor\}$  is the location in the subband  $k$ . And the wavelet coefficient of the current frame  $t$  can be represented by  $S_{t,l,k}(i, j) = f_{t,l,k}(i, j)$ .

For ME/MC in shift-invariant wavelet domain, the coefficients of each wavelet tree rooted in the lowest subband are rearranged to form a wavelet block. The purpose of the wavelet block is to provide a direct association between the wavelet coefficients and what they represent spatially in the image. Related coefficients at all scales and orientations are included in each wavelet block. The wavelet blocks in the search window  $w$  in the reference frame are compared to the current wavelet block, and a reference wavelet block that leads to the best match is selected. The sum of absolute difference (SAD) of the  $p$ th wavelet block for the displacement vector  $v$  is computed as follows:

$$\begin{aligned}
 SAD(p, v) = & \sum_{i=1}^{\frac{N}{2^L}} \sum_{j=1}^{\frac{N}{2^L}} |S_{t,L,0,p}(i, j) - S_{t',L,0,p+v}(i, j)| \\
 & + \sum_{l=1}^L \sum_{k=1}^3 \sum_{i=1}^{\frac{N}{2^l}} \sum_{j=1}^{\frac{N}{2^l}} |S_{t,l,k,p}(i, j) - S_{t',l,k,p+v}(i, j)| \quad (1)
 \end{aligned}$$

The optimum motion vector  $v^*$  of the  $p$ th wavelet block, which has minimum displacement error, is given by:

$$SAD(p, v^*) = \min_{v \in w} SAD(p, v) \quad (2)$$

## 2.2 Anisotropic Motion Model in Wavelet Domain

It has been observed that traditional 2D ME in spatial domain suffers from the aperture problem, due to the ill-posed nature and small observation

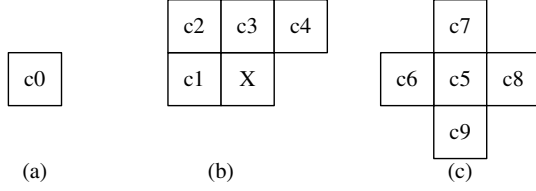
window, as shown in Fig.1(a). Aperture problem implies that the directional features in the spatial domain can only be uniquely determined along the normal flow direction. It is usually thought as a fundamental limitation of the traditional isotropic motion model in spatial domain. However, the aperture problem can in fact be exploited as an advantage of motion estimation in wavelet domain. Since wavelet transform structures the image into the subbands with different orientations, the subbands contain different edges with different normal flow directions. According to this property of wavelet transform, an anisotropic motion field representation in wavelet domain was proposed by Li et al. [8], as shown in Fig.1(b). The anisotropic property of the motion field refers to those attributes of the motion vector (search range and accuracy) that would match the subband orientation. Thus, the original 2D ME problem can be approximated by a 1D ME along the normal flow direction for the vertical/horizontal subbands.

## 3 Proposed Algorithm

### 3.1 Multiresolution-Spatio-Temporal Context

Due to the multiresolution nature of wavelet transform, we can easily obtain a hierarchical motion field at different scales. In the existing algorithms, the motion estimation is first performed in a coarser resolution to determine an initial MV candidate, and then the MV's of finer resolutions are refined based on the motion information obtained at coarser resolution. However, the coarse-scaled MV's for some blocks may not be accurate enough and could cause some errors which propagate along the hierarchical structure. For example, FIBME [7] method uses a coarse full search in subband LL to determine an initial search point. However, this algorithm tends to give a bad estimate of the motion displacement in case of large displacement. Therefore, using only multiresolution context is not enough for reducing the risk of getting trapped into a local minimum.

