

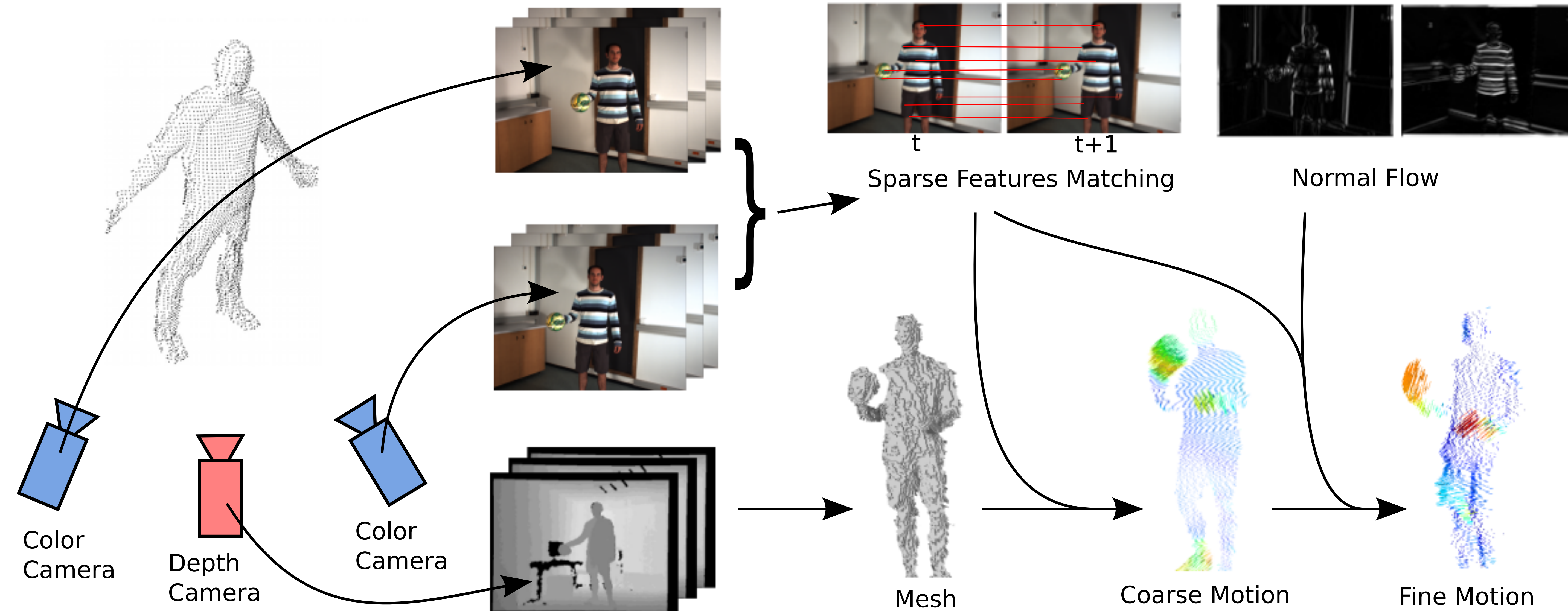
Scene Flow from Depth and Color Images

Antoine Letouzey, Benjamin Petit & Edmond Boyer

