33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

Affine Optical Flow Combined with Multiscale Image Analysis for Motion Estimation of the Arterial Wall from B-mode Ultrasound

Aimilia Gastounioti, *Student Member*, *IEEE*, Nikolaos N. Tsiaparas, Spyretta Golemati, *Member*, *IEEE*, John S. Stoitsis, *Member*, *IEEE*, and Konstantina S. Nikita, *Senior Member, IEEE*

Abstract—**This paper investigated the performance of affine optical flow (AFOF) in motion tracking of the arterial wall from B-mode ultrasound images and the effect of its combination with multiscale image analysis on the accuracy of the process. Multiscale AFOF (MAFOF) exploits the information obtained with AFOF from the approximation sub-images at different spatial resolution levels of the images, obtained using a 2D discrete wavelet transform. Both AFOF and MAFOF were evaluated through their application to synthetic image sequences of the common carotid artery. Multiscale image analysis increased the accuracy in motion tracking, with MAFOF yielding average displacement error reductions of 9% with respect to AFOF. The methods were also effectively applied to real ultrasound image sequences of the carotid artery. The results showed that MAFOF could be considered as a reliable estimator for arterial wall motion from B-mode ultrasound images.**

I. INTRODUCTION

Arterial wall motion during the cardiac cycle can be estimated from B-mode ultrasound by recording image sequences and subsequently applying a motion estimation (ME) algorithm. The two methods previously used for ME of the arterial wall are block matching (BM) and optical flow (OF) [1]. A comparative study of these methods was carried out by applying them to real B-mode ultrasound image sequences of the carotid artery [2] and, although peaks in waveforms produced with the two techniques occurred at the same time points, the shapes of the waveforms were different. Therefore, it was suggested to validate these algorithms, along with more sophisticated versions (e.g. affine OF (AFOF)), by means of simulated data experiments [1]. In another study [3] BM and Kalman filter-based extensions of BM were evaluated on noise-free and noisy synthetic image sequences, but the methods were not effective enough for increased noise levels.

Wavelet-based multiscale image analysis has recently emerged as a promising technique for several image processing tasks due to its flexibility in providing a unifying framework for decomposing images into their elementary constituents across scale and its ability to adapt to changing local image statistics and high background noise. This approach has been efficiently used for texture classification of atherosclerotic tissue from B-mode ultrasound [4]. Multiscale image analysis has also been combined with ME to improve video compression through the optimization of the ME process [5].

In this work, AFOF, which is an improved version of OF, was combined with wavelet-based multiscale image analysis. The proposed method, which will be referred to as multiscale AFOF (MAFOF), is based on the decomposition of the images of a sequence using a two-dimensional (2D) discrete wavelet transform (DWT) and it exploits the motion information obtained from different resolution levels of the images to produce an accurate estimation about the radial and longitudinal displacements of selected targets. AFOF and MAFOF were evaluated on both synthetic and real ultrasound image sequences of the common carotid artery.

II. METHODOLOGY

The combination of multiscale image analysis with ME was inspired by the multiresolution ME (MRME) scheme proposed by Zhang and Zafar [5]. In a MRME scheme, motion is first estimated at the lowest spatial resolution and the obtained information is then manipulated as the prediction at finer spatial resolutions. In the following, the basic principles of AFOF and DWT-based image decomposition are briefly described, and the proposed multiscale ME scheme is then presented.

A. Affine Optical Flow

The initial hypothesis of OF is that image intensities are approximately constant over motion for at least a short duration, which is a reasonable assumption for high frame rates. OF relies on first-order spatiotemporal derivatives of image intensities [6] and its accuracy increases when the intensity spatial gradient is not too small and varies from point to point. It estimates a pixel's velocity using the pixels in a local neighborhood, assuming that they share the same velocity. For these pixels it performs a least-squares minimization [6] with a low-pass weighted function that gives more weight to the pixels closer to the target.

AFOF is an improved version of OF, where velocity is not constant in the local neighborhood. In this case, velocity is parameterized by a six-dimensional vector and it permits not only translation, but also rotation and scaling of the local neighborhood across time [7].

