We present a novel approach for real-time joint reconstruction of 3D scene motion and geometry from binocular stereo videos. Our approach is based on a novel variational halfway-domain scene flow formulation, which allows us to obtain highly accurate spatiotemporal reconstructions of shape and motion. We solve the underlying optimization problem at real-time frame rates using a novel data-parallel robust non-linear optimization strategy. Fast convergence and large displacement flows are achieved by employing a novel hierarchy that stores delta flows between hierarchy levels. High performance is obtained by the introduction of a coarser warp grid that decouples the number of unknowns from the input resolution of the images. We demonstrate our approach in a live setup that is based on two commodity webcams, as well as on publicly available video data. Our extensive experiments and evaluations show that our approach produces high-quality dense reconstructions of 3D geometry and scene flow at real-time frame rates, and compares favorably to the state of the art.

HALFWAY-DOMAIN

The four captured images define the corners of the scene flow geometry. The five in-between frames define intermediate states of warping between the captured images based on the flow data. We define the scene flow as a combination of three flows (stereo, motion and difference) relative to the halfway domain in the middle.

SCENE FLOW OBJECTIVE

Our approach is based on energy minimization. The objective function takes into account both spatial alignment of the inputs warped to the halfway-domain as well as the validity of the flows.

\[ E(S) = E_{align}(S) + E_{reg}(S) \]

\[ E_{align}(S) = \sum_i V(x_i) \Phi \left( \frac{\partial_i^2 (x_i + s_i - m_i - d_i) - \partial_i^2 (x_i - s_i + m_i + d_i)}{\alpha} \right) \]

\[ E_{reg}(S) = \sum_i \omega_1 E_{external}(S) + \omega_2 E_{reg}(S) + \omega_3 E_{reg}(S) \]

RESULTS

Live reconstruction results:
(1280 x 720 @ 30Hz)

Results on the data of Valgaerts et al. 2010, 2012:
(1920 x 1088 @ 10Hz)

PERFORMANCE & CONVERGENCE

The runtime is linear in the image resolution and we obtain real-time performance for up to 0.9 MP.

Convergence behavior of our solver for the finest hierarchy level:

Error peaks correspond to the start of new non-linear Gauss-Newton steps.