

# MOTION INPAINTING BY AN IMAGE-BASED GEODESIC AMLE METHOD

*M. Oliver      L. Raad      C. Ballester      G. Haro*

Universitat Pompeu Fabra, Barcelona, Spain

## ABSTRACT

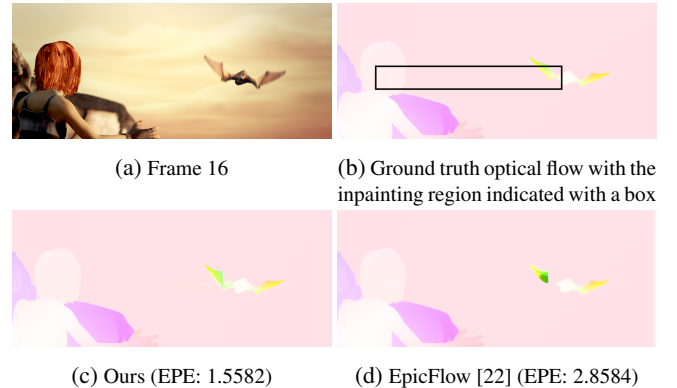
This work presents an automatic method for optical flow inpainting. Given a video, each frame domain is endowed with a Riemannian metric based on the video pixel values. The missing optical flow is recovered by solving the Absolutely Minimizing Lipschitz Extension (AMLE) partial differential equation on the Riemannian manifold. An efficient numerical algorithm is proposed using eikonal operators for nonlinear elliptic partial differential equations on a finite graph. The choice of the metric is discussed and the method is applied to optical flow inpainting and sparse-to-dense optical flow estimation, achieving top-tier performance in terms of End-Point-Error (EPE).

**Index Terms**— Optical flow inpainting, sparse to dense methods, partial differential equations, Absolute Minimal Lipschitz Extension, anisotropic methods.

## 1. INTRODUCTION

Optical flow or motion inpainting is a pervasive problem in many areas of computer vision which range from semantic video analysis and video editing to optical flow estimation in occlusion and disocclusion regions. Given a video, its optical flow is the vector field, defined on each point of the video domain, that describes the apparent motion of the objects through the frames. One of the capital difficulties of the optical flow estimation are the occlusions (regions that become not visible or emerge between one frame and the next one) where its estimation becomes impossible due to the lack of correspondence between the two consecutive frames. Therefore, the optical flow is partially missing in areas more or less large of the video whereas for most of the applications its completion is essential. For instance, in cinema post-production a completed optical flow is often needed for the elimination of unwanted (moving or static) objects. Other applications are automatic assistance of sensor-based optical flow estimation where the sensor acquisition usually produces large regions without optical flow data (e.g., in Kitty Vision benchmark [19]).

The motion inpainting problem has been addressed previously in the literature for different purposes. In order to inpaint the flow in the occlusion areas, Matsushita et al. [17] and Strobel et al. [26] extended the Telea-inpainting method [27]



**Fig. 1:** Some optical flow inpainting results for a frame of the temple\_3 sequence of MPI-Sintel benchmark.

to optical flow, i.e., they assume that the motion variation is locally small and propagate the optical flow according to a weighting function which depends on the Euclidean distance and the color difference among the interpolated pixel and its neighbours. Kondermann et al. [13] proposed a postprocess of the optical flow in order to improve it: the optical flow is retained at points where it is reliable and is then densified by minimizing the  $L^2$  norm of the spatio-temporal gradient of the flow. Berkels et al. [6] proposed to recover the optical flow in non-reliable regions using a TV-type anisotropic functional and a rotation-invariant regularizer was proposed by Palomares et al. [21]. On the other hand, Ince and Konrad [11] introduced a variational method for the joint estimation of optical flow and occlusions while extrapolating the optical flow in occluded areas by means of anisotropic diffusion based on the image gradients. Leordeanu et al. [15] and Revaud et al. [22] proposed a sparse-to-dense optical flow estimation method that takes as input an initial set of sparse matches. In a first stage, the flow is densified (completed) by fitting a local affine model, that uses edge-aware distances in the case of EpicFlow [22]. Then, the densified flow is refined by minimizing an optical flow energy functional.

