Joint Estimation of Motion, Structure and Geometry from Stereo Sequences

Levi Valgaerts[1], Andrés Bruhn[1], Henning Zimmer[1], Joachim Weickert[1], Carsten Stoll[2], and Christian Theobalt[2]

[1] Mathematical Image Analysis Group, Saarland University, Saarbrücken, Germany
{valgaerts,bruhn,zimmer,weickert}@mia.uni-saarland.de
[2] Max-Planck Institute for Informatics, Saarbrücken, Germany
{stoll,theobalt}@mpi-inf.mpg.de

Abstract. We present a novel variational method for the simultaneous estimation of dense scene flow and structure from stereo sequences. In contrast to existing approaches that rely on a fully calibrated camera setup, we assume that only the intrinsic camera parameters are known. To couple the estimation of motion, structure and geometry, we propose a joint energy functional that integrates spatial and temporal information from two subsequent image pairs subject to an unknown stereo setup. We further introduce a normalisation of image and stereo constraints such that deviations from model assumptions can be interpreted in a geometrical way. Finally, we suggest a separate discontinuity-preserving regularisation to improve the accuracy. Experiments on calibrated and uncalibrated data demonstrate the excellent performance of our approach. We even outperform recent techniques for the rectified case that make explicit use of the simplified geometry.

1 Introduction

For many tasks in computer vision, such as vehicle navigation, motion capture and dynamic rendering, it is essential to recover the three-dimensional displacement field of a scene. This so called *scene flow* represents the real 3D motion of objects – as opposed to optical flow that only describes the projection of this motion on the 2D image plane [23]. Since depth information is required to determine 3D motion, scene flow can not be computed without estimating the scene structure as well. In contrast to structure from motion, scene flow does not relate to a static world. Instead, objects in the scene are allowed to move freely and in a non-rigid fashion. Thus, for estimating scene flow, stereo sequences are required that provide two views per time instance.

Existing scene flow algorithms often treat stereo and motion independently. In fact, most of them rely on a sequential computation of the scene flow and structure [23, 19, 26, 20, 24]. However, to improve the quality of the estimation it is important that 3D motion and shape estimation are coupled. This can be achieved by exploiting the spatial and temporal dependencies in the image sequence [26, 12, 4, 18, 6]. Among those methods that solve for the scene flow and structure simultaneously, variational approaches play a major role. Some of these techniques parameterise the problem directly in 3D space [6]. Others are based on optical flow computation [26, 12, 18] and have consistently improved their results in the wake of increasing optical flow accuracy.

All of the afore mentioned methods have one aspect in common: they assume that the cameras have been calibrated beforehand. However, in order to deal with general stereo setups without requiring an explicit calibration step, it would be desirable to jointly estimate the scene flow, the scene structure *and* the stereo geometry.

In this paper we thus propose a variational scene flow method for *uncalibrated* stereo sequences. We do this by integrating the spatial and temporal information from two stereo pairs in a global energy functional while simultaneously estimating the unknown stereo geometry in consecutive time steps. Assuming that the internal camera parameters are known, our method allows to recover the dense scene structure and the dense scene flow up to a scale factor. Apart from this novel generalised model, we make two additional contributions: First, within the multiresolution framework required to handle large displacements, we introduce a tensor-based notation for linearised constraints. This notation allows to normalise these constraints such that deviations from the model can be interpreted as geometrical distances. Secondly, we propose a regularisation strategy that penalises discontinuities in the different displacement fields separately. This makes sense, since motion and depth continuities do not necessarily coincide. Our experiments clearly demonstrate the benefits of both contributions and show the favourable performance of our method compared to recent techniques for the rectified case.

Related Work. In the context of scene flow estimation, closely related to our work are the methods [26, 12, 18], which jointly compute spatial and temporal motion fields by minimising a single energy. In particular the method of Huguet and Devernay [12] uses similar data constraints as our approach. However, it applies a joint smoothness term to all displacement fields. A more adequate separate treatment of the smoothness term is proposed by Wedel *et al.* [25] who decouple the estimation of structure and motion to achieve real-time performance. However, in their case, the separate smoothness term does not yield more accurate results than their preceding work with joint regularisation [24]. All of the previous approaches are based on rectified sequences and do not consider a suitable constraint normalisation. Apart from these methods that parameterise the displacements in terms of image coordinates, there are also techniques that work directly in 3D space. Such techniques include methods based on reprojection errors [6], space carving and nonlinear optimisation [4], deformable meshes [9] and Markov Random Fields [13]. Moreover, all these methods rely on a previous calibration step, since they involve the use of projection matrices.

In the context of optical flow estimation, the work of Valgaerts *et al.* [22] and Zimmer *et al.* [27] are closest related to our approach. While the first one shows the benefit of jointly estimating dense displacements and the underlying stereo geometry, the second one proposes a normalisation of the data constraints to penalise a geometrically meaningful distance. In our approach we extend both ideas to scene flow and unify them by normalising both data and stereo constraints.

Paper Organisation. In Sect. 2 we derive our variational model for the uncalibrated case. Important issues like incremental computation and constraint normalisation are then discussed in Sect. 3. While Sect. 4 is dedicated to the alternating minimisation of the proposed energy, our results and a comparison to the literature are presented in Sect. 5. The paper concludes with a summary in Sect. 6.

2 A Scene Flow Model for Uncalibrated Stereo Sequences

In the following we consider the classical four-frame case depicted in Fig. 1. It consists of two consecutive image pairs of a stereo sequence: the left image $g_{1\mathrm{l}}(\boldsymbol{x})$ and the right image $g_{1\mathrm{r}}(\boldsymbol{x})$ at time t and the left image $g_{2\mathrm{l}}(\boldsymbol{x})$ and right image $g_{2\mathrm{r}}(\boldsymbol{x})$ at time $t+1$. Here $\boldsymbol{x} = (x, y)^\top$ denotes the location in a rectangular image domain $\Omega \subset \mathbb{R}^2$ that is assumed to be the same for all images. We furthermore assume that the sequence has been recorded by a single fixed stereo rig, i.e. there exists a common fundamental matrix F that describes the epipolar geometry [7] of the stereo pairs at time t and $t+1$.

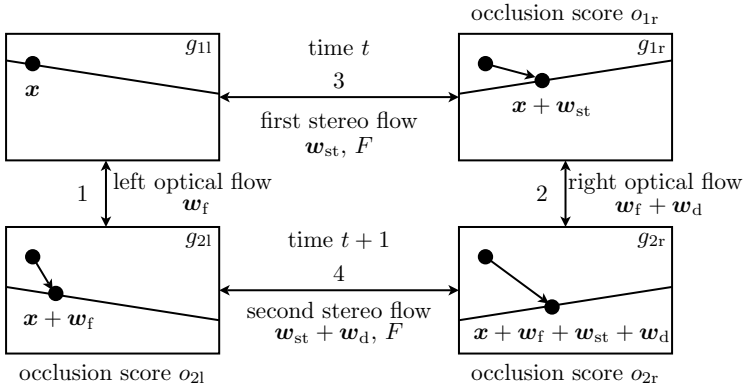

Fig. 1. The correspondences between the four frames of a binocular stereo sequence.

