Optical-flow estimation of dense motion field using robust techniques

S. ZIMERAS
University of the Aegean
Department of Statistics and Actuarial-Financial Mathematics
G.R.832 00, Karlovassi, Samos, GREECE

Abstract: - In recent years, interest in motion analysis has increased with advances in processing capabilities. The usual input in a motion analysis system is an image sequence, with a corresponding increase in the amount of processed data. A typical motion problem is to analyze the motion within 2D image data corresponding to a sequence of frames, of a 3D scene. In computer vision a number of techniques are available to estimate the optical flow; the more efficient (in terms of quality) are modulated as the minimization of a global objective function. This cost function includes an observation constraint and a smoothness term. Such models generally assume that the luminance is constant along its trajectory. This assumption is not valid in cases of spatial and temporal distortions as in fluid image sequence. As an extension, a new model is described based on the continuity equation of fluid mechanics and a smoothness function considering the divergence (div) and vorticity (curl) of the motion field. The proposed model is embedded in a multiresolution framework and the minimization is conducted with an efficient multigrid technique. In this paper, the performance of the proposed motion estimation technique is analyzed and compared to similar “standard” methods, using simulated and real satellite data (the last provided by EUMETSAT). Finally, measures as RMSE of the images, number of cuts and number of iterations for the minimization of the energy function are introduced to justify the improvement of the estimation technique.

Key-Words: - clouds motion, dense estimation techniques, optical flow, div-curl equations

1 Introduction

The main aspects of the motion field modeling are introduced by:

- Modeling the motions as a cost energy function including observation constraints and smoothness terms [5],
- Expressing the optical flow as a model, constant or affine [4],
- Introducing robustness to reduce the differences between the data and the optic flow model [7],
- Minimize the energy function using a general hierarchical optimization framework (multiresolution - multigrid) with an adaptive way [10, 11].

An effective way to extract informations about the movement of objects (in this case the shape of the clouds) is the use of the optical flow. An optical flow can be defined as the transformation of a 3D motion of objects and cameras to a 2D motion on the image via a suitable projection system. An efficient way to estimate the optical flow field, in terms of quality, is novelized as the minimization of a global objective function, which includes an observation constraint and a smoothness term [5]. Such models assume that the luminance function is constant along its trajectories. This assumption is not valid in cases of spatial and temporal distortion as in fluid image sequences. As an extension a new model is introduced based on the continuity equation of fluid mechanics and a smoothness function considering the divergence (div) and vorticity (curl) of the motion field. A general hierarchical optimization framework, which is both multiresolution and multigrid with an adaptive partition, is proposed for the optimization of the objective function. Results from the standard method and the new proposed method is presented, considering statistical measures like RSE, block numbers, and divergence and vorticity equation values.

2 Modeling of the motion field

Optical flow estimation aims at recovering the apparent velocity field \( \omega = \{ \omega_s, s \in S \} \) at each point of the rectangular pixel lattice \( S \) based on the luminance function \( f(t) = \{ f(s,t), s \in S \} \) at two
consecutive instants \( t \) and \( t+1 \). Assuming a temporal constancy of the brightness, the **optical constrained equation** (OFCE) is given by: \( \nabla f(s,t) + f_s(s,t) = 0 \), where \( \nabla f = (f_x, f_y)^T \) represents the spatial gradient of \( f \) and \( f_s(s,t) \) is the temporal partial derivative of luminance \( f \). The global estimation of the motion field can be achieved by optimizing the following cost energy function:

\[
U(\omega, f) = \sum_{s \in S} \| \nabla f(s,t) \omega_s + f_s(s,t) \|^2 + a \sum_{c < r_s \in C} \| \omega_s - \omega_r \|^2
\]

where \( S \) is the set of pixel grid, \( C \) are the possible cliques for the neighboring sites \( <s,r> \) (the 4 neighborhood system in this case), and \( a > 0 \) is a smoothing parameter controlling the balance between the two terms. The first term represents the interaction between the field (unknown variables) and the data (given variables), where the second term expresses the smoothness constraint. The disadvantages of this formulation are: (1) The OFCE is not valid in case of large displacements, because of the linearization; (2) The real field is not globally smooth, containing probably discontinuities that can not be novelized with the quadratic cost function. In case of long range motions, an **incremental displacement field** \( d\omega \) is searched by minimizing the energy function [10, 11]:

\[
U(d\omega, f) = \sum_{s \in S} \| \nabla f(s + \omega_s, t + dt) d\omega_s + f_s(s, t) \|^2 + a \sum_{c < r_s \in C} \| \omega_s + d\omega_r - (\omega_s + d\omega_r) \|^2
\]

where \( S \) is the set of pixel grid, \( C \) are the possible cliques for the neighborhood structure, \( f_s(s,t) = f(s,t) - f(s + \omega_s, t + dt) \) is the displacement frame difference, and \( a > 0 \) is a smoothing parameter controlling the balance between the two terms.

