OPTICAL FLOW AND DEPTH FROM MOTION FOR OMNIDIRECTIONAL IMAGES USING A TV-L1 VARIATIONAL FRAMEWORK ON GRAPHS

Luigi Bagnato, Pascal Frossard, Pierre Vandergheynst

Ecole Polytechnique Fédérale de Lausanne (EPFL)
Signal Processing Laboratory

ABSTRACT

This paper deals with the problem of efficiently computing the optical flow of image sequences acquired by omnidirectional (nearly full field of view) cameras. We formulate the problem in the natural spherical geometry associated with these devices and extend a recent TV-L1 variational formulation for computing the optical flow [1]. The discretization of differential operators occurring in this formulation turns out to be an extremely sensitive point, in particular for the TV part of our algorithm. We show that these difficulties can be very efficiently overcome using a graph-based formulation of TV denoising, which we solve by introducing a graph version of Chambolle's algorithm [2]. A slight modification of the original framework allows us to solve the depth from motion problem using the same techniques. In both cases, our graph-based algorithms provide computationally efficient solutions and significantly outperform naive implementations based on direct discretization of the operators, or on neglecting the influence of geometry.

Index Terms— Graphs, Manifold, Sparsity, Optical Flow

1. INTRODUCTION

Optical flow estimation is a traditional but key problem in vision systems, and it has attracted considerable attention since the seminal work of Horn and Schunck [3] (see for example [4] for a survey). The main objective of optical flow estimation methods is to compute a flow field that represents the motion in consecutive image frames. The main assumption in such problems is the brightness consistency, i.e., pixel intensity values do not change during motion between successive frames. However, optical flow estimation is a highly ill-posed inverse problem. Using pure image intensity-based constraints generally results in an under-determined system of equations, from which only normal flow can be computed. In order to solve this problem some kind of regularization is needed to obtain displacement fields that are physically meaningful. Some regularized variational formulations, more particularly those based on sparsity constraints, have been proved very effective. In [1], Zach et al. provide a real time optical flow algorithm with striking performance. They suggest to use a TV norm constraint to regularize the optical flow field and the L^1 norm to penalize deviations from the brightness consistency assumption, therefore handling outliers in a more graceful way than with the more traditional L^2 data fidelity term.

Even if most vision systems considered in optical flow estimation approaches are traditional planar cameras, omnidirectional imagers, such as catadioptric cameras, have recently sparked tremendous interest. Because of their (nearly) full field of view, these sensors are particularly attractive for applications like video surveillance and mobile robotics. Data captured by such devices, however, suffer from rather complex geometrical distortion (see Figure 3 for example). Current optical flow algorithms assume planar sensors (or equivalently small field of view) and cannot be adapted seamlessly to the particular geometry in omnidirectional images. It has been shown that processing omnidirectional images in their natural spherical geometry clearly leads to a more precise estimation of the optical flow, albeit at the price of increased computational complexity (see for example [5, 6] and references therein). For instance, the authors in [6] propose an optical flow variational formulation on Riemaniann manifolds: their approach is based on the minimization of a quadratic cost functional, following the classical approaches of [3, 7], but it is not very efficient from a computational point of view.

In this paper, we show that it is possible to extend very efficient variational approaches, while naturally handling the geometry of omnidirectional images. We propose an algorithm for $TV - L^1$ optical flow estimation on the 2-sphere, where the images are modeled as weighted graphs. The connections in the graph are given by the topology of the manifold and their weights by their geodesic distances. Our graph-based solutions are both very fast and precise, due to the efficient representation of the sensor geometry. Experimental results with both synthetic spherical images and natural images from a catadioptric sensor confirm the validity of our approach. To the best of our knowledge, this is the first time that graph-based techniques are applied to solve the optical flow estimation problem on a manifold. As a matter of fact the algorithm is general enough to be applied to other tasks, and we also illustrate its performance in a depth map estimation problem.