In contrast to previous variational methods that start out from a rectified stereo sequence [24, 12], our method assumes a general stereo geometry with unknown fundamental matrix. As a consequence, the stereo correspondences do not take on the form of a scalar valued disparity but of a 2-dimensional displacement field that we will refer to as *stereo flow*. In total, we consider four types of correspondences in our model: two optical flows between consecutive frames of the same camera (left, right) and two stereo flows between the left and right frame at the same time instance (t, $t+1$). Exploiting the dependencies in Fig. 1, these correspondences can be parameterised by six unknown functions with respect to the reference image $g_{1\mathrm{l}}(\boldsymbol{x})$: the first stereo flow $\boldsymbol{w}_{\mathrm{st}} = (u_{\mathrm{st}}, v_{\mathrm{st}})^\top$, the left optical flow $\boldsymbol{w}_{\mathrm{f}} = (u_{\mathrm{f}}, v_{\mathrm{f}})^\top$ and the difference flow $\boldsymbol{w}_{\mathrm{d}} = (u_{\mathrm{d}}, v_{\mathrm{d}})^\top$ that can be interpreted as a change in optical flow or a change in stereo flow. Moreover, we have seven degrees of freedom from the fundamental matrix F, which restricts points to lie on corresponding epipolar lines, as shown in Fig. 1. These degrees of freedom arise from the fact that F is a 3×3 matrix of rank 2 that is defined up to a scale factor. For given intrinsic camera parameters, knowing the fundamental matrix is sufficient to recover projection matrices (P_1, P_2) for the left and the right image sequence [10]. Together with the stereo flow $\boldsymbol{w}_{\mathrm{st}}$ at time t, these matrices allow to reconstruct a reference image point up to a scale in the camera coordinate system. To obtain a reconstruction at time $t+1$ and the scene flow relative to the cameras, the left optical flow $\boldsymbol{w}_{\mathrm{f}}$ and the flow change $\boldsymbol{w}_{\mathrm{d}}$ have to be known additionally.

Since we are interested in a joint computation of the 3D motion, structure and geometry, that are parameterised by $(\boldsymbol{w}_{\mathrm{f}}, \boldsymbol{w}_{\mathrm{st}}, \boldsymbol{w}_{\mathrm{d}})^\top$ and F, we propose to minimise a global energy functional that combines the spatial and temporal information of the different views while imposing geometric consistency. This functional has the form

$$\mathcal{E} = \int_\Omega \left(\mathcal{E}_{\mathrm{D}} + \mathcal{E}_{\mathrm{E}} + \mathcal{E}_{\mathrm{S}}\right) \, \mathrm{d}\boldsymbol{x} \ , \tag{1}$$

where $\mathcal{E}_{\mathrm{D}}$ is the data term that models the assumption that certain image features remain constant between the four frames, $\mathcal{E}_{\mathrm{E}}$ is the epipolar term that relates the stereo views by the unknown epipolar geometry, and $\mathcal{E}_{\mathrm{S}}$ is the smoothness term that assumes the solution to be piecewise smooth. In the following we will detail on the different terms.

2.1 Data Constraints

Let us now derive the four constraints that model the relation between the four input images w.r.t. the reference image. For simplicity, let us assume for the moment that the brightness of corresponding image points remains constant between all frames [11]. Following the enumeration of constraints in Fig. 1 we obtain the expressions

$$\mathcal{E}_{\mathrm{D1}} = \Psi\left(|g_{2\mathrm{l}}(\boldsymbol{x} + \boldsymbol{w}_{\mathrm{f}}) - g_{1\mathrm{l}}(\boldsymbol{x})|^2\right) \ , \tag{2}$$

$$\mathcal{E}_{\mathrm{D2}} = \Psi\left(|g_{2\mathrm{r}}(\boldsymbol{x} + \boldsymbol{w}_{\mathrm{f}} + \boldsymbol{w}_{\mathrm{st}} + \boldsymbol{w}_{\mathrm{d}}) - g_{1\mathrm{r}}(\boldsymbol{x} + \boldsymbol{w}_{\mathrm{st}})|^2\right) \ , \tag{3}$$

$$\mathcal{E}_{\mathrm{D3}} = \Psi\left(|g_{1\mathrm{r}}(\boldsymbol{x} + \boldsymbol{w}_{\mathrm{st}}) - g_{1\mathrm{l}}(\boldsymbol{x})|^2\right) \ , \tag{4}$$

$$\mathcal{E}_{\mathrm{D4}} = \Psi\left(|g_{2\mathrm{r}}(\boldsymbol{x} + \boldsymbol{w}_{\mathrm{f}} + \boldsymbol{w}_{\mathrm{st}} + \boldsymbol{w}_{\mathrm{d}}) - g_{2\mathrm{l}}(\boldsymbol{x} + \boldsymbol{w}_{\mathrm{f}})|^2\right) \ . \tag{5}$$

The first two terms correspond to an optical flow constraint between two time instances, while the last two terms arise from a stereo correspondence at consecutive time steps. As in [12] we choose to penalise all constraints separately since outliers for optical flow and stereo do not necessarily occur in the same location. As penalty function Ψ we choose the regularised L_1 norm $\Psi(s^2) = \sqrt{s^2 + \epsilon^2}$ with $\epsilon = 0.001$ as proposed e.g. in [2]. In our final model we include the gradient constancy assumption to cope with varying illumination and extend the expressions above to RGB colour images. Then the first term (2) becomes

$$\mathcal{E}_{\mathrm{D1}} = \Psi\left(\sum_{i=1}^{3}\left(\left|g^i_{2\mathrm{l}}(\boldsymbol{x} + \boldsymbol{w}_{\mathrm{f}}) - g^i_{1\mathrm{l}}(\boldsymbol{x})\right|^2 + \gamma\left|\nabla g^i_{2\mathrm{l}}(\boldsymbol{x} + \boldsymbol{w}_{\mathrm{f}}) - \nabla g^i_{1\mathrm{l}}(\boldsymbol{x})\right|^2\right)\right) \ , \tag{6}$$

where $\gamma \geq 0$ is a weighting factor, the symbol $\nabla = (\partial_x, \partial_y)^\top$ denotes the spatial gradient operator, and g^1, g^2, and g^3 represent the three RGB colour channels. The constraints $\mathcal{E}_{\mathrm{D2}}$, $\mathcal{E}_{\mathrm{D3}}$ and $\mathcal{E}_{\mathrm{D4}}$ are extended in the same way.

2.2 Occlusion Scores

In order to handle situations, where parts of the scene become occluded due to motion or a change of camera viewpoint, we additionally introduce occlusion scores. For instance,

the score $o_{1\mathrm{r}} : \Omega \to \{0,1\}$ takes on the value 1 for points in the reference image $g_{1\mathrm{l}}$ that are visible in $g_{1\mathrm{r}}$, and 0 otherwise. Once the fundamental matrix is known and the projection matrices (P_1, P_2) have been computed, the values of $o_{1\mathrm{r}}$ can be determined by projecting the reconstruction at time t back on the image plane using $P2$. Of all the points that reproject onto the same location, the one that lies closest to the optical centre of $P2$ will be marked as visible. This technique is also known as *Z-buffering*. The scores $o_{2\mathrm{l}}$ and $o_{2\mathrm{r}}$ for the image pairs $(g_{1\mathrm{l}}, g_{2\mathrm{l}})$ and $(g_{1\mathrm{l}}, g_{2\mathrm{r}})$ are determined analogously by reprojection on time $t+1$ with P_1 and P_2, respectively. The four data terms are multiplied by the occlusion scores to switch them off where the constancy assumptions can not be fulfilled. This yields the final data term

$$\mathcal{E}_{\mathrm{D}} = o_{2\mathrm{l}}\,\mathcal{E}_{\mathrm{D1}} + o_{1\mathrm{r}}\,o_{2\mathrm{r}}\,\mathcal{E}_{\mathrm{D2}} + o_{1\mathrm{r}}\,\mathcal{E}_{\mathrm{D3}} + o_{2\mathrm{l}}\,o_{2\mathrm{r}}\,\mathcal{E}_{\mathrm{D4}} \ . \tag{7}$$

Note that each term has to be multiplied by the occlusion scores of the images that occur in the according data constraint, since the reappearance of points in $g_{2\mathrm{r}}$ that are occluded in $g_{1\mathrm{r}}$ or $g_{2\mathrm{l}}$ is not noticed by the reference image.