INRIA Grenoble Rhône-Alpes, France

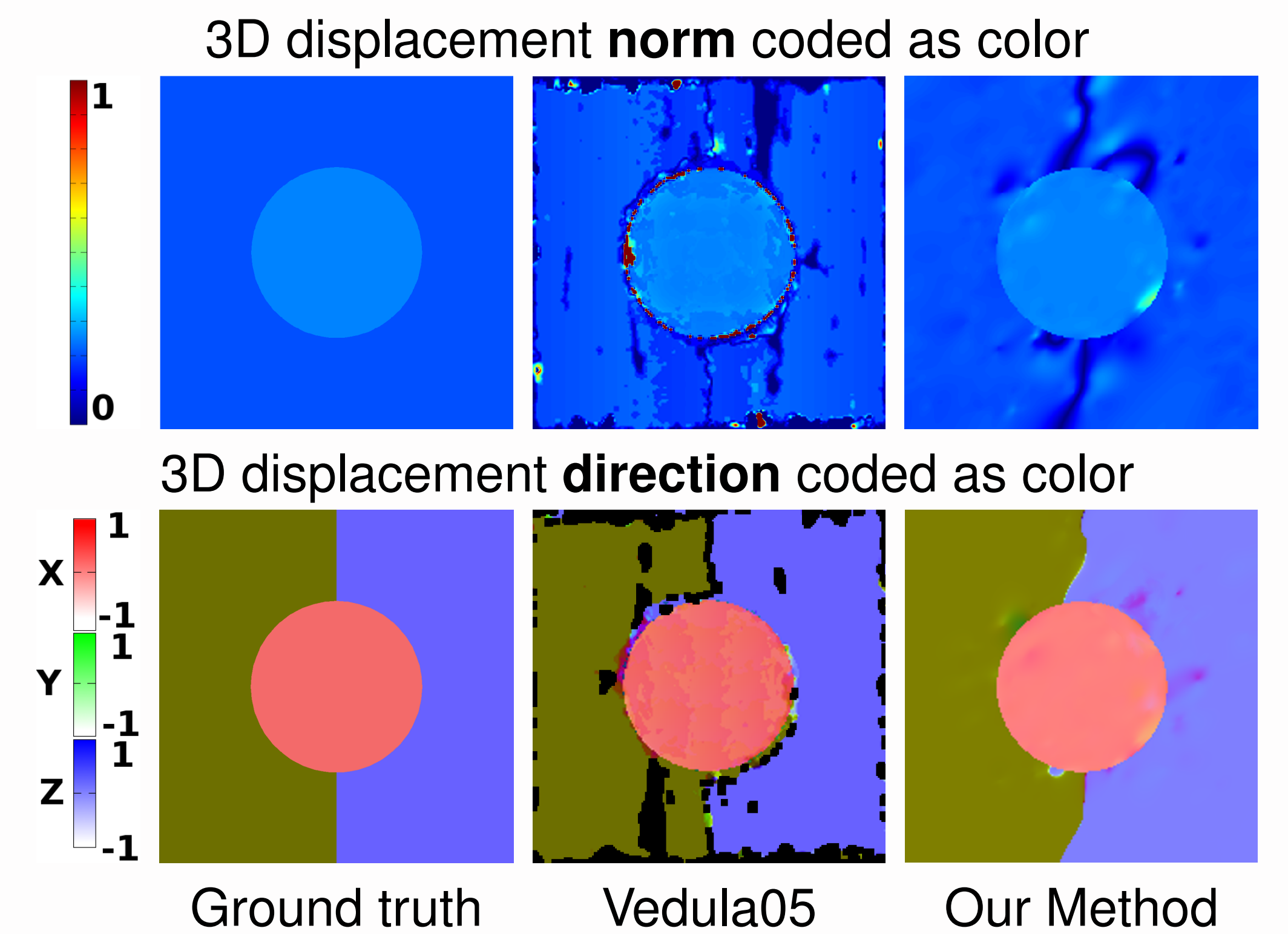
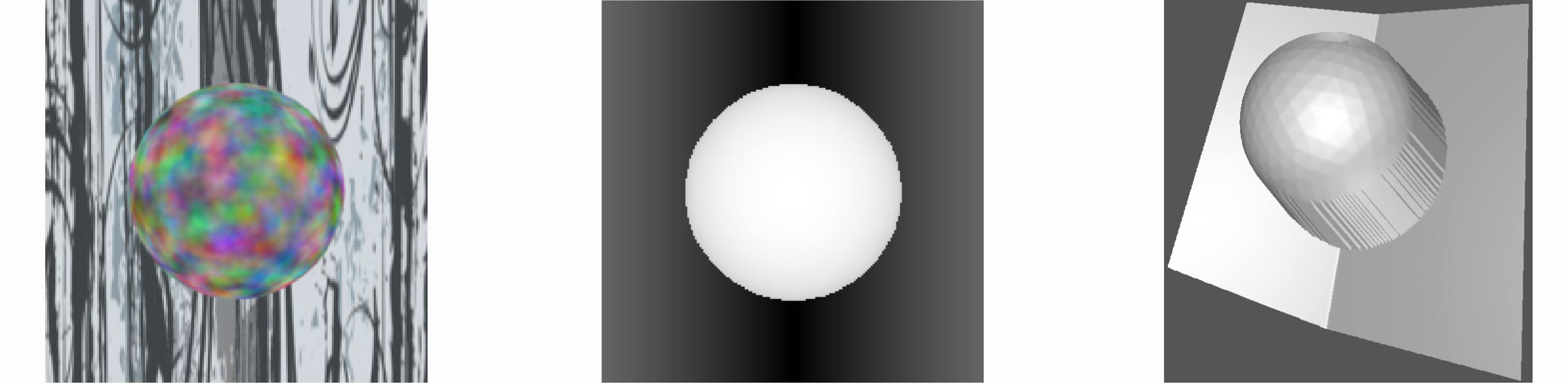


Overview



Comparison on Synthetic Data

Input: color and depth images and the computed meshed surface



Context & Motivation

Goal: Compute dense and accurate 3D motion of a scene.
Inputs: Calibrated **Depth** and **Color** image streams.
Output: 3D motion vector for each point of the scene visible in the depth camera.
Method: An unified framework which allows us to merge geometric information from depth sensor with intensity variation in color images. Depth map defines a surface domain over which visual cues can be consistently integrated in 3D. We cast the problem as an optimization :

$$\mathbf{E} = \mathbf{E}_{data} + \mathbf{E}_{smooth}$$

Geometric Constraints

\mathbf{E}_{smooth}

Regularization is required to diffuse sparse information and overcome the aperture problem arising from normal flow constraints. We choose to regularize motion in **3D**. Depth Images are converted into mesh surfaces using a basic triangulation. Laplacian smoothing (extension of Horn & Schunck's Method from 2D to 3D):

- global regularity makes more sens in 3D than it does in image domain
- prevents smoothing over object boundaries

Formulation & Resolution

Formulation :

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

- \mathbf{L} Laplacian of the surface
- \mathbf{A} and \mathbf{b} represent visual constraints
- \mathbf{V} Per-point 3D motion

More color camera stacks more lines in \mathbf{A} and \mathbf{b} .

Resolution :

Sparse linear solver or iterative method

Two Pass Algorithm :

1. Large motions are detected using only sparse features matching
2. Details are recovered by combining previous estimation and normal flow cues

Results on Real Data

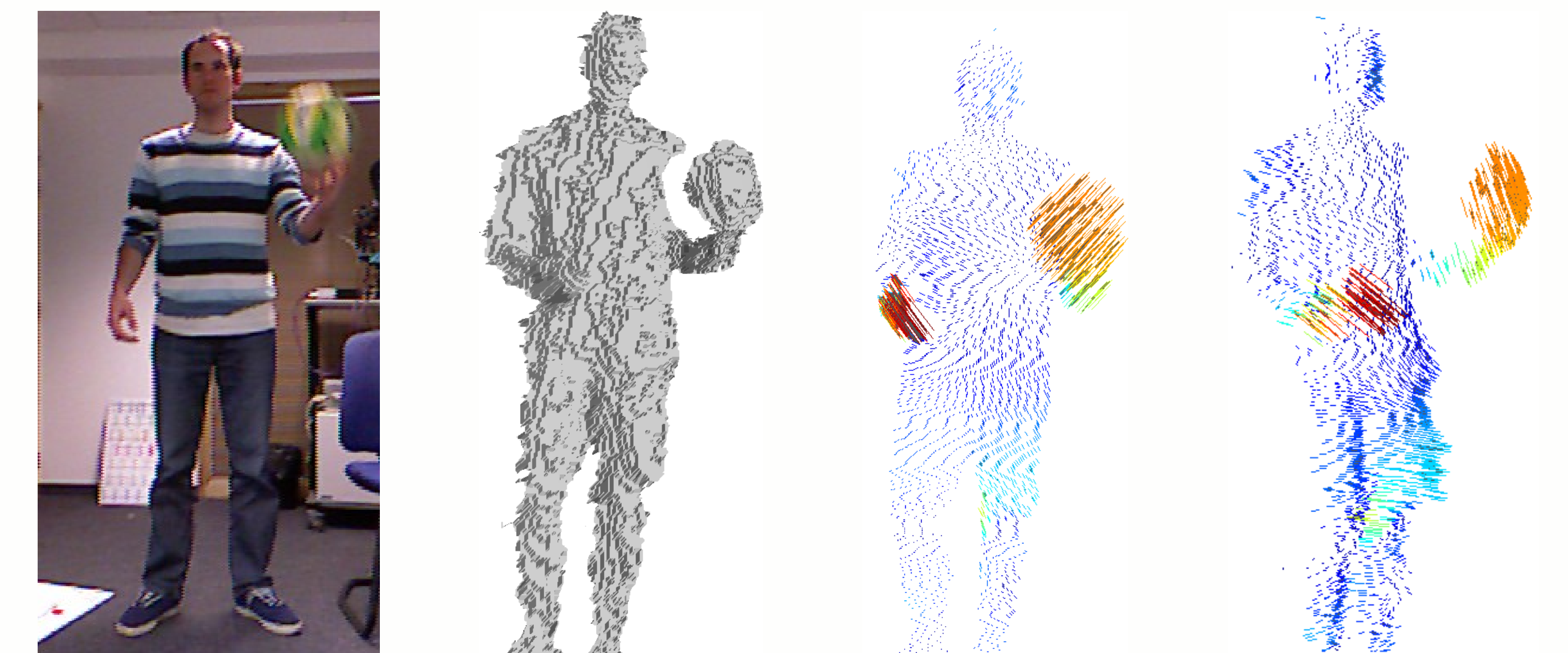


Figure: Kinect Device : depth camera and one color camera

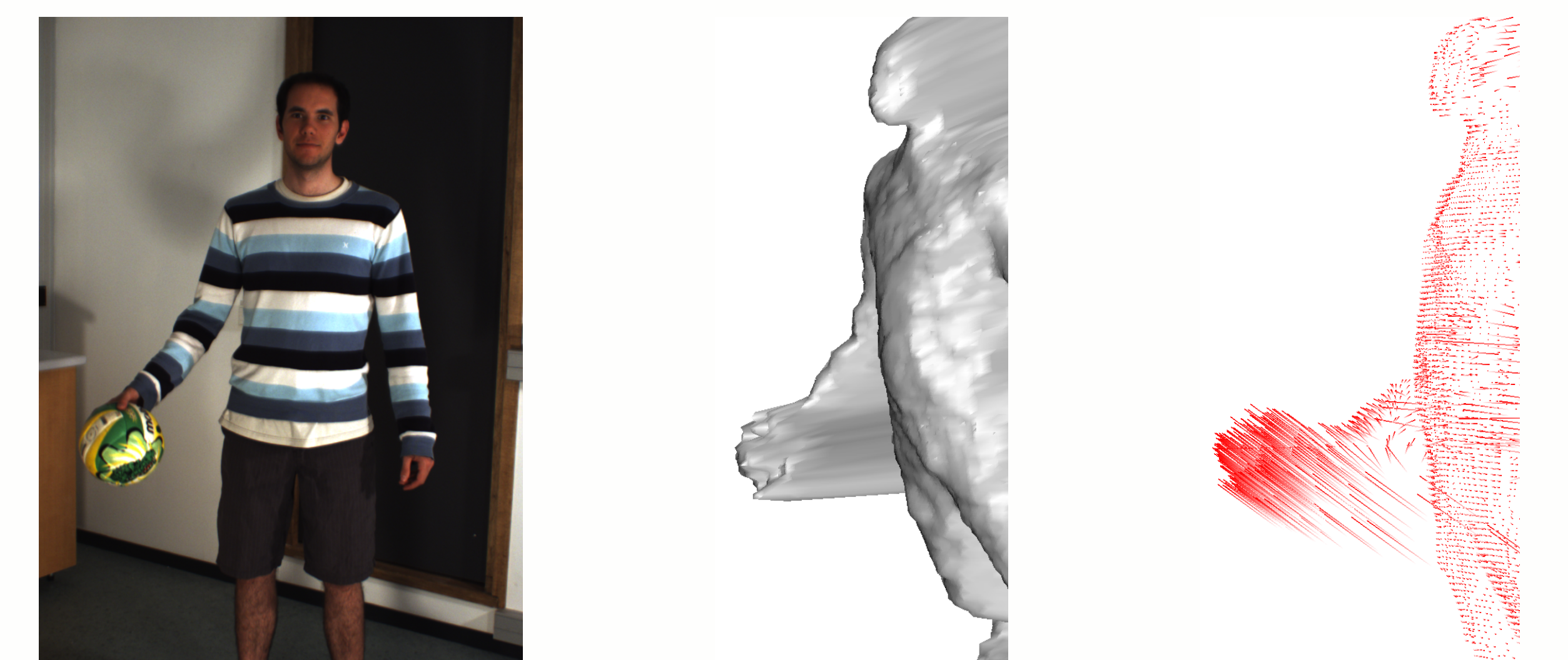


Figure: Time-of-Flight Camera and two color cameras

Visual Constraints

\mathbf{E}_{data}

We combine two kinds of visual cues to deal with small and large displacements. Both contribute a term to \mathbf{E}_{data}

Sparse Features Correspondances :

Sift matches between two consecutive color images gives reliable motion information that can handle arbitrary large motion. By backprojection on the associated depth map we get 3D anchor points.

Dense Normal Flow :

Gives motion information for each point of the scene visible but is only valid for small motion.