EFFICIENT MOTION ESTIMATION UTILIZING QUADRATURE FILTERS

Atanas Boev, Chavdar Kalchev, Atanas Gotchev, Tapio Saramäki and Karen Egiazarian

Institute of Signal Processing, Tampere University of Technology
P.O.Box 553, FI-33101 Tampere, Finland
firstname.familyname@tut.fi

ABSTRACT

This contribution introduces a computationally-efficient scheme for phase-based motion estimation. The local phase for consecutive dyadic scales and six different directions is retrieved through a complex-valued subband decomposition. It is obtained by a successive use of a recursive Hilbert transformer and recursive power-complementary half-band filter pairs. The so-called approximately linear-phase recursive half-band filter proposed by Renfors and Saramäki is used as a start-up filter for generating both the Hilbert transformer and the half-band filter pairs. Experiments with synthetic image sequences demonstrate that by properly designing the start-up filter, the proposed technique provides, with a considerably reduced number of computations, a performance similar to that in a recently introduced method.

1. INTRODUCTION

Motion estimation (ME) is the problem of determining correspondences or motion vectors in a sequence of images. It is a key image processing topic having many applications including, among others, video compression, stereo-optics, and surveillance. ME is often used in the analysis of moving three-dimensional (3D) objects. For example, it is applied in depth-from-motion methods by extracting dissimilarity information between multi-camera views [1]. In addition, ME is used in 3D object representations realized by depth ordering [2] and/or disparity compensation [3]. Another wide field of the 3D object analysis by ME is the so-called structure from motion [4]. It extracts 3D morphological information by analyzing the corresponding motion vectors.

In general, the image sequence (e.g. video) is formed by projecting 3D scenes onto the 2D plane of an imaging sensor at a certain time interval. The resulting 2D "optical flow" is often the only input available to the algorithm for estimating the "real" 3D motion [5].

Reconstructing the motion of a 3D object based on its 2D projection is an ill-posed inverse problem and for solving it some constraints should be applied [6]. The type of the constraints underlines the classes of ME algorithms. These classes include, for instance, *gradient-based, energy-based, block-matching*, and *phase-based*

algorithms [7]. A number of studies have favoured the use of phase-based ME as it copes better with illuminance changes and affine deformations in image sequences [7], [8]. The phase-based approach relies on equiphase surfaces and their evolution along time for getting the true motion field. The local phases are determined as the outputs of a set of complex-valued non-overlapping spatiotemporal Gabor filters with various scales and orientations [8]. The reported results are quite accurate at the price of a high computational burden caused by the convolutions with the Gabor filters. A substantial step for increasing the efficiency of algorithms of this kind was done by Magarey and Kingsbury in [9]. This improvement is based on the use of a complex-valued discrete wavelet transform (CDWT) that exploits short-length Gabor-like filters. In this approach, not only the local phase is obtained in a more efficient manner but also a hierarchical structure allowing multiresolution refinements is incorporated within the ME algorithm.

This contribution adopts the methodology from [9] and aims at further improving the filter structures for realizing the complex-valued subband transform. Our preliminary research, focused on a fixed-point DSP realization of ME, indicated that the most promising structures are based on the use of a combination of a Hilbert transformer and real-valued half-band filter pairs [10]. This earlier work is extended by properly designing such filters and filter pairs using various filter orders and by comparing the resulting performances in terms of the ME accuracy and computational cost. This paper is organized as follows: Section 2 briefly reviews the phase-based ME approach, as developed in [9]. In Section 3, a general complex-valued subband decomposition scheme is presented. It adheres to the adopted motion model as good as the CDWT does while offering better implementation flexibility. This section also deals with the design of half-band filters involved in the structure. Experiment with synthetic image sequences with known motion fields are included in Section 4, followed by conclusions made in Section 5.

2. MOTION ESTIMATION BASED ON CDWT

2.1. Motion model

Given a sequence of images (frames) u_i for $i = 0, 1, 2,\ldots$ at position $\mathbf{x}$, the *local translation model* [9]

$$u_{i-1}[\mathbf{x} + \mathbf{d}(\mathbf{x})] = u_i(\mathbf{x}) \qquad (1)$$

assumes that the changes in u_i are caused by the 2D projected motion only and not by changes in illumination. In Equation (1), $\mathbf{d}(\mathbf{x})$ is the translation to be estimated.

2.2. CDWT Algorithm

The core of the algorithm, as shown in Figure 1, is a multiscale subband transform using complex-valued Gabor-like wavelet filters. Every scale gives six complex-valued subbands discriminating the image into a set of the following six angular orientations: ± 15, ± 45 and ± 75 degrees. The efficiency is achieved by using short FIR filters (4 or 8 taps) and by decomposing the image into dyadic scales.
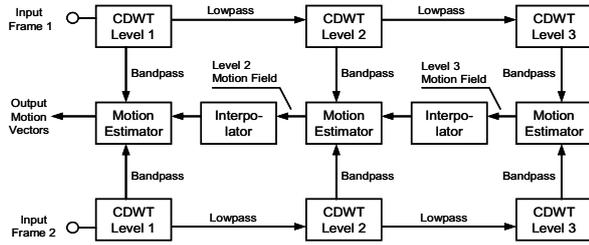
Figure 1. CDWT for ME [9].

The multiscale nature of the algorithm makes it suitable for a hierarchical, coarse-to-fine motion analysis.

A planar phase model expressed by an approximate relation between the subband component and its shifted version is used in this method as follows:

$$D^{(l,m)}(\mathbf{l} + \mathbf{f}) \approx D^{(l,m)}(\mathbf{l})e^{j\theta(\mathbf{f})}, \qquad (2)$$

where $D^{(l,m)}(\mathbf{l})$ denotes the complex subband pixel values in the l-th subband at scale m. The phase term is represented by $\theta(\mathbf{f}) = 2^m \left(\Omega^{(l,m)} \right)^T \mathbf{f}$, where Ω is a matrix of the center frequencies of the wavelet filter pairs. The following equiphase equation has to be met in the case of two time consecutive frames:

$$\phi_1^{(l,m)}(\mathbf{l} + \mathbf{f}) = \phi_2^{(l,m)}(\mathbf{l}), \qquad (3)$$

where ϕ_1 and ϕ_2 are the phases of the corresponding subband coefficients from Frame1 and Frame2, respectively, as shown in Figure 1. In order to solve the equation for $\mathbf{f}$, the phase term has to be expressed as follows:

$$2^m \left(\Omega^{(l,m)} \right)^T \mathbf{f} = \theta^{(l,m)}(\mathbf{l}), \qquad (4)$$

where

$$\theta^{(l,m)}(\mathbf{l}) = \angle \left[\frac{D_2^{(l,m)}(\mathbf{l})}{D_1^{(l,m)}(\mathbf{l})} \right]. \qquad (5)$$

Here, $D_1^{(l,m)}(\mathbf{l})$ and $D_2^{(l,m)}(\mathbf{l})$ are the subband coefficients from Frame1 and Frame2, respectively. This estimation is performed by the *Motion Estimator* blocks in Figure 1.

In order to generate a standard 8x8 block motion field, the subband transform is applied up to the level $m_{max} = 3$. The estimated displacements have a range of 4 pixels in both the vertical and horizontal coordinates for each 8x8 pixels block. Closed form expressions for computing the subband motion vectors are available in [11].

The *Interpolator* blocks interpolate the offset vector size by a factor of 2 in both directions in order to prepare the motion field for usage in the finer scale, if required [9].

3. COMPLEX-VALUED SUBBAND TRANSFORMS

The filters in the CDWT are of Hardy type, i.e., they remove the negative frequencies. Thus they produce an analytic signal at their outputs.

Another scheme, offering more flexibility and simplicity in the filter design is implied by the so-called mapping (projection)-based complex wavelet transform [12]. In this approach, the signal mapping onto the Hardy space is first implemented, followed by a real-valued discrete wavelet transform (DWT). In practice, the mapping is realized by an approximation based on the use of a Hilbert transformer [12]. As in the CDWT, the output subbands produced are complex-valued and discriminate image features in the same angular directions. However, the scheme is more general as it enables one to use arbitrary half-band filters provided that wavelet properties are not required specifically. This adds some freedom in the filter design, especially if a fixed-point implementation is envisaged. The projection stage realized by the Hilbert transformer is also flexible. It can be realized by modulating an arbitrary half-band filter realized as a parallel connection of two allpass filter components [12].