In order to solve this problem, we propose a solution to obtain smooth and consistent motion vectors by using the motion estimates previously obtained in the multiresolution-spatio-temporal context of a block before starting the search. This is based on the hypothesis that the motion field varies slowly hierarchically, spatially and temporally, and hence it is highly probable that the blocks closer



**Fig. 2.** Context template (a) multiresolution context, (b) spatial context, (c) temporal context

to the current block, in multiresolution, space and time, may show the same motion. With this assumption, instead of calculating all the possible motion vectors as the full search algorithm does, the previous calculated motion vectors that are hierarchically, spatially and temporally contiguous, can be used as the candidate predictors for the current block. Our proposed algorithm exploits the context information from both the hierarchically and spatio-temporally adjacent blocks to select a set of initial MV candidates. As shown in Fig. 2, in the proposed approach, the candidate predictors for the current block consist of one multiresolution context, four spatial contexts, and five temporal contexts. The set of the initial MV candidates can be expressed as  $\{c0, c1, \dots, c9\}$ , where  $c0$  is the initial MV candidate from the corresponding coarser scale;  $c1, \dots, c4$  are initial MV candidates from the left, top-left, top, and top-right spatially adjacent blocks of the corresponding scale in the current frame;  $c5, \dots, c9$  are initial MV candidates from the collocated, left, top, right, and bottom temporally adjacent blocks of the corresponding scale in the previous frame.

The best predictor (the MV with the minimum SAD) among the candidate predictors is sent to the anisotropic double cross search (ADCS) algorithm phase that refines the motion vector.

### 3.2 Anisotropic Double Cross Search Algorithm

The anisotropic motion model in wavelet domain, discussed in Section 2.2, is employed to develop an anisotropic double cross search algorithm. The proposed search algorithm initially considers all possible motion vector predictor candidates from multiresolution-spatio-temporal context and uses the best motion vector predictor candidate  $\{dx_p, dy_p\}$  as the center of the search. The initial best motion vector predictor candidate allows us to reduce the risk of getting trapped into a local minimum.

Then, the search starts from the two different subbands in different routes: one route is from HL subbands to LH subbands, that is, the horizontal displacement is first refined using 1D window searching along the normal flow direction in HL subbands, followed by vertical displacement refinement using 1D window searching along the normal flow direction in LH subbands; another route is from LH subbands to HL subbands, that is, the vertical displacement is first refined using 1D window searching along the normal flow direction in LH subbands, followed by horizontal displacement refinement using 1D window searching along the normal flow direction in HL subbands. During the 1D window searching refinement along the normal flow direction in each subband, only the coefficients in the corresponding subbands and LL subband are computed. Therefore, we define the sum of absolute difference from the corresponding subbands and LL subband as anisotropic SAD (ASAD), as follows:

$$ASAD_{l,h}(dx', dy') = SSAD_{L,LL}(dx', dy') + \sum_{r=l}^L SSAD_{r,h}(dx', dy') \quad (3)$$

$(l \in \{1, \dots, L\}; h \in \{LH, HL, HH\})$

where,

$$SSAD_{l,k}(dx, dy) = \sum_{i=1}^{\frac{N}{2^l}} \sum_{j=1}^{\frac{N}{2^l}} |S_{t,l,k}(i, j) - S_{t',l,k}(i + dx, j + dy)| \quad (4)$$

$(l \in \{1, \dots, L\}; k \in \{LL, LH, HL, HH\})$

For the first search route which is from HL subbands to LH subbands, in the first stage, only the coefficients in HL subbands and LL subband are computed for horizontal motion component  $dx'_{HL}$ :