2. VARIATIONAL OPTICAL FLOW

Variational methods are among the most successful approaches to calculate the optical flow between two consecutive image frames I_0 and I_1. For simplicity let us assume that I_0 and I_1 are defined on a two dimensional manifold $\mathbf{M}$ embedded in $\mathbb{R}^3$. We use the notation $\mathbf{x} = (x^1, x^2)$ for a local system of coordinates and $\mathbf{y} = (y^1, y^2, y^3)$ for a cartesian system of coordinates. If $\nabla_{\mathbf{M}}$ and G are respectively the gradient operator and the metric tensor on the manifold $\mathbf{M}$, then the following holds:

$$\nabla I_1(\mathbf{y}) = G^{-1}\nabla_{\mathbf{M}} I_1(\mathbf{x}). \tag{1}$$

Under the brightness consistency assumption, we have $I_0(\mathbf{y}) - I_1(\mathbf{y}+\mathbf{u}) = 0$, where $\mathbf{u}$ is the displacement field between the frames. We can linearize the brightness consistency constraint around $\mathbf{y}$ as:

$$I_1(\mathbf{y}) - (\nabla I_1(\mathbf{y}))^T\mathbf{u} - I_0(\mathbf{y}) = 0. \tag{2}$$

Using (1) we obtain the following data constraint equation:

$$I_1(\mathbf{x}) + \langle \mathbf{u}_{\mathbf{M}}, \nabla_{\mathbf{M}} I_1(\mathbf{x})\rangle - I_0(\mathbf{x}) = 0, \tag{3}$$

where $\mathbf{u}_{\mathbf{M}}$ is the optical flow on the manifold and the scalar product is computed in the tangent plane via

$$\langle \mathbf{u}_{\mathbf{M}}, \nabla_{\mathbf{M}} I_1(\mathbf{x})\rangle = (G^{-1}\nabla_{\mathbf{M}} I_1(\mathbf{x}))^T\mathbf{u}_{\mathbf{M}}. \tag{4}$$

A general variational formulation of the optical flow problem consists in finding $\mathbf{u}_{\mathbf{M}}$ that minimizes the following functional:

$$J = \int_\Omega \psi(\mathbf{u}_{\mathbf{M}}, \nabla_{\mathbf{M}} u^i_{\mathbf{M}})\,\mathrm{d}\Omega + \lambda\int_\Omega \rho(I_0, I_1, \mathbf{u}_{\mathbf{M}})\,\mathrm{d}\Omega, \tag{5}$$

where $u^i_{\mathbf{M}}$ refers to the coordinate i in $\mathbf{u}_{\mathbf{M}}$. We can identify two distinct terms: ψ is the regularization term containing the prior on the flow field $\mathbf{u}_{\mathbf{M}}$, ρ is the data fidelity function. A particular formulation with a total variation regularization and a robust L^1 norm for the data fidelity term has been shown to efficiently preserve discontinuities in the flow field and increase robustness against illumination changes, occlusions and noise, see [1]. In this case, both terms of Eq. (5) reads:

$$\psi(\mathbf{u}_{\mathbf{M}}, \nabla_{\mathbf{M}} u^i_{\mathbf{M}}) = \psi(\nabla_{\mathbf{M}} u^i_{\mathbf{M}}) = \sum_i |G^{-1}\nabla_{\mathbf{M}} u^i_{\mathbf{M}}|, \tag{6}$$

$$\rho(\mathbf{u}_{\mathbf{M}}) = I_1(\mathbf{x}) + \langle \mathbf{u}_{\mathbf{M}}, \nabla_{\mathbf{M}} I_1(\mathbf{x})\rangle - I_0(\mathbf{x}). \tag{7}$$