### 3 Robust estimators

To efficiently cope with the large deviations from the data model and from the prior model, **robust functions** are introduced and more precisely robust **M-estimators** [14]. An M-estimator \( \rho \) has the following properties (Fig. 1):

- \( \rho \) is increasing on \( \mathbb{R}^+ \)
- \( \phi(u) = \rho(\sqrt{u}) \) is strictly concave on \( \mathbb{R}^+ \)
- \( \lim \rightarrow \rho'(u) < \infty \)

Two robust estimators are introduced in the generic motion estimation; the first one within the data term and the second one within the regularization term. Fig. 1 shows the profile of two such estimators and their associate optimal weight functions; i.e. for this work the Leclerc estimator was used. The cost function \( U(d\omega, f) \) can be modified as:

\[
U(d\omega, f) = \sum_{s \in S} \rho_1 \| \nabla f(s + \omega_s, t + dt) d\omega_s + f_s(s, t) \|^2 + a \sum_{c < r_s \in C} \| \omega_s + d\omega_r - (\omega_s + d\omega_r) \|^2
\]

where \( \rho_1 \) and \( \rho_2 \) are the two robust Leclerc estimators. According to minimization aspects for the robust estimators [3, 7], the cost function, takes the form:

\[
U(d\omega, f) = \sum_{s \in S} \delta \| \nabla f(s + \omega_s, t + dt) d\omega_s + f_s(s, t) \|^2 + \phi(\delta_s) + a \sum_{c < r_s \in C} \| \omega_s + d\omega_r - (\omega_s + d\omega_r) \|^2 + \phi(\beta_s)
\]

where \( \delta_s \) are the weights of the data that controls the optical flow constrained equation, and \( \beta_s \) are the weights that control the velocity discontinuities.

### 4 Multiresolution/Multigrid approaches

In case of large displacements, a classical incremental **multiresolution procedure** (Fig. 2) is used [8]. For each instant \( t \) of the sequence, a pyramid of images \( (F) \) is derived by successive Gaussian smoothing and regular re-sampling by factor of 2. At coarsest level, displacements are reduced and cost function (2) can be used. For the next resolution levels, only one incremental \( d\omega^k \) is estimated to refine estimate \( d\omega \), obtained from the previous level. More informations about the process are given in [10, 11] papers. For a faster...
convergence of the minimization process and a better performance, a **multigrid approach** is applied (Fig. 2).

Fig. 2: Multigrid relaxation at a given resolution level k [10, 11]

The process consists to partition the image into lattice of size $2^k$ at the grid level $k$. The cost function (2) can then be expressed according to the partition and a parametric model is estimated for each patch as an increment on each pixel. The displacement increment estimated on a partition region depends on the total displacement on the neighborhood of this patch; that implies the field could be continuous between the pixels and in this case no block effects are appeared. The minimization of the cost function is achieved by using a deterministic approach (in this case the ICM algorithm was applied). A combine constant and affine model for the displacements $\Delta \omega^k_s$ was used with general form

$$d \omega^k_s = P_n(s) \Theta^k_n$$

where $x_s$ and $y_s$ stands for the coordinates of pixel $s$. When the grid level is changed, the partition of the grid is also changed (in an adaptive way). The number of blocks (during the subdivision) could be the criterion to measure the way the model fits the data or could be used as a prior knowledge for the structure of the particular application. Using this adaptive way of splitting the blocks, eventually there is a distinguish between the regions of interests, where the estimation must be accurate, and the regions where information are useless.

5 Extension of the standard method

The previous methodology was analyzed for the cases, where the luminance is constant along its trajectory. This assumption is not valid in cases of spatial and temporal distortions as in fluid image sequence. As an extension a dedicated model, proposed by [1, 2], is described. This model based on an observation constraints issuing from the continuity equation of fluid mechanics and a smoothness function taking into account the divergence (div) and vorticity (curl) of the motion field. The basic idea between div-curl method is to have a different penalization for $\text{div}(\omega)$ and $\text{curl}(\omega)$ in the smoothness terms, to encourage one or the other quantity. If we assume that the mass per unit volume remains the same during the transportation between two images, in this case the **equation of continuity** can be written as $\frac{\partial \omega}{\partial t} + \text{div}(\omega) = 0$.