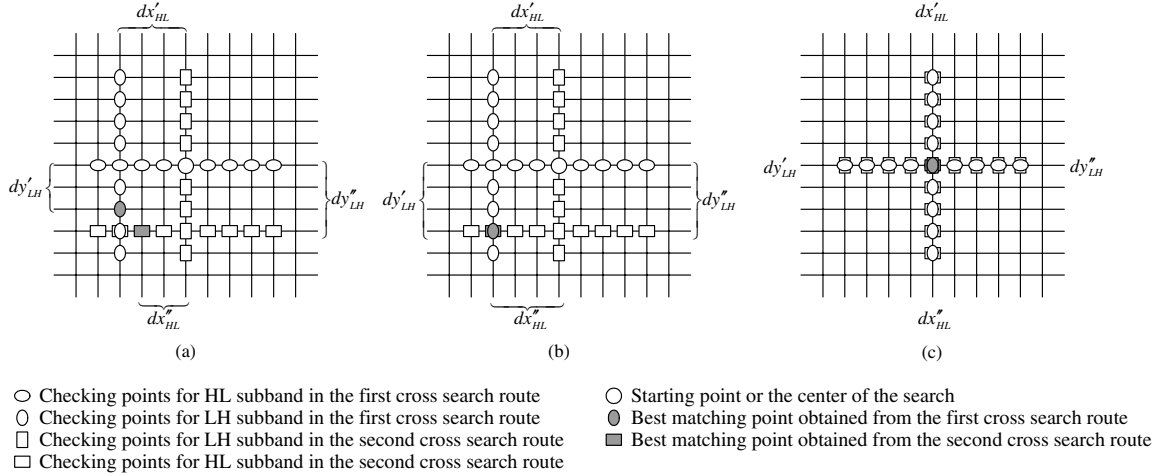
$$dx'_{HL} = \arg \min_{-W_l < dx'_{HL} < W_l} ASAD_{l,HL}(dx_p + dx'_{HL}, dy_p) \quad (5)$$

In order to improve the accuracy of estimation, the second stage takes  $\{dx_p + dx'_{HL}, dy_p\}$  as the starting point. Similarly, only the coefficients in LH subbands and LL subband are computed for vertical motion component  $dy'_{LH}$ :

$$dy'_{LH} = \arg \min_{-W_l < dy'_{LH} < W_l} ASAD_{l,LH}(dx_p + dx'_{HL}, dy_p + dy'_{LH}) \quad (6)$$

For the second search route which is from LH subbands to HL subbands, in the first stage, only the coefficients in LH subbands and LL subband are computed for horizontal motion component  $dy''_{LH}$ :

$$dy''_{LH} = \arg \min_{-W_l < dy''_{LH} < W_l} ASAD_{l,LH}(dx_p, dy_p + dy''_{LH}) \quad (7)$$



**Fig. 3.** Anisotropic double cross search algorithm, (a) Case I, (b) Case II, (c) Case III

Similarly, the second stage takes  $\{dx_p, dy_p + dy_{LH}''\}$  as the starting point and only the coefficients in HL subbands and LL subband are computed for vertical motion component  $dx_{HL}''$ .

$$dx_{HL}'' = \arg \min_{-W_l < dx_{HL}'' < W_l} ASAD_{l,HL}(dx_p + dx_{HL}'', dy_p + dy_{LH}'') \quad (8)$$

As it is highly likely that the target motion vector is within a smaller neighborhood of the predictors with the decreasing decomposition level, the initial search stepsize  $W_l$  is reduced to  $[-W/2^{2+L-l}, W/2^{2+L-l}]$  for the level  $l$ .

As illustrated in Fig.3, these two search routes form an anisotropic double cross search pattern. Therefore, there may exist two best matching points,  $\{dx_p + dx_{HL}', dy_p + dy_{LH}'\}$  and  $\{dx_p + dx_{HL}'', dy_p + dy_{LH}''\}$  for each cross search route. If the two best matching points from those two cross search routes are not the same, as shown in Fig.3(a) Case I, then the coefficients in HH subbands will be used to judge which one is the best matching point. And the winner is selected as the new center of the search; meanwhile the search stepsize keeps unchanged. If the two best matching points are the same but not the center of the search, as shown in Fig.3(b) Case II, then the best matching point is selected as the new center of the search; meanwhile the search stepsize is reduced by half until it is equal to 1. The final step is reached when the two best matching points from each cross search step are both located in the center of the search, as shown in Fig.3(c) Case III.

