Determining the optical flow using wavelet coefficients (Proceedings Paper)

Author(s): **A. Corghi; Riccardo Leonardi; J. M. Corridoni; Alberto Delbimbo**

Date: **10 January 1997**

ISBN: **9780819424358**

PDF
Member: **$18.00** | Non-member: **$18.00**

Hard Copy
Member: **$24.00** | Non-member: **$24.00**

Add to Cart

Proceedings Vol. 3024

Visual Communications and Image Processing '97, Jan Biemond; Edward J. Delp III, Editors, pp.315-1800
Date: **10 January 1997**
ISBN: **9780819424358**

Paper Abstract

The optical flow (OF) can be used to perform motion-based segmentation or 3D reconstruction. Many techniques have been developed to estimate the OF. Some approaches are based on global assumptions; others deal with local information. ALthough OF has been studied for more than one decade, reducing the estimation error is still a difficult problem. Generally, algorithms to determine the OF are based on an equation, which links the gradient components of the luminance signal, so as to impose its invariance over time. Therefore, to determine the OF, it is usually necessary to calculate the gradient components in space and time. A new way to approximate this gradient information from a spatio- temporal wavelet decomposition is proposed here. In other words, assuming that the luminance information of the video sequences be represented in a multiresolution structure for compression or transmission purposes, we propose to estimate the luminance gradient components directly from the coefficients of the wavelet

transform. Using a multiresolution formalism, we provide a way to estimate the motion field at different resolution levels. OF estimates obtained at low resolution can be projected at higher resolution levels so as to improve the robustness of the estimation to noise and to better locate the flow discontinuities, while remaining computationally efficient. Results are shown for both synthetic and real-world sequences, comparing it with a non multiresolution approach.

Out of print 11/14/07.

DOI: 10.1117/12.263244
Current SPIE Digital Library subscribers click here to download this paper.

DETERMINING THE OPTICAL FLOW USING WAVELET COEFFICIENTS

A. Corghi & R. Leonardi,

Signals & Communications Lab., Dept. of Electronics for Automation, *
University of Brescia, I-25123, E-mail: leon@bsing.ing.unibs.it

J.M. Corridoni & A. Delbimbo,

Dept. of Systems and Computer Science,
University of Florence, I-50139, E-mail:delbimbo@aguirre.ing.unifi.it

ABSTRACT

The optical flow (OF) can be used to perform motion-based segmentation or 3-D reconstruction. Many techniques have been developed to estimate the OF. Some approaches are based on global assumptions; others deal with local information. Although OF has been studied for more than one decade, reducing the estimation error is still a difficult problem. Generally, algorithms to determine the OF are based on an equation, which links the gradient components of the luminance signal, so as to impose its invariance over time.[1] *Therefore, to determine the OF, it is usually necessary to calculate the gradient components in space and time. A new way to approximate this gradient information from a spatio-temporal wavelet decomposition is proposed here. In other words, assuming that the luminance information of the video sequences be represented in a multiresolution structure for compression or transmission purposes, we propose to estimate the luminance gradient components directly from the coefficients of the wavelet transform. Using a multiresolution formalism, we provide a way to estimate the motion field at different resolution levels. OF estimates obtained at low resolution can be projected at higher resolution levels so as to improve the robustness of the estimation to noise and to better locate the flow discontinuities, while remaining computationally efficient. Results are shown for both synthetic and real-world sequences, comparing it with a non multiresolution approach.*

Keywords: Motion Estimation, Multiresolution, Optical Flow, Sequence of Images, Video, Wavelets.

1. INTRODUCTION

While motion analysis plays a dominant role in the computer vision area, it also becomes of significant interest to the video coding community. Motion analysis is typically performed in two stages: in the first one, an estimation of a 2-D motion field is performed; then, this estimate is generally used to infer information about 3-D motion and scene segmentation.[2] Various approaches exist to estimate the motion field: some techniques calculate a sparse motion field (i.e. block matching) by detecting correspondences between blocks in consecutive frames. On the contrary others determine an apparent dense motion field, i.e. a velocity vector for each pixel in a sequence of images, usually referred as OF. The main drawback of obtaining a sparse motion field lies in the fact that it does

This research has been supported in part by the Italian Ministery of the University and of the Scientific and Technological Research (MURST) and by the Italian National Council for Research (CNR).

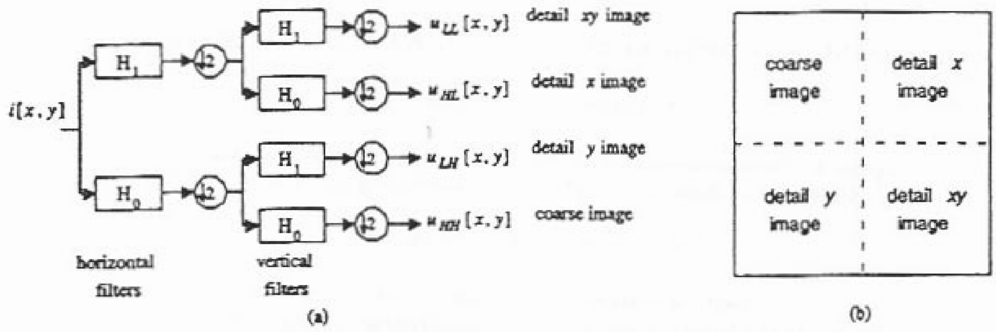

Figure 1: Single stage wavelet spatial decomposition filterbank; the first sub-stage operates along rows, while the second sub-stage operates along columns.