The div-factor can be found by the following equation: $\text{div}(\omega) = \omega \nabla f + f \nabla \omega$, where $\nabla f$ is the spatial gradient for the luminance function and $\omega$ is the velocity vector field. Combine both equations, the equation of continuity becomes:

$$\frac{\partial \omega}{\partial t} + \text{div}(\omega) = 0,$$

which is the first differential equation. For the standard cease (previous sections), the second term wasn’t considered. The solution of this equation for a stationary motion field between instant $t$ and $t+1$ becomes:

$$f(s+\omega,t+1)\exp[-\text{div}(\omega)] - f(s,t) = 0$$

This equation gives us a model for the intensity variation between times $t$ and $t+dt$. Using the first order linearization the first of the first member of the equation. For the standard cease (previous sections), the second term wasn’t considered. The solution of this equation for a stationary motion field between instant $t$ and $t+1$ becomes:

$$U_1(\omega, f) = \sum_{s \in S} \rho_s \exp(\text{div}\omega) \int f(s, \omega, t+1) \nabla d\omega_s + \nabla f(s, \omega, t+1) \nabla d\omega_s + f(s, \omega, t+1)] - f(s, t))$$

For the second term, it can be noticed that the minimization of $U_2(\omega) = a \int \left[\|\nabla f\|^2 + \|\nabla \omega\|^2\right]$ is equivalent to the minimization of $U_2(\omega) = \lambda \left[\text{div}(\omega) + \left\|\text{curl}\omega\right\|\right]$. Therefore a first order regularization penalizes both the curl and the divergence of the estimated flow field. Since high concentration of there quantities are most likely to be present in fluid motion we rather prefer to use a regularization which favor the apparition of this area

$$U_2(\omega) = \frac{a}{\rho} \sum_{s \in S} \left[\text{div}\omega_s + \text{div}\omega_s - \xi_s + \left\|\text{curl}\omega_s + \text{curl}\omega_s - \xi_s\right\|^2\right]$$

where $\text{div}(\omega_s)$ and $\text{curl}(\omega_s)$ denote the divergence and vorticity of the vector field $\omega$, respectively.
\( f(s,t) = f(s + \omega_x t + dt), \) \( \xi \) and \( \zeta \) are estimates of the div and curl functions and \( \lambda > 0 \) is a control parameter. A neighborhood structure for the calculation of the div and curl equations is considered with discrete forms:

\[
div(\omega_{i,j}) = \frac{1}{6} \left( [3u_{i,j} + u_{i,j-2} - 6u_{i,j-1} + 2u_{i,j+1}] + [3v_{i,j} + v_{i,j-2} - 6v_{i,j-1} + 2v_{i,j+1}] \right)
\]

and

\[
curl(\omega_{i,j}) = \frac{1}{6} \left( [3v_{i,j} + v_{i,j-2} - 6v_{i,j-1} + 2v_{i,j+1}] + [3u_{i,j} + u_{i,j-2} - 6u_{i,j-1} + 2u_{i,j+1}] \right)
\]

[1, 2, 13]. If \( \text{div}(\omega) = 0 \) and \( \text{curl}(\omega) = 0 \) then the modified equation comes the standard method.

A general hierarchical framework, which is both multiresolution and multigrid with an adaptive partition, is proposed for the optimization of the energy function [10].

### 6 Measures and parameters

The two algorithms are applied for a sequence of 3 images (512x512) using synthetic images, where the ground truth is known, and real satellite images. The considered measures for this investigation are:

- **Root Mean Squared Error (RMSE)**, between the second and the reconstructed image
- **Divergence and vorticity equations**, with forms
  \[
div(\omega) = \frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} \quad \text{and} \quad \text{div}(\omega) = \frac{\partial v}{\partial x} + \frac{\partial u}{\partial y}
\]
- **Mean and standard deviation of the angular error**, between the actual flow field and the estimated one with form
  \[
  \text{error} = \cos \left( \frac{u_{x_{\text{true}}} + vy_{x_{\text{true}}} + 1}{\sqrt{|u|^2 + 1}} \right)
\]
- **Number of iterations**
- **Number of cuts and number of blocks**