### 3.3 Analysis of the Computational Complexity

For LL subband or the coarsest scale, there is no candidate predictor from the multiresolution context because LL subband is the coarsest scale. Although the full search can be applied to LL subband because the LL subband has  $2^L \times 2^L$  fewer samples than the whole wavelet block, the speed-up ratio of such algorithm will be limited to the upper-bound speed-up ratio:

$$Speedup = \frac{(2W+1)^2 N^2}{(2W+1)^2 \frac{N^2}{2^L \times 2^L}} = 2^{2L} \quad (9)$$

Instead of full search in LL subband, a coarse full search in critically sampled LL subband is used in the FIBME method to determine an initial search point. According to the analysis of the computational complexity in the FIBME, the speed-up ratio of the FIBME approximates to  $6W$  with respect to the exhaustive full search algorithm.

Unlike the FIBME method, a different search strategy in LL subband is used in the proposed algorithm. Besides the candidate predictors from spatio-temporal context, we also check all the candidate points in the shifted LL subband where the median predictor is located. The best predictor among these candidate predictors is further refined by using diamond search algorithm [9] to obtain the best motion vector in the LL subband. Therefore, the computational complexity of the proposed algorithm in the LL subband is approximate to

$$O_{LL} = \left[ \left( 2 \times \frac{W}{2^L} + 1 \right)^2 + \epsilon_1 \right] \frac{N^2}{2^L \times 2^L} \approx \frac{W^2 N^2}{2^{4L-2}} \quad (10)$$

where  $\epsilon_1$  is the negligible computational cost that comes from the spatio-temporal context and the iterative procedure of diamond search algorithm.

For other scales, all the candidate predictors from the multiresolution-spatio-temporal context will be checked and the best predictor among these candidate predictors is further refined by using the proposed anisotropic double cross search algorithm, instead of diamond search algorithm, to obtain the best motion vector in the corresponding scale. Therefore, the computational complexity of the proposed algorithm in level  $l$  is approximate to

$$O_l = \left[ \left( 2 \times \frac{W}{2^{2+L-l}} \right) \times 4 + \epsilon_2 \right] \frac{N^2}{2^l \times 2^l} \approx \frac{WN^2}{2^{L+l-1}} \quad (11)$$

where  $\epsilon_2$  is the negligible computational cost that comes from the multiresolution-spatio-temporal context and the iterative procedure of anisotropic double cross search algorithm.

Therefore, the speedup of the proposed algorithm with respect to the full search algorithm (FSA) is approximate to

$$Speedup \approx \frac{(2W+1)^2 N^2}{\frac{W^2 N^2}{2^{4L-2}} + \sum_{l=1}^L \frac{WN^2}{2^{L+l-1}}} \approx 2^{L+1} W \quad (12)$$

## 4 Experimental Results

In order to compare the performance of the proposed algorithms with other methods, we use the first 150 frames of a large variety of video sequences for evaluation, including ten QCIF sequences, ten CIF sequences, and four 4CIF sequences. In the experiments, Daubechies 9/7 biorthogonal wavelet filter is employed for the wavelet transform, and the maximum wavelet decomposition level  $L = 3$  is applied. The wavelet block size is 16x16 pixels, while the search windows are [-15, 15] for QCIF, [-31, 31] for CIF, and [-63, 63] for 4CIF, respectively. Simulation results are reported in the following ways:

- 1) *PSNR* for the quality measure between the original and motion-compensated reconstructed frames
- 2) *MAD* for the distortion measure between the original and motion-compensated wavelet frames
- 3) *operation number per block* used to compute the distortion
- 4) *speed-up ratio* gained from the tested fast algorithms compared to FSA.

For comparison, we show in Table 1 the results obtained from the full search algorithm (FSA). To demonstrate the performance of the proposed algorithms, we compare the proposed MR-STC-ADCS algorithm with modified FMRME [6] and

**Table 1.** Experimental Results of Test Video Sequences With Full Search Algorithm

Format	Video	FSA		
		PSNR (dB)	MAD	Operation Number
QCIF (176x144)	Carphone	34.17	2.189	200246
	Claire	42.58	0.767	200246
	Foreman	32.56	2.736	200246
	Hall	36.40	1.755	200246
	Mobile	25.89	7.929	200246
	Mother	39.87	1.188	200246
	Salesman	39.26	1.261	200246
	Stefan	25.14	6.608	200246
	Suzie	35.89	1.841	200246
Tempete	28.04	4.832	200246	
CIF (352x288)	Akiyo	42.57	0.566	871659
	Bus	25.06	6.183	871659
	Coastguard	29.67	4.521	871659
	Container	38.33	1.532	871659
	Flower	25.47	6.345	871659
	Foreman	33.25	2.604	871659
	Mobile	23.86	8.452	871659
	Paris	30.91	2.365	871659
	Tempete	27.11	5.243	871659
Waterfall	35.12	2.719	871659	
4CIF (704x576)	City	31.42	3.816	3632446
	Crew	33.82	2.649	3632446
	Harbour	29.58	4.141	3632446
	Soccer	31.98	2.960	3632446

FIBME [7]. All these algorithms are implemented in shift-invariant wavelet domain. Although the original FMRME algorithm is implemented in critically-sampled wavelet domain, we modified the algorithm to work in shift-invariant wavelet domain for *fair* comparison. Table 2 presents the average PSNR, MAD, operation number per block, and speed-up ratio for different fast lossy algorithms compared to the FSA. On average, for the sequences examined in this test, MR-STC-ADCS is roughly 11.5 and 2.6 times faster whereas its PSNR is approximately 1.46 dB and 0.6 dB higher than the FMRME and FIBME algorithms, respectively; and its MAD is approximately 0.426 and 0.165 lower than the FMRME and FIBME algorithms, respectively. For the cases examined, MR-STC-ADCS is about 271 times faster than FSA for QCIF, 667 times for CIF, and 1313 times for 4CIF, while having an average PSNR loss of only 0.04 dB or an average MAD increase of only 0.018 compared to the FSA. It is obvious from the experimental results that the proposed algorithm is significantly faster than other algorithms while achieving a very close PSNR and MAD performance as that of FSA.

**Table 2.** Experimental Results of Video Sequences With Different Fast Lossy Algorithms

	Video	FMRME				FIBME				MR-STC-ADCS			
		PSNR diff.	MAD diff.	Operation Number	Speed up	PSNR diff.	MAD diff.	Operation Number	Speed up	PSNR diff.	MAD diff.	Operation Number	Speed up
Q C I F	Carphone	-1.34	+0.194	3966	50	-0.38	+0.091	1833	109	-0.02	+0.033	848	236
	Claire	-0.96	+0.037	3955	51	-0.15	+0.009	1812	111	0	+0.001	682	294
	Foreman	-1.17	+0.226	3963	51	-0.88	+0.211	1839	109	-0.05	+0.027	887	226
	Hall	-0.95	+0.079	3957	51	-0.07	+0.015	1801	111	-0.02	+0.003	697	287
	Mobile	-0.98	+0.487	3953	51	-0.11	+0.054	1780	112	0	+0.003	672	298
	Mother	-0.87	+0.073	3953	51	-0.18	+0.030	1795	112	-0.02	+0.005	715	280
	Salesman	-0.83	+0.065	3950	51	-0.13	+0.017	1766	113	-0.02	+0.002	651	308
	Stefan	-0.77	+0.599	3968	50	-0.10	+0.168	1851	108	-0.02	+0.015	840	238
	Suzie	-0.36	+0.081	3961	51	-0.19	+0.071	1823	110	0	+0.001	799	251
Tempete	-1.53	+0.817	3960	51	-0.40	+0.263	1810	111	-0.05	+0.013	679	295	
C I F	Akiyo	-0.62	+0.013	17226	51	-0.50	+0.016	3330	262	-0.02	+0.001	1156	754
	Bus	-1.80	+1.179	17350	50	-1.20	+0.719	3592	243	-0.05	+0.014	1429	610
	Coastguard	-2.63	+0.848	17321	50	-1.07	+0.325	3542	246	-0.03	+0.013	1342	650
	Container	-1.14	+0.107	17256	51	-0.14	+0.019	3404	256	0	+0.002	1220	714
	Flower	-1.74	+0.619	17309	50	-0.33	+0.148	3484	250	-0.01	+0.029	1410	618
	Foreman	-1.32	+0.225	17276	50	-1.25	+0.277	3528	247	-0.09	+0.020	1586	550
	Mobile	-1.66	+1.157	17267	50	-0.37	+0.303	3439	253	-0.01	+0.052	1278	682
	Paris	-1.34	+0.233	17231	51	-0.68	+0.118	3381	258	-0.07	+0.017	1231	708
	Tempete	-1.32	+0.717	17281	50	-0.39	+0.265	3494	249	-0.17	+0.022	1336	652
Waterfall	-1.44	+0.149	17233	51	-0.15	+0.017	3358	260	0	0	1182	737	
4 C I F	City	-2.66	+0.649	68573	53	-1.17	+0.249	7123	510	-0.01	+0.038	2585	1405
	Crew	-2.23	+0.392	68547	53	-1.70	+0.271	7487	485	-0.10	+0.053	3344	1086
	Harbour	-4.06	+1.054	68489	53	-2.07	+0.373	7116	510	-0.05	+0.024	2597	1399
	Soccer	-2.23	+0.671	68594	53	-1.83	+0.364	7166	507	-0.15	+0.023	2664	1364

