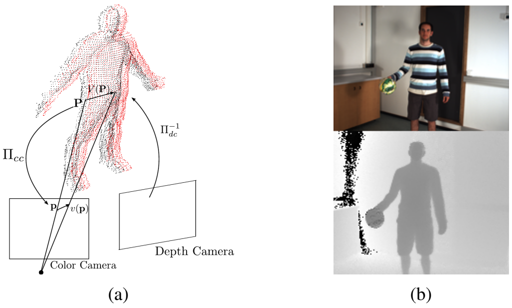
Scene Flow from Depth and Color Images

Antoine Letouzey
http://morpheo.inrialpes.fr/~letouzey
Benjamin Petit
http://morpheo.inrialpes.fr/~petit
Edmond Boyer
http://morpheo.inrialpes.fr/~boyer

Morpheo Team,
LJK / INRIA Grenoble Rhônes Alpes
Grenoble, France
{firstname}.{lastname}@inria.fr

In this paper we consider the problem of estimating a 3D motion field using multiple cameras. In particular, we focus on the situation where a depth camera and one or more color cameras are available, a common situation with recent composite sensors such as the Kinect. In this case, geometric information from depth maps can be combined with intensity variations in color images in order to estimate smooth and dense 3D motion fields. We propose a unified framework for this purpose, that can handle both arbitrary large motions and sub-pixel displacements. The estimation is cast as a linear optimization problem that can be solved very efficiently. The novelty with respect to existing scene flow approaches is that it takes advantage of the geometric information provided by the depth camera to define a surface domain over which photometric constraints can be consistently integrated in 3D. Experiments on real and synthetic data provide both qualitative and quantitative results that demonstrate the interest of the approach.



(a) (b)

Figure 1: (a) Projection of two consecutive surfaces onto the cameras, and (b) color and depth images.

Problem Formulation In order to estimate the 3D flow, we cast the problem as an minimization where data terms corresponding to photometric consistency constraints are combined with a regularization term that favors smooth motion fields:

$$\mathbf{E} = \mathbf{E}_{data} + \mathbf{E}_{smooth}. \tag{1}$$

Data terms enforce visual coherence of the computed displacement field while the regularization term imposes a deformation model with local rigidity constraints.

Visual Constraints As suggested by Xu *et al.* in their work on optical flow [3], we use two different kinds of photometric cues to deal with both large and small displacements. First we match sparse *visual features* (SIFT) between two consecutive color images. This information is not sensitive to the amplitude of the motion in the scene. For small details we use the well known *normal flow* information available at every pixels but only valid for small motion. Both cues contribute a term to $\mathbf{E}_{data}$ in equation 1.

Geometric Constraints The regularization stage is important for two reasons. First we need to propagate the sparse cues given by the visual features, and second the *aperture problem*, well known in optical flow estimation, extends from 2D to 3D. Hence the data term in our formulation is not sufficient to compute the scene flow. Unlike existing work [2], we choose to perform this regularization using 3D information given by the depth camera, instead of computing optical flow in the image domain and do a projection of this 2D flow on the depth maps. We extended Horn & Schunck's method [1] to 3D. This regularization enforces a global smoothness of the motion field in 3D. Therefore we do not suffer from 2D regularization-specific drawbacks, such as object boundaries and depth discontinuities oversmoothing. This geometric constraint yields the smoothing term in equation 1.

Formulation & Resolution We gather all the visual and geometric constraints into one single linear system of the following form :

$$\begin{bmatrix}\mathbf{L}\\ \mathbf{A}\end{bmatrix} V + \begin{bmatrix}\mathbf{0}\\ \mathbf{b}\end{bmatrix} = 0, \tag{2}$$

where $\mathbf{L}$ is the Laplacian matrix of the mesh associated to the depth map, V is a vector compounding the motion of all the scene points, and $\mathbf{A}$ and $\mathbf{b}$ stack all the motion constraints coming from data terms. The paper explains in details the construction of these matrices along with a discussion about Laplacian weights. This linear system is very sparse.

In practice, we propose a two-step algorithm. The first one handles large displacements, and the second recovers small motion details. This is done by adjusting the weight associated to each constraint and perform two consecutive resolutions of equation 2.

Results We tested our approach on both synthetic and real data. Figure 2 shows some results on real data. We used different setups with either one or two color cameras. We also tested two different depth camera types, a *time-of-flight* camera and a Kinect camera. Synthetic data allowed us to perform a numerical comparison between our method and the one proposed in [2]. The paper contains more details about setups and gives both quantitative and qualitative results.

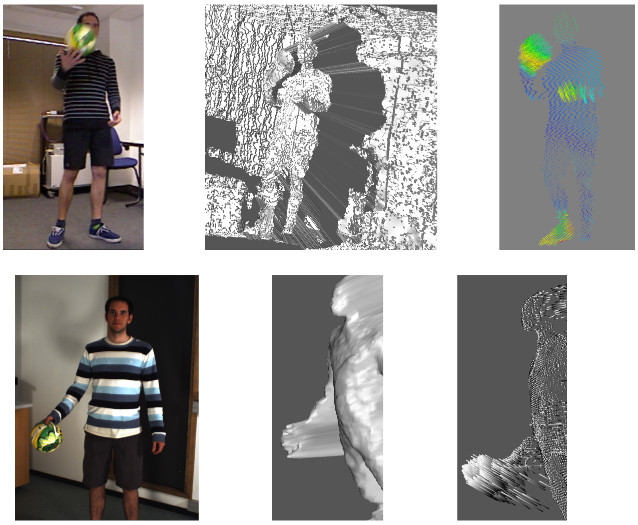


Figure 2: Two examples. Input data: the colour image (left) and the meshed surface (middle). Results: the 3D displacement field (right), color denotes 3D displacement norm.

Contribution (i) Following works on robust optical flow estimation [3], we take advantage of robust initial displacement values as provided by image features tracked over consecutive time instants. (ii) A linear framework that combines visual constraints with surface deformation constraints and allows for iterative resolution (variational approach) as well as coarse-to-fine refinement.

[1] B.K.P. Horn and B.G. Schunck. Determining Optical Flow. *Artificial Intelligence*, 1981.

[2] S. Vedula, S. Baker, P. Rander, R. Collins, and T. Kanade. Three-Dimensional Scene Flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2005.

[3] L. Xu, J. Jia, and Y Matsushita. Motion Detail Preserving Optical Flow Estimation. In *Computer Vision and Pattern Recognition*, 2010.