MRF MODELING FOR OPTICAL FLOW COMPUTATION FROM MULTI-STRUCTURE OBJECTS

V. Prinet *C. Cassisa* *FF Tang*

National Lab. of Pattern Recognition (NLPR) /
Sino-French Lab. in Computer Science, Automation and Applied Math. (LIAMA)
Institute of Automation, Chinese Academy of Sciences (CASIA).
{fftang,prinet}@nlpr.ia.ac.cn cassisa@liama.ia.ac.cn

ABSTRACT

We propose in this paper a new formulation of the equation of the optical flow enabling to compute global and local motions of multi-structure objects (flowers and petals, trees and leaves, ...). The displacement fields are computed using a Markovian Random Field (MRF) model. Local and global components of the vector flow are both explicitly retrieved. The minimization of the Gibbs energy is achieved with a down-scaling approach, in which we first analyze the motion of the compact object, the sub-structures' movement being retrieved in a second stage. We validate and demonstrate the efficiency of our approach on synthetic and real images for various applications.

Index Terms— Image motion analysis, stochastic fields

1. INTRODUCTION

The objective of this paper is to propose an efficient formulation and computational framework to estimate local and global motions of non-compact non-rigid objects. Modeling rigid and non-rigid objects movement using optical flow (OF) has been widely studied for various applications including tracking, surveillance [1, 2]. Analysis of deformation of non-compact material such as water or smoke has also recently raised much attention [3]. However, objects which are composed of sub-structures (leaves of trees, feathers of birds, petals of flowers, etc) do not fit into these categories. Existing approaches are unable to tackle efficiently the problem because the deformation of this type of object is both global (at the level of the entire object) and local (at the level of the sub-structure, usually of a few pixels size). Most recent works aim at tackling the problem of discontinuities [4], of noise [5], to develop hierarchical or multi-resolution computational approaches [6, 7].

We propose to generalize the constant intensity constraint of the optical flow, and to retrieve the optimal solution via a Markov Random Field (MRF) framework. The problem we address focuses on the multi-motion analysis of objects composed of sub-structures, for which both local and global components of the displacement vectors are modeled.

The rest of the paper is organized as follows: in section 2 recalling briefly the basic mathematical foundations of OF and MRF, we define the generalized constant intensity constraint (GCIC) and describe the energy function to be minimised. We propose a down-scaling approach, thus allowing to process the image at two different scales and details levels. Illustrations and analysis of the results on various synthetic and real data set are given in section 3. Finally, section 4 presents the conclusions.

2. THE APPROACH

While extremely simple, the two key points of the approach are (i) the formulation of a generalized constant intensity constraint (GCIC) and (ii) the down-scaling optimization scheme in a MRF framework.

2.1. General concept and definitions

We tackle the problem of one or several object(s) that are composed of multi-structures (trees, birds, flowers,..). Each object is animated by a global movement. Each object is itself composed of sub-objects or structures (such as leaves, feathers, petals, ...) which have a local motion independent from the global displacement of the object.

In order to differentiate the notion of global object and local structure, we define two graphs associated to different levels of resolution G :

$$G_l = (S_l, L_l) \qquad G_g = (S_g, L_g)$$

where $S_l = \{s_l\}$ represents the local nodes (indice l) defined either by the substructures of the objects or simply by the pixels of the image –i.e. $card(S_l) = dim(I)$–, while $S_g = \{s_g\}$ is represents global nodes (indice g) defined by the center of each object in motion in the image –i.e. $card(S_g) = number\,of\,objects$. L_l and L_g are two distinct levels of vertices, linking independently the neighborhood nodes s_l and s_g respectively. Note that this definition imposes $S_l \supset S_g$.

2.2. Modeling the GCIC

Considering objects animated of a double movement –local and global ones–, the velocity vector $v(i)$ at each point of the

image i is expressed by:

$$v(i) = v_g(i) + v_l(i) \quad v_l(i) << v_g(i) \tag{1}$$

where $v_g(i)$ is the global displacement of the object O at point i and $v_l(i)$ is the local displacement of the substructure composing O at point i. Since all the substructures belonging to a given object O should possess the same global movement we have:

$$v_g(i) = constant = v_g \ \forall i \in O \tag{2}$$

If the substructures are of one pixel size and have random independent movements, then local variations can be assimilated to white noise of zero mean and:

$$\sum_{i \in O} v(i) = N\, v_g \tag{3}$$

where N is the number of points within the object.

From equation (1), the first order generalized optical flow equation (so-called Generalised Constant Intensity Constraint) can be written:

$$\nabla I.\,(\mathbf{v}_g + \mathbf{v}_l) + \frac{\delta I}{\delta t} = 0 \qquad \mathbf{v}_l << \mathbf{v}_g \tag{4}$$

By averaging equation (4) over all the points of the object O and knowing from equation (2) and (3) that $\overline{\mathbf{v_g} + \mathbf{v_l}} = \overline{\mathbf{v_g}} + \overline{\mathbf{v_l}} = \overline{\mathbf{v_g}} = \mathbf{v_g}$, thus, for a given object and as a first approximation, we have:

$$(\overline{\nabla I})^T.\mathbf{v}_g + \overline{I_t} \approx 0 \tag{5}$$

with $\overline{a} = \frac{1}{N}\sum_{i=1}^{N} a_i$.

2.3. Energy terms of the MRF

The displacement field in the image is given by the $\mathbf{v}$ that satisfy equation (4). Knowing that the solution is undefined at points where $\mathbf{v}$ is orthogonal to ∇I -so called aperture effect-, some additional constraints –or regularization factor– should be added.

Using an MRF framework, the solution $\mathbf{v}$ will be given by the configuration of our random variable x, that maximizes a joint probability function taking the form of a Gibbs distribution:

$$P(\mathbf{x}, \mathbf{y}) = \frac{\mathbf{1}}{\mathbf{Z}} \mathbf{e}^{-\mathbf{E(x,y)}} \tag{6}$$

where Z is the normalization constant and

$$E(\mathbf{x}, \mathbf{y}) = -(\sum_{\mathbf{s} \in \mathbf{C_1}} \alpha_{\mathbf{d}} \mathbf{V_d}(\mathbf{y_s}, \mathbf{x_s}) + \sum_{\mathbf{s}, \mathbf{s'} \in \mathbf{C_2}} \alpha_{\mathbf{p}} \mathbf{V_p}(\mathbf{x_s}, \mathbf{x_{s'}}))) \tag{7}$$

where $\mathbf{y}$ is a function of the observed data. From a general point of view, C_1 and C_2 are single-site and pair-site cliques respectively; the Gibbs energy E is thus the summation of a data term $E_d = \sum \alpha_d V_d$, and a prior term $E_p = \sum \alpha_p V_p$ independent of the data but dependent of its neighborhood N_s; $\alpha_d = \alpha_d(s)$ and $\alpha_p = \alpha_p(s, s')$ are two weighting coefficients.

For our problem, the data term is naturally given from the GCIC (equation (4)). It should be satisfied at each site of the graph (S_l, L_l):

$$|E_d| = \sum_{s \in C_1} \alpha_d \left(\nabla I(s).\,(\mathbf{v_g}(\mathbf{s}) + \mathbf{v_l}(\mathbf{s})) + \frac{\delta I(s)}{\delta t} \right)^2 \tag{8}$$

with $C_1 = S_l$.

On the other hand, the prior term is defined such as to constrain the displacement flow to be continuous: amplitude and angular variations between neighborhood sites should be minimized, both at global and local levels. We have therefore:

$$|E_p| = \sum_{s,s' \in C_2} \beta_p \left(||\mathbf{v_l}(\mathbf{s}) - \mathbf{v_l}(\mathbf{s'})|| \right)^2 \tag{9}$$

$$+ \sum_{s,s' \in C_3} \gamma_p \left(||\mathbf{v_g}(\mathbf{s}) - \mathbf{v_g}(\mathbf{s'})|| \right)^2$$

$$+ \sum_{s,s' \in C_3} \lambda_p \left(\frac{\mathbf{v_g}(\mathbf{s}).\mathbf{v_g}(\mathbf{s'})}{||\mathbf{v_g}(\mathbf{s})||||\mathbf{v_g}(\mathbf{s'})||} - 1 \right)^2$$

where $C_2 = \{s_l, s'_l | s_l, s'_l \in S_l, s'_l \in N_l\}$ and $C_3 = \{s_g, s'_g |$ $s_g, s'_g \in S_g, s'_g \in N_g\}$. The neighborhoods N_l and N_g are defined within a 4-connectivity by the connected sites belonging to G_l and G_g respectively. The first two terms of equation (9) impose a smooth variation of the amplitude at local scale (first term) and global scale (second term). The last term constrains the motion's direction to change slowly between neighboring objects. The site dependent coefficients $\alpha_d, \beta_p, \gamma_p, \lambda_p$ are set such that : $\alpha_d \neq 0$ if $s \in O$, $\gamma_p, \lambda_p \neq 0$ everywhere, and

$$\beta_p = 0 \qquad if \quad s \in O, s' \in O' \tag{10}$$

$\beta_p \neq 0$ otherwise.

In other words, the smoothing between local vectorial components is to be applied only for intra-object and not inter-object neighborhood.

It is worth noticing that, in the case of several objects in the image, the model is properly defined only if the global movement between two neighborhood objects varies slowly. However, we will see that in practice we do not encounter this limitation, because the local term will serve to compensate the possible loss generated by the global smoothing.

2.4. Implementation

The optimal solution is retrieved by computing the $\mathbf{v}$ that minimizes $E(x, y)$. It is achieved without major difficulty with a ICM (iterated conditional mode) algorithm. However, since the random value x is a four dimensional vector which has components that can take any real value within a given range $[v_l^{min}, v_l^{max}, v_g^{min}, v_g^{max}]$, the direct optimization of equation

(7) using (8) would be extremely time consuming. In addition, the initialization of ICM algorithm would be twice as critical as traditional approaches (i.e. approaches that do not search for a local and a global component).

For these reasons, we choose to compute the displacement field via a down-scaling approach, following two steps :

- Initialization : compute the optimal solution of equation (7), using (5) and (9) as data term and prior term respectively. This enables to find a realisation $\hat{\mathbf{v}}_g^0$, which will serve as a first approximation. The process is extremely fast and enables to give a rough picture of the displacement flow.

- Estimation of $\hat{\mathbf{v}}_g, \hat{\mathbf{v}}_l$, by substituting equation (8) into (7) and using $\hat{\mathbf{v}}_g^0$ as initialization.

Note that the approach is based on the knowledge of the graph (S_g, L_g), which means that, theoretically, one should be able to localize the objects within the image, before running the optical flow. When feasible, a pre-processing segmentation step can be carried out for this purpose. In most cases however, we can perform arbitary segmentation using a regular grid, without affecting the quality of the results. We will illustrate several cases in the next section.

3. RESULTS ANALYSIS

Results analysis is performed on a set of very different synthetic and real images, thus illustrating various possible applications. All results presented in this section have been obtained with the values $(\alpha_d, \beta_p, \gamma_p, \lambda_p) = (1, 10, 50, 100)$ for forest sequence and $(1, 10, 10, 10)$ for the others. Video sequences have been previously smoothed in space and time with Gaussian filter of variance 1.5; the synthetic forest images, have been spatially smoothed only. The derivation mask we used is similar to [5, 8].

3.1. Synthetic data : forest

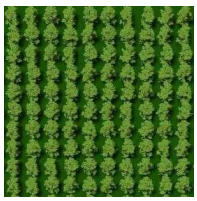
The initial motivation of this work was to analise the Earth ground displacement observed from forest trees aerial images. This phenomena can be observed on terrains of steep slopes and are characteristic of forthcoming landslides. Here we make use of synthetic images simulating a plantation of trees on a steep terrain, generated by the AMAP software [9]. Two sequences of two images each have been gene-rated. The first sequence shows a unidirectional displacement along the slope in the vertical direction, the second is a radial displacement. The position of the camera is unchanged between the two "acquisitions". One of the simulated image and the known displacements are illustrated on figure 1.

The objects we consider are the trees, the local structures being defined by the leaves. The images are particularly difficult to process because the intensity constancy is hardly verified at pixel level.

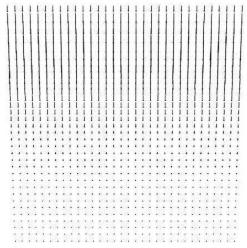
Figure 2 illustrates the results obtained from our approach (top), compared to those from Horn and Shunk (H&S) implemented by Barron [8, 10] (bottom). The computed displacement field, for transverse and radial motions, are both consistant with the ground truth. A contrario, H&S algorithm appears to be more sensitive to the texture effect and non uniform brightness variation of the leaves. Table 1 compares the average angle error between H &S and our approach.

In these images in which segmentation is rather easy to perform, we have analysed the impact of the pre-propressing step on the results. In figure 3, we compare the vector flow obtained with and without presegmentation. In case of presegmentation, each tree is treated as an individual object. Otherwise, the image is organised into a matrix of squared paches, each patch being associate to an object. Table 1 shows that the effect of the square size is negligeable, as expected.

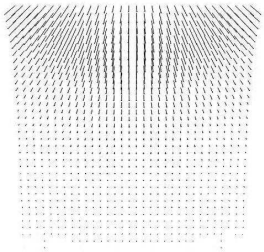
(a) Entire image



(b) Detail of (a).

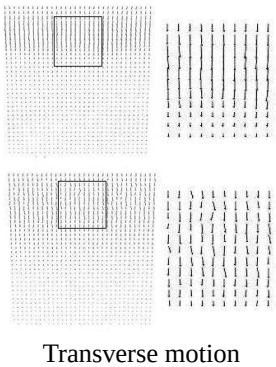


(c) Transverse displacement

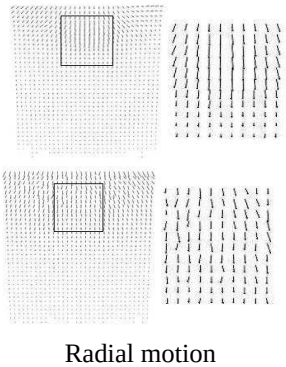


(d) Radial displacement.

Fig. 1. Simulated forest trees and ground truth motion.

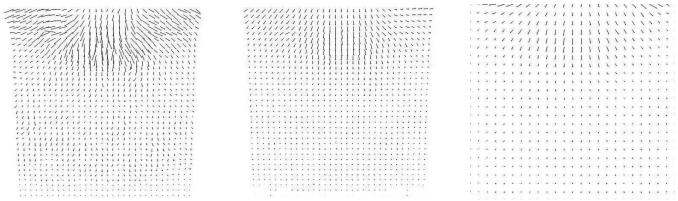


Transverse motion



Radial motion

Fig. 2. Results of GCIC (top) - Horn & Shunk (bottom).



$\mathbf{v_g}$ & segm. $(\mathbf{v_g} + \mathbf{v_l})$ & segm. $\mathbf{v_g}$ & grid 40x40

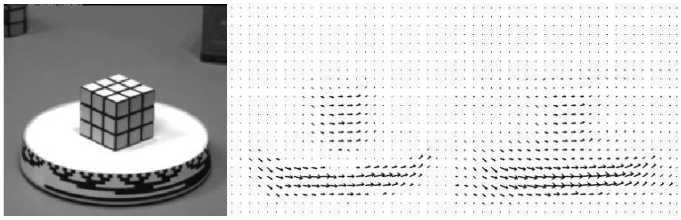
Fig. 3. Local vs global motion, with and without segmentation

Displacement	Tranversal	Radial
H & S	32.1	32.0
GCIC & seg	14.4	16.6
GCIC & grid 30x30	15.1	17.5
GCIC & grid 40x40	15.0	17.8
GCIC & grid 50x50	15.7	17.7

Table 1. Average angle error in degrees with presegmentation (seg) and with arbitrary patchs (grid size NxN).

3.2. Reference data set : synthetic and real data

Results on synthetic and real video sequences –data are downloaded from ftp.csd.uwo.ca/pub/vision– have been computed. For the synthetic sequences, we obtained the same range of average angle error than shown by [5].We illustrate here only one of them, i.e. the Magic Cube, given in figure 4. The computed displacement ($\mathbf{v} = \mathbf{v_g} + \mathbf{v_l}$) shows a clear smooth motion tracking of the cube and of the bottom part of the tray. The motion of the white part however has not been detected at all, due to the perfect uniformity of the gray level (constant white). Barron's result with Horn and Shunk approach is also illustrated for comparison.



(a) Original image (b) GCIC $(\mathbf{v_g} + \mathbf{v_l})$ (c) H & S

Fig. 4. Results on the Magic Cube sequence

3.3. Flower video

The last illustration is on a flower video sequence (downloaded data). The sequence visualizes the blooming of the flower. Each petal can be considered as an individual object. The global displacement is first computed in each patch grid of the image. Edge points detection is performed by a Canny filter for each frame of the sequence. The displayed result in figure 5 represents the $\mathbf{v_g}$ component of the displacement taken at each edge point. Result on the entire video provides clear cue of the flower blooming.

4. CONCLUSION

We proposed in this work an approach dedicated at first to the movement analysis of objects composed by sub-structures, i.e. animated by global and local motions. We illustrated that the introduction of a local displacement vector enables to handle efficiently the problem of discontinuity. With regard to forestry application, ongoing works are dedicated to textural movement analysis.

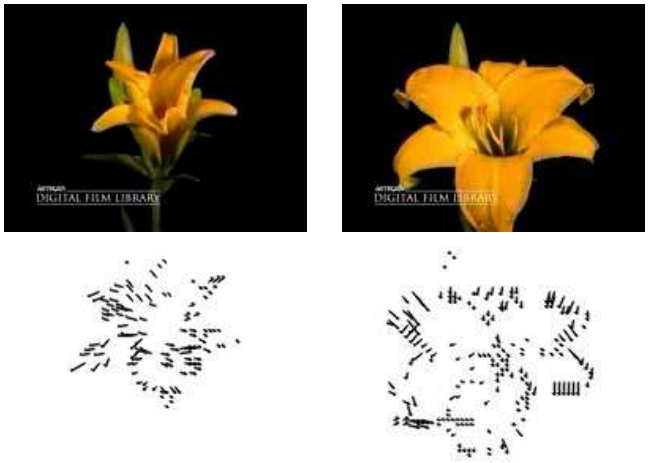


Fig. 5. Two flower frames and the associated computed $\mathbf{v_g}$

Acknowledgment The authors wish to thank Thierry Fourcaud for stimulating discussions, Guillaume Perrin for the creation of the forest images and Dominique Guyon for real data images.

5. REFERENCES

[1] F. Ranchin and F. Dibos, "Moving objects segmentation using optical flow estimation," *Conf. on Mathematics and Image Analysis (MIA)*, 2004.

[2] M. Ye and R. Haralick, "Two stages robust optical flow estimation," *Conf. on Computer Vision and Pattern Recognition*, vol. 2, pp. 623–628, 2000.

[3] E. Menin and P. Perez, "Hierarchical estimation and segmentation of dense motion field," *Int. J. Comput. Vision*, vol. 42, pp. 129–155, 2002.

[4] F. Heits and P. Bouthemy, "Multimodal estimation of discontinous optical flow using markov random fields," *IEEE Trans. Pattern Anal. and Mach. Intell.*, vol. 15, no. 12, pp. 1217–1232, 1993.

[5] S. Roy and V. Govindu, "Mrf solution for probabilistic optical flow formulation," *Int. Conf. on Comp. Vision*, 2000.

[6] E. Adachi and S. Horiuchi, "Multi-resolution optical flow estimation with local approximation," *Int. Conf. on Image Processing*, 2002.

[7] J.N. Provost, C. Collet, P. Rostaing, P. Perez, P. Bouthemy, "Hierarchical markovian segmentation of multispectral images for the reconstruction of water depth maps," *Int. J. on Computer Vision and Vision Understanding*, 2004.

[8] J.L. Barron, D.J.Fleet, S.S. Beauchemin, "System and experiment performance of optical flow techniques," *Int. J. Comput. Vision*, vol. 12, pp. 43–77, 1994.

[9] P. De Reffye,M. Jaeger,C. Edelin,J. Franon,C. Puech, "Plant models faithful to botanical structure and development," *Int. Conf. Computer Graphics*, vol. 22, pp. 151–158, 1988, Siggraph.

[10] B.Horn, B Shunck, "Determining optical flow," *Artif. Intell*, vol. 17, pp. 185–203, 1981.