The first part of the experiments (synthetic images) introduces the choosing parameter values for the cost function, and the second part used these "optimal" values for the investigation of the performance based on the real images. The pyramid levels for the multiresolution-multigrid process are \( k = 3 \) for the both process. Figure 3 and 4 illustrate the performance of the two algorithms using different cut-threshold values (0, 0.05, 0.1). Four different measures are considered: divergence, vorticity, angular mean and angular standard deviation. The process for the cutting blocks in general gives us a indication about the fitting of the proposed model (affine+constant in this case); three case can be considered:

- if the cut threshold is equal to 0 then the multigrid process applied everywhere (non adaptive case),
- if the cut threshold is large then the multigrid process did not apply at all,
- if the cut threshold is between \((0,k)\), where \(k\) is a upper bound value depending from the particular application, then the multigrid process applied only in the areas with high motions (adaptive case).

For all the cases the non-adaptive process (cut-threshold=0) performs better compared to the adaptive process (cut-threshold=0.05, 0.1); the divergence and vorticity values are close to the true values and the angular mean and standard deviations are smaller compared to the other cases. The non-adaptive cut-threshold suggests a lower bound value for performance of the algorithms.

Fig. 3. Performance of standard algorithm using different \( \alpha \) values for various cut-thresholds; from bottom to top cut-thresholds=0, 0.05, 0.1

Fig. 4. Performance of div-curl algorithm using different \( \lambda \) values for various cut-thresholds; from bottom to top cut-thresholds=0, 0.05, 0.1
Considering the adaptive process, the results based on the standard method indicate that the algorithm performs better using a cut-threshold=0.1 than the other case. A choice of the optimal value is a difficult part due to the case that different measures suggest different values. After extensive experiments a reasonable optimal value for parameter \( \alpha \) was found equals to 200. For the div-curl method the same results are applied as before. The cut-threshold=0.1 gives closer estimates compared to the true values (div and curl graphs) and performs better in the term of the proposed measures (angular mean and standard deviation graphs). After various experiments a reasonable optimal value for parameter \( \lambda \) was found equals to 200. Based on the previous extensive analysis, the values for the robust estimator are fixed to: \( \rho_1=49 \), \( \rho_2=0.5 \), \( \alpha=200 \) and \( \lambda=200 \) for the standard method and \( \rho_1,49 \), \( \rho_2=0.5 \) for the div-curl method [1, 2].

6.1 Synthetic sequence

The two methods are tested on two synthetic sequences; build by applying a known motion to satellite images. A fixed divergence and vorticity was applied on the overall image with values (div, curl) = (0.08, 0.06). The standard method was run for different \( \alpha \) values, while the div-curl method was run for different \( \lambda \) values, fixing the \( \alpha \) based on the standard result. The process suggests optimal values for standard method \( \alpha=200 \), and for div-curl method \( (\alpha, \lambda) = (200, 200) \); Table 1 presents the performance of the two methods for the synthetic motion experiments. From these results it is clear that the proposed div-curl method gives better results compared with the standard method with estimated div and curl values close to the ground truth; all the proposed measures (RMSE, angular mean and standard deviation criterion between the actual flow field and the estimated one following, number of iterations, cuts and blocks) are decreasing indicating that the process performs satisfactory [13].

<table>
<thead>
<tr>
<th>Methods</th>
<th>standard</th>
<th>div-curl</th>
</tr>
</thead>
<tbody>
<tr>
<td>(div, curl)</td>
<td>(0.0795, 0.057)</td>
<td>(0.0799, 0.059)</td>
</tr>
<tr>
<td>Angular (mean, std)</td>
<td>(1.904°, 4.048°)</td>
<td>(0.885°, 1.289°)</td>
</tr>
<tr>
<td>Number of iterations</td>
<td>226</td>
<td>169</td>
</tr>
<tr>
<td>Number of cuts</td>
<td>1252</td>
<td>1123</td>
</tr>
<tr>
<td>Number of blocks</td>
<td>7856</td>
<td>7469</td>
</tr>
</tbody>
</table>

Table 1: Performance of the standard and div-curl methods for synthetic motions

Fig. 5 gives the true and estimated optical flow fields for both methods; Fig. 5a illustrates the ground truth motion field using the fixed divergence and vorticity. The obtained motion fields show clearly the differences between the two approaches. As it was expected, the standard method (Fig. 5b) was indeed unable to match corresponding points and directions for the velocity fields, with the ground truth. Alternatively, the estimated flow field, obtained from the div-curl method (Fig. 5c), indicates certain stability, smoothness with motion directions close to the ground truth.