In this paper we use the Absolutely Minimizing Lipschitz Extension (AMLE) model in a Riemannian manifold in order to take advantage of the geometric information given by the video frames. Given a video and an incomplete motion field, we endow each 2D frame domain with a Riemannian metric based on the video values and propose to recover the missing

optical flow by solving the AMLE partial differential equation on the 2D Riemannian manifold from the known values on the boundary of the interpolation domain. Each of the coordinates of the optical flow is thus reconstructed with this metric-based anisotropic interpolation. The AMLE equation,  $\Delta_\infty u = 0$ , where  $\Delta_\infty u$  is also called the infinity Laplacian, was introduced by G. Aronsson in [2, 3] and uniqueness of viscosity solutions was proved in [12] (see also [4] for a review). The AMLE allows to interpolate the missing value of a function on isolated points and curves. It appeared as one of the interpolating operators satisfying a set of suitable axioms in [9]. This axiomatic approach was extended in [8, 25] to interpolate data given on a set of curves on a surface in  $\mathbb{R}^3$ . The AMLE on a manifold was applied in [14] for interpolating depth data in images or videos where large regions of incomplete depth information often appear. On the other hand, the classical AMLE has been used in [1] for the interpolation of digital elevation models.

The method proposed here is applied in two different scenarios: optical flow inpainting in large regions and densification of an optical flow from a sparse set of initial matches. The experiments show that in general our results outperform those of EpicFlow [22], which has become a reference method for optical flow estimation and a standard technique for post-processing an estimated and filtered optical flow.

The remainder of the paper is organized as follows. Section 2 presents the proposed approach, including the proposed numerical algorithm and its analysis. Section 3 presents experimental results for the mentioned applications. Finally, Section 4 concludes the paper.

## 2. THE MODEL

Given a video  $I(\mathbf{x}, t)$  defined on  $\Omega \times \{1, \dots, T\}$ , where  $\Omega \subset \mathbb{R}^2$  denotes the image frame domain and  $\{1, \dots, T\}$  is the set of discrete times, let  $\mathbf{v} = (v_1, v_2)$  be the optical flow of the video  $I$ , i.e., the apparent motion between a pixel  $\mathbf{x} \in \Omega \setminus \Omega_0(t)$  at time  $t$  and the corresponding one at time  $t + 1$ .

We assume that, at time  $t$ ,  $\mathbf{v}(\mathbf{x}, t)$  is unknown on a region  $\Omega_0(t) \subset \Omega$  whose boundary, denoted by  $\partial\Omega_0$ , consists of a finite union of smooth curves and possibly isolated points. In order to complete  $\mathbf{v}$  in  $\Omega_0(t)$  by an appropriate interpolation taking into account the objects in the video, we endow the whole domain  $\Omega$ , at each time  $t$ , with a metric  $g(t)$ . Let  $\mathcal{M}(t) = (\Omega, g(t))$  be the corresponding Riemannian manifold. We propose to complete  $\mathbf{v}$  in  $\Omega_0(t)$  with the motion field  $\mathbf{u} = (u_1, u_2)$  such that  $u_1$  and  $u_2$  are solutions, respectively, of the geodesic AMLE, given by the PDE

$$\Delta_{\infty, g} u_i = 0 \quad \text{in } \Omega_0(t) \quad \text{s.t.} \quad u_i|_{\partial\Omega_0(t)} = v_i, \quad (1)$$

for  $i = 1, 2$ , respectively. We also use Neumann boundary conditions on  $\partial\Omega$ . Here we have denoted

$$\Delta_{\infty, g} u_i := D_{\mathcal{M}}^2 u_i \left( \frac{\nabla_{\mathcal{M}} u_i}{|\nabla_{\mathcal{M}} u_i|}, \frac{\nabla_{\mathcal{M}} u_i}{|\nabla_{\mathcal{M}} u_i|} \right) \quad (2)$$

where  $\nabla_{\mathcal{M}} u_i$  and  $D_{\mathcal{M}}^2 u_i$  denote the gradient and the Hessian on the manifold. To simplify the notation, we have omitted the dependence on  $t$  of  $g$  and  $\mathcal{M}$ ; we will also drop the subindex  $i$  in what follows. When  $g$  is the Euclidean metric, the operator in (2) is known as the infinity Laplacian. The metric  $g$  is defined in this paper based on the local geometry and texture content and several definitions are tested. For instance,  $g$  can be given by affine covariant structure tensors [10] or taking into account spatial distances and photometric similarities as detailed in the following section, where we consider the interpolation problem (1) on finite graphs.

### 2.1. The geodesic AMLE on a finite graph

We solve the AMLE equation on the manifold with a numerical algorithm which is based on the eikonal operators for nonlinear elliptic PDEs on a finite graph, which was proposed by Manfredi et al. [16, 20]. In particular, we consider the discrete grid of  $\Omega$  as the nodes of a finite weighted graph  $G$ . If  $\mathbf{x}$  and  $\mathbf{y}$  are neighbouring pixels, we define its distance by one of the following simple possibilities

$$d_1(\mathbf{x}, \mathbf{y}) = \sqrt{(1 - \lambda) \|I(\mathbf{x}, t) - I(\mathbf{y}, t)\|^2 + \lambda \|\mathbf{x} - \mathbf{y}\|^2} \quad (3)$$

$$d_2(\mathbf{x}, \mathbf{y}) = (1 - \lambda) \|I(\mathbf{x}, t) - I(\mathbf{y}, t)\| + \lambda \|\mathbf{x} - \mathbf{y}\| \quad (4)$$

where  $\lambda \in [0, 1]$ . We also include

$$d_3(\mathbf{x}, \mathbf{y}) = (1 - \lambda) \|I(\mathbf{x}, t) - I(\mathbf{y}, t)\|^2 + \lambda \|\mathbf{x} - \mathbf{y}\|^2 \quad (5)$$

which is a semimetric, i.e., it does not satisfy the triangle inequality. We will use  $d(\mathbf{x}, \mathbf{y})$  to refer to anyone of them. Given a path  $\gamma = \{\gamma(i)\}_{i=0}^m$  on the graph  $G$  joining two points,  $\mathbf{x} = \gamma(0)$  and  $\mathbf{y} = \gamma(m)$ , its length is defined as usual by  $L^g(\gamma) = \sum_{i=0}^{m-1} d(\gamma(i), \gamma(i+1))$ . Given any two points  $\mathbf{x}$  and  $\mathbf{y}$  on the grid, then the geodesic distance  $d^g(\mathbf{x}, \mathbf{y})$  is

$$d^g(\mathbf{x}, \mathbf{y}) = \inf \{L^g(\gamma) : \gamma \text{ is a curve joining } \mathbf{x} \text{ and } \mathbf{y}\}.$$

This distance can be computed using Dijkstra's algorithm. In practice, for any pair of points  $\mathbf{x}, \mathbf{y}$  we approximate  $d^g(\mathbf{x}, \mathbf{y})$  by  $d(\mathbf{x}, \mathbf{y})$ , where  $d(\mathbf{x}, \mathbf{y})$  is defined by (3)-(5). Let us notice that more sophisticated metrics are possible.

Given a point  $\mathbf{x}$  on the grid, let  $\mathcal{N}(\mathbf{x})$  be a neighborhood of  $\mathbf{x}$ . Following [16], we consider the positive and negative eikonal operators given, respectively, by

$$\|\nabla u(\mathbf{x})\|_{\mathcal{M}}^+ = \max_{\mathbf{y} \in \mathcal{N}(\mathbf{x})} \frac{u(\mathbf{y}) - u(\mathbf{x})}{d^g(\mathbf{x}, \mathbf{y})}, \quad (6)$$

$$\|\nabla u(\mathbf{x})\|_{\mathcal{M}}^- = \min_{\mathbf{z} \in \mathcal{N}(\mathbf{x})} \frac{u(\mathbf{z}) - u(\mathbf{x})}{d^g(\mathbf{x}, \mathbf{z})}. \quad (7)$$

Then, the discrete infinity Laplacian corresponds to

$$\Delta_{\infty, g} u(\mathbf{x}) = \frac{\|\nabla u(\mathbf{x})\|_{\mathcal{M}}^+ + \|\nabla u(\mathbf{x})\|_{\mathcal{M}}^-}{2}. \quad (8)$$

We solve (1) with the following iterative discrete scheme

$$u^{k+1}(\mathbf{x}) = \frac{d^g(\mathbf{x}, \mathbf{z})u^k(\mathbf{y}) + d^g(\mathbf{x}, \mathbf{y})u^k(\mathbf{z})}{d^g(\mathbf{x}, \mathbf{z}) + d^g(\mathbf{x}, \mathbf{y})} \quad (9)$$

where  $\mathbf{y}$  and  $\mathbf{z}$  are the pixels providing the maximum and minimum in (6) and (7). This scheme is applied for  $\mathbf{x} \in \Omega_0$  keeping the values of  $v_1(\mathbf{x})$ , respectively  $v_2(\mathbf{x})$ , on the known region  $\Omega \setminus \Omega_0$  for all  $k$ .

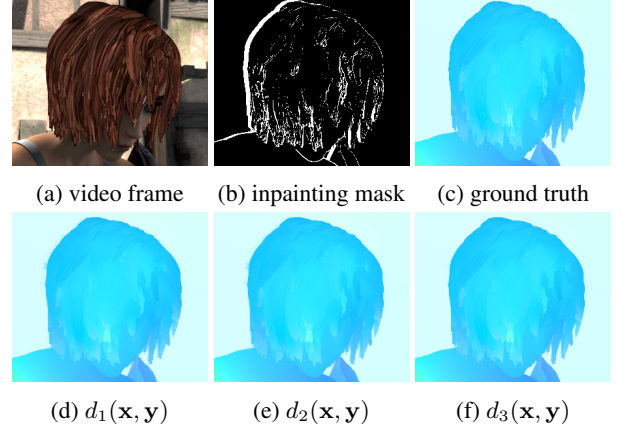
This scheme is embedded in a multiscale approach: the input optical flow and corresponding video frame are down-scaled to a set of scales and the solution is computed at each one using (9). The inputs are downsampled by a factor of 2 using  $2 \times 2$  block averages; bilinear interpolation is used for upsampling. At the coarsest level, the unknowns are initialized to zero; the other scales are initialized by upsampling the solution of the previous scale. The PDE defined in (1) has a unique solution [9], so the result of the multiscale approach is the same as in the single scale scheme. The multiscale pyramid provides just an initialization closer to the solution, leading to a faster convergence.

As mentioned in [20] the solution of the numerical scheme (9) converges to the solution of (1) when the local spatial resolution  $dx$  and the local directional resolution  $d\theta$  tend to 0. We have evaluated three different types of neighbourhoods  $\mathcal{N}(\mathbf{x})$ . Considering the square of radius  $dx$  centered in  $\mathbf{x}$ , in the first type of neighbourhood we keep for each possible direction in the square the point of lowest distance to  $\mathbf{x}$ . For the second type we keep all the points whose distance to  $\mathbf{x}$  is lower than  $dx$ . For the third type we keep only the points whose distance to  $\mathbf{x}$  is equal to  $dx$ . The three options have the same directional resolution  $d\theta$  but differ in the spatial resolution  $dx$ . We choose the first configuration with  $dx = 2$  for all the experiments of the paper, which implies that the cardinality of  $\mathcal{N}(\mathbf{x})$  equals 16.

To experimentally analyze the behavior of our method depending on the considered metric  $g$  we show in Fig. 2 some results using  $d_1$ ,  $d_2$  and  $d_3$ . We choose to work with  $d_3(\mathbf{x}, \mathbf{y})$  given by (5), which produces slightly better results. Indeed, when part of an edge is subjective or weak (i.e., the two regions separated by the edge are similar) the proposed metrics do not penalize completely the propagation of the optical flow from one region to another; (5) does a better job discriminating these cases. Fig. 2 shows an example, notice the small leak in the left middle part of the head which is reduced in (f).

### 3. APPLICATIONS

In this section we show results for different applications, namely, optical flow inpainting, both in occlusion areas and in large holes where different types of motion have to be recovered, and optical flow densification. All the results are compared to those of EpicFlow [22]. The authors in [22] propose to estimate the optical flow in two steps: the first step is an edge-preserving interpolation of a sparse optical flow



**Fig. 2:** Comparison of the three tested possibilities for the metric illustrated in an experiment of completion of the optical flow in the occlusion areas (white regions in image b).

generated with Deepmatching [23], the second step consists of a variational energy minimization that refines the dense flow achieved in the first step. We compare our results to those in their first step, since both of them are completions of an initial flow that use only the current frame, i.e., we are not estimating the optical flow given two consecutive frames.

We use the MPI-Sintel [7], Middlebury [5] and Kitti [19] datasets. Four experiments were performed for the first two datasets: optical flow inpainting in occlusion areas, densification of ground truth (gt) flow values at random positions (1%, 5% and 30% of values kept), densification of optical flow values given by Deepmatching [23], densification of gt values at the sparse positions given by Deepmatching (DM). For some frames of MPI-Sintel dataset, we also perform inpainting at manually added holes. In each case, we compute the End-Point-Error (EPE) excluding the given points, so as to not penalize EpicFlow interpolation which changes these values. Tables 1 and 2 report the EPE. Our method achieves lower error for all cases except for the densification of DM optical flow (sparse DM). This is probably due to the fact that the DM values are quantized values that also include errors and outliers. EpicFlow removes outliers and modifies the provided values which allows to compensate for the quantization and errors. Our method sticks to the given values propagating the error to the recovered optical flow. Depending on the application, one or the other behaviour could be preferred. We include in Fig. 4 an example of completion in large regions without motion information as in Kitti dataset [19].

Fig. 3 presents two different experiments and illustrates the behaviour of different interpolation methods: Total Variation (TV) regularization [24], rotation invariant regularization [21] (Rot-Inv), the interpolation step of EpicFlow [22], and the proposed method. Both type of experiments are done with the frame 10 of RubberWhale from the Middlebury dataset (image 3(a)), whose optical flow ground truth is available (3(b)). The first experiment is motion completion in differ-



**Fig. 3:** Comparison of different motion completion algorithms in two different cases: inpainting of large holes (holes in white in image (c) and results shown in the 2nd row) and flow densification from a sparse set of ground truth matches (shown in image (d) and results in 3rd row). The methods used are TV [24], R-Inv [21], EpicFlow (EF) [22] and ours. For the EpicFlow we consider only the interpolation step and overwrite the result with the ground truth flow outside the inpainting mask. Numbers in parenthesis show the EPE for each result.

ent large holes (shown in white in image 3(c)) and the corresponding results for the four evaluated methods are in the 2nd row. The second experiment is flow densification from a sparse set of matches (pixels where the motion is not known are shown in white in 3(d)) and the results can be seen in the 3rd row. One can observe that when no guide is used in the interpolation process (TV and Rotation-invariant methods) the discontinuities of the optical flow are not well recovered, since they are not aligned with the objects boundaries. This can be seen for instance, in the semi-circles in the second row of Fig. 3 where these methods fail, while the interpolation step of EpicFlow and our interpolation give satisfactory results. Our interpolation method cannot correctly inpaint the optical flow of rotating objects, as the wheel in the bottom-left. In this region the information provided by the metric-based guide is not useful (roughly homogeneous area). In such cases, our method reduces to a local isotropic average which can only solve translations. EpicFlow’s interpolation uses instead a local affine transformation, allowing to correctly recover rotations. In general, one can observe that visual results of our interpolation method are better (Fig. 1d).

#### 4. CONCLUSIONS

This paper presents a new method for optical flow completion which is based on the Absolutely Minimizing Lipschitz Extension equation (or the infinity Laplace equation) on the 2D

SINTEL	Ours - EPE	EF - EPE
occlusions	<b>5.4198</b>	6.8797
sparse 1%	<b>0.7061</b>	1.8532
sparse 5%	<b>0.4340</b>	1.4199
sparse 30%	<b>0.2241</b>	1.1212
sparse DM	4.4404	<b>4.1507</b>
sparse DM (gt)	<b>2.1360</b>	2.3802
hole	<b>1.7208</b>	1.9587

**Table 1:** Comparison of the EPE for EpicFlow and our method in different situations. The comparison is done over all the training Sintel dataset [7].

MIDDLEBURY	Ours - EPE	EF - EPE
sparse 1%	<b>0.1979</b>	0.3105
sparse 5%	<b>0.1053</b>	0.2426
sparse 30%	<b>0.0567</b>	0.1801
sparse DM	0.9216	<b>0.8112</b>
sparse DM (gt)	<b>0.2049</b>	0.2789

**Table 2:** Comparison of the EPE for EpicFlow and our method in different situations. The comparison is done in the Middlebury dataset on the optical flow corresponding to frames 10 and 11, where the groundtruth is available [5].



**Fig. 4:** A result on Kitti dataset [19] that contains large holes.

Riemannian manifold endowed with an appropriate metric, defined by the image frame, which acts as a guide for the resulting anisotropic diffusion. The proposed method has been analyzed in four different types of experiments: interpolation of sparse (ground truth or estimated) matches, i.e. optical flow densification, and motion completion both in occlusion areas and in large holes. The experimental results show how it outperforms the EpicFlow interpolation in practically all the situations. We have proposed three different simple metrics and as part of the future work we plan to study more sophisticated ones in order to further improve the results.

**Acknowledgements:** We thank Enric Meinhardt-Llopis for fruitful discussions and for kindly sharing his code [18]. This work is partially supported by MINECO/FEDER UE project, with reference TIN2015-70410-C2-1-R and by RISE project NoMADS.

#### 5. REFERENCES

- [1] A. Almansa, F. Cao, Y. Gousseau, and B. Rougé. Interpolation of digital elevation models using AMLE and related methods. *IEEE Trans. Geosci. Remote Sens.*, 2002.

- [2] G. Aronsson. Extension of functions satisfying Lipschitz conditions. *Ark. Mat.*, 1967.
- [3] G. Aronsson. On the partial differential equation  $u_x^2 u_{xx} + 2u_x u_y u_{xy} + u_y^2 u_{yy}^2 = 0$ . *Ark. Mat.*, 1968.
- [4] G. Aronsson, M. Crandall, and P. Juutinen. A tour of the theory of absolutely minimizing functions. *Bull. Am. Math. Soc.*, 2004.
- [5] S. Baker, D. Scharstein, J. Lewis, S. Roth, M. J. Black, and R. Szeliski. A database and evaluation methodology for optical flow. *Int. J. Comput. Vis.*, 2011.
- [6] B. Berkels, C. Kondermann, C. Garbe, and M. Rumpf. Reconstructing optical flow fields by motion inpainting. In *CVPR*, 2009.
- [7] D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black. A naturalistic open source movie for optical flow evaluation. In *ECCV*, 2012.
- [8] V. Caselles, L. Igual, and O. Sander. An axiomatic approach to scalar data interpolation on surfaces. *Numerische Mathematik*, 2006.
- [9] V. Caselles, J.-M. Morel, and C. Sbert. An axiomatic approach to image interpolation. *IEEE Trans. IP*, 1998.
- [10] V. Fedorov, P. Arias, R. Sadek, G. Facciolo, and C. Ballester. Linear Multiscale Analysis of Similarities between Images on Riemannian Manifolds: Practical Formula and Affine Covariant Metrics. *SIAM J Imaging Sci*, 2015.
- [11] S. Ince and J. Konrad. Occlusion-aware optical flow estimation. *IEEE Trans. IP*, 2008.
- [12] R. Jensen. Uniqueness of Lipschitz extensions: minimizing the sup norm of the gradient. *Arch. Ration. Mech. Anal.*, 1993.
- [13] C. Kondermann, D. Kondermann, and C. Garbe. Post-processing of optical flows via surface measures and motion inpainting. *Pattern Recognition*, 2008.
- [14] V. Lazcano. *Some Problems in Depth Enhanced Video Processing*. PhD. Thesis, 2016.
- [15] M. Leordeanu, A. Zanzfir, and C. Sminchisescu. Locally affine sparse-to-dense matching for motion and occlusion estimation. In *ICCV*, 2013.
- [16] J. J. Manfredi, A. M. Oberman, and A. P. Sviridov. Nonlinear elliptic partial differential equations and p-harmonic functions on graphs. *Differ Integr Equ*, 2015.
- [17] Y. Matsushita, E. Ofek, W. Ge, X. Tang, and H.-Y. Shum. Full-frame video stabilization with motion inpainting. *IEEE Trans. Pattern Anal. Mach. Intell*, 2006.
- [18] E. Meinhardt. AMLE from imscript. [https://github.com/mnhrdt/imscript/blob/master/src/misc/amle\\_recsep.c](https://github.com/mnhrdt/imscript/blob/master/src/misc/amle_recsep.c), 2017.
- [19] M. Menze and A. Geiger. Object Scene Flow for Autonomous Vehicles. In *CVPR*, 2015.
- [20] A. Oberman. A convergent difference scheme for the infinity Laplacian: construction of absolutely minimizing Lipschitz extensions. *Math. Comput.*, 2005.
- [21] R. P. Palomares, G. Haro, and C. Ballester. A rotation-invariant regularization term for optical flow related problems. In *ACCV*, 2014.
- [22] J. Revaud, P. Weinzaepfel, Z. Harchaoui, and C. Schmid. Epicflow: Edge-preserving interpolation of correspondences for optical flow. In *CVPR*, 2015.
- [23] J. Revaud, P. Weinzaepfel, Z. Harchaoui, and C. Schmid. Deepmatching: Hierarchical deformable dense matching. *International Journal of Computer Vision*, 120(3):300–323, 2016.
- [24] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: nonlinear phenomena*, 60(1-4):259–268, 1992.
- [25] O. Sander, V. Caselles, and M. Bertalmio. Axiomatic scalar data interpolation on manifolds. In *ICIP*, 2003.
- [26] M. Strobel, J. Diebold, and D. Cremers. Flow and color inpainting for video completion. In *GCPR*, 2014.
- [27] A. Telea. An image inpainting technique based on the fast marching method. *JGT*, 2004.