Universität des Saarlandes

Fachrichtung 6.1 – Mathematik

Preprint Nr. 263

Dense versus Sparse Approaches for Estimating the Fundamental Matrix

Levi Valgaerts, Andrés Bruhn,
Markus Mainberger and Joachim Weickert

Saarbrücken 2010

Fachrichtung 6.1 – Mathematik
Universität des Saarlandes

Preprint No. 263
submitted: April 30, 2010
revised: May 5, 2011

Dense versus Sparse Approaches for Estimating the Fundamental Matrix

Levi Valgaerts
Vision and Image Processing Group
MMCI Cluster of Excellence
Saarland University, Campus E1.1
66041 Saarbrücken
Germany
valgaerts@mmci.uni-saarland.de

Andrés Bruhn
Vision and Image Processing Group
MMCI Cluster of Excellence
Saarland University, Campus E1.1
66041 Saarbrücken
Germany
bruhn@mmci.uni-saarland.de

Markus Mainberger
Mathematical Image Analysis Group
Faculty of Mathematics and Computer Science
Saarland University, Campus E1.1
66041 Saarbrücken
Germany
mainberger@mia.uni-saarland.de

Joachim Weickert
Mathematical Image Analysis Group
Faculty of Mathematics and Computer Science
Saarland University, Campus E1.1
66041 Saarbrücken
Germany
weickert@mia.uni-saarland.de

Edited by
FR 6.1 – Mathematik
Universität des Saarlandes
Postfach 15 11 50
66041 Saarbrücken
Germany

Fax: + 49 681 302 4443
e-Mail: preprint@math.uni-sb.de
WWW: http://www.math.uni-sb.de/

Dense versus Sparse Approaches for Estimating the Fundamental Matrix

Levi Valgaerts Andrés Bruhn Markus Mainberger
Joachim Weickert

Abstract

There are two main strategies for solving correspondence problems in computer vision: sparse local feature based approaches and dense global energy based methods. While sparse feature based methods are often used for estimating the fundamental matrix by matching a small set of sophistically optimised interest points, dense energy based methods mark the state of the art in optical flow computation. The goal of our paper is to show that this separation into different application domains is unnecessary and can be bridged in a natural way. As a first contribution we present a new application of dense optical flow for estimating the fundamental matrix. Comparing our results with those obtained by feature based techniques we identify cases in which dense methods have advantages over sparse approaches. Motivated by these promising results we propose, as a second contribution, a new variational model that recovers the fundamental matrix and the optical flow simultaneously as the minimisers of a single energy functional. In experiments we show that our coupled approach is able to further improve the estimates of both the fundamental matrix and the optical flow. Our results prove that dense variational methods can be a serious alternative even in classical application domains of sparse feature based approaches.

1 Introduction

While correspondence problems are omnipresent in computer vision, it is surprising that their two main research directions have evolved without much interaction: On the one hand *sparse, feature based methods* have been developed for estimating the epipolar geometry from stereo and multiview data, and numerous statistical efforts have been spent to select the most useful set of sparse correspondences. On the other hand, *dense, energy based methods* have become the leading techniques for estimating the correspondences (optical flow) in image sequences. The goal of our paper is to show that these dense methods can also be very beneficial for estimating the epipolar geometry, and that combining epipolar geometry computation with energy minimisation also leads to better optical flow methods.
Let us first sketch feature based approaches for estimating the epipolar geometry. The epipolar geometry is the relation that underlies two stereo views and can be

described by a single entity, the *fundamental matrix* [33, 22]. For uncalibrated images it is possible to estimate the fundamental matrix solely from image correspondences by means of the *epipolar constraint*, which restricts corresponding points in the two views to lie on their respective *epipolar lines*. For feature based methods these correspondences are typically derived by matching a sparse set of characteristic image features, for example Förstner-Harris features [25, 30], KLT features [66], as well as SIFT or SURF features [43, 8]. The first such approach goes back to Longuet-Higgins [42], who introduced the 8-point algorithm to compute the *essential matrix*, the equivalent of the fundamental matrix for internally calibrated cameras. The 8-point algorithm excels in simplicity because of its linear nature but it has the disadvantage that the quantity being minimised has no geometrical interpretation. Weng *et al.* [86] and Luong and Faugeras [45] proposed nonlinear techniques involving geometrically meaningful measures such as the distance of a point to its epipolar line. Hartley and Zisserman [33] recommend a Maximum Likelihood (ML) estimation that minimises the distance of a point to the manifold determined by the parameterisation of the fundamental matrix. The ML estimate is optimal from a statistical viewpoint if a Gaussian error model is assumed [85] and is sometimes called the *Gold Standard algorithm*. In practice, a feature based estimation method has to be able to deal with false correspondences arising from the lack of geometrical constraints in the matching process. This has led to a multitude of robust extensions that can handle a relatively large amount of outliers: M-estimators [35], Least Median of Squares (LMedS) [59] and the numerous variants of the Random Sample Consensus (RANSAC) [23] number among such robust techniques. An overview of robust estimators in the context of fundamental matrix computation can be found in [76, 91, 70, 57]. In addition to feature based methods there exists a limited number of approaches that estimate the fundamental matrix directly from image information [65].

Clearly, the quality of feature based methods relies on the quality of the random sampling approach. However, one should not forget that also the features may suffer from well-known localisation errors due to their computation in scale-space; see e.g. [81, 90].

Now that we have discussed sparse, feature based methods for estimating the fundamental matrix, we review dense techniques for establishing correspondences within a global energy minimisation framework. Typical representatives are variational methods for computing the optical flow. They minimise an energy functional that models temporal constancy of image properties via a data term and regularity of the flow field via a smoothness term. The quadratic data term of the seminal model of Horn and Schunck [34] has been gradually extended with robust variants that combine several constancy assumptions in order to cope with noise, occlusions and illumination changes [14, 37, 52, 68, 87]. To respect discontinuities in the optical flow, smoothness terms have been proposed that take into

account image edges [53, 3], edges in the evolving flow field [63, 84] or both [93]. Similar extensions have been introduced in a probabilistic setting using discrete models [11, 49, 74] or by coupled systems of partial differential equations [56]. The minimisation of variational optical flow methods often proceeds by means of a gradient descent approach or by discretising the Euler-Lagrange equations.

The basic idea of using dense, energy based methods is not restricted to optical flow computation, it can also be applied to depth estimation from stereo pairs. If the epipolar geometry of a stereo image pair is known, the correspondence problem even reduces to a one dimensional search – the search along epipolar lines. Most successful stereo correspondence methods are either discrete or continuous optimisation techniques [61]. Discrete methods model the images and displacement field as Markov random fields and try to find the most probable displacement for the given data. Minimisation is usually done by means of graph cuts [39], belief propagation [38] or dynamic programming [40]. While discrete approaches allow a better optimisation of the energy by constraining the depth values to a discrete subset, continuous variational methods can be advantageous when smooth transitions are favoured. Variational methods either decompose the optical flow along the epipolar lines [2, 67], restrict the estimation process in horizontal direction after rectifying the images [9], or directly solve for the unknown depth in every pixel [72].

While ideas from optical flow computation have been widely adopted for dense stereo matching, it is astonishing that these concepts have rarely found their way into the estimation of the fundamental matrix. In fact, recent developments in variational optical flow (e.g. [14, 88, 54, 13, 93]) indicate that there are dense alternatives to feature based methods for reliable unconstrained correspondence computation. Apart from their high accuracy, variational methods offer two potential advantages for the computation of the fundamental matrix: (i) Due to the filling-in effect they provide a dense flow field, and thus a huge amount of correspondences, that can increase the robustness of the estimation process. (ii) They do not create gross outliers – i.e. arbitrary image locations that are wrongly matched across the whole image plane – because of the combination of robust data constraints and global smoothness assumptions.

However, not only the benefit of dense optical flow methods for estimating the epipolar geometry seems promising, but also the opposite direction: Knowing the epipolar geometry can have a stabilising effect on computing optical flow fields. First attempts along this line have been made by two-step approaches that feed a precomputed epipolar geometry into an optical flow method [2, 72, 67, 9, 83]. More recently it has been argued in two conference papers that it can be beneficial to couple the computation of epipolar geometry and optical flow fields within a joint energy functional [78, 82].

The goal of the present paper is to address these two issues. By presenting a first systematic juxtaposition of sparse and dense methods for correspondence problems, we come up with two contributions: First, we demonstrate that modern dense optical flow methods can serve as novel approaches for estimating the epipolar geometry with competitive quality. Secondly, we demonstrate that this quality can be further improved within a joint variational approach for simultaneous estimation of epipolar geometry and optical flow. This joint method also yields better results then an optical flow approach without epipolar geometry estimation.
While these key results are of a more general nature, we have to restrict our methodology to prototypical representatives. As a prototype for an accurate dense optical flow method we choose the approach of Brox *et al.* [14], sometimes also replaced by the recent method of Zimmer *et al.* [92]. For feature based approaches we consider two feature matching algorithms (KLT [66] and SIFT [43]), three random sampling algorithms (LMedS [59], LORANSAC [19] and DEGENSAC [20]), and two different distance measures (epipolar distance [22] and reprojection error [33]). This comes down to twelve different variants of feature based methods.
Our paper is organised as follows. In Section 2 we describe our method for estimating the fundamental matrix from optical flow. In Section 3 we give an overview of the selected feature based methods and we present a first experimental comparison in Section 4. Section 5 is dedicated to our novel variational model that couples the estimation of the epipolar geometry and the optical flow. After a second experimental comparison in Section 6 we conclude with a summary and outlook on future work in Section 7.

Related Work Early ideas that couple optical flow with the estimation of the epipolar geometry go back to the differential form of the epipolar constraint by Viéville and Faugeras [80] and Brooks *et al.* [12]. Based on these works, Ohta and Kanatani [55] and Kanatani *et al.* [36] have presented statistical evaluations of the estimation of the stereo geometry and the depth from monocular image sequences. These studies suffered, however, from inaccurate optical flow methods and a lack of absolute performance measures. Moreover, in a differential setting optical flow is regarded as the infinitesimal displacement field of an image sequences , thereby strictly separating *structure-from-motion* from *wide baseline stereo*. We do not make such a distinction since recent optical flow methods are able to cope with both small and large displacements.
Only the work of Strecha *et al.* [71] is known to the authors in which dense optical flow correspondences are used for the calibration of a stereo rig, but no quantitative results are reported. In our coupled model for optical flow and fundamental

matrix estimation a rigid motion model enters the functional in the form of a soft constraint with unknown fundamental matrix entries. In the optical flow model of Nir *et al.* [54] this happens via an explicit parameterisation of the displacements with unknown coefficients. Also close in spirit to our ideas are feature based ML methods that simultaneously estimate the fundamental matrix while correcting an initial set of point correspondences [85, 33]. These methods are, however, inherently sparse and differ by the distance measure that is being minimised. Another more recent sparse method that pairs the epipolar constraint with the brightness constancy assumption is that of Saragih and Goecke [60]. In an early work, Hanna [29] iteratively estimates camera motion and dense structure parameters in a joint refinement process by using the optical flow as an intermediate representation. Contrary to our approach, the problem is formulated in a differential setting and the optimisation performed locally. A simultaneous estimation of the fundamental matrix and the 3D surface is proposed by Schlesinger *et al.* [62] in an uncalibrated discrete setting. Despite a joint model formulation, proper initialisation is required to bootstrap the method.
Some preliminary results of our work have been presented at local conferences [46, 78]. In the present paper we consider a larger number of optical flow and feature based methods, including more recent ones. Last but not least, we carry out a significant number of additional experiments. They analyse more aspects and provide deeper insights in the real potential and the limitations of dense optical flow methods in connection with epipolar geometry estimation.

2 Fundamental Matrix Estimation from Optical Flow

In this section we propose our first contribution: a novel two-step method for estimating the fundamental matrix from dense optical flow. We first establish a dense correspondence set between the two input images by computing a displacement vector for every pixel. This results in a set of matches from which we then estimate the fundamental matrix with a modified version of the 8-point algorithm. We assume that the images have already been corrected for radial distortion. The 8-point algorithm, however, can be extended such that it simultaneously recovers the fundamental matrix and the radial distortion [24].

2.1 Variational Optical Flow Computation

We compute the optical flow with the variational method that was proposed by Brox *et al.* [14]. Although this is not one of the latest methods, it may serve as a popular prototype of modern variational optical flow techniques. We use a variant with spatial instead of spatio-temporal smoothing and briefly summarise it here.