In this study, 12×11 pixels local neighborhoods were used and a 2D Gaussian distribution was selected as a low-pass weighted function. Additionally, taking into

A. G. receives a scholarship from the Hellenic State Scholarships Foundation for Ph.D. Studies.

A. G., N. N. T., J. S. S., and K. S. N. are with the Biomedical Simulations & Imaging Laboratory, School of Electrical and Computer Engineering, National Technical University of Athens, Athens, Greece (e-mail: gaimilia@biosim.ntua.gr, ntsiapar@biosim.ntua.gr, stoitsis@biosim.ntua.gr, knikita@ece.ntua.gr).

S. G. is with the 1st Intensive Care Unit, Medical School, National & Kapodistrian University of Athens, Athens, Greece (e-mail: sgolemati@med.uoa.gr).

consideration that derivatives are strongly affected by noise, a preliminary step of image smoothing with a 2D Gaussian was implemented before the computation of spatiotemporal derivatives.

B. Discrete Wavelet Transform

The DWT of a signal is defined as its convolution with a set of lowpass $H[n]$ and highpass $G[n]$ half-band filters, followed by downsampling by two, where $G[n]=(-1)^{1-n}H[1-n]$ satisfy the following conditions among scaled versions of the functions $\varphi_{j,n}$ and $\psi_{j,n}$, $n \in Z$:

$$\phi_{j+1,t} = \sum_n H[n-2t] \cdot \phi_{j,n} \quad (1)$$

$$\psi_{j+1,t} = \sum_n G[n-2t] \cdot \phi_{j,n} \quad (2)$$

The functions $\varphi_{j,t}$ and $\psi_{j,t}$ consist of versions of the prototype scaling φ and wavelet ψ functions, discretized at level j and at translation t. They form an orthonormal set of vectors, a combination of which can completely define the signal allowing its analysis in a multiresolution scheme [8].

The 2D DWT of an image is defined as two successive DWTs, firstly on the rows of the image and then on the columns of the resulted image. The decomposition of the image yields four sub-images at the first level (j=1), namely an approximation sub-image A_j and the horizontal, vertical, and diagonal detail sub-images Dh_j, Dv_j, and Dd_j (Fig. 1). Each sub-image is the result of a convolution with two half-band filters; two lowpass filters for A_j, a lowpass and a highpass for Dh_j, a highpass and a lowpass for Dv_j, and two highpass filters for Dd_j. At the next, and each subsequent, level, only the approximation sub-image is further decomposed into new four sub-images. The total number of levels L depends on the size of the original image; the maximum value of L is equal to min (log_2N, log_2M), where N is the number of rows and M is the number of columns.

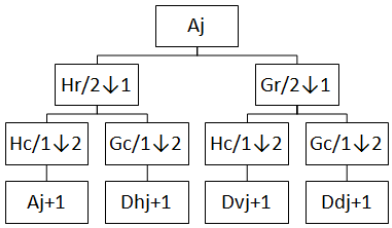

Fig. 1. Schematic diagram of the 2D DWT decomposition scheme for a given level of analysis. Note that, for j=0, A0 is the original image. Hr, Hc, Gr, Gc are the lowpass and highpass filters on the rows and columns of each subimage. The symbols '2↓1' and '1↓2' denote the downsampling procedure on the columns and rows, respectively.

Among a large number of wavelet families (Haar, Daubechies, symlets, coiflets and biorthogonal) Haar was used in this study due to its orthogonality and symmetry properties. Orthogonal filters conserve energy and maintain the same amount of energy noise at each level of decomposition. Symmetric filters are also important because they do not affect the output of the signal.

C. Multiscale Affine Optical Flow

MAFOF uses only one type of sub-images of each image, and it consists of the following steps (Fig. 2).

1) DWT at L levels: The images of a sequence are decomposed up to L levels using a 2D DWT. The lowest decomposition level corresponds to the highest spatial resolution level (original images), whereas the highest decomposition level corresponds to the lowest spatial resolution level (Fig. 3).