not provide enough information for any subsequent segmentation task. Besides techniques such as block matching often impose a constant velocity for all pixels within the boundary of the block of interest. Unfortunately, the evaluation of the OF at boundaries of moving objects is also a very difficult task; at these locations, the flow estimates are often very noisy or incorrect.
In what follows, section 2 provides a brief description of wavelet theory; section 3 then describes the most traditional ways of estimating an OF; in section 4, a new multiresolution OF estimation method is proposed. Finally some simulation results are provided in section 5.

2. SPATIO-TEMPORAL WAVELET REPRESENTATION OF A VIDEO SEQUENCE

In what follows, we assume without loss of generality, that a video sequence is simply characterized by its luminance component. A video sequence can be described by a multiresolution pyramid obtained from a spatio-temporal wavelet representation of the luminance signal.[3] This way, it is possible to transfer the signal domain space representation using a dyadic wavelet transform so as to segment the information in roughly octave frequency bands over space and time. A spatio-temporal wavelet representation of a video-sequence can be obtained by constructing a separable multiresolution description of the luminance function along every dimension (space/time). The decomposition at each stage is performed in two steps: a two dimensional spatial decomposition followed by a one-dimensional temporal decomposition. Figure 1 shows the step of a single stage spatial decomposition process (a) and the possible reorganization of the output coefficients (b). A recursive application of the spatial decomposition scheme, applying the filterbank in cascade on the coarse signal, leads to a multiresolution description of each frame. In order to distinguish the resolution level of the transform coefficients we will indicate it explicitly. After the spatial analysis, the information of each frame can be reorganized in frequency bands. It is important to note that the output coefficients do not belong solely to the frequency domain, but also to the space-time domain. The subsequent temporal decomposition stage operates on the one-dimensional signals formed by the pixels located at the same spatial positions in successive frames. This step splits the original sequence in different sub-sequences. Each one of them describes the corresponding signal as a function of time and frequency. The final coefficient reorganization after a single stage spatio-temporal wavelet representation is shown in Figure 2. A multiresolution representation can be further obtained by applying recursively the temporal decomposition to the low temporal frequency sub-sequence.
The spatial and temporal stages of a separable wavelet representation for a video sequence are totally independent, and so they can be performed in any order. For our purpose, a single temporal stage will be computed. Effectively,

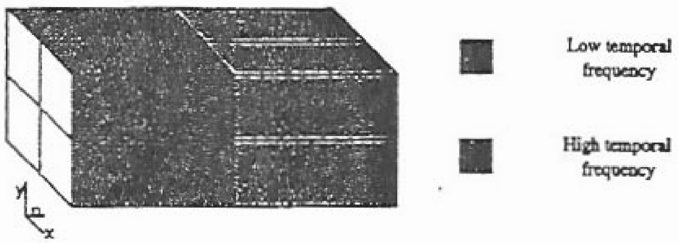

Figure 2: Data reorganization of a video sequence after one stage of a three-dimensional wavelet transform, with one spatial stage and one temporal stage.

it would be inappropriate for the purpose of estimating a proper temporal gradient (necessary to compute the OF) to substantially decimate in time any filtered version of the original sequence.

3. OPTICAL FLOW BASED MOTION ESTIMATION

The OF of a video sequence represents the distribution of the apparent velocities of brightness patterns in an image. It arises from the projection on the image plane of the objects present in a scene.

3.1. Problem statement

OF estimation is based on assuming temporal constancy of the luminance signal over time, as it is postulated by the optical flow constraint equation (OFCE)[1]:

$$\frac{d}{dt}Y(\vec{x},t) = \vec{\nabla}Y^T(\vec{x},t)\cdot\vec{u}(\vec{x},t) = 0 \quad (1)$$

where $Y(\vec{x},t)$ represents the luminance of the video sequence at the point (x,y) and time t, and $\vec{u}$ represents the velocity vector at coordinates $(\vec{x},t)$. Equation (1) can be rewritten as:

$$Y_x u_x + Y_y u_y + Y_t = (Y_x, Y_y)\cdot(u_x, u_y) + Y_t = 0 \quad (2)$$

Therefore, the components of the motion vector in the direction of the gradient (Y_x, Y_y) can be determined, while the component in the orthogonal direction cannot: this constrains $\vec{u}$ to belong to a line in the velocity space, usually referred as the constraint line (see Figure 3). Only a linear relationship between the two components can be calculated at every pixel location. The OFCE brings just a single constraint to the two unknowns u_x and u_y. It is an underconstrained problem: thus, it has to be supplemented with an additional constraint.