Let $g(x,y,t) : \Omega \times [0,\infty) \to \mathbb{R}$ be an image sequence and $\boldsymbol{x} = (x,y,t)^\top$ a location within the rectangular image domain $\Omega \subset \mathbb{R}^2$ at a time $t \geq 0$. We further assume that $g(x,y,t)$ is presmoothed by a Gaussian convolution of standard deviation σ and that the left and right image of the uncalibrated stereo pair are embedded in the sequence as two consecutive frames $g(x,y,t)$ and $g(x,y,t+1)$. The optical flow $\boldsymbol{w} = (u,v,1)^\top$ between the two frames is then found by minimising the energy functional

$$\begin{aligned}\mathscr{E}(\boldsymbol{w}) = \int_\Omega \Big(& \Psi\left(|g(\boldsymbol{x}+\boldsymbol{w}) - g(\boldsymbol{x})|^2 + \gamma \cdot |\nabla g(\boldsymbol{x}+\boldsymbol{w}) - \nabla g(\boldsymbol{x})|^2 \right) \\ & + \alpha \Psi\left(|\nabla \boldsymbol{w}|^2 \right) \Big) \,\mathrm{d}x\,\mathrm{d}y \ , \end{aligned} \tag{1}$$

where $\nabla = (\partial_x, \partial_y)^\top$ and $|\nabla \boldsymbol{w}|^2 := |\nabla u|^2 + |\nabla v|^2$ denotes the squared magnitude of the spatial flow gradient. The first term of $\mathscr{E}(\boldsymbol{w})$ is the data term. It models the constancy of the image brightness and the spatial image gradient along the displacement trajectories. These two constraints combined provide robustness against varying illumination, while their implicit formulation (no linearisation) makes it possible to deal with the large displacements that are usually present in wide baseline stereo. The second term in the functional is the smoothness term, which penalises deviations of the flow field from piecewise smoothness. For the function Ψ the regularised L_1 penaliser

$$\Psi(s^2) = \sqrt{s^2 + \varepsilon^2} \ , \tag{2}$$

is chosen, with $\varepsilon = 10^{-3}$ a small constant. In the case of the smoothness term this equals total variation (TV) regularisation.
The energy functional (1) is minimised via a warping strategy as described in [14]. The flow is incrementally refined on each level of a multi-resolution pyramid such that the algorithm does not easily get trapped in a local minimum. Moreover, we followed the multigrid framework suggested in [15] to speed up the computation of the resulting nonlinear systems of equations. To be able to use RGB colour images we consider a multichannel variant of energy (1) where the 3 colour channels are coupled in the data term as follows:

$$\int_\Omega \Psi\left(\sum_{i=1}^{3} |g_i(\boldsymbol{x}+\boldsymbol{w}) - g_i(\boldsymbol{x})|^2 + \gamma \cdot \sum_{i=1}^{3} |\nabla g_i(\boldsymbol{x}+\boldsymbol{w}) - \nabla g_i(\boldsymbol{x})|^2 \right) \mathrm{d}x\,\mathrm{d}y \ . \tag{3}$$

Once the optical flow $\boldsymbol{w}$ has been computed, we establish a set of matches by determining for every point $(x,y)^\top$ in the left image the corresponding point $(x+u, y+v)^\top$ in the right image. In practice we do this at the discrete pixel locations, resulting in a finite number of matches. We exclude points that are warped outside

the image domain by the optical flow because the data term cannot be evaluated in these regions, leading to less reliable correspondences.
While the method above serves as our baseline algorithm for obtaining dense optical flow fields, we will also use a more sophisticated optical flow method in our experiments in Section 6. It goes back to [92] and uses constraint normalisation in the data term and a specific anisotropic smoothness term that works in a way that is complementary to the data term.

2.2 The 8-point Algorithm of Longuet-Higgins

The fundamental matrix of a stereo pair is a 3×3 matrix of rank 2 that is defined up to a scaling factor. It can be computed from image correspondences by means of the epipolar constraint. The epipolar constraint between a given point $\tilde{\boldsymbol{x}} = (x, y, 1)^\top$ in the left image and its corresponding point $\tilde{\boldsymbol{x}}' = (x', y', 1)^\top$ in the right image can be rewritten in the form [22]

$$0 = \tilde{\boldsymbol{x}}'^\top F \, \tilde{\boldsymbol{x}} = \boldsymbol{s}^\top \boldsymbol{f} \quad , \tag{4}$$

with the two 9 dimensional vectors $\boldsymbol{s}$ and $\boldsymbol{f}$ defined as

$$\boldsymbol{s} = \big(xx', yx', x', xy', yy', y', x, y, 1\big)^\top , \tag{5}$$

$$\boldsymbol{f} = (f_{1,1}, f_{1,2}, f_{1,3}, f_{2,1}, f_{2,2}, f_{2,3}, f_{3,1}, f_{3,2}, f_{3,3})^\top \quad . \tag{6}$$

The tilde superscript indicates that we are using projective coordinates and the entries of the fundamental matrix F are denoted by $f_{i,j}$, with $1 \leq i, j \leq 3$. Since the 9 components of $\boldsymbol{f}$ are defined up to a scale factor, 8 point matches are in general sufficient to uniquely determine a solution from Eq. (4). In practice, point matches are not exact and the entries of F can be estimated more robustly from $n \geq 8$ correspondences by minimising the energy

$$\mathscr{E}(\boldsymbol{f}) = \sum_{i=1}^{n} (\boldsymbol{s}_i^\top \boldsymbol{f})^2 = \|S\,\boldsymbol{f}\|^2 \quad , \tag{7}$$

where S is an $n \times 9$ matrix with rows made up of the constraint vectors $\boldsymbol{s}_i^\top$, $1 \leq i \leq n$. Minimising the energy (7) is equivalent to finding a least squares solution to the overdetermined homogeneous system $S\,\boldsymbol{f} = \boldsymbol{0}$. We can avoid the trivial solution $\boldsymbol{f} = \boldsymbol{0}$ by imposing an explicit constraint on the Frobenius norm, such as $\|F\|_{\text{Frob}}^2 = \|\boldsymbol{f}\|^2 = 1$. The solution of this *total least squares (TLS)* problem is the eigenvector that belongs to the smallest eigenvalue of $S^\top S$. This simple algorithm is known in literature as the *8-point algorithm* [42] and can be implemented numerically with the help of the Jacobi method [27].

2.3 Dense Fundamental Matrix Estimation

If we use point matches that have been established by optical flow we expect that there will be outliers due to noise, occlusions and illumination changes that have not been modelled (e.g. reflections and transparencies). To account for this we estimate the fundamental matrix with a robust version of the 8-point algorithm. This is done by replacing the quadratic penalisation in the energy (7) by another function of the residual

$$\mathscr{E}(\boldsymbol{f}) = \sum_{i=1}^{n} \Psi\left((\boldsymbol{s}_i^\top \boldsymbol{f})^2\right) \ , \tag{8}$$

where $\Psi(s^2)$ is a positive, symmetric and in general convex function in s that grows sub-quadratically. We choose for our approach the regularised L_1 norm (2). Applying the method of Lagrange multipliers to the problem of minimising the energy (8) under the constraint $\|\boldsymbol{f}\|^2 = \boldsymbol{f}^\top \boldsymbol{f} = 1$ means that we are looking for critical points of the function

$$\mathscr{F}(\boldsymbol{f}, \lambda) = \sum_{i=1}^{n} \Psi\left((\boldsymbol{s}_i^\top \boldsymbol{f})^2\right) + \lambda(1 - \boldsymbol{f}^\top \boldsymbol{f}) \ . \tag{9}$$

Setting the derivatives of $\mathscr{F}(\boldsymbol{f}, \lambda)$ with respect to $\boldsymbol{f}$ and λ to zero yields the nonlinear problem

$$\begin{aligned} \mathbf{0} &= \left(\sum_{i=1}^{n} \Psi'\left((\boldsymbol{s}_i^\top \boldsymbol{f})^2\right) \boldsymbol{s}_i \boldsymbol{s}_i^\top - \lambda I \right) \boldsymbol{f}, & (10) \\ &=: \left(S^\top W(\boldsymbol{f}) S - \lambda I \right) \boldsymbol{f}, & (11) \\ 0 &= 1 - \|\boldsymbol{f}\|^2 \ . & (12) \end{aligned}$$

In the above formula W is an $n \times n$ diagonal matrix with positive weights $w_{i,i} = \Psi'\left((\boldsymbol{s}_i^\top \boldsymbol{f})^2\right)$. To solve this nonlinear system we follow a lagged iterative scheme in which we fix the symmetric positive definite system matrix $S^\top W S$ for the current estimate $\boldsymbol{f}^k$. This results in an eigenvalue problem that is solved in the same way as the standard 8-point algorithm to obtain the updated solution $\boldsymbol{f}^{k+1}$. By repeating this process, the solution is successively refined in a *reweighted total least squares* (RTLS) sense [76, 33, 22]. Because the calculation of the weights $w_{i,i}$ requires an estimate of the fundamental matrix and vice versa, we use the standard 8-point algorithm to obtain an initial estimate.

2.4 Data Normalisation and Rank Enforcement

The 8-point algorithm is not invariant to similarity transformations, such as translation, rotation and scaling of the image coordinates. At the same time the eigenvalue problem is poorly conditioned because of the different orders of magnitude of the projective coordinates. It is therefore essential that all points are expressed in a fixed coordinate frame prior to the application of the 8-point algorithm. Hartley [31] proposes a data *normalisation* that translates and scales the points $\tilde{\boldsymbol{x}}_i$ and $\tilde{\boldsymbol{x}}'_i$, $1 \leq i \leq n$, by the affine mappings T and T', such that $T\tilde{\boldsymbol{x}}_i$ and $T'\tilde{\boldsymbol{x}}'_i$ have the projective coordinate $(1,1,1)^\top$ on average. The epipolar constraint can be rewritten in terms of the normalised coordinates as

$$\tilde{\boldsymbol{x}}'^\top T'^\top \hat{F}\, T\, \tilde{\boldsymbol{x}} = \hat{\boldsymbol{s}}^\top\ \hat{\boldsymbol{f}}\ , \tag{13}$$

where $\hat{\boldsymbol{f}}$ is the vector notation of the fundamental matrix $\hat{F}$ of the transformed data. It has to be noted that the solution of the 8-point algorithm for the normalised data corresponds to a transformed energy and that a fundamental matrix for the original data can be recovered as $F = T'^\top \hat{F}\, T$.
A second issue posed by the method described here concerns the rank of F. The solution of the 8-point algorithm will in general not satisfy the singularity constraint, such that it is common to perform a *rank enforcement* step after the estimation. This can be done by replacing the solution with the closest rank 2 matrix using e.g. singular value decomposition (SVD) [77]. In our robust 8-point algorithm we use SVD to enforce the rank of the final estimate before the denormalisation step.

3 Feature Based Methods for Comparison

We compare the estimation of the fundamental matrix from optical flow with up to twelve variants of feature based techniques that are frequently encountered in literature. Based on the distance measure that is being minimised, we divide them in two different classes. The first class minimises the distance of a point to its corresponding epipolar line, while the second class minimises the so-called reprojection error in a Maximum Likelihood (ML) framework. We will refer to the first class of feature based technique as method class F1 and to the second class of feature based techniques as F2. In the following we will give a short overview of the different steps that make up these two classes.

3.1 Feature Extraction and Matching

Under *feature or interest point extraction* one traditionally understands the selection of a sparse set of image locations with distinctive neighbourhood information.

Classical examples of features are edges [48, 17] and corners [30]. Once a certain number of interest points has been extracted in both images, correspondences must be established between them, a process formally known as *feature matching*. To obtain a sparse set of feature correspondences for our comparison, we apply two widely used feature matching algorithms. The first one, known as the *Scale Invariant Feature Transform* (*SIFT*) [43], identifies locations of interest in scale space and associates with each of them a high dimensional descriptor vector. This vector representation is designed to be invariant with respect to scale and rotation and partially invariant with respect to affine distortions and illumination changes. As a result, a set of distinctive image features is obtained that can be matched correctly with high probability by means of a nearest neighbour search in descriptor space. Comparative studies by Mikolajczyk and Schmid [51] have repeatedly put forward SIFT as one of the most accurate local matching algorithms to date.
While SIFT has become a well accepted standard for stereo matching and object recognition, it may be outperformed in small-baseline scenarios by methods that are specifically tailored to small displacements. To account for these cases, we consider as a second feature matching algorithm the *Kanade-Lucas-Tomasi* tracker (*KLT*) [44, 75]. The KLT algorithm looks for local maxima of the eigenvalues of the structure tensor [25] and tries to detect the same features in the second image by minimising the intensity difference over a small local neighbourhood. While KLT feature extraction is closely related to the detection of other points of interest, such as Harris corners, the tracking mechanism basically solves a sparse optical flow problem.

