COMPLEX DISCRETE WAVELET TRANSFORM BASED MOTION ESTIMATION FOR VISION-BASED TRACKING OF TARGETS

Sener Yilmaz[1] and Mete Severcan[2]

[1]Image Processing Department, MGEO Division, ASELSAN Inc.
PO. Box: 30 Etlik, 06011, Ankara, Turkey
phone: + (90) 312-847-5300, fax: + (90) 312-847-5320, email: sener@mgeo.aselsan.com.tr
web: www.aselsan.com

[2]Department of Electrical and Electronics Engineering, The Middle East Technical University
Inonu Bulvari, 06531, Ankara, Turkey
phone: + (90) 312-210-2332, fax: + (90) 312-210-1261, email: severcan@metu.edu.tr
web: www.metu.edu.tr

ABSTRACT

For tracking applications, the estimation of the "true" motion vector is crucial. The Complex Discrete Wavelet Transform (CDWT) based motion estimation algorithm developed by Magarey and Kingsbury produced superior results for the estimation of the dense flow field. In this work, the use of the CDWT-based motion estimation algorithm for the vision-based tracking of targets has been evaluated. First, a comparison of the results of the CDWT-based ME algorithm with the results of the Lucas and Kanade's (LK) and Horn and Schunk's (HS) motion estimation algorithms is performed. Second, the tracking performances are compared for the cases of CDWT-based and LK-based flow fields. Lastly, the tracking performance of the proposed tracker is evaluated by using a number of test sequences and is compared to the Correlation and Mean Shift Tracker. It is observed that it can successfully track various different targets and is robust to changes of the target signature.

1. INTRODUCTION

Optical flow estimation methods have been used for tracking in several ways. The most popular flow estimation methods are the Lucas and Kanade's [1], the Horn and Schunk's [2], and the Fleet and Jepson's [3] methods. An extensive review of the performances of the optical flow methods can be found in [4].

In the recent years, the Discrete Wavelet Transform (DWT) has been used for motion estimation in a number of ways. Since the DWT is shift-variant [5,6,7,8], it cannot be directly used for motion estimation. Several modifications have been proposed to make the DWT shift-invariant. In [6], the Redundant Discrete Wavelet Transform (RDWT) is used. The RDWT is shift-invariant since the spatial sampling rate is fixed across scale. It is also called the "undecimated wavelet transform." Due to the lack of decimation it is highly redundant. In [7], the Overcomplete Discrete Wavelet Transform is proposed to overcome the shift-variant property of the DWT. These two methods, however, provide only invariance for integer-shifts. In [8], the low redundant Complex, directional Double-Density Wavelet Transform (CDDWT) is proposed.

In this work, the use of the Complex Discrete Wavelet Transform (CDWT) based motion estimation method is evaluated and used for target tracking. The CDWT is developed by Magarey and Kingsbury [9] for efficient phase-based motion estimation. The CDWT is approximately shift-invariant and has good directional selectivity. The CDWT-based motion estimation algorithm is robust and provides sub-pixel accuracy which is important for tracking.

The rest of this paper is organized as follows. Section 2 briefly reviews the CDWT-based motion estimation algorithm. Section 3 presents the proposed tracker. Section 4 gives the simulation results. We end with a discussion of the results attained in this work.

2. CDWT MOTION ESTIMATION

The Complex Discrete Wavelet Transform (CDWT) based motion estimation algorithm is developed by Magarey and Kingsbury [9] for the efficient implementation of the accurate and robust phase-based optical flow estimation method of Fleet and Jepson [3]. A detailed description of the method can be found in [9].

2.1 The Complex Discrete Wavelet Transform (CDWT)

The CDWT is similar to the DWT, but uses complex-valued kernels and has a mirror branch, which, on the overall results in a 4 to 1 redundancy. A 1-d filter pair is used which is a complex 4-tap filter that can be modelled as Gabor filters. The transform uses a dual-track structure in which the complex conjugates of the filter pair is used for row filtering in the lower track. In the first level, a prefilter is used to assure that the wavelet filters are exactly scaled versions of each other. The result is six complex detailed subimages and two lowpass images. The structure of the CDWT is shown in Figure 1.