3.1. Recursive half-band filters with approximately linear phase

The key interest in this paper is to generate both Hilbert transformers and half-band filter pairs in the proposed scheme for phase-phased motion estimation based on the use of the so-called approximately linear-phase recursive half-band filters proposed in [13]. The main motivation for using these filters is that they compare favorably with conventional linear-phase finite-impulse response (FIR) filters. Using these filters, a power-complementary filter pair consisting of a lowpass filter transfer function $H_{Lp}(z) = (1/2) [z^{-(n-1)} + A(z^2)]$ and a highpass transfer function $H_{Lp}(z) = (1/2) [z^{-(n-1)} - A(z^2)]$ can be realized as shown in Figure 2(a). Here, n is an even integer and $A(z^2)$ is an allpass filter of order n. It is worth pointing out that this $A(z^2)$ can be generated by replacing z^{-1} in an allpass filter transfer function $A(z)$ of order $n/2$ by z^{-2}. Furthermore, when implementing $A(z)$ as a cascade of first-order and second-order all-pass sections using the same structures as for wave digital filter (WDF) structures [14], $A(z^2)$ requires only $n/2$ multipliers. For this reason, the filters in Figure 2(a) are referred to as WDF filters. The corresponding WDF structure for the Hilbert transformer, as shown in Figure 2(b), is simply obtained by replacing z^{-1} in $A(z^2)$ by $-jz^{-1}$ resulting in $A(-z^2)$. This only changes the signs of certain coefficients of the allpass function. The Hilbert transformer and the half-

band filter pairs used in the proposed scheme for phase-phased motion estimation are generated based on the same lowpass filter of Figure 2(a). They have the following two basic advantages over their linear-phase FIR filter counterparts. First, they have much lower implementation complexities for providing the same or higher frequency selectivity. Second, these building blocks for the overall proposed scheme are very suitable for real-world realizations because of their high stability, high dynamic range, and low sensitivity to the coefficient quantization. Moreover, many power-efficient realization strategies for VLSI designs have been developed [1]. They are mainly targeted at applications demanding low-power designs used, e.g., in mobile devices.

A single stage of the subband transform is shown in Figure 2(c). For clarity, it has been named Complex wave digital filtering (CWDF). In the two-dimensional case, the filtering by rows and columns gives the following outputs: LL, LH, HL, and HH, where L and H stand for the low and high frequencies, respectively. The two-dimensional Hilbert transformer gives an analytic signal which is then decomposed into 8 imaginary and 8 real subbands. The 6 complex subbands, namely LH+, HL+, HH+, LH-, HL-, and HH- in Figure 2(c) are obtained by properly combining the real and imaginary outputs. These subbands contain the desired details extracted from the image.
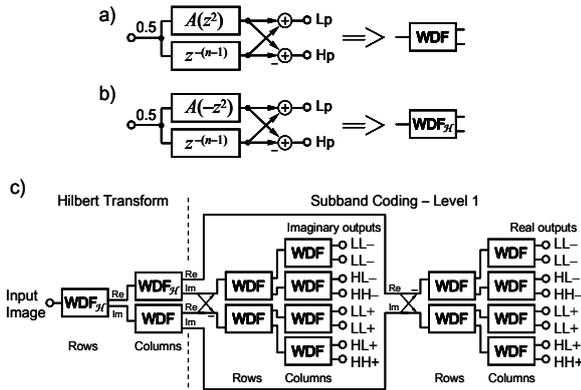
Figure 2. Complex wave digital filtering: (a) Half-band filter pair. (b) Hilbert transformer. (c) The complete CWDF transform.