3.2 Inlier Selection and Initialisation

The number of feature correspondences that is returned by a matching algorithm is generally controlled by thresholding a quality measure, such as the distance ratio between the first and the second nearest neighbour for SIFT descriptors and the smallest eigenvalue of the structure tensor for KLT features. Choosing the threshold less strict often guarantees a larger number of tentative correspondences, but at the same time increases the portion of false matches that have an adverse effect on the estimation of the fundamental matrix. Two robust techniques that are frequently used in computer vision to reduce the influence of such gross outliers are the *Random Sampling Consensus* (*RANSAC*) [23] and the *Least Median of Squares* (*LMedS*) [59]. In contrast to other robust methods that include as many correspondences as possible, RANSAC and LMedS repeatedly estimate the fundamental matrix from randomly sampled minimal data sets in a so-called *hypothesize-and-verify* framework. RANSAC ultimately selects the solution that is consistent with the largest number of correspondences by comparing the distance measure with a fixed inlier threshold t. LMedS, on the other hand, retains

the estimate for which the median of the squared distances is minimum over all samples.
In the last decade several extensions have been proposed to improve the performance of random sampling algorithms. Hereby the focus has been primarily on RANSAC due to its capability of dealing with a large proportion of mismatches. To incorporate the current state of the art, we consider in this work two such RANSAC extensions.
The first one has been proposed by Chum *et al.* [19] to reduce the influence of noise in the correspondence data and simultaneously achieve a speed-up over the theoretical number of samples that has to be drawn. It consists of performing a *local optimisation* (*LO*) step for each estimated fundamental matrix that has a larger support than all hypotheses generated so far. The LO-step comes down to applying a fixed number of *inner* RANSAC iterations that only draw samples from the current set of inliers. This will generally produce an improved hypothesis that will meet the termination criterion more rapidly.
The same authors propose a second extension to classical RANSAC for overcoming the ambiguity in the estimation process that can occur when the majority of points lie in a dominant plane. Chum *et al.* [20] showed that if five or more correspondences of the minimal sample are related by a planar homography, the estimated epipolar geometry can be wrong, yet consistent with a high number of correspondences. For such *degenerate configurations* their *DEGENSAC* algorithm simultaneously estimates a fundamental matrix and a homography and uses model selection to choose the correct solution. Concretely this is done by sampling correspondences that are outliers to the homography and estimating the fundamental matrix by means of the plane-and-parallax algorithm [33].
Both extensions described here are implemented as nested RANSAC loops and are easily combined for improved robustness.

3.3 Minimisation of a Geometric Distance Measure

Method Class F1: Minimisation of the Epipolar Distance. As a geometrically meaningful distance measure that does not depend on the scale of F, Luong and Faugeras [45] and Faugeras *et al.* [22] propose to minimise the squared *epipolar distance* over all n inliers.

$$\mathscr{E}_{\mathrm{F1}}(F) = \sum_{i=1}^{n} \left(d^2(\tilde{\boldsymbol{x}}_i', F\tilde{\boldsymbol{x}}_i) + d^2(\tilde{\boldsymbol{x}}_i, F^\top \tilde{\boldsymbol{x}}_i') \right) \quad . \tag{14}$$

Here $d(\tilde{\boldsymbol{x}}, \boldsymbol{l})$ denotes the Euclidean (geometric) distance between a point $\boldsymbol{x}$ and a line $\boldsymbol{l}$ in the image plane. The epipolar distance measures how far a point lies from the epipolar line of the corresponding point. The epipolar lines have to be

considered in both images, and definition (14) of the epipolar distance ensures that the measure is symmetric.
The epipolar distance equals a local weighing of the epipolar constraint. As proposed in [76], we minimise the energy (14) subject to the constraint $\|F\|^2_{\text{Frob}} = 1$, which comes down to an iteratively reweighted total least squares solution that is similar to the one of Eq. (11)-(12). We further reduce the effects of remaining outliers by including a statistical weighing of the epipolar distance by the tri-weight function as proposed by Huber in the context of M-estimators [35, 22, 76, 70]:

$$h(s) = \begin{cases} 1 & |s| \leq \sigma_{\text{r}} \\ \sigma_{\text{r}}/|s| & \sigma_{\text{r}} < |s| \leq 3\sigma_{\text{r}} \\ 0 & 3\sigma_{\text{r}} < |s| \end{cases} , \tag{15}$$

where the robust standard deviation σ_{r} is estimated via the error median as proposed by Rousseeuw and Leroy [59]. The fundamental matrix is initialised by the estimate provided by the random sampling algorithm and the rank of the final solution is enforced by SVD.

Method Class F2: Minimisation of the Reprojection Error. Hartley and Zisserman [33] propose to minimise the squared *reprojection error* over all n inliers:

$$\mathscr{E}_{\text{F2}}(P', \boldsymbol{X}_1, \ldots, \boldsymbol{X}_n) = \sum_{i=1}^{n} \left(d^2(\tilde{\boldsymbol{x}}_i, P\tilde{\boldsymbol{X}}_i) + d^2(\tilde{\boldsymbol{x}}'_i, P'\tilde{\boldsymbol{X}}_i) \right), \tag{16}$$

where $d(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{x}}')$ denotes the Euclidean distance between the two inhomogeneous points $\boldsymbol{x}$ and $\boldsymbol{x}'$. In the above definition P and P' are the 3×4 camera projection matrices for the left and the right image and $\boldsymbol{X}_i$, $1 \leq i \leq n$, are the 3D points that are reconstructed from the matching feature pairs. The two projections $P\tilde{\boldsymbol{X}}$ and $P'\tilde{\boldsymbol{X}}$ can be regarded as the most likely true positions of a given pair of points $\tilde{\boldsymbol{x}}$ and $\tilde{\boldsymbol{x}}'$ if the measurement errors are assumed to be independent and Gaussian distributed. The ML estimate of F is obtained as the rank 2 matrix that exactly satisfies the epipolar constraint

$$(P'\tilde{\boldsymbol{X}}_i)^\top F\,(P\tilde{\boldsymbol{X}}_i) = 0, \quad \forall i \ . \tag{17}$$

By choosing $P = (I, \boldsymbol{0})$, with identity matrix I, the fundamental matrix will be parameterised by the 12 entries of P', which automatically ensures the singularity constraint. Together with the 3 degrees of freedom for every 3D point, this brings the total number of variables to $3n + 12$. We minimise the highly nonlinear energy (16) over the motion parameters P' and the structure parameters $\boldsymbol{X}_i$, $1 \leq i \leq n$, by means of the iterative Levenberg-Marquardt algorithm [41, 47].

The Levenberg-Marquardt algorithm smoothly shifts between a gradient descent method, that can always be applied far from the minimum, and a Gauss-Newton method, that assures fast convergence in a small neighbourhood. This is achieved by effectively augmenting the Hessian of the error function with a factor that controls the transition between these two extremes. Because the Hessian has a sparse block structure, we used the sparse Levenberg-Marquardt algorithm described in [33] as the basis for our implementation. We additionally weigh the reprojection error by the tri-weight function (15) to reduce the effects of remaining outliers. The fundamental matrix is initialised by the estimate provided by the random sampling algorithm.

3.4 Inlier Refinement

After the fundamental matrix has been estimated by method classes F1 or F2, we reclassify the correspondences in inliers and outliers by a selection criterion that is similar to the one used in RANSAC. We chose the inlier threshold t based on the robust standard deviation [59] of the current set of inliers and reclassify and re-estimate the fundamental matrix until the number of inliers converges.

4 Evaluation of Optical Flow Based Fundamental Matrix Estimation

In our first experimental section we compare the performance of our dense optical flow based method with the two sparse feature based method classes F1 and F2. In different tests we compute the epipolar geometry of real-world image pairs that have been selected from several multiview stereo databases. All images in the databases have been calibrated by conventional techniques such that the ground truth fundamental matrix is known for each image pair. To assess the quality of the results, we evaluate the symmetric error between the estimated fundamental matrix and the ground truth according to Zhang [91] and Faugeras *et al.* [22]. This error measure is computed by using one matrix to randomly create a large number (100000) of correspondences and the other matrix to establish their epipolar lines. After the distances between the points and the lines have been computed, the roles of the two matrices are reversed to obtain a symmetric measure that describes the average deviation between two epipolar geometries in pixel units. In the following we denote the corresponding error by d_F.
For SIFT matching, we use the implementation of David Lowe [1] and only consider effective SIFT matches by removing feature pairs that have the same image

[1] available at http://www.cs.ubc.ca/~lowe/keypoints/

locations but different histogram orientations. To extract KLT features, we use publicly available code [2] that is based on the affine tracking algorithm described by Shi and Tomasi [66]. SIFT and KLT are applied to each image pair, but results are only listed for the best performing feature sets. Similarly, we either apply RANSAC or LMedS for the inlier selection, depending on which random sampling technique gives the best result in combination with the distance measures minimised by F1 and F2. For both random sampling algorithms we choose a minimal sample size of 7 and use the 7-point algorithm [33] to generate hypotheses for F. We additionally assure that the sampled points lie scattered enough over the whole image to avoid unstable estimates. For RANSAC we estimate the number of samples adaptively as described by Hartley and Zisserman [33] and use a fixed inlier threshold t between 0.5 and 1 pixels. This works well in practice because the distance measures minimised by F1 and F2 can be interpreted as geometrical distances in the image plane. If degeneracy is suspected, we apply the DEGENSAC variant. Both standard RANSAC and DEGENSAC are equipped with a LO-step for local model optimisation. The number of samples for LMedS normally requires an estimate of the proportion of outliers. Following Faugeras *et al.* [22], choosing an iteration number of 2000 allows us to deal with the maximum percentage (50%) of outliers. Due to the random nature of both RANSAC and LMedS, we run the feature based methods F1 and F2 for 100 consecutive times with constant settings and present the average of the error d_F over all test runs.
Our evaluation includes indoor sequences that have been captured under lab conditions, as well as outdoor sequences with varying illumination and large relative motion. All images are corrected for radial distortion. The indoor image pairs depict objects against a homogeneous black background, which we exclude from the estimation by detecting the object silhouette with a Chan-Vese segmentation technique [18]. For a fair comparison with the feature based methods, we run our optical flow based method for a fixed set of *default* parameters, that are given by $\alpha = 20.0$, $\gamma = 20.0$ and $\sigma = 0.9$. To visualise the epipolar geometries, we draw the epipolar lines that are estimated for the default settings of our optical flow based method and the epipolar lines that are estimated from a *representative set of inliers* for the feature based methods. By a representative set of inliers, we denote a set of feature correspondences for which the error d_F lies close to (i.e. does not deviate more than 0.1 pixels from) the average error of F1 or F2. We only draw the epipolar lines for a set of meaningful points in the left and the right image. For the left image these are 8 feature points from the representative set of inliers. For the right image these are either the corresponding features or the pixel locations warped by the optical flow.

[2] available at http://www.ces.clemson.edu/ stb/klt/

Table 1: Overview of the settings for the feature extraction. We list the type of feature used (SIFT/KLT), the ratio of distances of the first and second neighbour in SIFT descriptor space (*ratio*) and the total number of matches (*# match*). For KLT we used the standard settings provided in the publicly available code.

Image Pair		feature extraction		
sequence	*frames*	*type*	*ratio*	*# match*
DinoRing	24 - 25	KLT	-	737
Entry-P10	1 - 0	SIFT	0.80	979
TempleRing	13 - 14	SIFT	0.90	627
Herz-Jesu-P25	5 - 6	SIFT	0.90	945
City-Hall	1 - 2	SIFT	0.90	1502

Table 2: Overview of the settings of the feature based estimation methods techniques F1 and F2. We list the type of random sampling algorithm (*randsam*), the applied RANSAC threshold t (*thresh*) and the average number of inliers (# inl) over 100 test runs.

Image Pair		F1			F2		
sequence	*frames*	*randsam*	*thresh*	*# inl*	*randsam*	*thresh*	*# inl*
DinoRing	24 - 25	LORANSAC	0.5	587	LORANSAC	0.5	585
Entry-P10	1 - 0	DEGENSAC	1.0	749	DEGENSAC	1.0	743
TempleRing	13 - 14	LORANSAC	0.8	468	LORANSAC	0.8	464
Herz-Jesu-P25	5 - 6	LMedS	-	663	LMedS	-	667
City-Hall	1 - 2	LORANSAC	1.0	1096	LORANSAC	1.0	1092

Table 3: Overview of the error d_F for our optical flow based method and the average error for the feature based methods F1 and F2 over 100 test runs. The best results are highlighted in bold face.