2.2 CDWT-based Motion Estimation

The motion estimation algorithm has a hierarchical structure and proceeds from coarse to fine resolution levels. At each level, motion is estimated for each subpel and the resultant flow field is propagated to the next resolution level by scaling the flow vectors and warping the transform coefficients of the reference image accordingly. For estimating the motion of each subpel, a quantity called the "subband squared difference" is minimized. This quantity is obtained by the sum of absolute differences of the values of the subpels in the six detailed subimages. This corresponds to a quadratic surface whose minimum gives the desired displacement. These surfaces are accumulated through the levels to obtain the "cumulative squared difference". The result of the algorithm is a real-valued motion estimate for each pixel in the image. The structure of the motion estimation algorithm is shown in Figure 2.

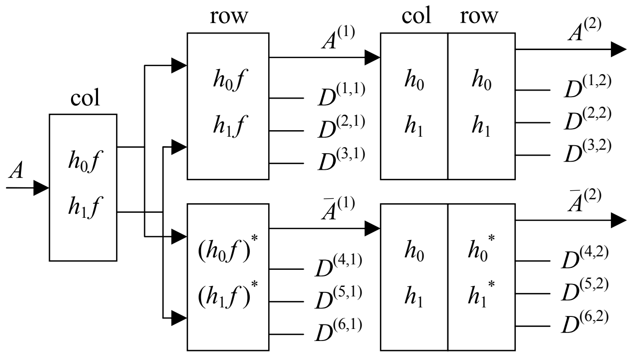

Figure 1. 2-D CDWT (2 levels shown).

3. THE TRACKING ALGORITHM

The proposed tracking algorithm uses the dense flow field generated by the CDWT-based motion estimation algorithm. The aim is to track any kind of target selected by the operator. The target can be rigid or non-rigid and can change pose, size and shape during tracking. Camera motion can also be present in the sequences. The usage of the flow field is therefore suitable to cope with all these problems.

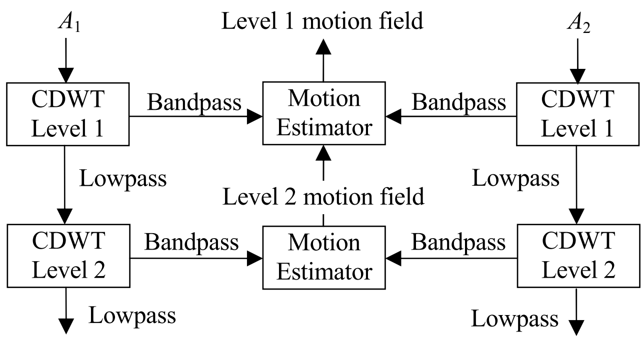

Figure 2. CDWT-based motion estimation algorithm.

In the proposed tracking algorithm, the dense flow field within the target gate is evaluated and used in computing the displacement of the target. This flow information is also used for adapting the track gate to the changes of the target size. The accuracy of the optical flow estimator is therefore crucial for the success of the tracking algorithm. This is the main reason why the CDWT-based motion estimator is selected for this purpose.

4. SIMULATION RESULTS

Simulations have been performed to test different aspects of the algorithm. First, the quality and suitability of the flow field generated by the CDWT-based motion estimation algorithm have been evaluated. Second, simulations replacing the CDWT-generated flow in the tracking algorithm with the Lucas and Kanade's flow are performed. Lastly, the proposed tracking algorithm is compared with the Correlation Tracker and the Mean Shift Tracker [10].