Unfortunately the resulting functional J is not strictly convex and poses severe computational difficulties. Following [1], we propose a convex relaxation and we rewrite the functional as:

$$J = \int_\Omega \psi(\nabla_{\mathbf{M}} u^i_{\mathbf{M}}) + \frac{1}{2\theta}|\mathbf{u}_{\mathbf{M}} - \mathbf{v}|^2 + \lambda\rho(\mathbf{v})\mathrm{d}\Omega, \tag{8}$$

where $\mathbf{v}$ is an auxiliary variable that should be as close as possible to $\mathbf{u}_{\mathbf{M}}$. The minimization must now be performed with respect to both the variables $\mathbf{u}_{\mathbf{M}}$, $\mathbf{v}$ and the solution can be then obtained by an iterative two step procedure:

1. For $\mathbf{u}_{\mathbf{M}}$ fixed, solve:

$$\min_{\mathbf{v}} \left\{ \int_\Omega \frac{1}{2\theta}|\mathbf{u}_{\mathbf{M}} - \mathbf{v}|^2 + \lambda\rho(\mathbf{v})\mathrm{d}\Omega \right\}. \tag{9}$$

2. For $\mathbf{v}$ fixed, solve:

$$\min_{\mathbf{u}_{\mathbf{M}}} \left\{ \int_\Omega \psi(\nabla_{\mathbf{M}} u^i_{\mathbf{M}}) + \frac{1}{2\theta}|\mathbf{u}_{\mathbf{M}} - \mathbf{v}|^2\mathrm{d}\Omega \right\}. \tag{10}$$

The minimization in the first step is straightforward since the functional does not depend on the derivatives of $\mathbf{v}$, and the solution can be found pointwise (see Section 3). The minimization in Eq. (10) corresponds to the total variation image denoising model, for which Chambolle proposed an efficient fixed point algorithm [2]:

$$\begin{aligned} v^i &= u^i_{\mathbf{M}} - \theta div\mathbf{p}_i && i \in \{1,2\},\\ \mathbf{p}_i^{n+1} &= \frac{\mathbf{p}_i^n + \tau\nabla(div\mathbf{p}_i^n - u^i_{\mathbf{M}}/\theta)}{1+\tau|\nabla(div\mathbf{p}_i^n - u^i_{\mathbf{M}}/\theta)|} && i \in \{1,2\}. \end{aligned} \tag{11}$$

Note that in this equation, with an abuse of notation, ∇ is a generic discrete differential operator while div is its adjoint operator: $div = -\nabla^*$. The success of this algorithm depends strongly on a stable discretization of the gradient operator $\nabla_{\mathbf{M}}$ on the manifold, which is not always straightforward. In spherical coordinates, a simple discretization obtained from finite differences reads:

$$\begin{aligned} \nabla_\theta f(\theta_i,\phi_j) &= \frac{f(\theta_{i+1},\phi_j) - f(\theta_i,\phi_j)}{\Delta\theta},\\ \nabla_\phi f(\theta_i,\phi_j) &= \frac{1}{\sin\theta_i}\left(\frac{f(\theta_i,\phi_{j+1}) - f(\theta_i,\phi_j)}{\Delta\phi}\right). \end{aligned} \tag{12}$$

It contains a $(\sin\theta)^{-1}$ term that induces very high values around the poles (i.e. $\theta \simeq 0$ and $\theta \simeq \pi$) and can cause numerical instability. In the next section, we propose to define discrete differential operators on weighted graphs (i.e. discrete manifold) as a general way to deal with geometry in a coordinate-free fashion and obtain a stable solution to the variational problem.

3. GRAPH BASED FRAMEWORK

We represent our manifold (i.e., the imaging surface) as a weighted graph where the vertices represent image pixels and edges define connections between pixels (i.e., the topology of the surface). A weighted graph $\Gamma = (V, E, w)$ consists of a set of vertices V, a set of vertices pairs $E \subseteq V \times V$, and a weight function $w : E \mapsto \mathbb{R}$ satisfying $w(u,v) > 0$ and $w(u,v) = w(v,u)\ \forall (u,v) \in E$. We define our differential operators over Γ as in [8]:

$$\begin{aligned} &\text{Gradient:} && (\nabla^w f)(u,v) = \sqrt{w(u,v)}\left(\frac{f(u)}{\sqrt{d(u)}} - \frac{f(v)}{\sqrt{d(v)}}\right)\\ &\text{Divergence:} && (div^w F)(u) = \sum_{u\sim v}\sqrt{\frac{w(u,v)}{d(v)}}\,(F(v,u) - F(u,v)) \end{aligned} \tag{13}$$

where $d : V \mapsto \mathbb{R}$ is the degree function defined as: $d(v) = \sum_{u \sim v} w(u, v)$. The weight $w(u, v)$ is typically defined as a decreasing function of the geodesic distance between the vertices u and v.

Now that discrete operators are settled, we can give the complete algorithm to solve the problem described in Section 2. The solution of the first step of the algorithm (Eq. (9)) can be obtained by the following thresholding scheme:

$$\mathbf{v} = \mathbf{u_M} + \begin{cases} \lambda\theta G^{-1}\nabla_{\mathbf{M}} I_1 & \text{if} \rho(\mathbf{u_M}) < -\lambda\theta |G^{-1}\nabla_{\mathbf{M}} I_1|^2 \\ -\lambda\theta G^{-1}\nabla_{\mathbf{M}} I_1 & \text{if} \rho(\mathbf{u_M}) > \lambda\theta |G^{-1}\nabla_{\mathbf{M}} I_1|^2 \\ -\frac{\rho(\mathbf{u_M}) G^{-1}\nabla_{\mathbf{M}} I_1}{|G^{-1}\nabla_{\mathbf{M}} I_1|^2} & \text{if} |\rho(\mathbf{u_M})| < \lambda\theta |G^{-1}\nabla_{\mathbf{M}} I_1|^2. \end{cases} \tag{14}$$

The second step is performed by substituting the definitions of the operators in Eq. (13) in the iterative algorithm described in Eq. (11). Note that similar approaches for TV denoising on graphs have been proposed in [9, 10] in the framework of non-local regularization.

4. EXPERIMENTAL RESULTS

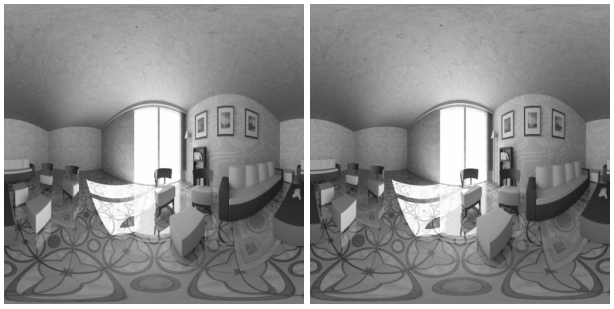
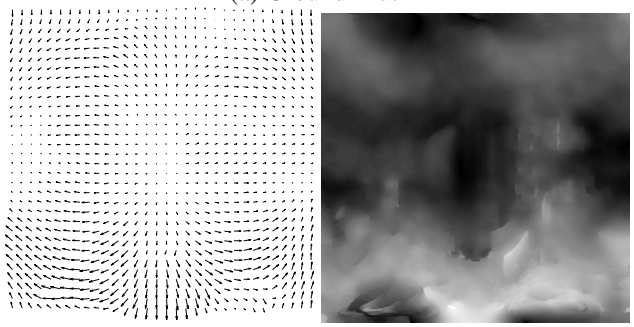


Fig. 1. Synthetic spherical sequence.