3.2. Additional constraints

Many different approaches exist to add an additional constraint to the OF estimation problem so as to cope with the underconstraint. We may distinguish mainly two classes: the local techniques and the global ones.
If every pixel were to move independently, the solution of the estimation problem would be impossible. Usually, the scene consists of opaque objects moving in front of a fixed background. Apart from partially occluded objects and some artificial image sequences, the image of a projected scene surface does not change abruptly with time,

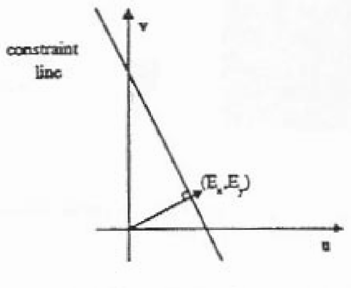

Figure 3: Optical flow constraint line: it is orthogonal to the gradient direction

if the relative motion between the scene and the image plane is not too large, or, equivalently, if the frame rate is high enough. Except at moving object boundaries, neighboring pixels have similar velocities. Therefore, Equation (1) can be retained as valid for a local neighborhood $(x + \delta x, y + \delta y)$ of the actual pixel location. The OFCE can thus be extended as[4]:

$$\vec{\nabla} Y^T(\vec{x} + \delta\vec{x}, t) \cdot \vec{u}(\vec{x} + \delta\vec{x}, t) = 0 \tag{3}$$

The main drawback of this kind of approach (i.e. a *local technique*) is that it is ill-conditioned on particular regions of the image plane, i.e. regions of approximately uniform brightness and regions containing flow discontinuities. On the contrary it is possible to formulate a global technique. Horn and Schunck suggest to minimize a functional that includes two terms: a distance measure to the solution from the constraint line, and a measure to establish the departure from a smoothness condition. Imposing a smoothness criterion to the entire image domain introduces however a correlation among the field vectors that has no physical reason.
In what follows we will limit ourselves to the usage of local techniques.

3.3. A least square solution

Based on the local approach described in the previous paragraph, Otte and Nagel[4] propose to assume that the OF either varies linearly (linear assumption) or is constant in the neighborhood of any given pixel $\vec{x}_0$. In both cases, equation (3) becomes by Taylor series expansion of the motion vector components:

$$\vec{\nabla} Y^T(\vec{x}, t) \cdot [\vec{u}(\vec{x}_0, t) + \vec{\nabla}_{\vec{x}}^T u_x(\vec{x}_0, t) \cdot (\vec{x} - \vec{x}_0) + \vec{\nabla}_{\vec{x}}^T u_y(\vec{x}_0, t) \cdot (\vec{x} - \vec{x}_0)] = 0 \tag{4}$$

Under the constant hypothesis the terms corresponding to the gradient of each vector component is zero and the the OFCE simply becomes equivalent to:

$$\vec{\nabla} Y^T(\vec{x}, t) \cdot \vec{u}(\vec{x}_0, t) = 0 \tag{5}$$

Now, considering a square (spatial) region around the actual point $\vec{x}_0$ and sampling the gradient in n^2 points, it is possible to define, using equation (4)—(5), an overconstrained system of n^2 equations and 6—2 unknowns, respectively. This linear system can be solved using the least square method.[5,6] In both cases, the computational complexity of the pseudo-inverse solution, varies linearly with respect to the number of pixels in the neighborhood (i.e. the number of equations in the system). Under the linear assumption, the formulation allows to estimate the spatial variations of the OF too,[7] but it requires a greater computational load. Moreover the linear hypothesis must hold true over a larger neighborhood, as there are six unknowns to the problem rather than two. Although quite insensitive to noise, this technique brings artifacts when the examined neighborhood is crossing a moving object boundary, e.g. it comprises two different objects, moving in different directions so that they identify two different constraint lines. As we will show later, the new multiresolution strategy has the advantage of decreasing this type of artifact.

4. MULTIRESOLUTION STRATEGY FOR OPTICAL FLOW ESTIMATION

Generally, a multiresolution representation of signals is constructed by successive approximations. For example, using a Laplacian pyramid, a signal is divided into a coarse version and a series of difference signals between this coarse version and versions of itself at higher resolution.[3] Using a multiresolution strategy a given task is typically performed by successive approximation steps, starting from the coarse version of the information, then using recursively an initial estimate at a given resolution level to obtain an estimate at the next higher resolution level till the full resolution is reached.

For the task of interest here, a wavelet decomposition of the video sequence will be used mainly for two purposes: a rough estimation of the luminance gradient at different resolution levels and a pyramidal structure to implement the task at hand.

4.1. Gradient estimation using wavelet coefficients

Both equations (4) and (5) require, to be solved, the gradient information of the luminance component both along the space and time dimensions. Suppose that the luminance information be represented for compression purposes using a wavelet decomposition: we claim that the luminance gradient can be estimated directly from wavelet coefficients. For simplicity let us assume that a single stage spatio-temporal wavelet decomposition has been applied to an image sequence: this leads to two groups of sub-sequences whose length is half the original one. Each group of sub-sequences is composed of four sub-sequences with spatial information organized as shown in Figure 2. Using this signal representation it is possible to build a downsampled estimate (by a factor 2 in time and space) of the motion field at full resolution. An approximation of the gradient information used in the overconstrained system represented either by equation (4) or (5), can be directly obtained from the wavelet decomposition: the luminance gradient component along the x—y direction is estimated from the detail x—y image at resolution -1 contained in the low temporal frequency sub-sequence; while the temporal gradient component of the luminance is estimated from the coarse image at resolution -1 in the high temporal frequency sub-sequence.