## 5 Conclusion

In this paper, we proposed a fast lossy MR-STC-ADCS algorithm which is based on the anisotropic property of motion field in shift-invariant wavelet domain. Experimental results demonstrate the superiority of the proposed fast lossy MR-STC-ADCS algorithm against other fast lossy algorithms in terms of speed-up and visual quality.

## References

1. S. Zafar, Y.Q. Zhang, B. Jabbari, "Multiscale video representation using multiresolution motion compensation and wavelet decomposition," *IEEE Journal on Selected Areas in Commun.*, Vol.11(1), pp.24-35, Jan. 1993
2. S. Kim, T. Aboulnasr, and S. Panchanathan, "Adaptive multiresolution motion estimation techniques for wavelet-based video coding," *Proc. of SPIE Visual Commun. Image Process.*, vol. 3309, pp. 965-974, Jan. 1998
3. H.W. Park and H.S. Kim, "Motion estimation using low-band-shift method for wavelet-based moving-picture coding," *IEEE Trans. Image Process.*, Vol.9(4), pp.577-587, Apr. 2000
4. Y. Andreopoulos, A. Munteanu, G. Van der Auwera, J.P.H. Cornelis, P. Schelkens, "Complete-to-overcomplete discrete wavelet transforms: theory and applications," *IEEE Trans. Signal Process.*, Vol.53(4), pp.1398-1412, Apr. 2005
5. K.Y. Wong, W.C. Siu, K.C. Hui, "Fast motion estimation for wavelet-based video coding," *Proc. of Int. Symp. Intelligent Multimedia, Video and Speech Processing*, 2004, pp.398-401, Oct. 2004
6. M.K. Mandal, E. Chan, X. Wang and S. Panchanathan, "Multiresolution motion estimation techniques for video compression," *Optical Engineering*, Vol.35, pp.128-136, 1996
7. D. Maestroni, A. Sarti, M. Tagliasacchi, S. Tubaro, "Fast in-band motion estimation with variable size block matching," *Proc. of Int. Conf. Image Process.*, 2004. Vol. 4, pp.2287-2290, Oct. 2004
8. X. Li and S. Lei, "Efficient motion field representation in the wavelet domain", *Proc. of Int. Conf. Image Process.* 2002, vol.3, pp.257-260, June 2002
9. S. Zhu and K.K. Ma, "A new diamond search algorithm for fast block matching motion estimation", *Proc. of Int. Conf. Information, Communications and Signal Processing*, vol.1, pp.292-296, 1997