2.3 Epipolar Constraints

Let us now model the geometric relation between the left and right images of the stereo pairs $(g_{1\mathrm{l}}, g_{1\mathrm{r}})$ and $(g_{2\mathrm{l}}, g_{2\mathrm{r}})$. To this end we introduce two terms that relate the unknown flows and the fundamental matrix F via the respective epipolar constraints [16]:

$$\mathcal{E}_{\mathrm{E1}} = \Psi\left(\left((\boldsymbol{x}+\boldsymbol{w}_{\mathrm{st}})_{\mathrm{h}}^{\top} F\,(\boldsymbol{x})_{\mathrm{h}}\right)^2\right) \text{ , and} \tag{8}$$

$$\mathcal{E}_{\mathrm{E2}} = \Psi\left(\left((\boldsymbol{x}+\boldsymbol{w}_{\mathrm{f}}+\boldsymbol{w}_{\mathrm{st}}+\boldsymbol{w}_{\mathrm{d}})_{\mathrm{h}}^{\top} F\,(\boldsymbol{x}+\boldsymbol{w}_{\mathrm{f}})_{\mathrm{h}}\right)^2\right) \ . \tag{9}$$

Here the subscript h denotes the use of homogeneous coordinates, i.e. $(\boldsymbol{x})_{\mathrm{h}} = (x, y, 1)^{\top}$. Both terms $\mathcal{E}_{\mathrm{E1}}$ and $\mathcal{E}_{\mathrm{E2}}$ are soft constraints that penalise deviations of a point from its epipolar line. The use of Ψ increases the robustness of the estimation of F with respect to outliers. While the first epipolar term can be modelled completely in accordance with [22], the second epipolar constraint is much more complicated: Although it is linear in $\boldsymbol{w}_{\mathrm{st}}$ and $\boldsymbol{w}_{\mathrm{d}}$, it is quadratic with respect to the left optical flow $\boldsymbol{w}_{\mathrm{f}}$. This makes the minimisation of the corresponding energy difficult. To nevertheless obtain a linear expression in all flows we thus propose to introduce an auxiliary variable $\boldsymbol{w}_{\mathrm{a}} = (u_{\mathrm{a}}, v_{\mathrm{a}})^{\top}$, which is assumed to be close to $\boldsymbol{w}_{\mathrm{f}}$, and split up the epipolar constraint such that $\boldsymbol{w}_{\mathrm{f}}$ and $\boldsymbol{w}_{\mathrm{a}}$ take on symmetric roles. In this way we can approximate term (9) via

$$\begin{aligned}\mathcal{E}_{\mathrm{E2}} = \Psi\Big(\quad & \tfrac{1}{2}\left((\boldsymbol{x}+\boldsymbol{w}_{\mathrm{f}}+\boldsymbol{w}_{\mathrm{st}}+\boldsymbol{w}_{\mathrm{d}})_{\mathrm{h}}^{\top} F\,(\boldsymbol{x}+\boldsymbol{w}_{\mathrm{a}})_{\mathrm{h}}\right)^2 \\ & + \tfrac{1}{2}\left((\boldsymbol{x}+\boldsymbol{w}_{\mathrm{a}}+\boldsymbol{w}_{\mathrm{st}}+\boldsymbol{w}_{\mathrm{d}})_{\mathrm{h}}^{\top} F\,(\boldsymbol{x}+\boldsymbol{w}_{\mathrm{f}})_{\mathrm{h}}\right)^2\Big) + \mu\left(|\boldsymbol{w}_{\mathrm{f}}-\boldsymbol{w}_{\mathrm{a}}|^2\right) \ ,\end{aligned} \tag{10}$$

where μ is the weight of the additional similarity term that is required to couple $\boldsymbol{w}_{\mathrm{a}}$ and $\boldsymbol{w}_{\mathrm{f}}$. Introducing the weights β_1 and β_2 we obtain the final epipolar term

$$\mathcal{E}_{\mathrm{E}} = \beta_1\,\mathcal{E}_{\mathrm{E1}} + \beta_2\,\mathcal{E}_{\mathrm{E2}} \ . \tag{11}$$

To avoid the trivial solution we additionally impose the constraint $\|F\|_{\mathrm{Frob}}^2 = 1$ on the Frobenius norm of the fundamental matrix F as proposed in [14].

2.4 Smoothness Constraints

Let us finally detail on the design of the smoothness term. Its task is to regularise the problem in locations where the remaining terms do not guarantee a unique solution (aperture problem) or to fill in information in the presence of outliers, e.g. occlusions. Because there often exists an overlap between the discontinuities of $\boldsymbol{w}_{\mathrm{f}}$, $\boldsymbol{w}_{\mathrm{st}}$ and $\boldsymbol{w}_{\mathrm{d}}$, the authors of [12] suggested a joint piecewise smoothness assumption on all flows. With our method, however, we want to cover the general case where the flow and stereo discontinuities do not necessarily coincide, e.g. for different in-plane motions. Therefore we propose a separate penalisation of the flow gradients:

$$\mathcal{E}_{\mathrm{S1}} = \Psi\left(|\nabla \boldsymbol{w}_{\mathrm{f}}|^2\right) \text{ , } \mathcal{E}_{\mathrm{S2}} = \Psi\left(|\nabla \boldsymbol{w}_{\mathrm{st}}|^2\right) \text{ , and } \mathcal{E}_{\mathrm{S3}} = \Psi\left(|\nabla \boldsymbol{w}_{\mathrm{d}}|^2\right) \quad , \tag{12}$$

with $|\nabla \boldsymbol{w}_*|^2 := |\nabla u_*|^2 + |\nabla v_*|^2$, where $*$ stands for f, st or d. The penalisation via the subquadratic function Ψ, as defined before, equals total variation (TV) regularisation [21]. This gives rise to the smoothness term

$$\mathcal{E}_{\mathrm{S}} = \alpha_1 \, \mathcal{E}_{\mathrm{S1}} + \alpha_2 \, \mathcal{E}_{\mathrm{S2}} + \alpha_3 \, \mathcal{E}_{\mathrm{S3}} \quad , \tag{13}$$

where α_1, α_2, α_3 are positive weights that balance the smoothness assumptions for the three displacement fields.