Now, if K spatial stages of the wavelet transform are applied to construct the multiresolution representation, the spatial gradient information is extracted first from the signals at spatial resolution $-K$. All gradient information is downsampled by a factor 2^K along the spatial directions and by a factor 2 in the time direction.

In order to show the relationship between the luminance gradient and the results of the wavelet decomposition, let us consider for simplicity, a one-dimensional case. Suppose $x[n] \in L_2(Z)$ is the input signal whose gradient information has to be estimated. The ideal bandlimited differentiator for discrete-time signals has a frequency response of the type:

$$H_{id}(e^{j2\pi f}) = j2\pi f, \qquad |f| < \frac{1}{2}. \tag{6}$$

In order to take into account the decimation factor of the tree structured filterbank implementing the wavelet transform, it is necessary to identify the aliasing band, that is the frequency interval that could generate aliasing as a result of the decimation. The aliasing bandwidth is directly proportional to the resolution level, i.e.:

$$\text{resolution } -K \Rightarrow \text{aliasing bandwidth } \frac{2^{-K}}{2} ;$$

The horizontal and vertical detail information provided by the wavelet transform at a given resolution level must be compared with the output of an ideal differentiator applied to a decimated version of the original signal. The corresponding ideal frequency response of the decimation and ideal differentiator cascade becomes:

$$H_{id}(e^{j2\pi f}) = \begin{cases} j2\pi f & \text{for } |f| < \frac{2^{-K}}{2} \\ 0 & \text{for } \frac{2^{-K}}{2} \leq |f| < \frac{1}{2} \end{cases} \tag{7}$$

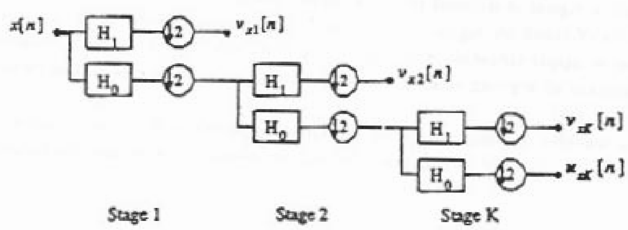

Figure 4: Multiresolution filterbank using the wavelet transform.

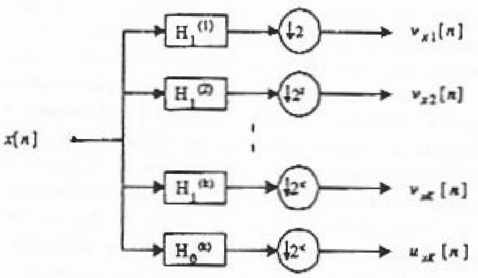

Figure 5: Equivalent multiresolution filterbank using the wavelet transform.

The multiresolution analysis filterbank, used for the wavelet transform is shown in Figure 4. Consider the input/output path that links $x[n]$ to $v_{xK}[n]$. On the basis of the previous discussion, the output signal to this path represents an approximation of the gradient information at resolution $-K$. Moreover the equivalent filter along this path can be obtained as a product of matrices whose rows contain the impulse responses $h_1[n]$ and $h_0[n]$ of all analysis filters along the path[3]:

$$\mathbf{H}_1^{(K)} = \mathbf{H}_0^{K-1} \cdot \mathbf{H}_1 \tag{8}$$

All other paths can be analyzed in the same way. An equivalent system can be designed as indicated in Figure 5. Since FIR filters are generally used in the wavelet transform, it is not possible to guarantee that the frequency response of the equivalent filters is linear and not aliased like the ideal model described by the equation (7). Figure 6 shows the magnitude of the frequency response for the equivalent filters of a biorthogonal wavelet basis for resolutions -1 to -4. Since these functions are continuous in the frequency domain, it is not possible to avoid aliasing. Besides, the more the frequency response has a linear behavior in the band of interest, the more aliasing is produced, as the decreasing part of the frequency response is shifted to higher frequencies. Different wavelet bases produce different equivalent multiresolution filters and thus different gradient approximations. There are three critical parameters, referred to the frequency response of the equivalent filters, to evaluate the wavelet transform performance: 1) the energy fraction contained in the aliasing band; 2) the linearity of the response magnitude in the band of interest; 3) the phase linearity.

In tables 4.1. and 4.1., the simulation results for a set of wavelet basis at different resolutions levels are reported. While in the band of interest, the behavior of all filters is quite similar, in the aliasing band there are large differences which do not significantly affect however the corresponding motion fields. In practical systems there is no possibility to design a FIR filter with the features of the ideal bandlimited differentiator. Any FIR filter for the gradient estimation at different resolution levels cannot avoid aliasing and, at the same time, be linear in the interest band. A compromise must thus be accepted; however, we have experimented that this does not