3.2. Filter design

The design of the start-up filter for both the Hilbert transformer and the half-band filter pair can be accomplished very fast by using the design technique described in [13]. It enables one to solve very fast the following problem: Given n in the structure of Figure 2(b) for the lowpass filter and its stopband edge angle $\omega_s > \pi/2$, minimize the attenuation in the resulting stopband region. Based on the use of this algorithm, start-up lowpass filters with n varying from 4 to 12 have been designed in order to compete with the 8-tap Gabor-like filter and the 4-tap rotation invariant (RI) filter as described in [11]. The stopband edge that appeared to be the most appropriate for our application was $\omega_s = 0.55\pi$. Finally, for each direction we obtain the 6 filter kernels

shown in Figure 3. The magnitude responses of the filters down to level 4 are shown in Figure 4.
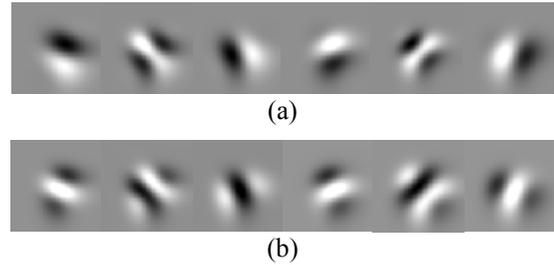
Figure 3. Two-dimensional WDF kernels. (a) Real part. (b) Imaginary part.
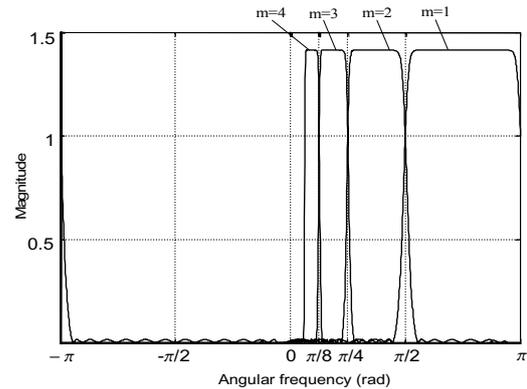
Figure 4. Magnitudes of the one-dimensional WDF frequency responses. The subband level is denoted by *m*.

4. EXPERIMENTAL RESULTS

When comparing the performances of various filter structures, the experiments were performed only for the coarsest level motion estimation case, even though a structure producing better results in *non-refining* algorithm will also lead to a better accuracy in the *refining* case.

Experiments were performed with three synthetic image sequences from the University of Western Ontario database - "Translating tree", "Diverging tree", and "Yosemite". Each sequence has the frame size of 128×128 pixels, the length is 20 frames, and for each sequence 19 true motion fields were provided for reference.

Two ME algorithms were compared in terms of the computational complexity and estimation accuracy. For the CDWT-based ME, two filter pairs were involved, namely, the 4-tap Rotation Invariant (RI) pair and the 8-tap Gabor-like pair.

The computational complexity was estimated both in terms of the number of basic arithmetic operations and the number of instruction cycles. The basic arithmetic operations under consideration were additions and multiplications. Table 1 shows the results for the number of basic arithmetic operations as well as their relative value compared to the 8-tap filter Gabor-like pair taken as a reference.

The estimation accuracy was evaluated by applying the angular measure of error [11]. This measure uses the angular difference between the two 3D vectors which

connect the current frame block to its true position and its estimated position in the next frame. Therefore, the error was measured in degrees.

Table 1. Computational complexity for different sub-band decompositions

Transform Algorithm		Calculations			
		Additions	Multiplications	Total	%
CDWT	8-tap Gabor-like	335872	761856	1097728	100
	4-tap RI	167936	380928	548864	50
CWDF: WDF filters with $\omega_s = 0.55\pi$	$n = 12$	622080	207360	829440	75
	$n = 10$	518400	172800	691200	62
	$n = 8$	414720	138240	552960	50
	$n = 6$	311040	103680	414720	37
	$n = 4$	207360	69120	276480	25

The errors obtained for all motion blocks were averaged. A strip of 17 pixels was excluded from the border to avoid errors caused by boundary effects. The mean error estimate was averaged again in the temporal domain for the 19 motion fields. Table 2 shows a comparison of the errors measured for the algorithms tested.