3 Linearisation and Normalisation

Substituting all data, epipolar and smoothness terms into (1) we obtain an energy functional that is rather complicated. Moreover, it is non-convex, since the unknown flows appear implicitly in the arguments of the data term. A common strategy to resolve this problem is to perform an incremental computation of the unknowns within a coarse-to-fine multiscale approach. This can either be done by a fixed point iteration on the Euler-Lagrange equations [2] or by a series of energies that approximate the original model on every resolution level [17]. In the following we stick to the second strategy and discuss how the corresponding energy for each level can be derived. Assuming that solutions $\boldsymbol{w}_{\mathrm{f}}$, $\boldsymbol{w}_{\mathrm{st}}$, $\boldsymbol{w}_{\mathrm{d}}$ and $\boldsymbol{w}_{\mathrm{a}}$ are available from a coarser scale, we aim at expressing the total energy in terms of the increments $d\boldsymbol{w}_{\mathrm{f}} = (du_{\mathrm{f}}, dv_{\mathrm{f}})$, $d\boldsymbol{w}_{\mathrm{st}} = (du_{\mathrm{st}}, dv_{\mathrm{st}})$, $d\boldsymbol{w}_{\mathrm{d}} = (du_{\mathrm{d}}, dv_{\mathrm{d}})$, and $d\boldsymbol{w}_{\mathrm{a}} = (du_{\mathrm{a}}, dv_{\mathrm{a}})$. This allows us to introduce a tensor notation which offers two advantages: (i) The convexity of the resulting energy functional in the flow increments becomes explicit, and (ii) a normalisation strategy can be applied that makes deviations from the model assumptions interpretable in a geometric way.

3.1 Linearisation in the Data Term

Let us first discuss the differential form of the data term by the example of the simplified data constraint from expression (3). Using a first order Taylor expansion to linearise this expression with respect to all increments we obtain the approximation

$$\begin{aligned} & g_{2\mathrm{r}}(\boldsymbol{x} + \boldsymbol{w}_{\mathrm{f}} + d\boldsymbol{w}_{\mathrm{f}} + \boldsymbol{w}_{\mathrm{st}} + d\boldsymbol{w}_{\mathrm{st}} + \boldsymbol{w}_{\mathrm{d}} + d\boldsymbol{w}_{\mathrm{d}}) - g_{1\mathrm{r}}(\boldsymbol{x} + \boldsymbol{w}_{\mathrm{st}} + d\boldsymbol{w}_{\mathrm{st}}) \\ \approx \quad & g_{2\mathrm{r}} + \partial_x g_{2\mathrm{r}} \cdot (du_{\mathrm{f}} + du_{\mathrm{st}} + du_{\mathrm{d}}) + \partial_y g_{2\mathrm{r}} \cdot (dv_{\mathrm{f}} + dv_{\mathrm{st}} + dv_{\mathrm{d}}) \\ & - g_{1\mathrm{r}} - \partial_x g_{1\mathrm{r}} \cdot (du_{\mathrm{st}}) - \partial_y g_{1\mathrm{r}} \cdot (dv_{\mathrm{st}}) \ . \end{aligned} \tag{14}$$

Rearranging the terms and using the following abbreviations

$$g_{2z} = g_{2\mathrm{r}}(\boldsymbol{x} + \boldsymbol{w}_{\mathrm{f}} + \boldsymbol{w}_{\mathrm{st}} + \boldsymbol{w}_{\mathrm{d}}) - g_{1\mathrm{r}}(\boldsymbol{x} + \boldsymbol{w}_{\mathrm{st}}) \,, \tag{15}$$

$$g_{2\mathrm{r}x} = \partial_x g_{2\mathrm{r}}(\boldsymbol{x} + \boldsymbol{w}_{\mathrm{f}} + \boldsymbol{w}_{\mathrm{st}} + \boldsymbol{w}_{\mathrm{d}}) \,, \quad g_{2xz} = \partial_x g_{2z} \,, \tag{16}$$

$$g_{2\mathrm{r}y} = \partial_y g_{2\mathrm{r}}(\boldsymbol{x} + \boldsymbol{w}_{\mathrm{f}} + \boldsymbol{w}_{\mathrm{st}} + \boldsymbol{w}_{\mathrm{d}}) \,, \quad g_{2yz} = \partial_y g_{2z} \;\; , \tag{17}$$

we can rewrite the linearised term in (14) as inner product

$$\boldsymbol{g}_2^\top \boldsymbol{d} = g_{2\mathrm{r}x} du_{\mathrm{f}} + g_{2\mathrm{r}y} dv_{\mathrm{f}} + g_{2xz} du_{\mathrm{st}} + g_{2yz} dv_{\mathrm{st}} + g_{2\mathrm{r}x} du_{\mathrm{d}} + g_{2\mathrm{r}y} dv_{\mathrm{d}} + g_{2z} \;\; , \tag{18}$$

where the two vectors are defined as $\boldsymbol{g}_2 := (g_{2\mathrm{r}x}, g_{2\mathrm{r}y}, g_{2xz}, g_{2yz}, g_{2\mathrm{r}x}, g_{2\mathrm{r}y}, g_{2z})^\top$ and $\boldsymbol{d} := (du_{\mathrm{f}}, dv_{\mathrm{f}}, du_{\mathrm{st}}, dv_{\mathrm{st}}, du_{\mathrm{d}}, dv_{\mathrm{d}}, 1)^\top$. The equation $\boldsymbol{g}_2^\top \boldsymbol{d} = 0$ can be seen as a multidimensional extension of the classical optical flow constraint [11]. Inserting it as squared argument into the penaliser Ψ yields the robustified quadratic form

$$\mathcal{E}_{\mathrm{D2}} = \Psi\left((\boldsymbol{g}_2^\top \boldsymbol{d})^2\right) = \Psi\left(\boldsymbol{d}^\top J_2 \, \boldsymbol{d}\right) \;\; , \tag{19}$$

where $J_2 = \boldsymbol{g}_2 \, \boldsymbol{g}_2^\top$ is a 7×7 matrix that provides coupling between all increments. By analogy to the motion tensor notation in optical flow estimation [3], we denote J_2 as *scene flow tensor*. The linearisation of the three remaining data constraints is carried out accordingly, and results in the 7×7 scene flow tensors J_1, J_3 and J_4. Missing dependencies between the variables give rise to zero tensor entries. Including the gradient constancy assumption and extending it to RGB colour images as in equation (6) is straightforward and leads to a weighted sum of the corresponding tensors [27].

3.2 Treatment of the Epipolar Term

The first epipolar term $(\boldsymbol{x}+\boldsymbol{w}_{\mathrm{st}}+d\boldsymbol{w}_{\mathrm{st}})_{\mathrm{h}}^\top F(\boldsymbol{x})_{\mathrm{h}}$ is already linear in the increment $d\boldsymbol{w}_{\mathrm{st}}$. As in the case of the data terms we can thus define the vector $\boldsymbol{d}_1 = (du_{\mathrm{st}}, dv_{\mathrm{st}}, 1)^\top$ and write the argument of the first epipolar term (8) as a quadratic form

$$\mathcal{E}_{\mathrm{E1}} = \Psi\left(\boldsymbol{d}_1^\top E_1 \, \boldsymbol{d}_1\right) \;\; . \tag{20}$$

The corresponding epipolar tensor E_1 of size 3×3 is defined as $(a_1, b_1, q_1)^\top (a_1, b_1, q_1)$, where a_1 and b_1 are the coefficients of the epipolar line $\boldsymbol{l} = F(\boldsymbol{x})_{\mathrm{h}}$, and q_1 is the scaled distance of the point $\boldsymbol{x}$ to this line [22]. However, care has to be taken with respect to symmetry when introducing the flow increments in the second epipolar term (10). The expanded differential variant of its argument reads

$$\begin{aligned}
&\tfrac{1}{4}\big((\boldsymbol{x} + \boldsymbol{w}_{\mathrm{f}} + d\boldsymbol{w}_{\mathrm{f}} + \boldsymbol{w}_{\mathrm{st}} + d\boldsymbol{w}_{\mathrm{st}} + \boldsymbol{w}_{\mathrm{d}} + d\boldsymbol{w}_{\mathrm{d}})_{\mathrm{h}}^\top \, F\,(\boldsymbol{x} + \boldsymbol{w}_{\mathrm{a}})_{\mathrm{h}} \big)^2 \\
+&\tfrac{1}{4}\big((\boldsymbol{x} + \boldsymbol{w}_{\mathrm{a}} + d\boldsymbol{w}_{\mathrm{a}} + \boldsymbol{w}_{\mathrm{st}} + d\boldsymbol{w}_{\mathrm{st}} + \boldsymbol{w}_{\mathrm{d}} + d\boldsymbol{w}_{\mathrm{d}})_{\mathrm{h}}^\top \, F\,(\boldsymbol{x} + \boldsymbol{w}_{\mathrm{f}})_{\mathrm{h}} \big)^2 \\
+&\tfrac{1}{4}\big((\boldsymbol{x} + \boldsymbol{w}_{\mathrm{a}} + d\boldsymbol{w}_{\mathrm{a}})_{\mathrm{h}}^\top \, F^\top\,(\boldsymbol{x} + \boldsymbol{w}_{\mathrm{f}} + \boldsymbol{w}_{\mathrm{st}} + \boldsymbol{w}_{\mathrm{d}})_{\mathrm{h}} \big)^2 \\
+&\tfrac{1}{4}\big((\boldsymbol{x} + \boldsymbol{w}_{\mathrm{f}} + d\boldsymbol{w}_{\mathrm{f}})_{\mathrm{h}}^\top \, F^\top\,(\boldsymbol{x} + \boldsymbol{w}_{\mathrm{a}} + \boldsymbol{w}_{\mathrm{st}} + \boldsymbol{w}_{\mathrm{d}})_{\mathrm{h}} \big)^2 \;\; ,
\end{aligned} \tag{21}$$

where we have additionally included the last two terms with the transposed fundamental matrix to ensure a symmetrical treatment of the left and right flow increments. This is

required since as opposed to the first epipolar constraint variations can occur in both the left and the right image position. Since all terms of expression (21) are linear in the increments, the second epipolar term can be written as

$$\begin{aligned}\mathcal{E}_{\mathrm{E2}} = \quad & \Psi\left(\tfrac{1}{4}\,\boldsymbol{d}_2^\top E_2\,\boldsymbol{d}_2 + \tfrac{1}{4}\,\boldsymbol{d}_3^\top E_3\,\boldsymbol{d}_3 + \tfrac{1}{4}\,\boldsymbol{d}_4^\top E_4\,\boldsymbol{d}_4 + \tfrac{1}{4}\,\boldsymbol{d}_5^\top E_5\,\boldsymbol{d}_5\right) \\ & +\mu\left(|\boldsymbol{w}_{\mathrm{f}} + d\boldsymbol{w}_{\mathrm{f}} - \boldsymbol{w}_{\mathrm{a}} - d\boldsymbol{w}_{\mathrm{a}}|\right)^2 \quad , \end{aligned} \tag{22}$$

where we have defined the following vectors:

$$\boldsymbol{d}_2 = (du_{\mathrm{f}} + du_{\mathrm{st}} + du_{\mathrm{d}}, dv_{\mathrm{f}} + dv_{\mathrm{st}} + dv_{\mathrm{d}}, 1)\,, \quad \boldsymbol{d}_3 = (du_{\mathrm{a}}, dv_{\mathrm{a}}, 1)\,, \tag{23}$$

$$\boldsymbol{d}_4 = (du_{\mathrm{a}} + du_{\mathrm{st}} + du_{\mathrm{d}}, dv_{\mathrm{a}} + dv_{\mathrm{st}} + dv_{\mathrm{d}}, 1)\,, \quad \boldsymbol{d}_5 = (du_{\mathrm{f}}, dv_{\mathrm{f}}, 1)\ . \tag{24}$$

As in the case of the first epipolar tensor, the entries of the other epipolar tensors $E_i = (a_i, b_i, q_i)^\top (a_i, b_i, q_i)$, for $2 \leq i \leq 5$, are related to the coefficients of the epipolar lines.

3.3 Constraint Normalisation

In [27] the authors demonstrate that the linearised brightness constancy assumption for optical flow can be interpreted geometrically as a weighted distance of the estimated flow to the line described by the optical flow constraint. Equivalently, the multidimensional brightness constancy constraint in (18) can be considered as the weighted distance of the scene flow to the hyperplane described by $\boldsymbol{g}_2^\top \boldsymbol{d} = 0$. To obtain the actual distance to the hyperplane we have to normalise the constraint by dividing it by the magnitude of the hyperplane normal. Since the last entry of $\boldsymbol{d}$ is constant, this normal vector is given by the first six components of $\boldsymbol{g}_2$, i.e. $\boldsymbol{n} = (g_{2\mathrm{rx}}, g_{2\mathrm{ry}}, g_{2xz}, g_{2yz}, g_{2\mathrm{rx}}, g_{2\mathrm{ry}})^\top$. Now it becomes explicit why it is desirable to penalise the actual distance to the hyperplane: Unlike the original constraint this distance does not scale with the magnitude of the derivatives contained in $\boldsymbol{g}_2$. This prevents overweighting at unreliable structures such as noise or occlusions that typically manifest themselves in large image gradients. The corresponding normalised quadratic form is given by

$$\frac{1}{|\boldsymbol{n}|^2 + \zeta^2}\,(\boldsymbol{g}_2^\top \boldsymbol{d})^2 \;=\; \boldsymbol{d}^\top \left(\frac{J_2}{\sum_{i=1}^{6} (J_2)_{ii} + \zeta^2}\right) \boldsymbol{d} \;=\; \boldsymbol{d}^\top \widehat{J}_2\, \boldsymbol{d}\ , \tag{25}$$

where $\zeta = 0.1$ is a constant that avoids division by zero, and $\widehat{J}_2$ denotes the normalised version of J_2. We apply the same normalisation strategy to the remaining data terms. For the extension to the gradient constancy and colour images we refer to [27] .

Our normalisation idea is, however, not restricted to the scene flow tensors only. By normalising the epipolar tensors as well we obtain a widely used geometrical error measure from computer vision: the distance to the epipolar lines [16]. Analogously to (25), we can derive the normalisation factor for the epipolar tensors. It reads

$$|\boldsymbol{n}_i|^2 + \zeta^2 \;=\; \textstyle\sum_{j=1}^{2} (E_i)_{jj} + \zeta^2 \;=\; a_i^2 + b_i^2 + \zeta^2\ . \tag{26}$$

Division by this factor then results in the normalised epipolar tensors $\widehat{E}_i$, for $1 \leq i \leq 5$.

4 Minimisation and Numerical Solution

By combining all terms derived in Sect. 3, we obtain the following differential form of our energy that has to be minimised at each level of the coarse-to-fine approach:

$$\begin{aligned}
&\mathcal{E}(d\boldsymbol{w}_{\mathrm{f}}, d\boldsymbol{w}_{\mathrm{st}}, d\boldsymbol{w}_{\mathrm{d}}, d\boldsymbol{w}_{\mathrm{a}}, F) = \\
&\int_{\Omega} \Bigg(o_{2\mathrm{l}}\, \Psi\Big(\boldsymbol{d}^{\top} \widehat{J}_1\, \boldsymbol{d}\Big) + o_{1\mathrm{r}}\, o_{2\mathrm{r}}\, \Psi\Big(\boldsymbol{d}^{\top} \widehat{J}_2\, \boldsymbol{d}\Big) + o_{1\mathrm{r}}\, \Psi\Big(\boldsymbol{d}^{\top} \widehat{J}_3\, \boldsymbol{d}\Big) + o_{2\mathrm{l}}\, o_{2\mathrm{r}}\, \Psi\Big(\boldsymbol{d}^{\top} \widehat{J}_4\, \boldsymbol{d}\Big) \\
&\quad + \beta_1\, \Psi\Big(\boldsymbol{d}_1^{\top} \widehat{E}_1\, \boldsymbol{d}_1\Big) + \beta_2\, \Psi\Big(\tfrac{1}{4}\, \boldsymbol{d}_2^{\top} \widehat{E}_2\, \boldsymbol{d}_2 + \tfrac{1}{4}\, \boldsymbol{d}_3^{\top} \widehat{E}_3\, \boldsymbol{d}_3 + \tfrac{1}{4}\, \boldsymbol{d}_4^{\top} \widehat{E}_4\, \boldsymbol{d}_4 + \tfrac{1}{4}\, \boldsymbol{d}_5^{\top} \widehat{E}_5\, \boldsymbol{d}_5\Big) \\
&\quad + \alpha_1\, \Psi\Big(|\nabla(\boldsymbol{w}_{\mathrm{f}} + d\boldsymbol{w}_{\mathrm{f}})|^2\Big) + \alpha_2\, \Psi\Big(|\nabla(\boldsymbol{w}_{\mathrm{st}} + d\boldsymbol{w}_{\mathrm{st}})|^2\Big) + \alpha_3\, \Psi\Big(|\nabla(\boldsymbol{w}_{\mathrm{d}} + d\boldsymbol{w}_{\mathrm{d}})|^2\Big) \\
&\quad + \beta_2\, \mu\, \Big(|\boldsymbol{w}_{\mathrm{f}} + d\boldsymbol{w}_{\mathrm{f}} - \boldsymbol{w}_{\mathrm{a}} - d\boldsymbol{w}_{\mathrm{a}}|^2\Big)\Big) \mathrm{d}\boldsymbol{x} \;\;,\;\; \text{with} \;\; \|F\|_{\mathrm{Frob}}^2 = 1 \;\; . \qquad (27)
\end{aligned}$$

Note that this energy is convex in the flow increments $d\boldsymbol{w}_{\mathrm{f}}$, $d\boldsymbol{w}_{\mathrm{st}}$, $d\boldsymbol{w}_{\mathrm{d}}$ and the auxiliary variable $d\boldsymbol{w}_{\mathrm{a}}$, since only squared arguments and convex penaliser functions are used. In order to minimise it under the given constraint $\|F\|_{\mathrm{Frob}}^2 = 1$, we follow [22] and use the method of the Lagrange multipliers. We thus obtain the Lagrangian

$$\mathcal{L}(d\boldsymbol{w}_{\mathrm{f}}, d\boldsymbol{w}_{\mathrm{st}}, d\boldsymbol{w}_{\mathrm{d}}, d\boldsymbol{w}_{\mathrm{a}}, F, \lambda) = \mathcal{E}(d\boldsymbol{w}_{\mathrm{f}}, d\boldsymbol{w}_{\mathrm{st}}, d\boldsymbol{w}_{\mathrm{d}}, d\boldsymbol{w}_{\mathrm{a}}, F) + \lambda(1 - \boldsymbol{f}^{\top} \boldsymbol{f}) \;\; , \quad (28)$$

where λ is the Lagrangian multiplier, and $\boldsymbol{f}$ is a vector that contains all 9 entries of F. This formulation suggests an alternating minimisation with two steps:

(i) Minimising the Lagrangian with respect to the flow increments leads to the corresponding Euler-Lagrange equations. By discretising them via finite difference approximations, one ends up with a nonlinear system of equations due to the robust function Ψ. To ensure fast convergence, we solve this system with a bidirectional multigrid framework based on a nonlinear point coupled Gauß-Seidel solver [3]. In the coarse-to-fine pyramid we use a downsampling factor of $\eta = 0.9$, while the images are warped onto the reference image using Coons patches based on bicubic interpolation [5].

(ii) Differentiation of the Lagrangian with respect to the fundamental matrix results in an eigenvalue problem [22] that is nonlinear due to Ψ and the normalisation weights (26). To solve this eigenvalue problem we apply a reweighted total least squares method in which the weights and the arguments of Ψ are fixed iteratively. We would like to point out that this step of the minimisation estimates the fundamental matrix from the *dense* correspondences of both stereo pairs.

The alternating computation of the flow increments and the fundamental matrix works as follows: The Euler-Lagrange equations are solved with a current estimate of the fundamental matrix. Using the newly computed flows, the fundamental matrix is updated by solving the eigenvalue problem. We extract a pair of camera matrices and perform a dense scene reconstruction by triangulation [10]. After recomputing the occlusion scores, the Euler-Lagrange equations are then solved again. This iterative process is repeated until convergence. We initialise the occlusion scores with 1 and compute the first iteration with disabled epipolar constraints.

Table 1. Evaluation of different methods on the rectified sphere sequence. Runtime on Intel Core2 1.86 GHz: ∼420 seconds. Parameters: $\alpha_1 = 2$, $\alpha_2 = 1.5$, $\alpha_3 = 0.3$, $\beta_1 = \beta_2 = 0.1$, $\gamma = 0.1$, $\mu = 1$.

Method	RMSE			AAE
	$(u_\mathrm{f}, v_\mathrm{f}, u_\mathrm{d}, v_\mathrm{d})$	$(u_\mathrm{f}, v_\mathrm{f})$	$(u_\mathrm{st}, v_\mathrm{st})$	$(u_\mathrm{f}, v_\mathrm{f})$
Our method initialised with [8]	**1.76**	**0.63**	3.8	1.17
Our method	1.78	**0.63**	5.5	**1.16**
Wedel *et al.* [24] with ground truth	2.40	0.65	–	1.40
Wedel *et al.* [24] (87%)	2.45	0.66	**2.9**	1.50
Huguet and Devernay [12]	2.51	0.69	3.8	1.75
Wedel *et al.* [24] (100%)	2.55	0.77	10.9	2.76

5 Experiments

We evaluate the performance of our method on synthetic stereo sequences with ground truth and on real world images. To assess the quality we compute the root mean square error RMSE of the scene flow $(u_\mathrm{f}, v_\mathrm{f}, u_\mathrm{d}, v_\mathrm{d})$, the optical flow $(u_\mathrm{f}, v_\mathrm{f})$ and the stereo flow $(u_\mathrm{st}, v_\mathrm{st})$, as well as the absolute angular error AAE of the optical flow, see [24]. As a quality measure for the fundamental matrix we use the error d_F according to [7]. It is determined by using the estimated fundamental matrix to randomly create a large number (100,000) of correspondences and the ground truth fundamental matrix to establish their epipolar lines. After computing the average distance between all points and lines, the roles of the matrices are reversed to obtain a symmetric measure in pixel units.

In a first experiment we consider the synthetic sphere sequence of Huguet and Devernay [12] (*http://devernay.free.fr/vision/varsceneflow/*), which is composed of four 512×512 images of a textured sphere with rotating hemispheres. Despite the fact that this sequence is rectified, and thus constitutes a special case with vanishing vertical components of the stereo flow, it is a good benchmark for comparison against existing techniques. Additionally it requires to estimate large stereo displacements which pose a challenge to variational methods. In this context we follow the idea of [24] and [12], and initialise $(u_\mathrm{st}, v_\mathrm{st})$ with a dedicated method for large displacements. To this end, we use a variant of the recent optical flow technique of [1] with constraint normalisation and SIFT matches [15] as prior. For consistency we also included results for initialisation with the belief propagation algorithm of [8], as used by Huguet and Devernay. However, this initialisation is only applicable for rectified images.

Table 1 compares our results with those of the variational methods of Huguet and Devernay [12] and Wedel *et al.* [24] and lists the errors computed within the sphere. With a substantial improvement in the RMSE for $(u_\mathrm{f}, v_\mathrm{f}, u_\mathrm{d}, v_\mathrm{d})$ and in the AAE we consistently outperform the other approaches for the scene flow, although these methods are specifically tailored to the rectified case. The lower RMSE of the method of Wedel *et al.* for $(u_\mathrm{st}, v_\mathrm{st})$ is due to the fact it uses sparse stereo correspondences that do not provide results in occluded regions. However, the accuracy of their estimated scene flow is significantly lower than ours. This even holds if they use *ground truth* for the stereo correspondences. The good performance of our method is also reflected in the accurate estimation of the stereo geometry: We obtain a subpixel precision of $d_F = 0.019$.

In a second experiment we evaluate the performance for a general stereo geometry. To this end we generated a synthetic sequence of four frames with ground truth (available at *http://www.mia.uni-saarland.de/valgaerts/eccv10/sceneflow*). It is similar to the one of the previous experiment: A textured sphere with rotating hemispheres is positioned against a plane in the background as shown in Fig. 2. To demonstrate the benefits of the different design steps in our model we start from a variant that performs a joint regularisation of the flows as in [12] and does not include constraint normalisation. We then refine the model by subsequently adding the normalisation and the separate regulariser. Table 2 lists the progressively improving results. The errors are computed in the non-occluded regions of the whole image domain. The AAE is not listed because it is not defined for the zero flow in the background. In Fig. 2 the computed flow fields are shown together with the obtained occlusion scores, the 3D reconstruction and the scene flow. As one can see, the estimated displacements resemble the ground truth very well. Again, this is confirmed by a subpixel precision of $d_F = 0.021$ for the stereo geometry.

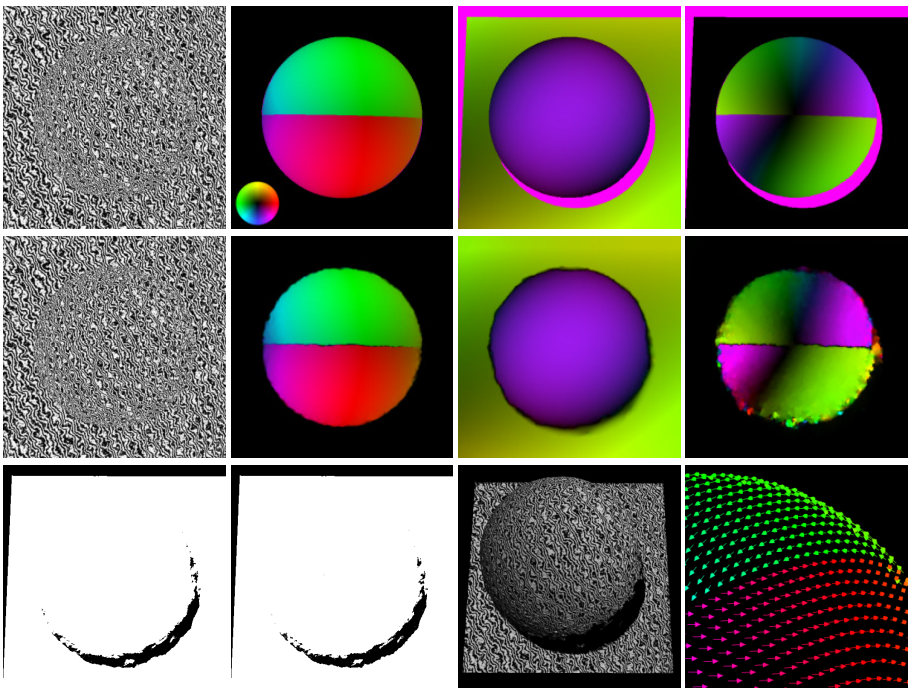

Fig. 2. Results for the general sphere sequence (image size 512×512). **Top Row: (a)** Left frame at first time step. **(b) + (c) + (d)** Ground truth of left optical flow, first stereo flow and flow change. Colour encodes the direction, brightness the magnitude (see colour circle). Occlusions are coloured pink. **Middle Row: (e)** Left frame at second time step. **(f) + (g) + (h)** Estimated left optical flow, first stereo flow and flow change. **Bottom Row: (i) + (j)** Estimated occlusion scores $o_{1\mathrm{r}}$ and $o_{2\mathrm{r}}$. **(k)** Estimated scene reconstruction. **(l)** Estimated scene flow. Runtime: ∼420 seconds. Parameters: $\alpha_1 = 1.5$, $\alpha_2 = 2$, $\alpha_3 = 0.8$, $\beta_1 = \beta_2 = 0.03$, $\gamma = 0.1$, $\mu = 1$.

Table 2. Evaluation of different variants of our method on the general sphere sequence.

Method	RMSE $(u_\mathrm{f}, v_\mathrm{f}, u_\mathrm{d}, v_\mathrm{d})$	RMSE $(u_\mathrm{f}, v_\mathrm{f})$	RMSE $(u_\mathrm{st}, v_\mathrm{st})$
joint regularisation	0.67	0.64	2.08
joint regularisation + normalisation	0.63	**0.59**	1.86
separate regularisation + normalisation	**0.61**	**0.59**	**1.61**

Fig. 3. Results for real world sequences (image size 470×340). **Top Row: (a) + (b)** *Smiling*, left frames at consecutive time steps. **(c) + (d)** *Closing Mouth*, left frames at consecutive time steps. **Bottom Row: (e)** Reconstruction and overlayed scene flow for *Smiling*. Increasing magnitude from green to red. **(f)** Close-up *Smiling*. **(g)** Close-up *Closing Mouth*. Runtime: $\sim$260 seconds. Parameters: $\alpha_1 = 15$, $\alpha_2 = 20$, $\alpha_3 = 15$, $\beta_1 = \beta_2 = 0.5$, $\gamma = 30$, $\mu = 1$.

For our last experiment we have recorded two uncalibrated stereo sequences to test the performance of our method on real world data. The results are shown in Fig. 3 for the sequences *Smiling* and *Closing Mouth*. As one can verify in both cases the 3D structure and the motion of the face are captured well and look very realistic. We emphasise that these two results are obtained from only four frames. Additional real world results can be found at *http://www.mia.uni-saarland.de/valgaerts/eccv10/sceneflow*.

6 Conclusions

We have presented a general approach for the dense estimation of scene flow, scene structure and geometry from uncalibrated stereo sequences. Our contributions are threefold: (i) We generalise the classical four-frame case to arbitrary stereo setups by embedding epipolar constraints into a joint energy functional with data and smoothness terms. (ii) We introduce a tensor notation which allows us to normalise the data and stereo constraints such that they become geometrically interpretable. (iii) We present a separate robustfication of the smoothness terms to handle scenarios where flow discontinuities do not coincide. Our evaluation has demonstrated that the proposed approach is not only more general than existing methods but also more accurate: Even without explicitly knowing the stereo geometry, we outperform recent techniques that have been specifically designed for the rectified case. Furthermore, the stereo geometry is estimated with sub-pixel precision and reconstructions for real world data show that scene structure and motion are determined with high quality. This clearly demonstrates the benefit of a joint computation of flow, structure and geometry.

Acknowledgements. We gratefully acknowledge partial funding by the Deutsche Forschungsgemeinschaft (*WE 2602/6-1*), and by the International Max-Planck Research School. We thank Pascal Gwosdek, Jennifer Metzger and Sebastian Volz for their help.

References

1. Brox, T., Bregler, C., Malik, J.: Large displacement optical flow. In: Proc. 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. pp. 41–48. IEEE Computer Society Press, Miami, FL (2009)
2. Brox, T., Bruhn, A., Papenberg, N., Weickert, J.: High accuracy optic flow estimation based on a theory for warping. In: Pajdla, T., Matas, J. (eds.) Computer Vision – ECCV 2004, Lecture Notes in Computer Science, vol. 3024, pp. 25–36. Springer, Berlin (2004)
3. Bruhn, A., Weickert, J.: Towards ultimate motion estimation: Combining highest accuracy with real-time performance. In: Proc. Tenth International Conference on Computer Vision. vol. 1, pp. 749–755. IEEE Computer Society Press, Beijing, China (2005)
4. Carceroni, R.L., Kutulakos, K.N.: Multi-view scene capture by surfel sampling: From video streams to non-rigid 3d motion, shape and reflectance. International Journal of Computer Vision 49(2-3), 175–214 (2002)
5. Coons, S.A.: Surfaces for computer aided design of space forms. Tech. Rep. MIT/LCS/TR-41, Massachusetts Institute of Technology, Cambridge, MA (1967)
6. Courchay, J., Pons, J.P., Monasse, P., Keriven, R.: Dense and accurate spatio-temporal multi-view stereovision. In: Zha, H., Taniguchi, R., Maybank, S. (eds.) Proc. Ninth Asian Conference on Computer Vision. Lecture Notes in Computer Science, China (2009)
7. Faugeras, O., Luong, Q.T., Papadopoulo, T.: The Geometry of Multiple Images. MIT Press, Cambridge, MA (2001)
8. Felzenszwalb, P.F., Huttenlocher, D.P.: Efficient belief propagation for early vision. International Journal of Computer Vision 40(1), 41–54 (2006)
9. Furukawa, Y., Ponce, J.: Dense 3d motion capture from synchronized video streams. In: Proc. 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. IEEE Computer Society Press, Anchorage, AK (2008)

10. Hartley, R., Zisserman, A.: Multiple View Geometry in Computer Vision. Cambridge University Press, Cambridge, UK (2000)
11. Horn, B., Schunck, B.: Determining optical flow. Artificial Intelligence 17, 185–203 (1981)
12. Huguet, F., Devernay, F.: A variational method for scene flow estimation from stereo sequences. In: Proc. Eleventh International Conference on Computer Vision. IEEE Computer Society Press, Rio de Janeiro (2007)
13. Isard, M., MacCormick, J.: Dense motion and disparity estimation via loopy belief propagation. In: Narayanan, P.J., Nayar, S.K., Shum, H.Y. (eds.) Proc. Seventh Asian Conference on Computer Vision. Lecture Notes in Computer Science, vol. 3852, pp. 32–41 (2006)
14. Longuet-Higgins, H.C.: A computer algorithm for reconstructing a scene from two projections. Nature 293, 133–135 (1981)
15. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision 60(2), 91–110 (2004)
16. Luong, Q.T., Faugeras, O.D.: The fundamental matrix: theory, algorithms, and stability analysis. International Journal of Computer Vision 17(1), 43–75 (1996)
17. Mémin, E., Pérez, P.: Dense estimation and object-based segmentation of the optical flow with robust techniques. IEEE Transactions on Image Processing 7(5), 703–719 (1998)
18. Min, D.B., Sohn, K.: Edge-preserving simultaneous joint motion-disparity estimation. In: Proc. 18th International Conference on Pattern Recognition. pp. 74–77. Hong Kong (2006)
19. Patras, I., Alvertos, N., Tziritas, G.: Joint disparity and motion field estimation in steroscopic image sequences. In: Proc. 13th International Conference on Pattern Recognition. vol. 1, pp. 359–362. Vienna, Austria (1996)
20. Pons, J.P., Keriven, R., Faugeras, O.D.: Multi-view stereo reconstruction and scene flow estimation with a global image-based matching score. International Journal of Computer Vision 72(2), 179–193 (2007)
21. Rudin, L.I., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D 60, 259–268 (1992)
22. Valgaerts, L., Bruhn, A., Weickert, J.: A variational model for the joint recovery of the fundamental matrix and the optical flow. In: Rigoll, G. (ed.) Pattern Recognition. Lecture Notes in Computer Science, vol. 3663, pp. 314–324. Springer, Berlin (2008)
23. Vedula, S., Baker, S., Rander, P., Collins, R.T., Kanade, T.: Three-dimensional scene flow. IEEE Transactions on Pattern Analysis and Machine Intelligence 27(3), 475–480 (2005)
24. Wedel, A., Rabe, C., Vaudrey, T., Brox, T., Franke, U., Cremers, D.: Efficient dense scene flow from sparse or dense stereo data. In: Forsyth, D., Torr, P., Zisserman, A. (eds.) Computer Vision – ECCV 2008, Lecture Notes in Computer Science, vol. 5302, pp. 739–751. Springer, Berlin (2008)
25. Wedel, A., Vaudrey, T., Meissner, A., Rabe, C., Brox, T., Franke, U., Cremers, D.: An evaluation approach for scene flow with decoupled motion and position. In: Cremers, D., Rosenhahn, B., Yuille, A.L., Schmidt, F.R. (eds.) Statistical and Geometrical Approaches to Visual Motion Analysis. Lecture Notes in Computer Science, vol. 5604, pp. 46–69. Springer, Berlin (2008)
26. Zhang, Y., Kambhamettu, C.: On 3d scene flow and structure estimation. In: Proc. 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. pp. 778–785. IEEE Computer Society Press, Kauai, HI (2001)
27. Zimmer, H., Bruhn, A., Weickert, J., Valgaerts, L., Salgado, A., Rosenhahn, B., Seidel, H.P.: Complementary optic flow. In: Cremers, D., Boykov, Y., Blake, A., Schmidt, F.R. (eds.) Energy Minimization Methods in Computer Vision and Pattern Recognition – EMMCVPR, Lecture Notes in Computer Science, vol. 5681, pp. 207–220. Springer, Berlin (2009)