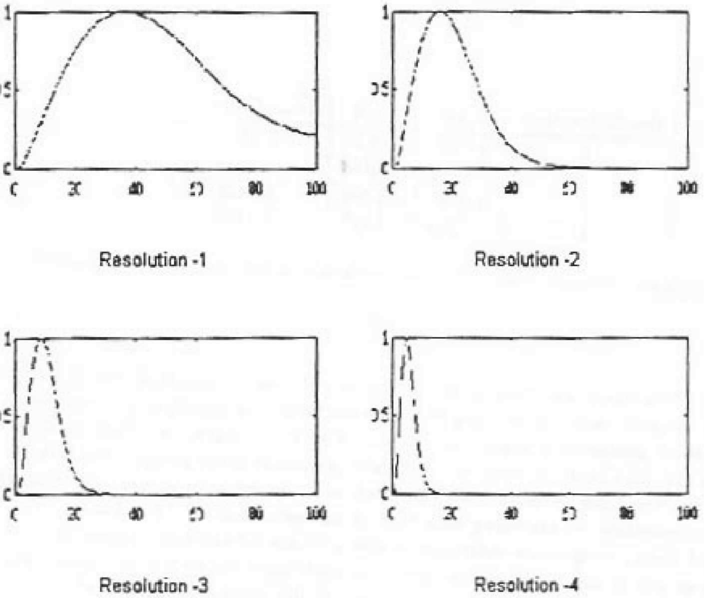

Figure 6: Equivalent filters frequency magnitude for the biorthogonal wavelet (resolution -1 to -4).

significantly affect the performance of the OF estimation approach we will describe in the next paragraph.

4.2. Multiresolution Optical Flow Estimation

The multiresolution approach used for the evaluation of the OF comprises of two main steps: a rough estimation of the motion field at low spatial resolution and subsequent refinements obtained by using higher resolution wavelet coefficients.
From the wavelet representation of the luminance signal, we can recover an estimate of the gradient information directly from the spatial detail information coefficients. Let us assume that the luminance component is stored using a wavelet transform with K spatial stages and 1 temporal stage. The first OF estimation is computed at spatial resolution $-K$, by solving in the least square sense equation (4) or (5). The obtained motion field is

Resolution level	Haar	Lin4	DaubechiesD8
-1	0.6718	0.3925	0.8580
-2	0.6803	0.2649	0.8626
-3	0.6665	0.2286	0.8513
-4	0.6524	0.2065	0.8279

Table 1: Aliasing energy for different wavelet bases normalized with respect to the total signal energy.

Resolution level	Haar	Lin4	DaubechiesD8
-1	0.0140	0.0821	0.1397
-2	0.0271	0.1374	0.1400
-3	0.0286	0.1399	0.1428
-4	0.0299	0.1447	0.1447

Table 2: Euclidean distance from the linear behavior in the band with no-aliasing.

downsampled by a factor 2^K in space and by a factor 2 in time. By back-projecting this initial guess to higher resolution levels only the velocity vector at moving object boundaries are improperly estimated. These can be refined using the information available at higher resolution. For this purpose, we first must identify the OF discontinuities, which are the most likely to occur in the neighborhood of moving object boundaries. An iterative refinement procedure of the OF close to these discontinuities is provided. Let us consider the refinement step from resolution $-j-1$ to resolution $-j$ assuming that the OF at resolution $-j-1$ is known. The flow discontinuities extraction is based upon a magnitude information and a phase information. These two contributions are considered separately so as not to influence one another. The difference between the actual vector components and the average components on a small neighborhood are used as discontinuity indicators[8]:

$$\Delta M = \frac{|\vec{u}|}{\sum_{\vec{u}_i \neq \vec{u}} |\vec{u}_i|} \tag{9}$$

$$\Delta\varphi = arg(\vec{u}) - \sum_{\vec{u}_i \neq \vec{u}} (arg(\vec{u}_i)) \tag{10}$$

Once areas of discontinuity have been detected, the OF at the corresponding locations is initialized by a bilinear interpolation of the surrounding neighborhood. Then the whole OF is spatially upsampled by a factor 2 using again a bilinear interpolation. The corresponding motion field is then averaged for discontinuity areas, with the estimate $u_0^{(-j)}$ that can be obtained at resolution level $-j$, as follows[9]:

$$\vec{u}^{(-j)} = \frac{\vec{u}^{(-j-1)} + \vec{u}_0^{(-j)}}{2}, \tag{11}$$

To obtain an estimate of gradient information at resolution $-j$, the coarse signal representation at that level is reconstructed in the discontinuity areas, by feeding the lower resolution information into the synthesis filter pair associated to the inverse wavelet transform. This procedure is iterated from resolution level $-K+1$ to -1. Unfortunately, once we reach the full spatial resolution, no more spatial gradient information is readily available from the wavelet coefficients. Estimation of the luminance gradient along the discontinuity areas is then provided by using the low temporal frequency sub-sequence. Remember that the wavelet decomposition with no spatial stages and a single temporal stage produce just two sub-sequences related to the temporal frequency, no spatial operations being performed on the sequence itself.