4.1 Optical Flow Estimation Results

The accuracy of the flow field is evaluated by using virtual and real image sequences presented in [4]. The flow field is compared with Lucas and Kanade's (LK) and Horn and Schunk's (HS) flow estimation algorithms.

4.1.1 Error Measures

Two error measures have been used. The angular measure [4] is defined by

$$\theta_E = \arccos\left(\frac{uu_c + vv_c + 1}{\sqrt{u^2+v^2+1}\sqrt{u_c^2+v_c^2+1}}\right) \quad (1)$$

and the magnitude measure is defined by

$$m_E = \left|\sqrt{u^2+v^2} - \sqrt{u_c^2+v_c^2}\right|. \quad (2)$$

Here, u and v represent the components of the flow vector and the subscript c indicates the components of the correct flow.

4.1.2 Test Sequences

One virtual and one real image sequence will be presented in this paper. The sequences are shown in Figure 3.

Figure 3. Yosemite (left) and SRI (right) Test Sequences.

Yosemite Sequence: The motion of the clouds is translational and 2 pixels to the right. The rest of the flow is divergent, with speeds of about 5 pixels per frame in the lower left corner. The correct flow is shown in Figure 4 (upper-left).

SRI Sequence: The motion is translational in the fronto-parallel plane. The camera translates parallel to the ground plane, perpendicular to its line of sight, in front of clusters of trees. Velocities are as large as 2 pixels per frame.

4.1.3 Results

The comparison results of the Yosemite sequence is given in Table 1. The dense flow fields generated by the algorithms are shown in Figure 4.

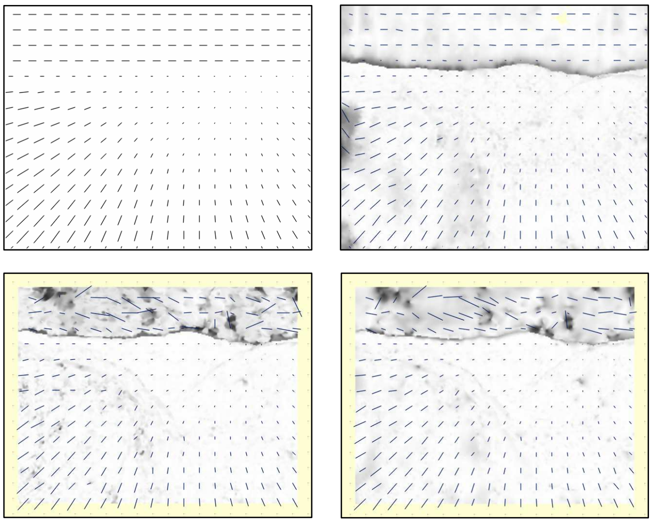

Figure 4. Yosemite Sequence flow diagrams. Correct flow (top-left), CDWT-based ME (top-right), LK (bottom-left), and HS (bottom-right).

In Table 1, the threshold column indicates the value used for thresholding out the flow vectors having confidence less than this value. The ratio of the remaining ones to the total number of pixels is given as the density. The errors are computed on this thresholded pixels only. The number of frames used for computing the flow is also given in the table. The last row gives the error of the CDWT-based method for the pixels at locations thresholded according to the LK method with threshold 1.0.

Table 1. Yosemite Sequence Error Results.

Method	Threshold	Density (%)	Angle Error	Magnitude Error	No. of Frames
LK	1.0	32,2	4,4875	0,0952	15
	0.0	81,0	9,9308	0,4968	15
HS	5.0	26,6	5,4825	0,1407	15
	0.0	81,0	9,5685	0,4548	15
CDWT-based ME	0.95	49,1	6,8730	0,2390	2
	0.00	92,8	7,7314	0,2950	2
	LK	32,1	6,2745	0,1378	2

Investigation of the table reveals that, although the CDWT-based method is not as accurate as the LK, taking into account the whole image, i.e. for zero thresholds, it is observed that the CDWT-based method is more accurate. This can also be seen by the flow diagrams shown in Figure 2. The angle errors are also shown as the background of the flow diagrams as grey levels where darker regions indicate more error. In the upper region where there is only cloud motion, the CDWT-based flow is highly smooth and accurate where the other algorithms produced irrelevant results.