Table 2. Motion estimation accuracy in terms of the mean angular error [in degrees]

Transform Algorithm		"Translating tree"	"Diverging tree"	"Yosemite"
CDWT	8-tap Gabor-like	3.66	6.74	10.7
	4-tap RI	6.42	8.91	13.3
CWDF: WDF filters with $\omega_s = 0.55\pi$	$n = 12$	3.82	6.38	11.2
	$n = 10$	5.95	8.13	12.7
	$n = 8$	7.10	10.6	13.6
	$n = 6$	8.35	13.2	15.0
	$n = 4$	9.02	13.3	17.7

As seen from the tables, the WDF filters compete successfully with those in the CDWT, especially when the filter order increases. Thus, a lower cost scheme providing a good performance is achievable.

5. CONCLUSIONS

A phase-based scheme for ME has been proposed. It combines the following two ideas: a) obtaining the information about the local phase through a multiscale and multidirectional complex-valued transform resembling the highly acclaimed Gabor transform and b) involving half-band quadrature filters with good frequency characteristics in this transform. When seeking for the best filter structure, both the subband filters and Hilbert transformer were implemented using WDF structures with the same order and they were designed to have a very good stopband attenuation while also having a very good passband behavior. The resulting structure ensures fast processing and a stable performance. A further improvement can be expected if the Hilbert transformer is improved at the price of slightly more computations.

6. REFERENCES

[1] A. Alatan and L. Onural, "Estimation of depth fields suitable for video compression based on 3-D structure and motion of objects," *IEEE Trans. Image Processing*, vol. 7, no. 6 , June 1998

[2] L. Bergen and F. Meyer, "A novel approach to depth ordering in monocular image sequences," *Proc. IEEE Int. Conf. Computer Vision and Pattern Recognition,* vol.2, pp. 536-541, June 2000,

[3] X. Tong and R. Gray, "Interactive rendering from compressed light fields," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 13, no. 11, pp. 1080-1091, Nov. 2003.

[4] T. Jebara, A. Azarbayejani, and A. Pentland, "3D structure from 2D motion," *IEEE Signal Processing Magazine*, vol. 16., no. 3, pp. 66-84, May 1999.

[5] D. Fleet, *Measurements of Image Velocity*. Kluwer Academic Publishers, 1992.

[6] B. Horn and B. Schunk, "Determining optical flow," *Artificial Intelligence,* vol. 17, no. 1, pp. 185-203, August 1981.

[7] J. Barron, D. Fleet, and S. Beauchemin, "Performance of optical flow techniques," *Int. Journal Computer Vision*, vol. 12, no. 1, pp. 43-77, January 1994.

[8] D. Fleet and A. Jepson, "Computation of component image velocity from local phase information," *Int. Journal Computer Vision*, vol. 5, pp. 77-104, 1990.

[9] J. Magarey and N. Kingsbury, "Motion estimation using a complex-valued wavelet transform," *IEEE Trans. Signal Processing*, vol. 46, no. 4, pp. 1069-1084, April 1998.

[10] C. Kalchev, A. Boev, A. Gotchev, K. Egiazarian, and J. Astola, "Motion estimation algorithms based on complex halfband filters for OMAP platform," , *Proc. SPIE*, vol. 5684, to appear.

[11] J. Magarey, *Motion Estimation using Complex Wavelets.* PhD dissertation, Cambridge University, U.K., February 1997.

[12] F. Fernandes, I. Selesnick, R. van Spaendonck, and C. Burrus, "Complex wavelet transforms with allpass filters," *Signal Processing*, vol. 83, no. 8, pp. 1689-1706, August 2003.

[13] M. Renfors and T. Saramäki, "Recursive Nth-band digital filters – Part I: design and properties", *IEEE Trans. Circuits and Systems*, vol. 34, no. 1, pp. 24-39, Jan. 1987.

[14] A. Fettweiss, "Wave digital filters: theory and practice," *Proc. IEEE*, vol. 74, no. 2, pp. 270-327, February 1986.

[15] J. Chung and K. Parhi, *"Pipelined Lattice and Wave Digital Recursive Filters"*, Kluwer Academic Publishers, Boston, May 1996.