Image Pair	Our Method	F1	F2
DinoRing	**0.717**	3.865	3.429
Entry-P10	**2.448**	3.530	4.611
TempleRing	**0.151**	0.810	0.881
Herz-Jesu-P25	3.227	**1.139**	3.021
City-Hall	7.349	1.236	**1.159**

4.1 Low Texture

In our first experiment we compute the epipolar geometry of frames 24 and 25 of the *DinoRing* [3] multiview data set [64]. Both images have a resolution of 640×480, with a maximum displacement of 26 pixels, while the depicted scene is characterised by the absence of texture. For default settings, our optical flow based method obtains an error d_F of 0.717, which is well within sub-pixel precision. This result is listed in Table 3, together with the errors obtained by the feature based methods. The settings for the feature based estimation methods F1 and F2 are summarised in Table 2 for all image pairs in this section.
Whereas optical flow based methods benefit from the filling-in effect of the smoothness term in homogeneous regions, insufficient texture often poses a challenge to feature extraction, such as SIFT, which is unable to find a sufficient amount of features in the DinoRing images. The number of features tracked by the KLT algorithm for this sequence is much larger, but their quality is insufficient to render the results for F1 and F2 sub-pixel precise. We can conclude from Table 3 that the KLT features suffer from poor localisation, as the average performance of the feature based techniques is worse than our method. The flow field for the default settings of our method is shown in Fig. 1 (a). For the visualisation we use the colour code depicted in Fig. 2 (a), where colour encodes the direction of the flow and brightness the magnitude. In the estimated optical flow we can distinguish occlusion artifacts near the tail of the dinosaur model but their influence on the fundamental matrix estimation is reduced by the proposed robust L_1 penalisation. Fig. 1 further shows a set of inliers and the corresponding epipolar geometry representing the best average feature based result.

4.2 Near-Degeneracy and Repetitive Structures

A scenario that frequently occurs in stereo vision is that the majority of correspondences lie in the same plane, such as in the case of a dominant plane or when features are primarily extracted on a planar surface. A random sampling algorithm that is based on the 7- or 8-point algorithm can then produce a consensus set of coplanar inliers [20, 26]. This set is called *degenerate* because it does not provide enough constraints to uniquely compute the fundamental matrix [33]. For degenerate configurations, it is crucial that a sufficient number of out-of-plane inliers are selected to overcome this ambiguity and this requires to take special care in the case of sparse feature based methods. Optical flow, on the other hand, will likely include a larger amount of out-of-plane correspondences due to its dense and global nature. This will then provide the necessary constraints for the

[3] available at http://vision.middlebury.edu/mview/data/

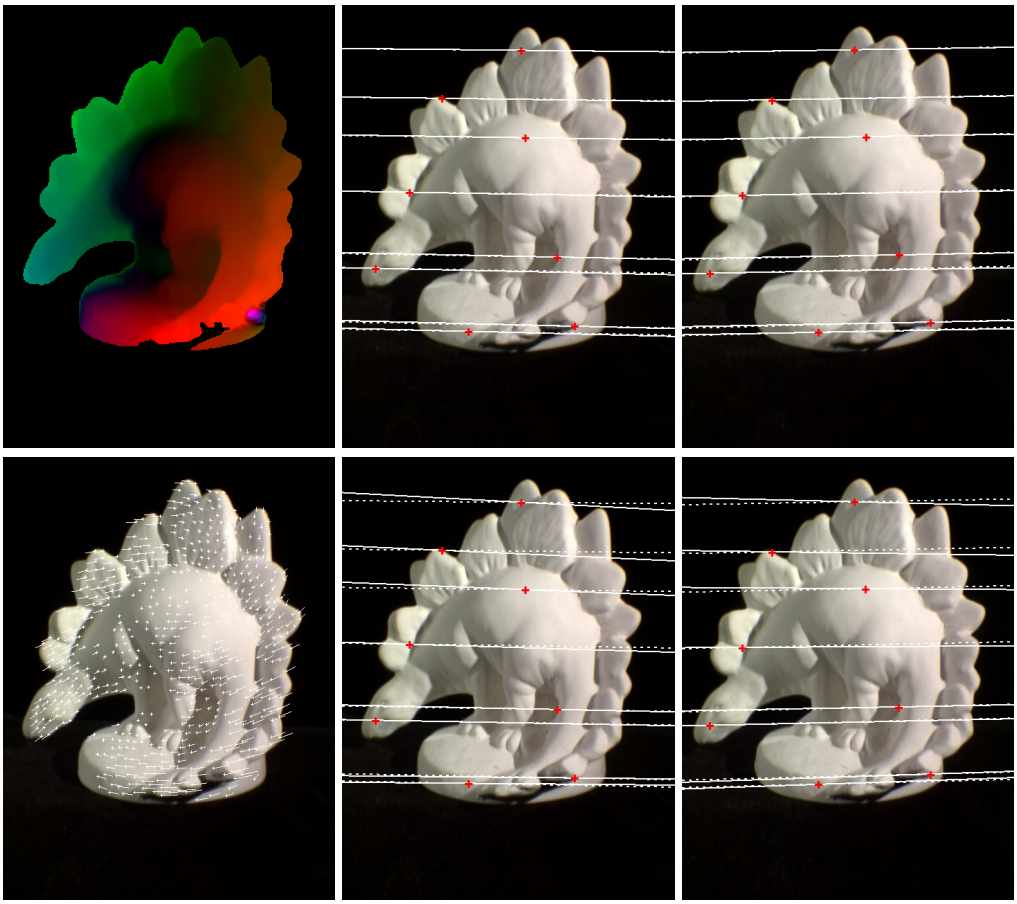

Figure 1: Results for DinoRing within the object silhouette. **Top Row: (a)** The optical flow between frames 24 and 25. **(b) + (c)** The epipolar geometry estimated from the optical flow for frames 24 and 25. Points are depicted as red crosses, their corresponding estimated epipolar lines as full white lines and their corresponding ground truth lines as dotted white lines. **Bottom Row: (d)** A representative set of 587 inliers for F2. The correspondences are drawn on frame 24 as lines connecting the matched features. **(e) + (f)** The epipolar geometry estimated from these inliers.

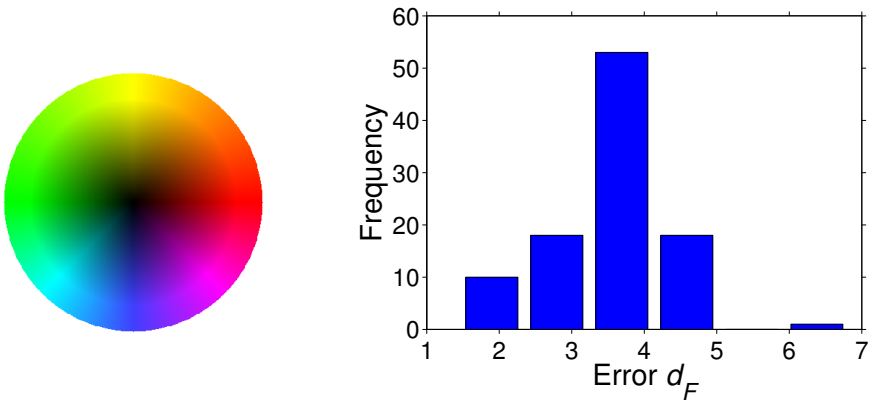

Figure 2: **Left: (a)** Colour circle. **Right: (b)** Histogram for the Entry-P10 data set for F1 in combination with SIFT–LORANSAC.

8-point algorithm in a natural way. We illustrate this intrinsic robustness by estimating the epipolar geometry of frames 1 and 0 of the *Entry-P10* [4] multiview data set [73]. This image pair depicts the facade of a building with a balcony as the only out-of-plane element. To ensure realistic image sizes for our optical flow based technique, we test all estimation methods on 640×427 versions of the original 3072×2048 images.

If we take a look at the histogram of d_F for the method F1 in Fig. 2, we can clearly distinguish a pronounced mode that is centred between 3 and 4. This mode corresponds to about 50% of the 100 LORANSAC test runs that predominantly select the inliers within the plane of the facade. The middle row of Fig. 3 shows such a degenerate set of inliers and the corresponding estimated epipolar lines. These are wrongly estimated as the vanishing lines of the facade. A similar sensitivity to degeneracy was also observed for LMedS. Applying the robust DEGENSAC algorithm improves the performance of the feature based methods only slightly, as the results in Table 3 show. The reason for this, is the large amount of in-plane outliers that arise from mismatched repetitive structures such as windows. These outnumber the out-of-plane inliers such that model verification based on the amount of support tends to fail, even when the planar homography is estimated correctly. As an illustration, a set of non-degenerate inliers for F2 is depicted in the bottom row of Fig. 3, together with the outliers and the planar homography estimated by DEGENSAC. For default settings, our optical flow based method performs better than the feature based methods, but the error is not sub-pixel due to the disturbing occlusion in the lower right corner. The flow field and the corresponding estimated epipolar geometry are shown in the top row of Fig. 3. The

[4] available at http://cvlab.epfl.ch/~strecha/multiview/denseMVS.html

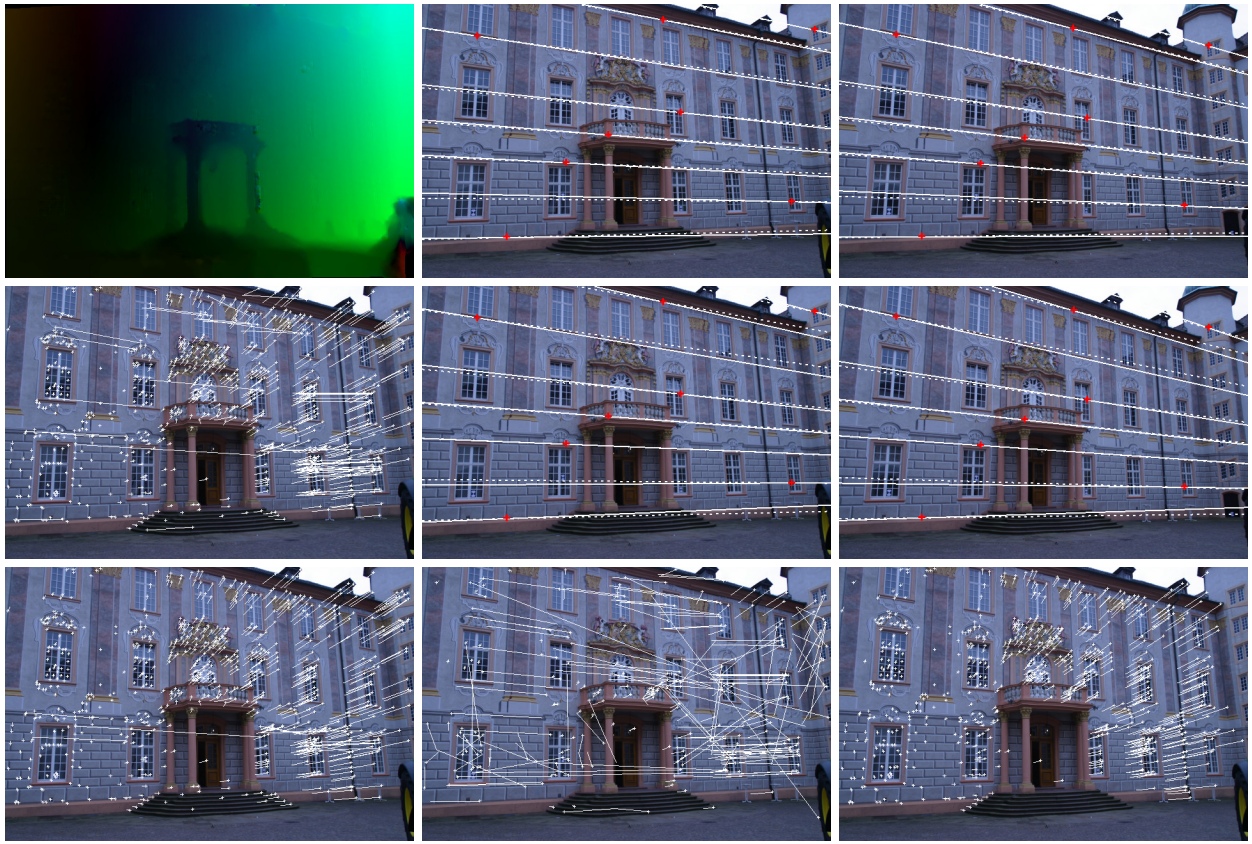

Figure 3: Results for Entry-P10. **Top Row: (a)** The optical flow between frames 1 and 0. Pixels that are warped outside the image by the optical flow are colored black. **(b) + (c)** The epipolar geometry estimated from the optical flow for frames 1 and 0. **Middle Row (d)** A set of 763 degenerate inliers for F1. **(e) + (f)** The epipolar geometry estimated from these inliers for frames 1 and 0. **Bottom Row (g)** A set of 728 non-degenerate inliers for F2. **(h)** The corresponding outliers. **(i)** The 706 inliers with respect to the planar homography.

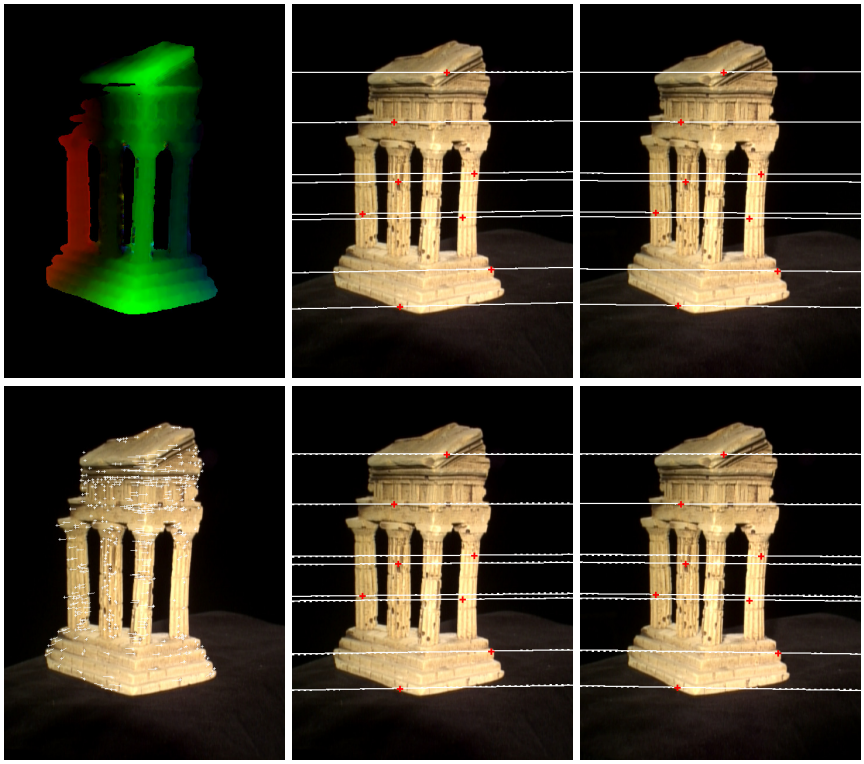

Figure 4: Results for TempleRing within the object silhouette. **Top Row: (a)** The optical flow between frames 13 and 14. **(b) + (c)** The epipolar geometry estimated from the optical flow. **Bottom Row: (d)** A representative set of 466 inliers for F1. **(e) + (f)** The epipolar geometry estimated from these inliers.

balcony is clearly visible and the estimated epipolar lines are closer to ground truth than the feature based results. In the next section we will improve upon this result and obtain a sub-pixel error.

4.3 Sufficient Texture and No Degeneracy

For the remainder of our comparison we have selected three image pairs that do not suffer from degeneracy or a lack of texture. First we compute the fundamental matrix for frames 13 and 14 of the *TempleRing* [5] sequence. In Table 3 we observe that all estimation methods achieve sub-pixel precision, but that our optical flow based method performs best with an error of only 0.151 pixels for default parameters. This is well below the average error of both feature based methods.

[5] available at http://vision.middlebury.edu/mview/data/

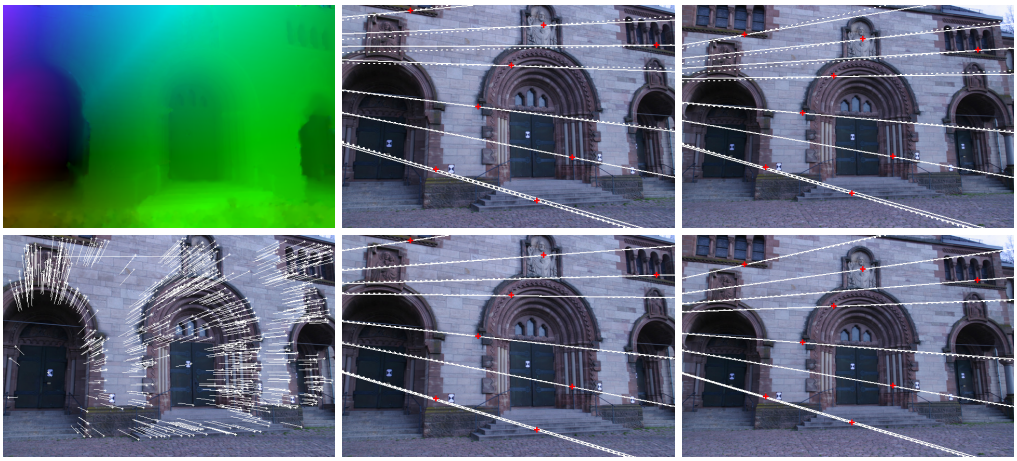

Figure 5: Results for Herz-Jesu-P25. **Top Row: (a)** The optical flow between frames 5 and 6. **(b) + (c)** The epipolar geometry estimated from the optical flow. **Bottom Row (d)** A representative set of 664 inliers for F1. **(e) + (f)** The epipolar geometry estimated from the inliers.

Figure 6: Results for City-Hall. **Top Row: (a)** The optical flow between frames 1 and 2. **(b) + (c)** The epipolar geometry estimated from the optical flow. **Bottom Row (d)** A representative set of 1089 inliers for F2. **(e) + (f)** The epipolar geometry estimated from the inliers.

Fig. 4 shows the estimated optical flow within the model silhouette and the corresponding epipolar geometry. It can be observed that the epipolar lines practically coincide with the ground truth. A representative set of inliers and the corresponding feature based result are shown as well.
Next we consider 640×427 versions of frames 5 and 6 of the *Herz-Jesu-P25* [5] sequence. Contrary to the Entry-P10 data set, the entrances of the building provide sufficient out-of-plane correspondences to avoid degeneracy. In Table 3 we observe that our optical flow based method achieves an error of more than 3 pixels for default parameter settings and is outperformed by both feature based methods. The optical flow and the corresponding epipolar geometry are shown in Fig. 5, together with a set of inliers that is representative for the best average feature based result. For outdoor sequences like this one, matching SIFT correspondences was overall more accurate than tracking KLT features due to the large apparent motion. It can be seen in the cobbled stone region of the scene that the large change in viewpoint also makes matching more difficult for optical flow, causing a deterioration in the flow field at the bottom of the image. This leads to an undesirable parameter sensitivity for our method and explains the larger error for fixed default parameters.
We conclude this section with the recovery of the epipolar geometry of frames 1 and 2 of the *City-Hall* [6] sequence [72]. Despite being scaled down to a resolution of 640×427, the image pair contains large displacements of more than 85 pixels. The large apparent motion and the occlusion on the left side of the building distort the optical flow, which is reflected by the large error of 7.3 for our method. Both feature based methods perform similar to each other with an average error close to one pixel. The estimated optical flow, a representative set of inliers and the corresponding epipolar geometry are shown in Fig. 6.

4.4 Intermediate Conclusions

The previous experiments have shown that the dense estimation of the fundamental matrix from optical flow can be competitive with classical sparse techniques. The advantage of dense estimation methods becomes especially apparent in situations where the sparse features are not well localised or when the inclusion of a small number of out-of-plane correspondences is crucial in overcoming the degeneracy problem. On the other hand, we have seen that the optical flow based method is more sensitive to large displacements and occlusions that are present in wide baseline stereo images.

[6] available at http://cvlab.epfl.ch/data/strechamvs/

5 A Joint Variational Model

So far we have fed a dense optical flow method into a classical approach for estimating the fundamental matrix. Let us now investigate how we can achieve further improvements by coupling optical flow computation and fundamental matrix estimation in a joint model where they influence each other in a beneficial way. To this end we look at the epipolar constraint not only as a means of fitting the fundamental matrix to a given set of correspondences, but also as an additional restriction on the correspondence search.
In this section we present an intuitive way of coupling the computation of the fundamental matrix and the optical flow by minimising a single functional for both unknowns. Their simultaneous solution will ensure a scene structure that is most consistent with the camera motion, and vice versa, resulting in a higher overall accuracy and a lower parameter sensitivity.

5.1 Integrating the Epipolar Constraint

In order to jointly estimate the optical flow and the fundamental matrix, we propose to extend the optical flow model of Eq. (1) with an extra term as follows:

$$\begin{aligned}\mathscr{E}(\boldsymbol{w},\boldsymbol{f}) = \int_\Omega \Big(&\Psi\left(|g(\boldsymbol{x}+\boldsymbol{w}) - g(\boldsymbol{x})|^2 + \gamma \cdot |\nabla g(\boldsymbol{x}+\boldsymbol{w}) - \nabla g(\boldsymbol{x})|^2\right) \\ &+ \alpha\,\Psi\left(|\nabla \boldsymbol{w}|^2\right) + \beta\,\Psi\Big((\boldsymbol{s}^\top\ \boldsymbol{f})^2\Big)\Big)\,\mathrm{d}x\,\mathrm{d}y\ , \end{aligned} \tag{18}$$

and impose the explicit constraint $\|\boldsymbol{f}\|^2 = 1$. While the first two terms in $\mathscr{E}(\boldsymbol{w},\boldsymbol{f})$ are identical to the original model, the third term has been newly introduced to penalise deviations from the epipolar constraint $\boldsymbol{s}^\top \boldsymbol{f} = 0$. The vectors $\boldsymbol{s}$ and $\boldsymbol{f}$ are defined as in Eq. (5) and (6), but this time $\boldsymbol{s}$ is a function of $\boldsymbol{x}$ and $\boldsymbol{w}$. The regularised L_1 penaliser Ψ reduces the influence of outliers in the computation of F and the weight β determines to what extent the epipolar constraint will be satisfied in all points. The constraint on the Frobenius norm of F avoids the trivial solution. An extension of our functional to RGB-images is obtained by replacing the data term by its multichannel variant (3).

5.2 Minimisation

To minimise the functional $\mathscr{E}(\boldsymbol{w},\boldsymbol{f})$ with respect to u, v and $\boldsymbol{f}$, subject to the constraint $\|\boldsymbol{f}\|^2 = 1$, we use the method of Lagrange multipliers. We are looking for critical points of

$$\mathscr{F}(\boldsymbol{w},\boldsymbol{f},\lambda) = \mathscr{E}(\boldsymbol{w},\boldsymbol{f}) + \lambda(1 - \boldsymbol{f}^\top \boldsymbol{f})\ , \tag{19}$$

i.e. tuples $(u^*, v^*, \boldsymbol{f}^*, \lambda^*)$ for which the functional derivatives of the Lagrangian $\mathscr{F}$ with respect to u and v and the derivatives of $\mathscr{F}$ with respect to $\boldsymbol{f}$ and λ vanish.

Optical Flow. The Euler-Lagrange equations of the optical flow components u and v are obtained by setting

$$\frac{\partial}{\partial u}\mathscr{F}(\boldsymbol{w}, \boldsymbol{f}, \lambda) = 0 \quad \text{and} \quad \frac{\partial}{\partial v}\mathscr{F}(\boldsymbol{w}, \boldsymbol{f}, \lambda) = 0 \ . \tag{20}$$

To derive them in more detail we write the argument of the epipolar term as a scalar product involving the optical flow:

$$\begin{aligned} \boldsymbol{s}^\top \boldsymbol{f} &= \begin{pmatrix} x+u \\ y+v \\ 1 \end{pmatrix}^\top F \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} && (21) \\ &= \begin{pmatrix} u \\ v \\ 0 \end{pmatrix}^\top F \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} + \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}^\top F \begin{pmatrix} x \\ y \\ 1 \end{pmatrix} && (22) \\ &= au + bv + q \ . && (23) \end{aligned}$$

Here a and b denote the first two coefficients of the epipolar line $F\tilde{\boldsymbol{x}}$ of a point $\tilde{\boldsymbol{x}} = (x, y, 1)^\top$ in the left image,

$$a = (F\tilde{\boldsymbol{x}})_1 \quad \text{and} \quad b = (F\tilde{\boldsymbol{x}})_2 \ , \tag{24}$$

while the quantity q can be interpreted as the distance of $\tilde{\boldsymbol{x}}$ to this epipolar line up to a scale factor:

$$q = \tilde{\boldsymbol{x}}^\top F\, \tilde{\boldsymbol{x}} \ . \tag{25}$$

With the help of formula (23) we can easily derive the contributions of the epipolar term in $\mathscr{E}(\boldsymbol{w}, \boldsymbol{f})$ to the Euler-Lagrange equations. The partial derivatives of its integrand $\Psi\left((\boldsymbol{s}^\top \boldsymbol{f})^2\right)$ with respect to u and v are

$$\begin{aligned} \frac{\partial}{\partial u}\Psi\left((\boldsymbol{s}^\top \boldsymbol{f})^2\right) &= 2\Psi'\left((\boldsymbol{s}^\top \boldsymbol{f})^2\right)(a^2 u + ab v + aq), && (26) \\ \frac{\partial}{\partial v}\Psi\left((\boldsymbol{s}^\top \boldsymbol{f})^2\right) &= 2\Psi'\left((\boldsymbol{s}^\top \boldsymbol{f})^2\right)(ab u + b^2 v + bq) \ . && (27) \end{aligned}$$

The contributions from the data term and the smoothness term remain unchanged with respect to the original model. Thus, we obtain the final Euler-Lagrange equations of u and v by adding the right hand sides of equations (26) and (27) to the

Euler-Lagrange equations given in [14]:

$$\begin{aligned} 0 \;=\;& \Psi'\left(g_z^2+\gamma\left(g_{xz}^2+g_{yz}^2\right)\right)\left(g_x g_z+\gamma(g_{xx}g_{xz}+g_{xy}g_{yz})\right) \\ & -\alpha\,\mathrm{div}\left(\Psi'\left(|\nabla u|^2+|\nabla v|^2\right)\nabla u\right) \\ & +\beta\,\Psi'\left((\boldsymbol{s}^\top \boldsymbol{f})^2\right)\left(a^2 u+a\,b\,v+a\,q\right), \end{aligned} \tag{28}$$

$$\begin{aligned} 0 \;=\;& \Psi'\left(g_z^2+\gamma\left(g_{xz}^2+g_{yz}^2\right)\right)\left(g_y g_z+\gamma(g_{yy}g_{yz}+g_{xy}g_{xz})\right) \\ & -\alpha\,\mathrm{div}\left(\Psi'\left(|\nabla u|^2+|\nabla v|^2\right)\nabla v\right) \\ & +\beta\,\Psi'\left((\boldsymbol{s}^\top \boldsymbol{f})^2\right)\left(a\,b\,u+b^2 v+b\,q\right) \quad . \end{aligned} \tag{29}$$

Here we have made use of the same abbreviations for the partial derivatives and the temporal differences in the data term as in [14]:

$$g_* \;=\; \partial_* g(\boldsymbol{x}+\boldsymbol{w}), \tag{30}$$

$$g_z \;=\; g(\boldsymbol{x}+\boldsymbol{w})-g(\boldsymbol{x}), \tag{31}$$

$$g_{*z} \;=\; \partial_* g(\boldsymbol{x}+\boldsymbol{w})-\partial_* g(\boldsymbol{x}) \quad , \tag{32}$$

where $*$ stands for either x, y, xx, xy or yy. The subscript z indicates the occurrence of a temporal difference in contrast to a temporal derivative.

Fundamental Matrix. To solve for the fundamental matrix we have to set

$$\nabla_{\boldsymbol{f}}\,\mathscr{F}(\boldsymbol{w},\boldsymbol{f},\lambda)=\boldsymbol{0} \quad \text{and} \quad \frac{\partial}{\partial\lambda}\mathscr{F}(\boldsymbol{w},\boldsymbol{f},\lambda)=0 \quad , \tag{33}$$

where $\nabla_{\boldsymbol{f}}$ stands for the gradient operator $(\partial_{f_{1,1}},\ldots,\partial_{f_{3,3}})^\top$. To differentiate the Lagrangian $\mathscr{F}$ with respect to $\boldsymbol{f}$, we only have to consider the newly introduced epipolar term since neither the data term nor the smoothness term depends on $\boldsymbol{f}$. Equations (33) then give rise to the eigenvalue problem

$$\boldsymbol{0} \;=\; \left(\int_\Omega \Psi'\left((\boldsymbol{s}^\top \boldsymbol{f})^2\right)\boldsymbol{s}\boldsymbol{s}^\top \mathrm{d}x\,\mathrm{d}y-\lambda I\right)\boldsymbol{f}, \tag{34}$$

$$=:\; (M-\lambda I)\,\boldsymbol{f}, \tag{35}$$

$$0 \;=\; 1-\|\boldsymbol{f}\|^2 \quad . \tag{36}$$

Note that we were able to switch the order of differentiation and integration because $\boldsymbol{f}$ is a constant over the domain Ω. The system matrix M is symmetric

positive definite and its entries are the integral expressions

$$m_{i,j} = \int_\Omega \Psi' \left((\boldsymbol{s}^\top \boldsymbol{f})^2 \right) s_i s_j \, \mathrm{d}x \, \mathrm{d}y \ , \tag{37}$$

with $1 \leq i, j \leq 9$ and s_i being the i-th component of $\boldsymbol{s}$.

5.3 Solution of the System of Equations

The system of equations (20) and (33) is solved by iterating between the optical flow computation and the fundamental matrix estimation. The Euler-Lagrange equations (28) and (29) are first solved for $\boldsymbol{w}$ with a current estimate of the fundamental matrix. Using the computed optical flow, we then compose the system matrix M and solve the eigenvalue problem (35)-(36) for $\boldsymbol{f}$. Due to the constraint (36) the solution will always be of unit norm. The new estimate of the fundamental matrix will in turn be used to solve the Euler-Lagrange equations again for the optical flow. This process is repeated until convergence. To initialise the fundamental matrix we compute it in the first iteration step from pure optical flow as proposed in the previous section. Our model does not explicitly enforce the singularity constraint of F and therefore its rank is not enforced in the iterative process. The Euler-Lagrange equations are solved by a coarse-to-fine warping strategy in combination with a multigrid solver [15], while Equation (35) is solved as a series of linear eigenvalue problems as described in Section 2. In practice we exclude points from the estimation process that are warped outside the image by the optical flow.

5.4 A Joint Model with Data Normalisation

Data normalisation, as discussed in Section 2.4, dramatically improves the conditioning of the eigenvalue problem and is essential for obtaining an accurate estimate of the fundamental matrix. It consists of replacing each point $\tilde{\boldsymbol{x}} = (x, y, 1)^\top$ in the left image and its corresponding point $\tilde{\boldsymbol{x}}' = (x+u, y+v, 1)^\top$ in the right image by the transformed points $T\tilde{\boldsymbol{x}}$ and $T'\tilde{\boldsymbol{x}}'$. The normalisation transformations T and T' are composed of a translation and a scaling such that the normalised coordinates are of the same order.
It is highly desirable that this normalisation step is integrated into our energy functional (18). To this end we express the epipolar term in function of the normalised fundamental matrix with the help of Eq. (13). This leads to a joint variational

model with data normalisation:

$$\mathscr{E}(\boldsymbol{w},\hat{\boldsymbol{f}}) = \int_{\Omega} \Big(\Psi\left(|g(\boldsymbol{x}+\boldsymbol{w}) - g(\boldsymbol{x})|^2 + \gamma \cdot |\nabla g(\boldsymbol{x}+\boldsymbol{w}) - \nabla g(\boldsymbol{x})|^2\right) + \alpha\,\Psi\left(|\nabla \boldsymbol{w}|^2\right) + \beta\,\Psi\Big(\left(\hat{\boldsymbol{s}}^{\top}\ \hat{\boldsymbol{f}}\right)^2\Big)\Big)\,\mathrm{d}x\,\mathrm{d}y\ , \tag{38}$$

By imposing the constraint $\|\hat{\boldsymbol{f}}\|^2 = 1$ and by applying the method of Lagrange multipliers we obtain a similar eigenvalue problem as (35) - (36), which can be solved for the normalised fundamental matrix $\hat{F}$. For the computation of the optical flow from energy (38), however, we have to take into consideration that the vector $\hat{\boldsymbol{s}}$ is now not only a function of $\boldsymbol{x}$ and $\boldsymbol{w}$, but also of the normalisation transformations T and T'. It is important to note here that in the approach of Hartley [31] each normalisation transformation depends on the set of points that has to be normalised. Because the set of correspondence points in the left image consists of the pixels of the rectangular image domain Ω, T is a constant mapping that only depends on the image size. The transformation T', on the other hand, normalises the warped pixel coordinates and thus depends on the optical flow $\boldsymbol{w}$. To avoid derivatives of T' with respect to u and v in the Euler-Lagrange equations, we replace T' with a constant transformation. As a result, the Euler-Lagrange equations do not change under the normalisation step and thus remain the same as those presented in Eq. (28) and (29). We further assume that the normalising transformations for the left and right correspondences are similar, such that we can choose $T' = T$. Experiments have shown that this approximation has only a minor influence on the results compared to the approach of Hartley [31]. The solution of the system of equations is done iteratively, as explained in Section 5.3, by solving the eigenvalue problem for $\hat{F}$ and using the fundamental matrix $F = T^{\top}\,\hat{F}\,T$ to solve the Euler-Lagrange equations for $\boldsymbol{w}$.

6 Evaluation of the Joint Method

In our second experimental section we assess the performance of our joint variational method by evaluating the fundamental matrix estimation and the optical flow computation separately. We recover the epipolar geometry of the image pairs of Section 4 and present results for the afore mentioned fixed default settings ($\alpha = 20.0$, $\gamma = 20.0$ and $\sigma = 0.9$). To judge the quality of the optical flow computation, we use stereo pairs from the Middlebury optical flow database for which the ground truth is publicly available. We evaluate the estimated optical flow by means of the average angular error (AAE) [7] and the average endpoint error (AEE) [6].

Table 4: Overview of the error d_F for 30 iterations of our joint variational method and the average error for the feature based methods F1 and F2 over 100 test runs. The best results are highlighted in bold face.

Image Pair	Our Method	F1	F2
DinoRing	**1.175**	3.865	3.429
Entry-P10	**0.645**	3.530	4.611
TempleRing	**0.274**	0.810	0.881
Herz-Jesu-P25	**0.502**	1.139	3.021
City-Hall	**1.002**	1.236	1.159

6.1 Fundamental Matrix

We first demonstrate the convergence behaviour of our iterative minimisation strategy. To this end we recover the epipolar geometry of the Herz-Jesu-P25 and City-Hall image pairs with our joint estimation method for the default parameter settings. The first row of Fig. 7 and Fig. 8 shows the estimated epipolar lines after the first iteration step. These geometries correspond to the error d_F that has been listed before in Table 3 for the respective image pairs. The second row of these figures shows how these initial estimates are readjusted after 30 iterations to almost coincide with the ground truth. These geometries correspond to the errors that can be found in Table 4. We additionally observe that the simultaneous recovery of the optical flow and the epipolar geometry has led to a visual improvement of both flow fields. This is most apparent in the occluding side of the building in the City-Hall sequence.

We found empirically that the value of the weight β mainly has an influence on the convergence speed and to a much lesser extend on the final error. In combination with the default settings we used $\beta = 40$, for which we obtained convergence within 10 iteration steps for all image pairs. If we compare the results of Table 4 with those of Table 3, we see that our joint estimation method has improved the accuracy substantially for all outdoor sequences. At the same time our method performs better on average than both feature based methods for all image pairs. For 3 out of 5 data sets our results are within sub-pixel precision. We make the remark here that our joint method converges towards an optical flow that is most consistent with the estimated fundamental matrix. While this will generally result in a more accurate epipolar geometry than the purely optical flow based estimates of the previous section, the errors for DinoRing in Table 3 and Table 4 illustrate that this is not always true.

To demonstrate the scalability of our results, we present the error d_F for the full

(3072×2048) and half (1536×1024) resolution images of the Strecha data base. These are shown in Table 5, together with the outcome for the quarter size (640×427) versions. We compare our results with those achieved by the feature based methods, for which we chose the same settings as in Table 2. We see that the errors for the full resolution images scale well for our method. For the half resolution versions we even obtain sub-pixel precision. For the feature based methods, the inlier ratios of the different sequences stayed roughly the same for all image sizes. The absolute number of inliers for Entry-P10 and Herz-Jesu-P25 (≈ 3270 and ≈ 1520 respectively for full resolution) did, however, not scale with the image size. The good results of F1 for the full resolution Herz-Jesu-P25 sequence can therefore mainly be attributed to an increased precision of the feature locations. For the half resolution images of the City-Hall sequence, the absolute inlier count scaled significantly with the image size (≈ 4304) and this helped in achieving a lower feature based error.
Table 5 additionally lists run time information for our joint method. For the quarter size images, one iteration of our alternating minimisation requires about 30 s, which is almost exclusively spent on the optical flow computation. For the half size images this grows to approximately 170 s. The total run time of the feature based methods with adaptive RANSAC ranges from less than 10 s for quarter size images to more than 150 s for half size images, depending on the amount of features and the minimised distance measure. Here, the run time is dominated by feature extraction and matching. In order to speed up the run time of our method, several options exist. First of all, we require significantly less than 30 iterations to converge to an accurate solution. This allows us to reduce the number of iterations drastically. Secondly, variational optical flow computation can be parallelised efficiently on recent graphical hardware. According to [28], the run time for a GPU implementation of the advanced method of Zimmer *et al.* [93] is less than 1 s for quarter size images and less than 2.5 s for images of half size. If we assume convergence within 10 iterations, the total run time of a parallel implementation of our baseline method would thus be around 10 s and 30 s, respectively. For the full size images, run times are prohibitively large and memory requirements make the execution on current GPUs infeasible.

6.2 Optical Flow

In a second set of experiments we provide evidence that the concept of simultaneously recovering the correspondences and the epipolar geometry can improve the optical flow computation. First we use our joint estimation method to compute the optical flow between frames 8 and 9 of the *Yosemite sequence* [7] without clouds.

[7] available at http://www.cs.brown.edu/people/black/images.html

Table 5: Overview of the error d_F for the quarter (640×427), half (1536×1024) and full (3072×2048) resolution versions of the outdoor image pairs for 30 iteration steps of our joint variational method and for 100 test runs of the feature based methods F1 and F2. The best results are highlighted in bold face. The corresponding run time (in seconds) of our method on a machine with a 1862MHz Intel Core2 CPU is given in the last column.

Image Size	Image Pair	Our Method	F1	F2	Run Time
quarter	Entry-P10	**0.645**	3.530	4.611	880
	Herz-Jesu-P25	**0.502**	1.139	3.021	
	City-Hall	**1.002**	1.236	1.159	
half	Entry-P10	**0.400**	7.062	10.094	5190
	Herz-Jesu-P25	**0.977**	1.940	4.393	
	City-Hall	0.657	**0.591**	0.600	
full	Entry-P10	**1.957**	10.142	19.165	$>$ 8 h
	Herz-Jesu-P25	2.404	**2.149**	7.997	
	City-Hall	**1.305**	2.011	2.102	

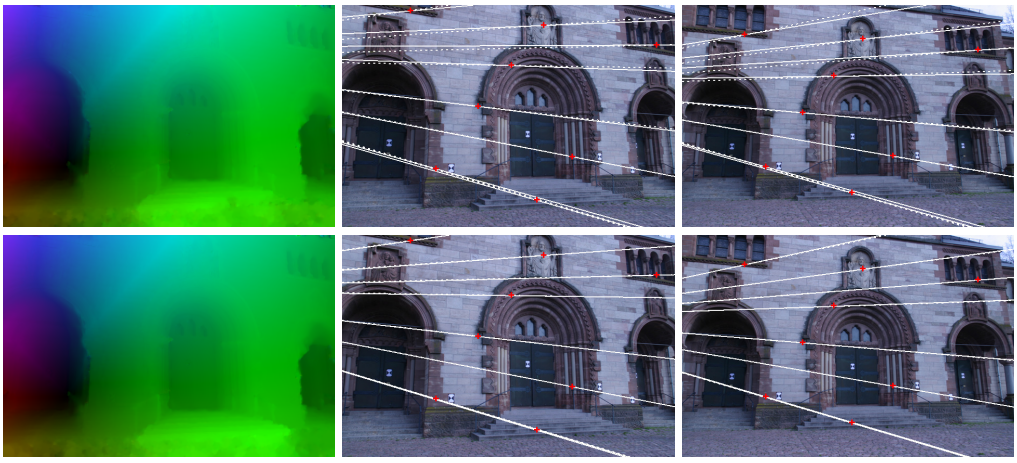

Figure 7: Results for Herz-Jesu-P25 **Top Row: (a)** The optical flow between frames 5 and 6 after 1 iteration step. **(b) + (c)** The corresponding epipolar geometry. **Bottom Row (d)** The optical flow between frames 5 and 6 after 30 iteration steps. **(e) + (f)** The corresponding epipolar geometry.

Figure 8: Results for City-Hall. **Top Row: (a)** The optical flow between frames 1 and 2 after 1 iteration step. **(b) + (c)** The corresponding epipolar geometry. **Bottom Row (d)** The optical flow between frames 1 and 2 after 30 iteration steps. **(e) + (f)** The corresponding epipolar geometry.

This classical sequence of size 316×252 actually depicts a static scene captured by a moving camera and therefore forms a stereo pair. Table 6 shows that we are able to improve the AAE from 1.59° for standard optical flow to 1.15°. This ranks us among the best methods with spatial regularisation published so far [8]. For this experiment all parameters have been optimised with respect to the AAE of the optical flow in the first iteration step and β has been set to 50. The sky region has not been excluded from the computation, and pixels that are warped outside the image are included in the evaluation of the AAE. Our result is similar to the one presented by Nir *et al.* [54], which is not surprising since a rigid motion model enters the functional of both methods. It has to be noted that methods with spatio-temporal smoothness terms give lower errors in general. In Fig. 9 we show the results for the estimated optical flow and the corresponding epipolar geometry for 15 iteration steps.

In a final experiment we evaluate our methodology on four image pairs of the Middlebury optical flow benchmark [5]. Frames 10 and 11 of the synthetic *Urban2*, *Urban3*, *Grove2* and *Grove3* [9] training set deal with rigid stereo motion for which the ground truth is publicly available. In Table 7 we show the influence of including the epipolar constraint in pure optical flow by collecting the AAE and the AEE of the estimated flow fields. The first column shows the errors for the

[8] Better results are reported in the technical report [10] which did not yet undergo the process of peer reviewing.

[9] all available at http://vision.middlebury.edu/flow/data/

Table 6: Results for the Yosemite sequence without clouds compared to other 2D methods.

Method	AAE
Brox *et al.* [14]	1.59°
Mémin and Pérez [50]	1.58°
Roth and Black [58]	1.47°
Bruhn *et al.* [16]	1.46°
Amiaz *et al.* [4]	1.44°
Nir *et al.* [54]	1.15°
Our method	**1.15°**

Table 7: Influence of including the epipolar term (+ET) in optical flow for the four stereo image pairs of the Middlebury optical flow training set. The results in the first two columns are presented for the default settings ($\alpha = 20.0$, $\gamma = 20.0$, $\sigma = 0.9$ and $\beta = 40.0$). The results in the last two columns are presented for the fixed settings given in [92] ($\alpha = 400.0$, $\gamma = 20.0$, $\sigma = 0.5$ and $\beta = 5.0$).

Image Pair	[14] AAE / AEE	[14] + ET AAE / AEE	[92] AAE / AEE	[92] + ET AAE / AEE
Grove2	2.67 / 0.19	2.53 / 0.17	2.19 / 0.16	2.13 / 0.14
Grove3	6.78 / 0.69	5.89 / 0.64	5.84 / 0.59	5.61 / 0.57
Urban2	2.66 / 0.32	2.20 / 0.29	2.46 / 0.26	2.15 / 0.24
Urban3	5.26 / 0.61	4.96 / 0.56	3.40 / 0.44	3.11 / 0.39

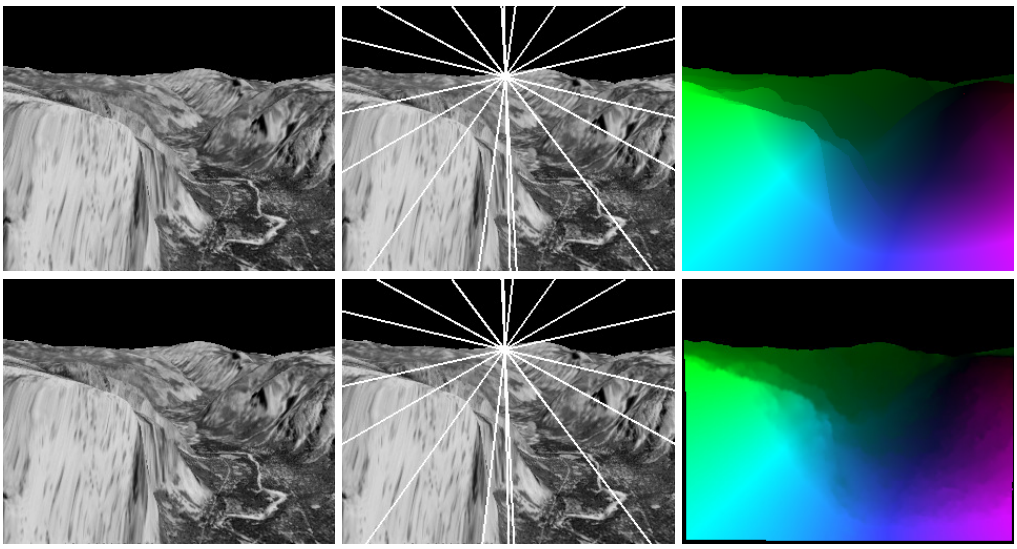

Figure 9: Results for the Yosemite sequence without clouds. **Top Row: (a)** Frame 8. **(b)** Estimated eipolar lines in frame 8. **(c)** Ground truth optical flow. **Bottom Row: (b)** Frame 9. **(d)** Estimated eipolar lines in frame 9. **(f)** Estimated optical flow (settings: $\alpha = 19.1$, $\gamma = 2.1$, $\sigma = 0.9$ and $\beta = 50.0$). Pixels (apart from the sky region) that are warped outside the image are colored black.

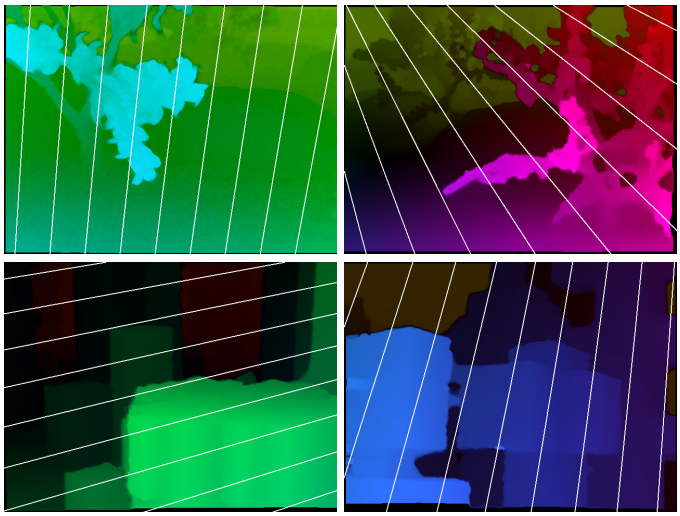

Figure 10: The flow fields between frames 10 and 11 of the Middlebury training sequences. The estimated epipolar lines for frame 10 have been overlaid. **Top Left: (a)** Grove2. **Top Right: (b)** Grove3. **Bottom Left: (c)** Urban2. **Bottom Right: (d)** Urban3.

optical flow method of Brox *et al.* [14], while the second column lists the errors for our joint model that adds the epipolar term to the functional. The results for both methods are obtained for the default settings with $\beta = 40$. Our joint method outperforms pure optical flow for all tested image pairs, even for Urban2 where the motion of a small car does not fulfill the epipolar constraint. To illustrate that our idea can even improve the performance of more recent state of the art optical flow techniques, we additionally provide results for the method of Zimmer *et al.* [92] without and with the additional epipolar term proposed in (18). The original method without epipolar term is currently one of the top ranking methods in the Middlebury benchmark and has been briefly sketched at the end of Subsection 2.1. For *fixed* settings ($\alpha = 400.0$, $\gamma = 20.0$, $\sigma = 0.5$ and $\beta = 5.0$), the joint variant of [92] improves the AAE and AEE over pure optical flow for all four sequences, leading to some of the best results published so far. The corresponding estimated flow fields and epipolar lines are shown in Fig. 10.

6.3 Automatic 3D Reconstruction

To conclude, we present reconstruction results for some of the images that were used previously. To perform a dense reconstruction of the depicted scene, we extract the left and right camera projection matrices P and P' from the estimated fundamental matrix F and use them to triangulate the back-projected rays for each pixel. If no additional information about the cameras or the scene is available, a reconstruction is only possible up to a projective transformation of 3D space [21, 32]. One type of information that is often at hand in practice are the internal camera parameters, such as the focal length and the principal point. If these parameters are known, we can extract the essential matrix from F and determine the relative pose and orientation of the second camera with respect to the first one [42, 33, 22]. This way we obtain a reconstruction that is up to scale.
By simultaneously solving for the dense optical flow and the epipolar geometry of *two* images, we can associate with each pixel of the left image a 3D point in space, a so-called *range image*. Contrary to multiview reconstruction [64, 73], range images are not a complete representation of the depicted scene and can therefore not be evaluated against multiview ground truth. Although the methodology presented in this paper could serve as a basis for an uncalibrated multiview system that integrates these range images [89], we restrict ourselves for now to a visual assessment of the results.
In Fig. 11 we present the reconstruction from frames 5 and 6 of the Herz-Jesu-P25 data set as an untextured mesh. The ground part is left out for visualisation purposes. Many details are visible and discontinuities in the depth are accurately recovered. This is also true for the reconstruction from frames 1 and 2 of the City-Hall sequence in Fig. 12. We do not display points that are warped outside the im-

Figure 11: Untextured reconstruction from frames 5 and 6 of the Herz-Jesu-P25 data set.

age by the optical flow. Fig. 12 (b) and (c) show a close-up of the middle section of the building. Fine details can clearly be distinguished and the statues are easily recognisable in the untextured surface. For both pairs we used the 1536×1024 versions of the original images, which comes down to more than 1.5 million reconstructed points. The settings for our joint estimation method are those from Table 5 and for the reconstruction we used the provided internal camera parameters. In a final experiment we reconstruct a face from a stereo pair that we have recorded with two Point Grey Flea cameras. The images are shown in Fig. 13 (a) and (b) and are of size 280×430. We do not perform a full calibration of the stereo rig but only use the focal length and an approximation of the principal point for our reconstruction. The result in Fig. 13 (c)-(e) is obtained by replacing the TV-regularisation of our original model by an anisotropic flow-driven one [84]. This results in a better smoothing between and along flow discontinuities. The facial expression is captured very well with only a slight degeneration of the reconstruction near specularities on the nose and the eyes. The background is excluded from the fundamental matrix estimation.