A full spatial resolution OF has thus been obtained. However it has been decimated by a factor 2 in time. To recover the missing samples, we can just use a shifted-by-one wavelet decomposition, the wavelet transform $\tilde{X}[n]$ associated with the original signal shifted by one position in time, $x[n-1]$. It can be proved that it is possible to obtain the coefficients of $\tilde{X}[n]$ from those of $X[n]$, the wavelet transform of $x[n]$, using two couples of translation filters. Once again, for simplicity, assume that $x[n]$ is a one-dimensional signal with N samples; $X[n]$ is its wavelet transform and contains in the first $N/2$ coefficients the coarse type information, while the last $N/2$ coefficients represent the detail information. $\tilde{X}[n]$, can be written in function of $X[n]$ as:

$$\tilde{X}[n] = \begin{cases} \varphi_0[n] * X[n] + \varphi_1[n] * X[n+N/2] & 0 \leq n < N/2 \\ \varphi_2[n] * X[n] + \varphi_3[n] * X[n+N/2] & N/2 \leq n < N \end{cases} \tag{12}$$

with

$$\begin{aligned}\varphi_0[n] &= f(h_0[n] * g_0[n+1]) \\ \varphi_1[n] &= f(h_0[n] * g_1[n+1]) \\ \varphi_2[n] &= f(h_1[n] * g_0[n+1]) \\ \varphi_3[n] &= f(h_1[n] * g_1[n+1])\end{aligned} \quad (13)$$

where $h_0[n]$ and $h_1[n]$ are the analysis filter pair impulse responses while $g_0[n]$ and $g_1[n]$ are the synthesis filter pair impulse responses, for a single stage wavelet decomposition.
By computing the shifted decomposition just for the points of interest, it is possible to determine a new OF which is temporally interleaved with the old one, obtaining this way the OF at full spatio-temporal resolution associated to the sequence of images.

5. EXPERIMENTAL SIMULATIONS

In the experimental simulations a comparison between the multiresolution strategy and the traditional method has been performed. In the traditional approach the OF has been estimated directly at the full resolution by solving in the least square sense equations (4) and (5). In the multiresolution approach, the OF has been calculated from a video sequence to which a wavelet decomposition with 3 spatial stages and 1 temporal stage has been applied. The neighborhood size to compute the least square solution of the OF been set to 7×7 for all resolution levels. This allows to consider large areas at lower resolution levels so as to properly initialize the OF estimation for the macroscopic objects. By increasing the resolution while maintaining fixed the neighborhood size, the algorithm gradually focuses on smaller and smaller areas, so that the OF estimation is improved next to its discontinuities. Such discontinuities are extracted using the simple metric described in the previous section. Threshold values have been set to 5 for the vector magnitude variation and $\pi/9$ for its phase variation.
Two series of tests have been carried out to evaluate on one hand the computational load and on the other the quality of multiresolution OF.
In Figures 7 (constant assumption) and 8 (linear assumption) a comparison between the theoretical number of operations required by the two approaches is presented. In all experiments a spatio-temporal decomposition using the Haar basis was selected for simplicity. Other simulation have demonstrated that the filter length does not substantially influence the overall computational efficiency. The number of operations was estimated while varying the size of the square neighborhood for which the least square solution to the OF was computed. As pointed out by the graphs of Figures 7 and 8, the multiresolution approach is more efficient when the neighborhood size is increased. For little neighborhood sizes in the constant assumption case, the number of operations is quite similar, while in the linear case the multiresolution strategy is always more efficient.
The algorithm performance are compared in Figures 9 to 10. The PSNR between the original sequence and the compensated sequence has been chosen as an evaluation parameter.[10] The motion compensated error sequence is generated from the OF to the original sequence. The PSNR (measured in dB) is defined for each frame as the ratio between the peak signal power (255^2) and the mean square error, i.e. the power of the displaced frame difference signal $f[m,n] - f_c[m,n]$):

$$PSNR = \frac{S}{N} = 10\log_{10}\left[\frac{255^2}{\frac{1}{MN}\cdot\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}|f[m,n]-f_c[m,n]|^2}\right] \quad (14)$$

All experiments were carried out on standard sequences for the evaluation of the OF. In this paper the results about a natural sequence (taxi) and an artificial sequence (sine) are shown. Moreover Figure 11 provides a needle diagram for one frame of the sequence taxi, while in Figure 12 the result of a motion compensation process is shown for a frame of the same sequence. Finally Table 6. summarizes the execution time in seconds on a Silicon Graphics Indigo 200 MHz workstation for the computation of the OF associated with a sequence of 10 frames at a spatial resolution 128×160.

	Constant Assumption		Linear Assumption	
Sequence Name	Wavelet	Traditional	Wavelet	Traditional
Taxi	18.9	26.54	35.79	140.75
Sine	22.45	26.82	36.98	140.15

Table 3: Execution times.

6. CONCLUSIONS AND FUTURE WORK

In this paper a new multiresolution approach to estimate the OF of a video sequence has been proposed. If the video information has been represented for compression or transmission purposes in a multiresolution wavelet structure, it is possible to obtain directly from the wavelet coefficients, the gradient information used to solve the OFCE. The multiresolution strategy allows to improve an initial guess of the motion field at coarse resolution by back-projecting it on higher resolution levels. The OF is refined only at the moving object boundaries, determined by a discontinuity measure based on the magnitude and direction of the velocity vectors. This technique decreases the presence of noisy vectors at OF discontinuities and can serve as a good initial estimate for a motion field segmentation result.