For the SRI sequence, the flow fields generated by the CDWT-based method and the LK are shown in Figure 5. This is a real sequence, therefore no correct flow information is present.

Investigation of the flow fields reveals that the flow generated by the CDWT-based method is more accurate and smooth than the one generated by LK's method.

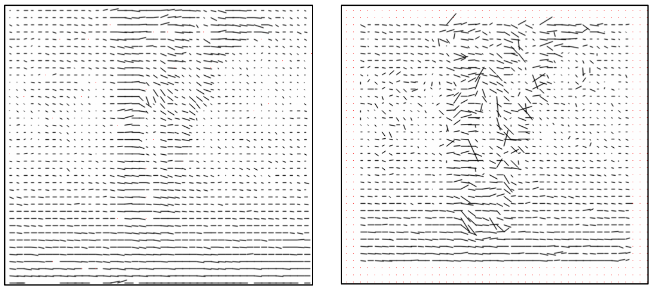

Figure 5. SRI Sequence flow diagrams. CDWT-based ME (left) and LK (right).

4.2 Tracking Results

The suitability of the flow generated by the CDWT-based motion estimation method is evaluated in three different ways. Firstly, the flow method, secondly, the proposed tracking algorithm is evaluated. In the last part, other properties of the tracking algorithm is evaluated.

4.2.1 Flow Comparison

The first comparison is performed by changing only the flow estimation method while leaving the rest of the tracking algorithm the same. The LK algorithm is used for this comparison. Two results will be presented here. Representative frames of the sequences are shown in Figure 6.

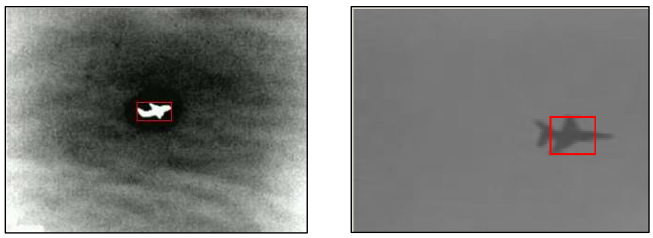

Figure 6. Pursaklar-2 (left) and Air32 (right) Sequences.

The first sequence includes severe camera motion while aiming the plane. The tracking algorithm maintained track until the 295th frame while using the flow field generated by the CDWT-based method and until the 165th frame while using the flow field generated by LK.

In the second sequence, there is limited camera motion, but there are extensive maneuvers taken by the target. For this sequence, the tracker maintained track until the 531st frame using the flow field generated by the CDWT-based method and until 347th frame while using the one generated by LK.

4.2.2 Track Comparison

The second evaluation is performed by comparing the track performance of the proposed CDWT-based tracking algorithm with the Correlation Tracker and the Mean Shift Tracker [10].

The sequence presented here is shown in Figure 7. There is slight camera motion throughout the sequence and the car at the centre of the frame moves away from the camera.

The CDWT-based tracking algorithm maintained track until the 725th frame, the Correlation Tracker until the 39th, and the Mean Shift Tracker until the 371st frame. Since target motion is very small compared to the distance between pixels, the target drifts from the gate of the Correlation Tracker. The similarity of the background clutter statistics to the target statistics makes it a difficult case for the Mean Shift Tracker. The CDWT-based tracker, however, follows the flow information and maintains track successfully.

Figure 7. Pursaklar-1a Sequences.

4.2.3 *CDWT-based Tracker Evaluation*

The proposed tracking algorithm is based on the dense flow field which provides valuable information for the tracker. The changes occurring on the target are monitored using this information. Figure 8 demonstrates such a case.