2) Target selection: A pixel is selected as a target in the first image of the sequence of original images. The initial position of the target at lower spatial resolution levels is computed as follows: Considering that the size of the sub-images at level j is twice the size of sub-images at level (j+1), the position of a target at level j can be found by scaling its position in the original image by a factor of 2^{-j}.

Steps (3)-(5) are repeated for every subsequent image of the sequence:

3) AFOF at level j=L: AFOF is initially performed at the highest decomposition level L and the radial, rad_L, and longitudinal, $long_L$, positions of the target are estimated.

4) Coarse-to-fine transition: For every lower decomposition level j, with $0 \leq j < L$, it is assumed that the positions of the target, rad_j and $long_j$, are twice the estimated positions, rad_{j+1} and $long_{j+1}$, at the previous (higher decomposition) level. Afterwards, measurements of rad_j and $long_j$, namely $zrad_j$ and $zlong_j$, are obtained by performing AFOF at level j. The final estimates are produced by the average of these two approaches, as shown in (3).

$$\begin{bmatrix} rad_j \\ long_j \end{bmatrix} = 0.5*2*\begin{bmatrix} rad_{j+1} \\ long_{j+1} \end{bmatrix} + 0.5*\begin{bmatrix} zrad_j \\ zlong_j \end{bmatrix} \quad (3)$$

This step finally, i.e. when j=0, leads to rad_0 and $long_0$ which correspond to the final motion estimates at the lowest decomposition level of the current image.

5) Update lower spatial resolution levels: Following the procedure described in step (2), the final motion estimates rad_j and $long_j$ for $0 < j \leq L$ are updated by appropriately scaling the positions rad_0 and $long_0$. This step is necessary for the implementation of steps (3-4) for the next image, because the weighted window is centered on the position of the target in the previous image.

III. Synthetic Data Experiments

MAFOF was optimized and, both AFOF and MAFOF were validated by applying them to four synthetic 87-image sequences of the common carotid artery, corresponding to three cardiac cycles. The first synthetic sequence (S_0) was created by distorting a real ultrasonic B-mode image according to a mathematical motion model [9]. Two additional sequences were created by adding noise levels of 25 decibels (S_{25}) and 15 decibels (S_{15}), respectively, to the first sequence. The fourth synthetic sequence (S_F) was constructed using the Field II software package and the same mathematical motion model [9]. Fig. 4 presents examples of the first images of the synthetic sequences and the corresponding first level approximation and detail sub-images. 176 pixels for sequences S_0, S_{25}, and S_{15}, and 196 pixels for the sequence S_F, were selected as targets. Performance was assessed by means of the warping index for total displacements [3], which represents an overall estimate of the error for all interrogated targets and all images of the sequence.

The optimization procedure first included the type of sub-images, which was used in MAFOF (Table I). This investigation showed that MAFOF produced the highest warping indices when the vertical or the diagonal detail sub-images were used. No considerable differences were observed in the performance of MAFOF, when either the approximation or the horizontal detail sub-images were selected. However, performance was maximized in the first case and, as a result, MAFOF was implemented using the approximation sub-images. The experimentation with different values of L, varying between 1 and 3, then showed that best results were achieved for L=1. The selected parameters were independent of the noise levels, and can be held constant in motion tracking of the arterial wall.

According to Table I, the warping indices increased with increasing noise levels for both AFOF and MAFOF. Additionally, all versions of MAFOF produced lower errors than AFOF for all sequences, which suggests that multiscale image analysis increased the accuracy of ME independent of the selected sub-images. The most effective version of MAFOF, i.e. when the approximation sub-images were used and L=1, yielded error reductions of 7%, 7%, 11%, and 10%, compared to AFOF, for the sequences S_0, S_{25}, S_{15}, and S_F, respectively. The statistical significance of these results was validated through Wilcoxon rank-sum tests (p-values: 0.028, 0.015, 0.019, and 0.003 for S_0, S_{25}, S_{15}, and S_F, respectively).