6.2 Real sequence

In this case, the optical flow between the two images is calculated considering the previous optical flow calculation; parameter values for both methods were chosen based on the synthetic sequence results. Fig. 6 shows us a sequence of 3 METEOSAT images at the time intervals 100=\( (t) \) sec, 101=\( (t+7.5) \) sec, 102=\( (t+15) \) sec. The resulting measures for different sequences using both methods are presented in Table 2. Analysis of the performance, suggests that the results for the real case are similar compared to the
results for the synthetic case. For the div-curl method, RMSE values are better suggesting that the algorithm gives satisfactory results. The number of cuts and blocks are increased, indicating that there are same areas where the standard methods could not identify the motion directions [13].

![Image](image1)

Fig. 6: METEOSAT sequence - resolution 512x512; left to right 100, 101, 102

![Image](image2)

Table 2: Performance of the standard and div-curl methods for the real sequences

<table>
<thead>
<tr>
<th></th>
<th>100-101 standard</th>
<th>100-101 div-curl</th>
<th>101-102 standard</th>
<th>101-102 div-curl</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of iterations</td>
<td>207</td>
<td>205</td>
<td>204</td>
<td>207</td>
</tr>
<tr>
<td>Number of cuts</td>
<td>19477</td>
<td>21194</td>
<td>19643</td>
<td>23376</td>
</tr>
<tr>
<td>Number of blocks</td>
<td>74819</td>
<td>79970</td>
<td>75317</td>
<td>86516</td>
</tr>
<tr>
<td>RMSE</td>
<td>7.478</td>
<td>7.357</td>
<td>7.479</td>
<td>7.381</td>
</tr>
</tbody>
</table>

![Image](image3)

Fig. 7: Velocity fields for sequence 100-101; (a): standard, (b): div-curl method

For the benefit of the comparison the divergence and vorticity maps are calculated and the respectively images are given in Fig 8 and 9. Visual comparison between the graphs suggest that the div-curl method gives better results compare to the standard method; the high spot colors indicates the areas with the higher divergence and vorticity. Particularly, the improvement of the standard method can be seen at the vorticity maps; the div-curl method presents areas where the standard method could not find them.

![Image](image4)

Fig. 8: Divergence maps for sequence 100-101; left to right: standard, div-curl method

For the benefit of the comparison the divergence and vorticity maps are calculated and the respectively images are given in Fig 8 and 9. Visual comparison between the graphs suggest that the div-curl method gives better results compare to the standard method; the high spot colors indicates the areas with the higher divergence and vorticity. Particularly, the improvement of the standard method can be seen at the vorticity maps; the div-curl method presents areas where the standard method could not find them.

![Image](image5)

Fig. 7: Velocity fields for sequence 100-101; (a): standard, (b): div-curl method

The resulting velocity fields are illustrated in Fig 7. Some differences can be spot between the two images, especially concerning the directions of the motion fields. The vectors, demonstrating the divergence and vorticity, appeared clearer for the div-curl method compared to the standard one; especially in the down left corner where a kind of vorticity can be spotted.
Fig. 9: Vorticity maps for sequence 100-101; left to right: standard, div-curl method

7 Conclusion

In this work an analysis of the multi-grid/multi-resolution process is explained and the performance of this algorithm was analyzed. Different measures of the algorithm are presented as well as comparisons between different approaches. An extension of the previous methodology was defined with the previous results. Analysis was taken place for whole the image sequence and for the part of the image sequence. As a general result is that, when the frame of image increased, the performance of the algorithm becomes better. Experiments have taken place for different image dimensions to justify the above remark. The previous analysis, assumes that the luminance function is constant along its trajectories. This assumption is not valid in cases of spatial and temporal distortions as in fluid image sequence. As an extension, a dedicated model has been considered [1, 2] based on the continuity equation of fluid mechanics and a smoothness function considering the divergence (div) and vorticity (curl) of the motion field. Comparison between the extension approach and the standard approach, show us that the dedicated method works better; the regions, where small motions appear, are identified better, due to the introduction of the two new parameters inside the model.

8 Acknowledgements

The author would like to thank Dr. E. Memin, Dr. P. Perez and Dr. T. Copperti for their usefully help and comments during the development of this work. Also many thanks to EUMETSAT for providing the image data. This work has been supported by a EUMETSAT Fellowship, EUM/CO/99/699/KTH.

References: