HIGHER ORDER POLYNOMIALS, FREE FORM DEFORMATIONS AND OPTICAL FLOW ESTIMATION

[§,∞]*Konstantinos Karantzalos* [§] *Nikos Paragios*

[§]Atlantis, C.E.R.T.I.S.
Ecole Nationale des Ponts et Chaussees, France

[∞] School of Rural and Survey Engineering,
National Technical University of Athens, Greece

ABSTRACT

In this paper, we propose a novel technique to represent and recover optical flow through free form deformations. Such a technique is based on representing the motion field using regular connected grids according to higher order polynomials, a compromise between dense motion estimation and parametric motion models. Optical flow is determined through the deformation of the grid - derived from the optimization of a cost function - and consequently of the underlying image structures towards satisfying the constant brightness constraint. Smoothness conditions are implicity accounted for through the free form deformation approach. Promising results demonstrate the potentials of our approach.

1. INTRODUCTION

Optical flow estimation is a fundamental component of motion analysis often used as a primitive cue in higher level tasks of computational vision. Recognition, 3D reconstruction are some examples. Given two images that refer to the same scene from different positions of the observer, 2D motion estimation consists of recovering a displacement field that explains the real 3D motion of the scene patches in the image domain. Constant brightness assumption is the most common hypothesis to recover such estimates.

Scenes are assumed to be Lambertian and observations intensity-wise in the 2D image plane assumed to remain constant. Let us consider a 3D scene patch $\mathcal{P}$ and its projection ($\mathcal{P}_1$, $\mathcal{P}_2$) in two 2D images f and g within a sequence. The constant brightness assumption refers to the following condition:

$$f(\mathcal{P}_1) = g(\mathcal{P}_2)$$

where the motion vector $(\mathbf{u}, \mathbf{v})$ for the pixel $(\mathbf{x}) = \mathcal{P}_1$ that is to be estimated refers to $(\mathbf{u}(\mathbf{x}), \mathbf{v}(\mathbf{x})) = \mathcal{P}_2 - \mathcal{P}_1$. Such a condition is necessary but not sufficient and results in an ill-posed problem. The number of constraints in the image (one per pixel) is inferior to the number of unknowns to be recovered.

Starting from the pioneering work of Horn and Schunck [6] a variety of models have been proposed for optical flow estimation [2] according to the intensity conservation law or minimizing.

$$E(\mathbf{u}, \mathbf{v}) = \iint \big(f(\mathbf{x}) - g(\mathbf{x} + (\mathbf{u}(\mathbf{x}), \mathbf{v}(\mathbf{x})))\big)^2 d\mathbf{x}$$

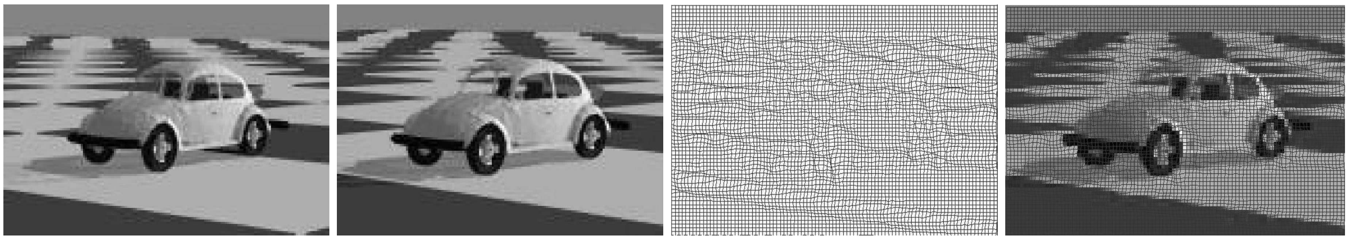
In order to account for the ill-poseness of the motion estimation problem, Lucas and Kanade [9] have proposed the estimation of a motion field that is constant within blocks while later on additional constrains were introduced refereing to a smoothness assumption

$$E(\mathbf{u}, \mathbf{v}) = \alpha \iint \big(|\nabla \mathbf{u}(\mathbf{x})| + |\nabla \mathbf{v}(\mathbf{x})|\big) d\mathbf{x} + \iint \big(f(\mathbf{x}) - g(\mathbf{x} + (\mathbf{u}(\mathbf{x}), \mathbf{v}(\mathbf{x})))\big)^2 d\mathbf{x}$$

of the optical flow field that have led to piecewise smooth results [14]. Despite the use of smoothness constraints, dense optical flow remained a challenging task mostly because of the aperture problem. Robust statistical methods [7] were considered to account for outliers. Parametric motion models [3, 10] was an alternative to dense motion estimation, that consist of a small number of parameters applied either to the entire image or to sizable image areas (motion decomposition in layers [1, 13]). Examples of such models $(\mathcal{A})$ consist of rigid, similarity and affine transformations, and are determined through the optimization of appropriate cost functionals:

$$E(\mathcal{A}) = \iint \big(f(\mathbf{x}) - g(\mathcal{A}(\mathbf{x}))\big)^2 d\mathbf{x}$$

One can claim that parametric motion models are efficient representations of optical flow, a good compromise between low complexity and reasonable flow estimates that suffer at the object boundaries. Higher order local polynomials and free form deformations is a more general class of parametric models that can be used to describe motion. Such models



(a) (b) (c) (d)

Fig. 1. Car sequence (a) first image f, (b) second image g, (c) deformed grid, (d) deformed grid overlaid to second image.

can account for local/global motion, inherit smoothness and are robust to the presence of noise. In this paper, we represent motion using regular connected grids, an excellent alternative to dense estimation and parametric motion models.

This paper is organized as follows. In Section 2 the essence of free form deformations is presented while in Section 3 an FFD variational framework for motion estimation is being described as well as its minimization process. The multiresolution incremental variant of the approach follows in Section 4, while discussion is part of Section 5.

2. FREE FORM DEFORMATIONS

Optical flow estimation is equivalent with recovering a pixel-wise deformation field $T(\mathbf{\Delta P}; \mathbf{x}, \mathbf{y})$ that creates visual correspondences between the images f and g at a pixel level. Such deformation field $T(\mathbf{\Delta P}; \mathbf{x}, \mathbf{y})$ can be recovered either using standard optical flow constraints or through the use of warping techniques like the free form deformations method [12], [11], which is a popular approach in graphics, animation and rendering [5].

Opposite to optical flow techniques, FFD techniques support smoothness constraints, exhibit robustness to noise and are suitable for modelling large and small non-rigid deformations. Furthermore, under certain conditions, it can support a dense registration paradigm that is continuous and guarantees a one-to-one mapping.

The essence of FFD is to deform an object by manipulating a regular control lattice P overlaid on its volumetric embedding space (figure 1). We consider an Incremental Cubic B-spline Free Form Deformation (FFD) to model the local transformation T. To this end, dense registration is achieved by evolving a control lattice P according to a deformation improvement $[\delta P]$. The inference problem is solved with respect to - the parameters of FFD - the control lattice coordinates.

Let us consider a regular lattice of control points

$$P_{m,n} = (P^x_{m,n}, P^y_{m,n});\ m = 1, ..., M, \quad n = 1, ..., N$$

overlaid to an image $I = \{(x, y) | 1 \leq x \leq X, 1 \leq y \leq Y\}$ in the embedding space that encloses the source structure. Let us denote the initial configuration of the control lattice as P^0, and the deforming control lattice as $P = P^0 + \delta P$. Under these assumptions, the incremental FFD parameters are the deformations of the control points in both directions (x, y):

$$\mathbf{\Delta P} = \{(\delta P^x_{m,n}, \delta P^y_{m,n})\};\ (m, n) \in [1, M] \times [1, N]$$

The motion of a pixel (x, y) given the deformation of the control lattice from P^0 to P, is defined in terms of a tensor product of Cubic B-spline:

$$\begin{aligned} T(\mathbf{\Delta P}; (x, y)) &= ((x, y)) + \delta T(\mathbf{\Delta P}; (x, y)) \\ &= \sum_{k=0}^{3} \sum_{l=0}^{3} B_k(u) B_l(v) (P^0_{i+k,j+l} + \delta P_{i+k,j+l}) \end{aligned}$$

where $k = \lfloor \frac{x}{X} \cdot M \rfloor - 1,\ l = \lfloor \frac{y}{Y} \cdot N \rfloor - 1$. The terms of the deformation component refer to (i) $\delta P_{i+l,j+l}$, $(k, l) \in [0, 3] \times [0, 3]$ consists of the deformations of pixel (x, y)'s (sixteen) adjacent control points, (ii) $\delta T(\mathbf{\Delta P}; \mathbf{x}, \mathbf{y})$ is the incremental deformation at pixel (x, y), and (iii) $B_k(u)$ is the k^{th} basis function of a Cubic B-spline:

$$B_0(u) = (1 - u)^3/6, \quad B_1(u) = (3u^3 - 6u^2 + 4)/6$$
$$B_2(u) = (-3u^3 + 3u^2 + 3u + 1)/6, \quad B_3(u) = u^3/6$$

with $u = \frac{x}{X} \cdot M - \lfloor \frac{x}{X} \cdot M \rfloor$ ($B_l(v)$ is defined in a similar fashion with $v = \frac{y}{Y} \cdot N - \lfloor \frac{y}{Y} \cdot N \rfloor$)

3. OPTICAL FLOW ESTIMATION

Optical flow estimation is now equivalent with finding the best lattice P configuration such that the overlaid structures (images) coincide. One can consider the Sum of Squared Differences (SSD) as the data-driven term to recover the deformation field $T(\mathbf{\Delta P}; (x, y))$;

$$E_{data}(\mathbf{\Delta P}) = \iint_{\Omega} \rho(r)\ dxdy$$

where $\rho(r) = r^2, \quad r = f(x, y) - g(T(\mathbf{\Delta P}; (x, y)))$. The use of such a technique to represent motion introduces in an implicit form some smoothness constraint that can deal with a limited level of deformation. In order to account for outliers and noise, one can replace the error-two norm with more appropriate robust metrics [7]. In order to further preserve the regularity of the recovered motion flow, one can consider an additional smoothness term on the deformation

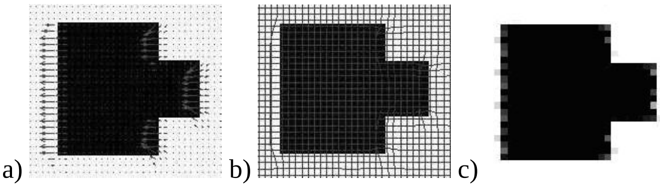


Fig. 2. A binary test image with a global one pixel movement to the left (a) grid's flow overlaid to second image g, (b) FFD's grid overlaid to second image g, (c) reconstructed from the final flow image $\hat{f} \approx f$.

field δP. We consider a computationally efficient smoothness term:

$$E_{sm}(\mathbf{\Delta P}) = \iint \left(\left|\left| \frac{\partial \delta T(\mathbf{\Delta P}; x, y)}{\partial x} \right|\right|^2 + \left|\left| \frac{\partial \delta T(\mathbf{\Delta P}; x, y)}{\partial y} \right|\right|^2 \right) \, dxdy$$

Such smoothness term is based on a classic error norm that has certain known limitations. Within the proposed framework, an implicit smoothness constraint is also imposed by the Spline FFD. Therefore there is no need for introducing complex and computationally expensive regularization components (figures 2, 3).

4. INCREMENTAL, MULTI-SCALE FFD OPTICAL FLOW

The data-driven term and the smoothness constraints term can now be integrated to recover the motion field:

$$E(\mathbf{\Delta P}) = E_{data}(\mathbf{\Delta P}) + \alpha E_{sm}(\mathbf{\Delta P})$$

where α is the constant balancing the contribution of the two terms. The calculus of variations and a gradient descent method can be used to optimize such objective function. A minimiser must fulfill the Euler-Lagrange equation:

$$\frac{\partial E(\mathbf{\Delta P})}{\partial \mathbf{\Delta P}^{[x]}_{(m,n)}} = \frac{\partial E_{data}(\mathbf{\Delta P})}{\partial \mathbf{\Delta P}^{[x]}_{(m,n)}} + \alpha \frac{\partial E_{sm}(\mathbf{\Delta P})}{\partial \mathbf{\Delta P}^{[x]}_{(m,n)}}$$

where $\mathbf{\Delta P}^{[x]}_{(m,n)}$ refers to the horizontal motion component and the $\mathbf{\Delta P}^{[y]}_{(m,n)}$ to the vertical one. Similar flow like the one earlier presented can be recovered for $\mathbf{\Delta P}^{[y]}_{(m,n)}$. One can further develop this flow:

$$\frac{\partial E_{data}(\mathbf{\Delta P})}{\partial \mathbf{\Delta P}^{[x]}_{(m,n)}} = \iint_\Omega 2r \frac{\partial r}{\mathbf{\Delta P}^{[x]}_{(m,n)}} dxdy, \ \ where$$

$$\frac{\partial r}{\mathbf{\Delta P}^{[x]}_{(m,n)}} = \frac{\partial (T(\mathbf{\Delta P}; x, y))}{\partial \mathbf{\Delta P}^{[x]}_{(m,n)}} * g_x(T(\mathbf{\Delta P}; x, y))$$

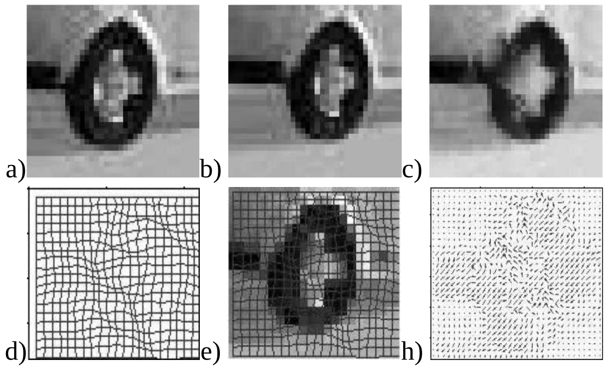


Fig. 3. Zoom on the wheel of the car sequence: (a) first frame f, (b) second frame g, (c) reconstructed image from grid's flow, (d) deformed grid, (e) deformed grid overlaid to second image (h)grid's flow.

Similar approach can be considered for the smoothness term (in the $[x]$ direction):

$$\frac{\partial E_{sm}(\mathbf{\Delta P})}{\partial \mathbf{\Delta P}^{[x]}_{(m,n)}} = 2 \iint \frac{\partial \delta T(\mathbf{\Delta P}; x, y)}{\partial x} * \frac{\frac{\partial \delta T(\mathbf{\Delta P}; x,y)}{\partial x}}{\partial \mathbf{\Delta P}^{[x]}_{(m,n)}} \, dxdy + 2 \iint \frac{\partial \delta T(\mathbf{\Delta P}; x, y)}{\partial y} * \frac{\frac{\partial \delta T(\mathbf{\Delta P}; x,y)}{\partial y}}{\partial \mathbf{\Delta P}^{[x]}_{(m,n)}} \, dxdy$$

The flow consists of a data-driven update component and a diffusion term that constraints the parameters of the free form deformation to be locally smooth. While such a model can be quite efficient it still suffers from the aperture problem. One can consider additional constraints to the constant brightness assumption similar to the one recently introduced in [4]:

$$E(\mathbf{\Delta P}) = \alpha \iint_\Omega \left(f(x,y) - g(T(\mathbf{\Delta P}; x, y))\right)^2 dxdy + \beta \iint_\Omega \left(\left|\left|\nabla f(x,y) - \nabla g(T(\mathbf{\Delta P}; x, y))\right|\right|\right)^2 dxdy + \gamma \iint \left(\left|\left| \frac{\partial \delta T(\mathbf{\Delta P}; x, y)}{\partial x} \right|\right|^2 + \left|\left| \frac{\partial \delta T(\mathbf{\Delta P}; x, y)}{\partial y} \right|\right|^2 \right) \, dxdy$$

like a gradient preservation assumption, a constraint that improves the estimation of the optical flow on the object boundaries where parametric motion models fail.

A straightforward application of the FFD manipulation cannot always guarantee the successful motion estimation between the two images. One reason for this is that we limit the maximum displacement of a control point to approximately a half of the spacing between control points in order to make the deformation function one-to-one [8]. Multilevel incremental free-form deformation (MIFFD) technique overcomes the drawbacks of the straightforward method, since it can handle both large and small non-rigid deformations. The overall dense deformation field for motion estimation is defined by the incremental deformations from all levels. Let $P^1, ..., P^K$ denote a hierarchy of control

Fig. 4. MIFFD (4 levels) for an IKONOS satellite stereo pair. From the g image (first from the left), the four reconstructed images from the estimated flow at each level are shown, until the first image f (last one) is approximated.

point meshes at different resolutions. Each control mesh P^k and the associated spline-based FFD defines a transformation $T^k(\mathbf{\Delta P}; x, y)$ at each level of resolution and the total deformation $\delta T(x, y)$ for a pixel (x, y) in a hierarchy of K levels is: $\delta T(x, y) = \sum_{k=0}^{K} \delta T^k(\mathbf{\Delta P^k}; x, y)$.

The hierarchy of control lattices can have arbitrary number of levels, but typically 3-4 levels are sufficient to handle both large and small deformations (figures 4, 5).

5. DISCUSSION - CONCLUSIONS

In this paper we have proposed a novel algorithm for motion estimation using local polynomial parametric models. Such models close the gap between dense motion estimation and global motion models, can guarantee a one-to-one correspondence between images, encode in implicit fashion smoothness constraints and support a multiresulotion implementation. In addition we have considered a criterion that also accounts for discontinuities by forcing motion estimates to be consistent on the gradient space too.

One can consider numerous extensions of the method. The use of FFD that also encode the structure of the image is a prominent one. The grid that was considered to represent motion has a fixed topology and the motion of each image pixel is reproduced using the same number of neightbouring elements that are distributed according to the same topology. Examples consist of modelling/connecting pixels that are part of a line and forcing motion estimates to be consistent along such a structure. Non-regular, image/content-based grids according to higher order polynomials could encode the image structure and better account for the motion field.

Last, but not least the use of a 3D deformation grid can be considered to account for motion decomposition in layers. In the case of depth discontinuities, one can assume that each layer can be represented using an FFD, introduce costly connections between layers that heavily penalize pixels that belong to more than one layer and recover the geometric structure of the scene as well its decomposition in layers of constant depth.

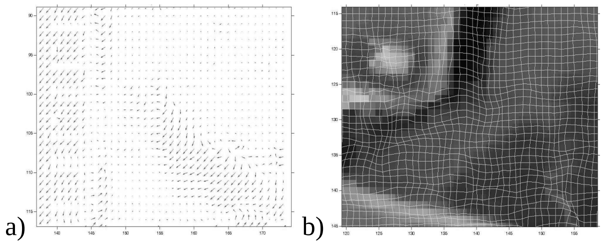


Fig. 5. Zoom on the lower right building's part of the IKONOS (figure 4) image (a) grid's flow, (b) deformed grid overlaid to second image.

6. REFERENCES

[1] S. Ayer and H. Sawhney. Layered Representation of Motion Video Using Robust Maximum-Likelihood Estimation of Mixture Models and MDL Encoding. In *IEEE International Conference in Computer Vision*, pages 777–784, Caibridge, USA, 1995.

[2] J. Barron, D. Fleet, and S. Beauchemin. Performance of Optical Flow Techniques. *International Journal of Computer Vision*, 12:43–77, 1994.

[3] M. Black and A. Jepson. Estimating optical flow in segmented images using variable-order parametric models with local deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18:973–986, 1996.

[4] T. Brox, A. Bruhn, and J. Weickert. High accuracy optical flow estiation based on a theory for warping. In *European Conference on Computer Vision*, pages 158–163, 2004.

[5] P. Faloutsos, M. van de Panne, and D. Terzopoulos. Dynamic Free-Form Deformations for Animation Synthesis. *IEEE Transactions on Visualization and Computer Graphics*, 3:201–214, 1997.

[6] B. Horn and B. Schunck. Determinating Optical Flow. *Artificial Intelligence*, 17:185–203, 1981.

[7] P. Huber. *Robust Statistics*. John Wiley & Sons, 1981.

[8] S. Lee, K.-Y. Chwa, and S.Y. Shin. Image metamorphosis with scattered feature constraints. *IEEE Transactions on Visualization and Computer Graphics*, 2:337–354, 1996.

[9] B. Lucas and T. Kanade. An Iterative Image Registration Technique with an Application to Stereo Vision. In *International Joint Conference on Artificial Intelligence*, pages 674–679, 1981.

[10] J-M. Odobez and P. Bouthemy. Robust multiresolution estimation of parametric motion models. *Journal of Visual Communication and Image Representation*, 6:348–365, 1995.

[11] D. Rueckert, L.I. Sonda, C. Hayes, D. Hill, M. Leach, and D. Hawkes. Nonrigid registration using free-form deformations: Application to breast MR images. *IEEE Transactions on Medical Imaging*, 18:712–721, 1999.

[12] T. Sederberg and S. Parry. Free-Form Deformation of Solid Geometric Models. In *ACM SIGGRAPH*, volume 4, pages 151–160, 1986.

[13] J. Wang and E. Adelson. Representing Moving Images with Layers. *IEEE Transactions on Image Processing*, 3:625–638, 1994.

[14] J. Weickert and C. Schn *o*rr. Variational optic flow computation with a spatio-temporal smoothness constraint. *Journal of Mathematical Imaging and Vision*, 14:245–255, 2001.