Future works on this approach attempt to study techniques that allow to refine the OF even in areas that do not seem to present at first discontinuities, by using a motion-compensated error measure as an additional indicator. This should allow to determine whether a refinement should also be provided for the smooth areas of the OF at higher resolution. Efforts are also being carried out to obtain good motion field segmentation results, using this multiresolution approach.

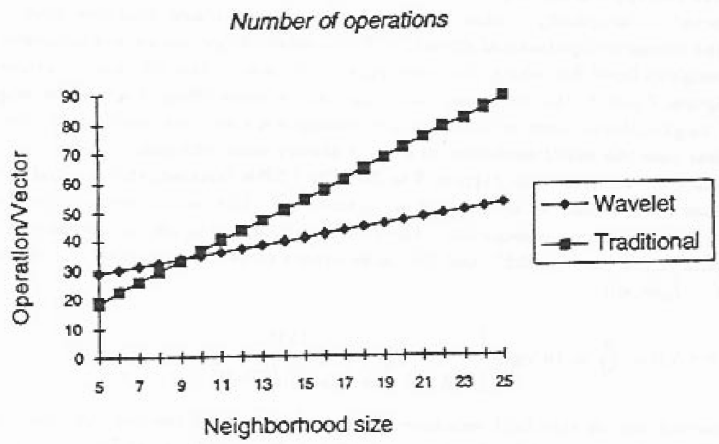

Figure 7: Number of operations per vector required by the multiresolution strategy under the constant assumption.

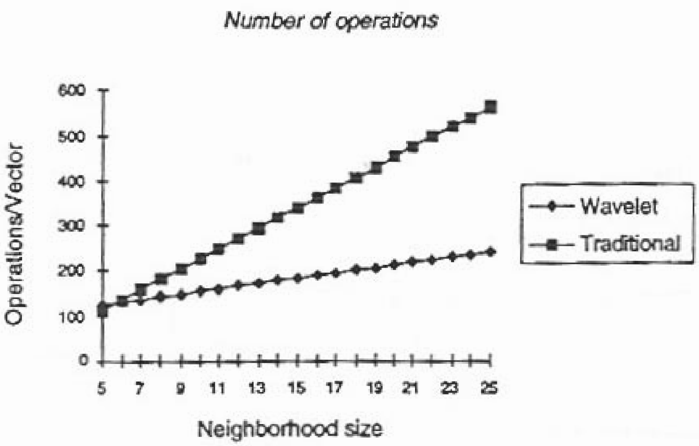

Figure 8: Number of operations per vector required by the multiresolution strategy under the linear assumption.

7. REFERENCES

[1] B.K.P Horn and B.G.Schunck *Determining optical flow*, Artificial Intelligence, 17, 185-203 (1981).

[2] G.Adiv, *Determining three-dimensional motion and structure from optical flow generated by several moving objects*, IEEE Transaction on Pattern Analysis and Machine Intelligence, PAMI-7(4), 384-401 (Jul. 1985).

[3] M.Vetterli and J.Kovacevic', *Wavelet and subband coding*, Prentice Hall, 1995.

[4] M.Otte and H.H.Nagel, *Optical Flow Estimation: Advances and Comparisons*, Artificial Intelligence, 33, 299-324 (1987).

[5] H.H.Nagel, *Analyse und Interpretation von Bildfolgen*, Informatik-Spektrum, 8, 178-200 & 213-328 (1985).

[6] J.K.Kearney, W.B. Thompson, D.L.Boley, *Optical Flow Estimation: An Error Analysis of Gradient-Based Methods with Local Optimization*, IEEE Transaction on Pattern Analysis and Machine Intelligence, PAMI-9(2), 229-244 (Mar. 1987).

[7] M.Campani and A.Verri, *Computing Optical Flow from an Overconstrained System of Linear Algebraic Equations*, Proc. Third Int. Conf. on Computer Vision ICCV'90, Osaka, Japan, 22-26 (Dec. 4-7 1990).

[8] B.G.Schunck, *Image flow segmentation and estimation by constraint line clustering*, IEEE Transaction on Pattern Analysis and Machine Intelligence, PAMI-11(10), 1010-1027 (Oct. 1989).

[9] J.Astola, P.Haavisto and Y.Neuvo, *Vector median filters*, Proceedings of the IEEE, 78(4), 678-689 (Apr. 1990).

[10] A.M.Eskicioglu and P.S.Fisher, *Image Quality Measures and Their Performances*, IEEE Transaction on Communications, COM-43(12), 2959-2965 (Dec. 1995).

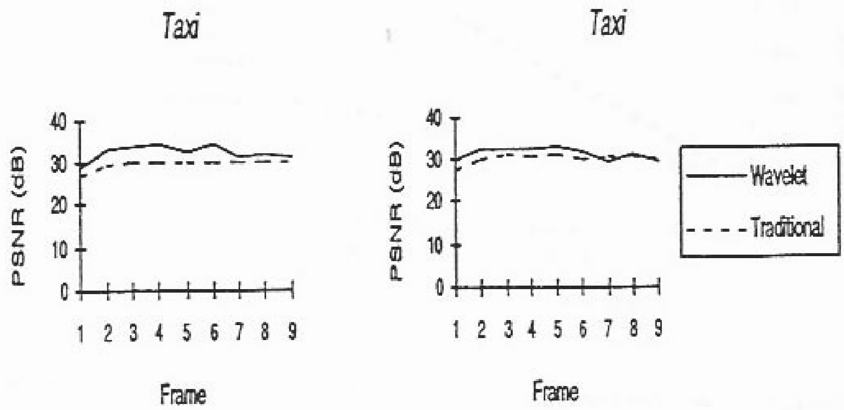

Figure 9: PSNR comparison between multiresolution and traditional approaches under the constant (a) and linear (b) assumptions on the first ten frames of the taxi sequence.

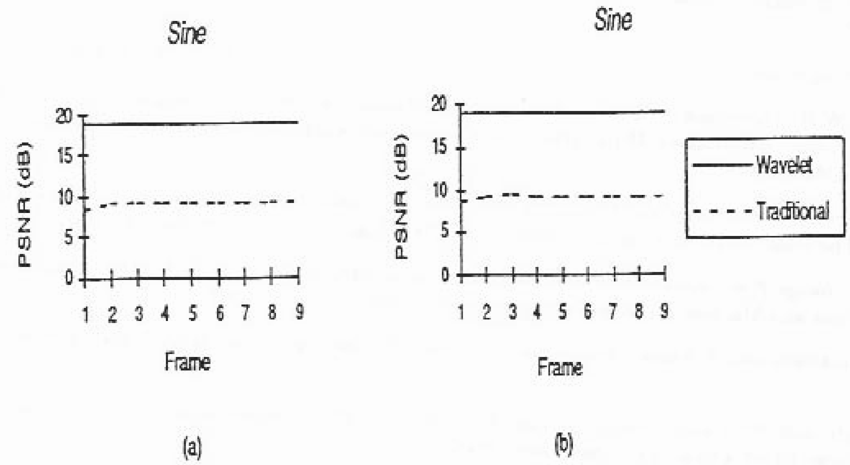

Figure 10: PSNR comparison between multiresolution and traditional approaches under the constant (a) and linear (b) assumptions on the first ten frames of the sine sequence.

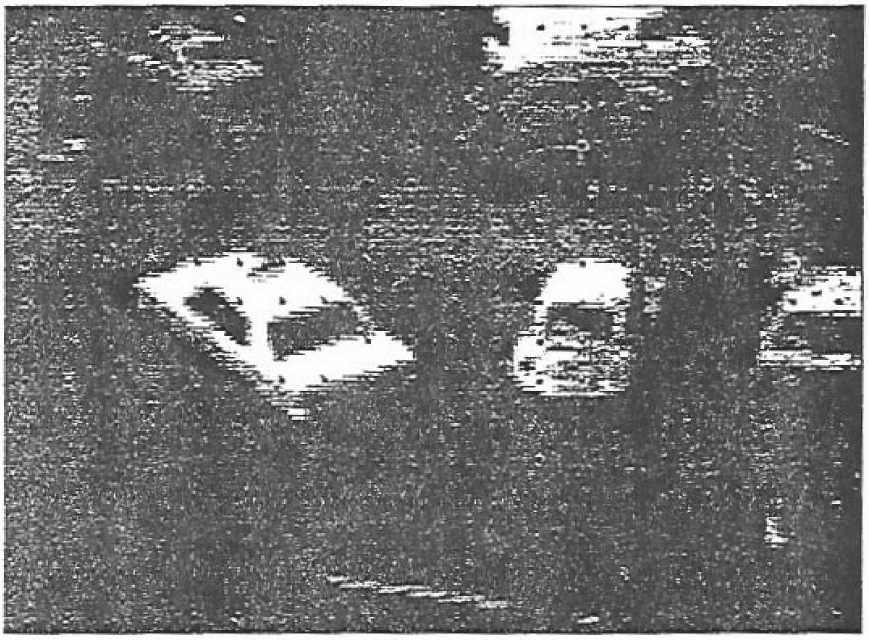

Figure 11: Needle diagram of a taxi sequence frame.

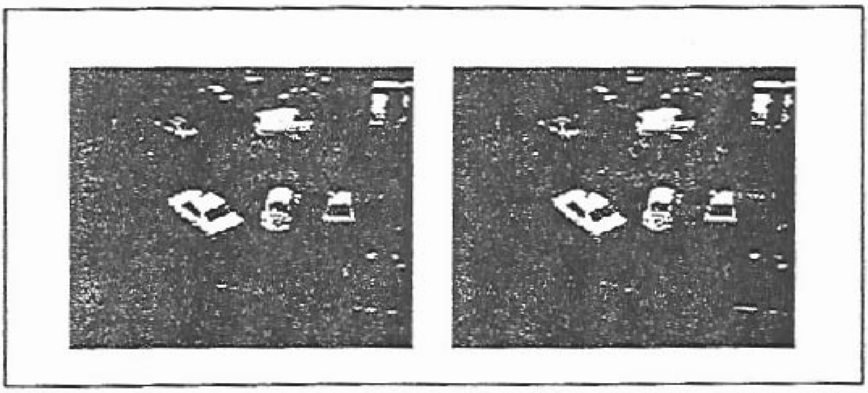

Figure 12: A comparison between two frames respectively extracted from the original sequence taxi and from the compensated one; the recovered zones have been left in black.