In terms of the computational cost, tracking one target in an 87-image sequence required 30 s for AFOF and 45 s for MAFOF, using a Pentium(R) Dual-Core CPU T4400 at 2.20 GHz.

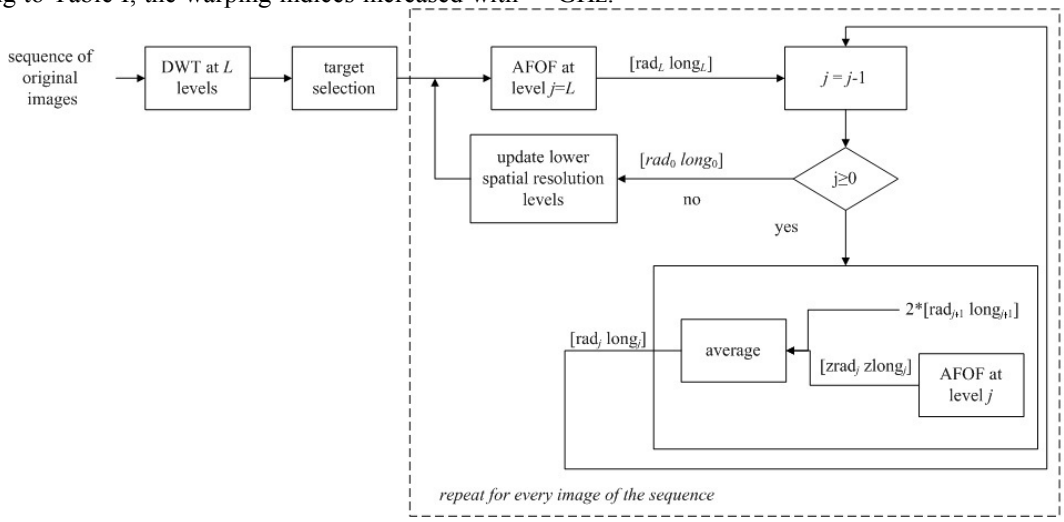

Fig. 2. Block diagram of MAFOF.

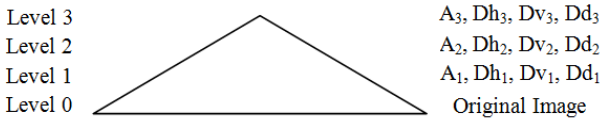

Fig. 3. The pyramid structure of DWT image decomposition.

IV. Real Data Experiments

AFOF and MAFOF (as optimized above) were applied to two real ultrasound image sequences of the carotid artery; one of a young normal subject and one of an elderly patient with an atherosclerotic plaque on the posterior wall. The sequences were recorded with an ATL (Advanced Technology Laboratory) Ultramark 4 Duplex scanner at a rate of 25 frames/s for approximately 4 s (2-3 cardiac cycles). The first frame of each sequence is shown in Fig. 5 (a, c).

TABLE I
WARPING INDICES FOR TOTAL DISPLACEMENTS IN PIXELS FOR AFOF AND DIFFERENT VERSIONS OF MAFOF FOR THE SEQUENCES S0, S25, S15, AND SF

	S_0	S_{25}	S_{15}	S_F
AFOF	1.07	2.02	2.81	3.25
MAFOF (A)	0.99	1.88	2.50	2.91
MAFOF (Dh)	1.01	1.88	2.52	2.93
MAFOF (Dv)	1.04	1.95	2.57	3.01
MAFOF (Dd)	1.06	1.98	2.62	3.08

Fig. 5 (b, d) illustrates examples of radial and longitudinal displacements for two blocks ('ANT' and 'POST') located in opposite wall-lumen interfaces (Fig. 5 (a, c)), using the two methods. The displacements in each direction were calculated by subtracting the positions of the block at end diastole, which was identified from the minimum radial distance between ANT and POST.

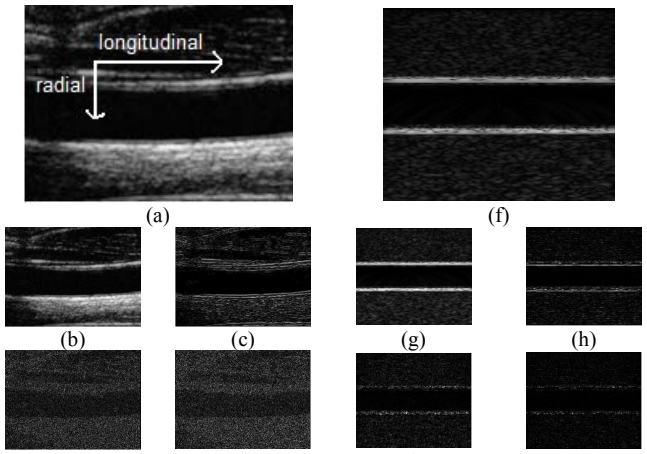

Fig. 4. Examples of images of the common carotid artery wall in synthetic sequences (a) S_{25} and (f) S_F. First level (b, g) approximation and (c, h) horizontal, (d, i) vertical, and (e, j) diagonal detail sub-images of the synthetic images.

Both methods detected the periodic vessel motion, especially in the radial direction, which is caused by the

periodic function of the heart. As we can see, the two methods produce very similar waveforms, probably because of low noise levels where their performances do not significantly differ.

V. Discussion

This study investigated the performance of AFOF in motion tracking of the arterial wall from B-mode ultrasound images and the effect of multiscale image analysis on the accuracy of the process. According to the synthetic data experiments, MAFOF was more accurate than AFOF, which could be associated with the fact that the wavelet-based multiresolution approach provides information about motion structures at different resolutions and scales [5]. The resolution is linked to frequency content, whereas scale to mapping of large or small objects of the image. As the level of decomposition increases the original image passes through a series of low-pass and high-pass filters and its resolution is reduced. However, scale is increased due to downsampling. Thus, increased levels of decomposition refer to sub-images where objects are relatively large (large scale) and their motion can be detected.

However, the optimization process showed that the approximation sub-images were more suitable to MAFOF, probably because they result from the convolution with two lowpass filters. Consequently, they contain a large percentage of the total energy of the original image, maintaining a large amount of the information which is necessary for the ME process. The convolution with lowpass filters also causes the reduction of noise levels, because both Gaussian and speckle noise (which may exist in an ultrasound image) are high-frequency components of the images. On the other hand the detail sub-images, which result from the convolution with one or two highpass filters, lack information and they mainly maintain fine texture, edges and noisy characteristics of the original image. This is more obvious in the diagonal detail sub-images which result from the convolution with two highpass filters.

The optimization process also showed that performance was maximized when the images were decomposed up to one level. This might be associated with (a) the limitation (at that level) of the shift-variant property of the DWT because of the decimation process (lowpass filtering followed by downsampling), and (b) the reduction of the information which is maintained in higher decomposition levels.

In the context of real data experiments, AFOF and MAFOF produced the expected periodic waveforms when applied to real ultrasound image sequences of the carotid artery, which, along with the synthetic data experiments, enhanced the reliability of both methods.

In conclusion, wavelet-based multiscale image decomposition combined with AFOF proved to be an effective computational tool for motion analysis of the arterial wall from B-mode ultrasound. Further investigations using additional wavelet families in combination with alternative shift-invariant wavelet decomposition schemes would be interesting future perspectives in this line of work.

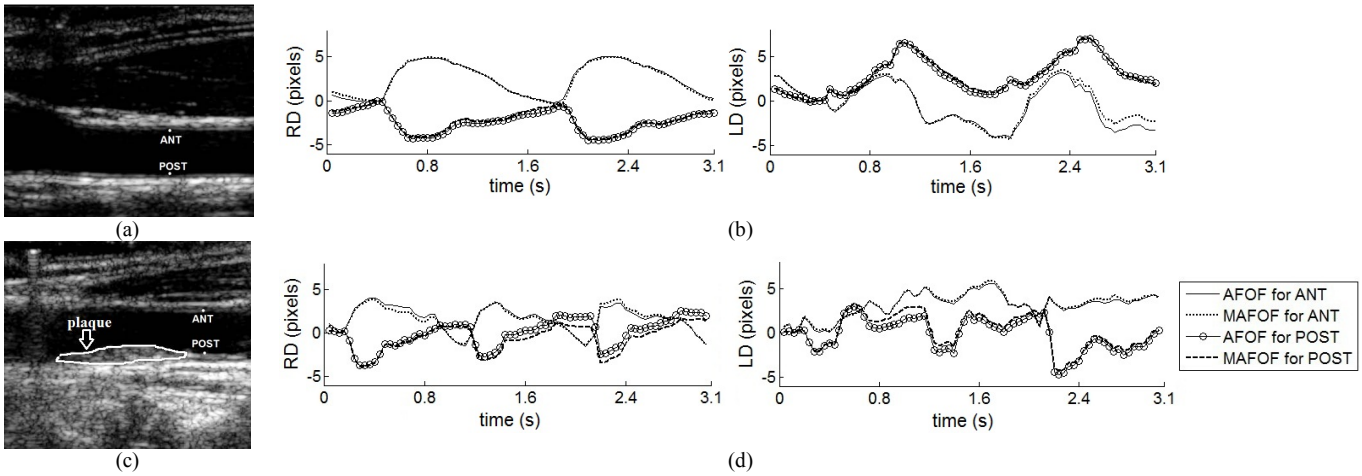

Fig. 5. (a, c) The first frame of an ultrasound image sequence of (a) a young normal subject and (c) an elderly patient with an atherosclerotic plaque on the posterior wall. (b, d) Examples of radial (RD) and longitudinal (LD) displacements of ANT and POST for (b) the young normal subject and (d) the elderly patient, using AFOF and MAFOF.

References

[1] S. Golemati, J. S. Stoitsis, and K. S. Nikita, "Motion analysis of the carotid artery wall and plaque using B-mode ultrasound," *Vasc. Disease Prev.*, vol. 4, no. 4, pp. 296–302, Nov. 2007.

[2] J. Stoitsis, S. Golemati, A. K. Dimopoulos, K. S. Nikita, "Analysis and quantification of arterial wall motion from B-mode ultrasound images: Comparison of block matching and optical flow," *Conf. Proc. IEEE Eng. Med. Biol. Soc. 2005*, vol. 5, pp. 4469–4472.

[3] A. Gastounioti, S. Golemati, J. S. Stoitsis, and K. S. Nikita, "Kalman-filter-based block matching for arterial wall motion estimation from B-mode ultrasound," *Conf Proc. IEEE IST 2010*, pp. 234–239.

[4] N. N. Tsiaparas, S. Golemati, I. Andreadis, J. S. Stoitsis, I. Valavanis, and K. S. Nikita, "Comparison of multiresolution features for texture classification of carotid atherosclerosis from B-mode ultrasound," *IEEE Trans. Inf. Technol. Biomed.*, vol. 15, no. 1, pp. 130–137, Jan. 2011.

[5] Y. Zhang, and S. Jafar, "Motion-compensated wavelet transform coding for color video compression," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 2, no. 3, pp. 285–296, Sept. 1992.

[6] S. S. Beauchemin, and J. L. Barron, "The computation of optical flow," *ACM Computing Surveys*, vol. 27, no. 3, pp. 433–467, Sept. 1995.

[7] D. Fleet, and Y. Weiss, *Handbook of Mathematical Models in Computer Vision*, Springer, Heidelberg, pp. 246–247, 2005.

[8] I. Daubechies, *Ten Lectures on Wavelets,* Society for Industrial and Applied Mathematics, Pennsylvania, 1992.

[9] J. S. Stoitsis, S. Golemati, V. Koropouli, and K. S. Nikita, "Simulating dynamic B-mode ultrasound image date of the common carotid artery," *Conf Proc. IEEE IST 2008.*