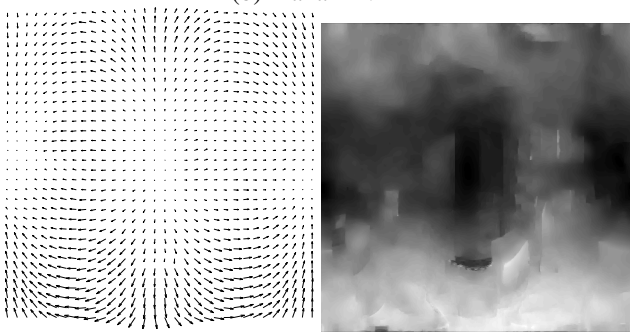
We tested our algorithm on both synthetic and natural video sequences. The synthetic sequence represented in Figure 1 is a spherical rendering from a 3D model of the interior of a living room. A sphere is conveniently represented in a spherical coordinates system using the zenith angle $\theta \in [0, \pi]$ and the azimuth $\phi \in [0, 2\pi[$. In the image plane, vertical coordinates correspond to θ and horizontal to ϕ, such that the top of the image correspond to the north pole and the bottom to the south pole. In the sequences the spherical camera undergoes rigid translational motion, and for all of them we have, of course, ground truths.

We compare the performance of the proposed algorithm, called *GrH-TVL1*, with the one of a similar algorithm, called *Planar-TVL1*, which uses planar differential operators. Or, in other words, we consider the images in Figure 1 as planar images. The optical flow estimation results are shown in Figure 2: *Planar-TVL1* misbehaves near the poles (top and bottom of the image) where the spherical equiangular grid presents the highest distortion. We also provide comparisons with the ground truth information in Table 1, which confirms that *GrH-TVL1* is by far more accurate than *Planar-TVL1*.

(a) Ground Truth



(b) Planar-TVL1



(c) GrH-TVL1

Fig. 2. Optical flow estimation for the synthetic sequence: vector representation (left), module of the optic flow field (right).

Error	Module (SSE)	Angle (AAE)
Planar-TVL1	9.5354	0.2839
GrH-TVL1	2.02	0.1509

Table 1. Comparison. Sum of Square Error (SSE) for module and average angular error (AAE) with respect to ground truth.

The second video sequence has been captured with a catadioptric sensor and represents an object moving in a simple scene (see Figure 3). The catadioptric images are first projected on a sphere before optical flow estimation with *GrH-*

TVL1. The results are visually illustrated in Figure 4, where we see that the optical flow estimation permits to drastically reduce the residual energy after motion compensation.

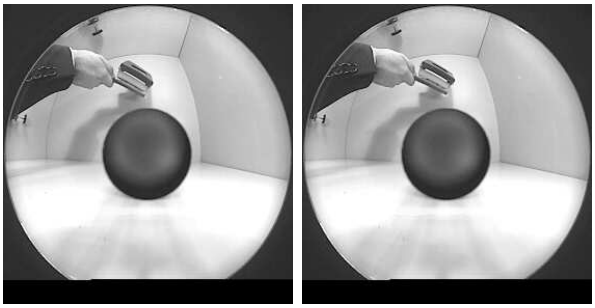


Fig. 3. Catadioptric sequence

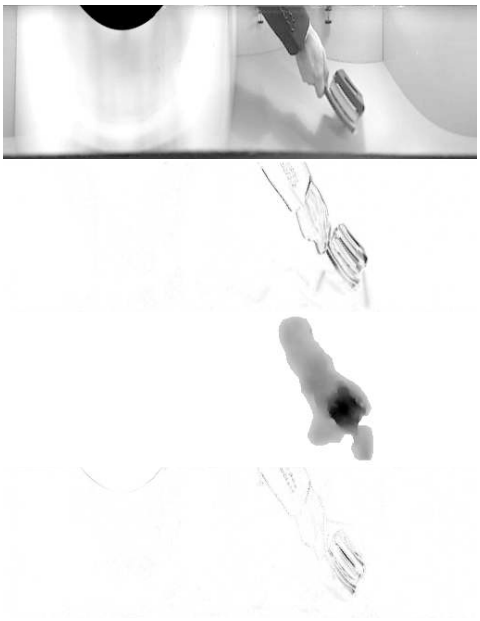
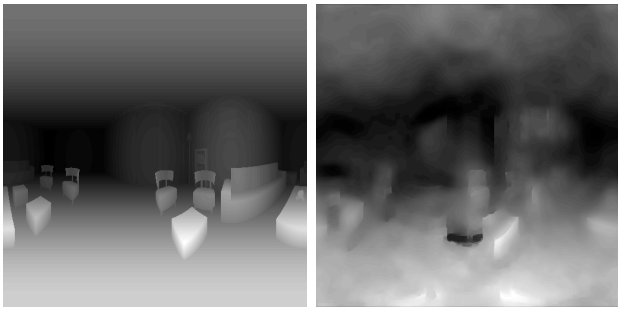


Fig. 4. Optical flow estimation from catadioptric images. From Top to bottom: image after projection on the sphere; frame difference $I_1 - I_0$; optical flow field (module); image residual $I_1(\mathbf{x} + \mathbf{u}) - I_0$, after motion compensation.

Finally, we illustrate the genericity of our graph-based algorithm, and present its application to a depth map estimation problem. We adapt our algorithm by simply rewriting the functional (Eq. (8)) as a function of $D(\phi, \theta)$, which is the inverse of the depth map:

$$\begin{aligned} J &= \int_\Omega |G^{-1}\nabla_{\mathbf{M}} D)| + \frac{1}{2\theta}(D - Z)^2 \\ &+ \lambda |I_1(\mathbf{x}) + D\langle \mathbf{t}, \nabla_{\mathbf{M}} I_1\rangle - I_0(\mathbf{x}))| \mathrm{d}\Omega, \end{aligned} \quad (15)$$

where $\mathbf{t}$ is the projection of the true 3D motion field on the spherical surface of the camera and Z is again an auxiliary variable. The depth estimation results illustrated in Figure 5 demonstrate that our graph-based method is able to efficiently solve the difficult problem of depth map estimation.



Ground Truth Estimated Depth Map

Fig. 5. Depth map estimation in the synthetic sequence.

5. CONCLUSIONS

We presented a graph-based TV-L^1 variational framework on Riemaniann manifolds. We applied the proposed approach to optical flow and depth estimation problems. Experimental results on both synthetic and natural omnidirectional images demonstrated the efficiency of the proposed method.

6. REFERENCES

[1] C. Zach, T. Pock, and H. Bischof, "A duality based approach for realtime tv-l^1 optical flow," *Pattern Recognition*, pp. 214–223, 2007.

[2] A. Chambolle, "An algorithm for total variation minimization and applications," *J. Math. Imaging. Vis.*, vol. 20, no. 1-2, pp. 89–97, Jan 2004.

[3] B. Horn and B. Schunck, "Determining optical flow," *Artifical Intelligence*, pp. 185–203, 1981.

[4] S. S. Beauchemin and J. L. Barron, "The computation of optical flow," *ACM Computing Surveys (CSUR)*, vol. 27, no. 3, pp. 433–466, 1995.

[5] K. Daniilidis, A. Makadia, and T. Bulow, "Image processing in catadioptric planes: Spatiotemporal derivatives and optical flow computation," in *OMNIVIS '02: Proceedings of the Third Workshop on Omnidirectional Vision*, Washington, DC, USA, 2002, p. 3, IEEE Computer Society.

[6] A. Imiya, A. Torii, and H. Sugaya, *Optical Flow Computation of Omni-Directional Image*, pp. 143–162, Springer Netherlands, 2006.

[7] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *Proc. of Imaging Understanding Workshop*, 1981, pp. 121–130.

[8] D. Zhou and B. Scholkopf, "A regularization framework for learning from graph data," *Proc. of ICML Workshop on Statistical Relational Learning and Its ...*, Jan 2004.

[9] G. Gilboa and S. Osher, "Nonlocal operators with applications to image processing," Tech. Rep. 07-23, UCLA CAM Report, July 2007.

[10] G. Peyré, S. Bougleux, and L. D. Cohen, "Non-local regularization of inverse problems," in *Proc. 10th European Conference on Computer Vision (ECCV'08)*, Marseille (France), October 2008, pp. 57–68.