The car in the sequence is approaching towards the camera. Slight camera motions are also present. The initial target gate can be seen in the left frame and the final gate can be seen in the right frame. It is clearly seen that the target gate has been adapted appropriately. This feature is very important for target tracking.

Figure 8. The 1st (left) and 736th (right) frames of the Pursaklar-1b Sequence. Demonstration of target size adaptation of the CDWT-based tracking algorithm.

5. CONCLUSION

In this paper, the use of the CDWT-based motion estimation algorithm for target tracking purposes has been evaluated. The dense flow field generated by the CDWT-based motion estimation algorithm has been compared with the Lucas and Kanade's and Horn and Schunk's flow estimation methods. Both virtual and real sequences have been used for this comparison. The CDWT-based method, although not as precise as the others, has produced a denser and smoother flow field than the other two methods. Especially for regions where only low frequency components were present, the CDWT-based method was the only one producing reliable results.

A tracking algorithm using the flow generated by the CDWT-based motion estimation algorithm is also proposed and evaluated by changing the flow estimation method with the Lucas and Kanade's algorithm. For most of the sequences the performances were similar. However, in some cases, using the Lucas and Kanade's flow resulted in better tracking while in others using the CDWT-based flow resulted in a better track.

The proposed tracking algorithm using the CDWT-based generated flow has also been compared with the Correlation Tracker and the Mean Shift Tracker. It is observed that the proposed tracker were successful in maintaining track for highly cluttered environments where the target shape is also changing dramatically.

Among the other properties of the proposed tracker, the most important one is the use of the information contained in the flow field which is utilized for adapting to target changes.

Taking into account all these aspects and evaluations the CDWT-based motion estimation method proved to be suitable for vision-based target tracking. The proposed tracking algorithm making use of the CDWT-based motion estimator is successful in tracking vision-based targets.

REFERENCES

[1] B. Lucas and T. Kanade, "An Iterative Image Registration Technique with an application to Stereo Vision," *Proc. of 7th Int. Conf. on Artifi. Intell.*, pp. 674-679, Auğ 1981.

[2] B. K. P. Horn and B. G.. Schunck., "Determining Optical Flow," 1980.

[3] D. J. Fleet and A. D. Jepson, "Computation of Component Image Velocity from Local Phase Information," *Int. J. of Computer Vision*, vol. 5, no. 1, pp. 77-104, 1990.

[4] J. L. Barron, D. J. Fleet, and S. S. Beauchemin, "Performance of Optical Flow Techniques," *lnt. J. of Computer Vision*, vol. 12, no. 1, pp. 43-77, 1994.

[5] N. G. Kingsbury, "Complex Wavelets and Shift Invariance," *Proc. of IEE Colloquium on Time-Scale and Time-Frequency Analysis and Applications*, 2000.

[6] S. Cui, Y. Wang, and J. E. Fowler, "Mesh-Based Motion Estimation and Compensation in the Wavelet Domain using a Redundant Transform," *Proc. of the IEEE Conf. on Image Processing*, vol. 1, pp. 693-696, 2002.

[7] N. Sebe, C. Lamba, and M. S. Lew, "An Overcomplete Discrete Wavelet Transform for Video Compression," *ICME*, 2002.

[8] F. C. A. Fernandes, R. L. C. v. Spaendonck, and C. S. Burrus, "A New Framework for Complex Wavelet Transforms," *IEEE Trans. on Signal Processing*, 2002.

[9] J. Magarey and N. Kingsbury, "Motion Estimation Using a Complex-Valued Wavelet Transform," *IEEE Trans. on Signal Processing*, vol. 46, no. 4, pp. 1069-1084, 1998.

[10] D. Comaniciu, V. Ramesh and P. Meer "Real-time Tracking of Non-rigid Objects using Mean Shift," *IEEE Conf. on Computer Vision and Pattern Recognition*, vol. 2, pp. 142-149, 2000.