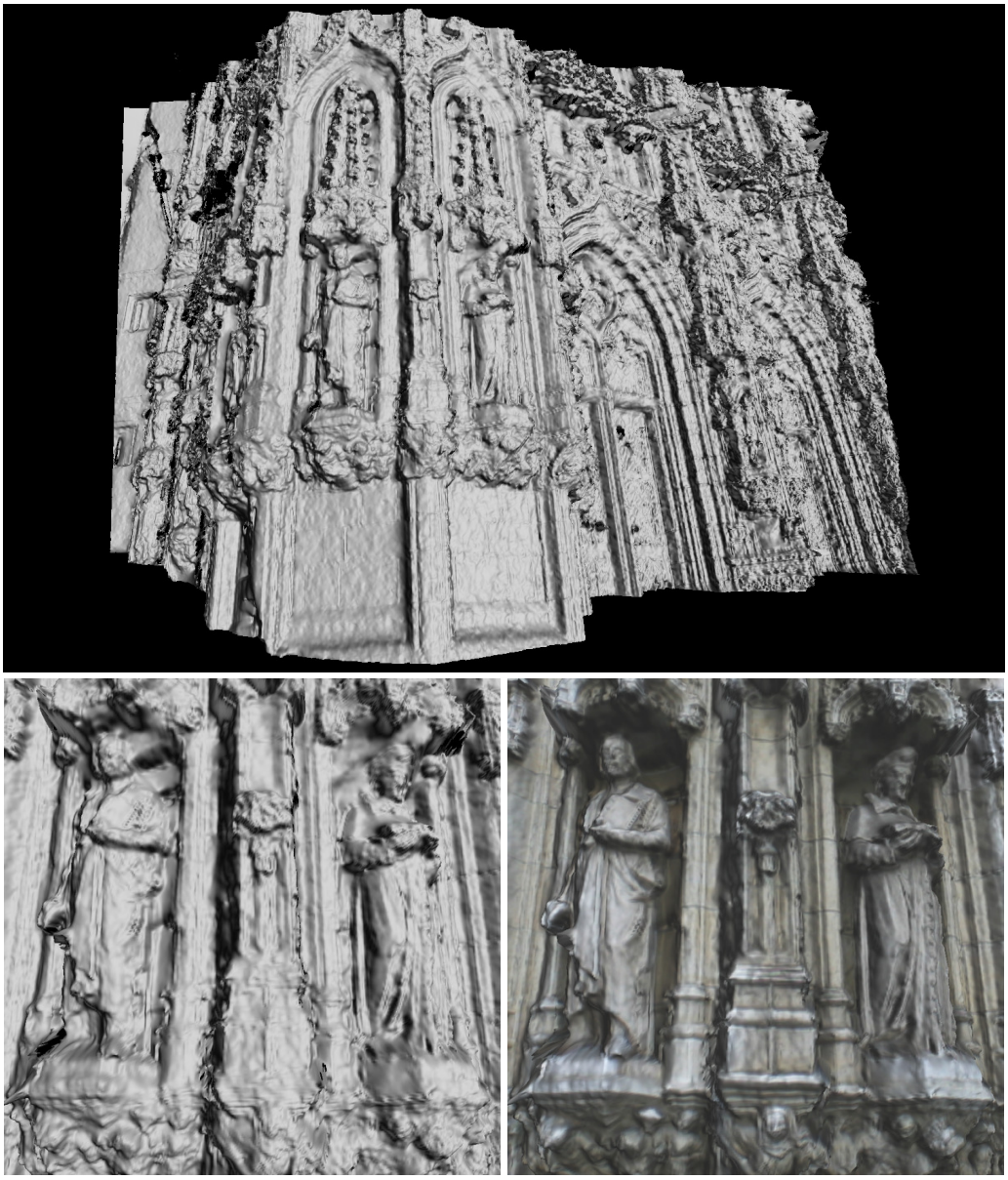

Figure 12: Reconstruction from frames 1 and 2 of the City-Hall data set. **Top (a)** Untextured reconstruction. **Bottom Left: (b)** Untextured close-up. **Bottom Right: (c)** Textured close-up.

Figure 13: Face reconstruction from 2 frames. **Top Left: (a)** Left frame. **Top Middle: (b)** Right frame. **Top Right: (c)** Untextured frontal view. **Bottom Left: (d)** Untextured side view. **Bottom Right: (e)** Textured side view.

6.4 Limitations of Dense Methods

In this paper we show the advantages of dense methods for the estimation of the fundamental matrix, particularly in challenging cases with low-texture or near-degenerate configurations.
From our experiments, however, it is clear that large changes in view point can pose a problem to dense matching algorithms due to the large displacements and induced occlusions. While some of these effects are counterbalanced by including a coupling between optical flow and fundamental matrix estimation, for very wide baseline image pairs sparse feature based methods still offer advantages. Recent optical flow methods have been proposed to overcome this limitation by either integrating feature matches [13] or by applying a more global search [69]. Also motion parallax, induced by sudden and large changes in depth, forms a problem for traditional optical flow, since large jumps in the displacement field are difficult to capture by global methods. Feature matching, in contrast, does not suffer from parallax because it is essentially a local process that does not enforce spatial consistency.
Occlusions can form a second challenge. They are usually present in dense flow fields, but hardly pose a problem in feature matching where disappearing interest points will generally not be matched in the next frame. For our joint method we experienced that small occlusions only have a very limited influence because they are down weighted by the robust epipolar term. Additionally these regions are filled in by the smoothness term in accordance with the estimated stereo geometry. In wide baseline scenarios, occlusions will have a larger impact such that their explicit detection [1, 79] should be considered in combination with the afore mentioned techniques for large displacement optical flow.
In contrast to wide baseline images and occlusions, illumination changes are less problematic for dense estimation methods. While the SIFT descriptor is invariant under multiplicative and additive illumination changes by using normalised gradient information, similar concepts can be used by optical flow methods, e.g. by using photometric invariants [52] or normalised cross correlation [68, 87].
Since our joint variational model relies on image sequences that allow a stable estimation of the fundamental matrix, the application is restricted to rigid scenes which are not dominated by moving objects.

7 Conclusions and Future Work

We have explored a new application field for dense optical flow techniques: the robust estimation of the fundamental matrix. Variational optical flow methods incorporate a global smoothness constraint that ensures filling-in in the absence of

texture and dense correspondences in the case of degeneracy. Our experiments demonstrate that in these scenarios optical flow based fundamental matrix estimation clearly outperforms the widely-used feature based methods. In these scenarios we recommend to favour dense over sparse methods for estimating the fundamental matrix.

As a second contribution we have also shown that epipolar geometry helps to improve the computation of dense optical flow. A simultaneous estimation of the fundamental matrix and the optical flow leads to higher accuracy and better stability than their separate estimation. To this end, we have proposed a novel coupled energy formulation and an iterative solution strategy. This allows us to obtain estimates of the fundamental matrix that are competitive to and more stable than those of well-established feature based methods. Additionally, the accuracy of the optical flow improves significantly when applied to rigid scenes.

It is interesting to analyse the reasons why a dense approach that incorporates *all* correspondences can be competitive with fairly sophisticated strategies that single out only the *very best* correspondences. Our explanation for this observation is as follows: In those cases where feature based methods produce a mismatch, its influence on the final result is severe. Hence they require involved robustification methods such as RANSAC and its numerous variants. Dense methods, on the other hand, incorporate smoothness terms that prevent individual outliers. Furthermore, the accuracy of the fundamental matrix estimation benefits from the error averaging when exploiting thousands of correspondences. This accuracy will improve even further with ongoing, rapid progress in dense optical flow estimation.

Our short term goals for improving the present approach include, among others, the incorporation of occlusion handling, the segmentation of the scene into homogeneous motion regions and extensions to multiview settings. In the long run we hope that our paper helps to pave the road towards a much larger class of novel, more robust computer vision approaches based on dense correspondences.

Acknowledgements We gratefully acknowledge partial funding by the Deutsche Forschungsgemeinschaft under grant *WE 2602/6-1* and the *Cluster of Excellence "Multimodal Computing and Interaction"*. We thank Henning Zimmer for providing his optical flow algorithm and Pascal Gwosdek for his help.

References

[1] L. Alvarez, R. Deriche, T. Papadopoulo, and J. Sánchez. Symmetrical dense optical flow estimation with occlusions detection. In A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, editors, *Computer Vision – ECCV 2002*, vol-

ume 2350 of *Lecture Notes in Computer Science*, pages 721–736. Springer, Berlin, 2002.

[2] L. Alvarez, R. Deriche, J. Sánchez, and J. Weickert. Dense disparity map estimation respecting image derivatives: a PDE and scale-space based approach. *Journal of Visual Communication and Image Representation*, 13(1/2):3–21, 2002.

[3] L. Alvarez, J. Esclarín, M. Lefébure, and J. Sánchez. A PDE model for computing the optical flow. In *Proc. XVI Congreso de Ecuaciones Diferenciales y Aplicaciones*, pages 1349–1356, Las Palmas de Gran Canaria, Spain, September 1999.

[4] T. Amiaz, E. Lubetzky, and N. Kiryati. Coarse to over-fine optical flow estimation. *Pattern Recognition*, 40(9):2496–2503, 2007.

[5] S. Baker, S. Roth, D. Scharstein, M. J. Black, J. P. Lewis, and R. Szeliski. A database and evaluation methodology for optical flow. In *Proc. 2007 IEEE International Conference on Computer Vision*, Rio de Janeiro, Brazil, October 2007. IEEE Computer Society Press.

[6] S. Baker, D. Scharstein, J. P. Lewis, S. Roth, M. J. Black, and R. Szeliski. A database and evaluation methodology for optical flow. Technical Report MSR-TR-2009-179, Microsoft Research, Redmond, WA, December 2009.

[7] J. L. Barron, D. J. Fleet, and S. S. Beauchemin. Performance of optical flow techniques. *International Journal of Computer Vision*, 12(1):43–77, February 1994.

[8] H. Bay, A. Ess, T. Tuytelaars, and L. J. Van Gool. Speeded-up robust features (SURF). *Computer Vision and Image Understanding*, 110(3):346–359, 2008.

[9] R. Ben-Ari and N. Sochen. Variational stereo vision with sharp discontinuities and occlusion handling. In *Proc. Eleventh International Conference on Computer Vision*, Rio de Janeiro, October 2007. IEEE Computer Society Press.

[10] D. Bitton, G. Rosman, T. Nir, A. M. Bruckstein, A. Feuer, and R. Kimmel. Over-parameterized optical flow using a stereoscopic constraint. Technical Report CIS-2009-18, Computer Science Department, Technion - Israel Institute of Technology, Israel, November 2009.

[11] M. J. Black and P. Anandan. Robust dynamic motion estimation over time. In *Proc. 1991 IEEE Conference on Computer Vision and Pattern Recognition*, pages 292–302, Maui, HI, June 1991. IEEE Computer Society Press.

[12] M. J. Brooks, W. Chojnacki, and L. Baumela. Determining the ego-motion of an uncalibrated camera from instantaneous optical flow. *Journal of the Optical Society of America A*, 14:2670–2677, 1997.

[13] T. Brox, C. Bregler, and J. Malik. Large displacement optical flow. In *Proc. 2009 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 41–48, Miami, FL, June 2009. IEEE Computer Society Press.

[14] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optic flow estimation based on a theory for warping. In T. Pajdla and J. Matas, editors, *Computer Vision – ECCV 2004*, volume 3024 of *Lecture Notes in Computer Science*, pages 25–36. Springer, Berlin, 2004.

[15] A. Bruhn, J. Weickert, T. Kohlberger, and C. Schnörr. A multigrid platform for real-time motion computation with discontinuity-preserving variational methods. *International Journal of Computer Vision*, 70(3):257–277, December 2006.

[16] A. Bruhn, J. Weickert, and C. Schnörr. Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods. *International Journal of Computer Vision*, 61(3):211–231, 2005.

[17] J. Canny. A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:679–698, 1986.

[18] T. F. Chan and L. A. Vese. Active contours without edges. *IEEE Trans. Image Processing*, 10(2):266–277, February 2001.

[19] O. Chum, J. Matas, and S. Obdrzalek. Enhancing RANSAC by generalized model optimization. In K.-S. Hong and Z. Zhang, editors, *Proc. Sixth Asian Conference on Computer Vision*, volume 2 of *Lecture Notes in Computer Science*, pages 812–817, January 2004.

[20] O. Chum, T. Werner, and J. Matas. Two-view geometry estimation unaffected by a dominant plane. In *Proc. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 772–779, San Diego, CA, June 2005. IEEE Computer Society Press.

[21] O. Faugeras. What can be seen in three dimensions with an uncalibrated stereo rig? In Giulio Sandini, editor, *Computer Vision – ECCV 1992*, volume 588 of *Lecture Notes in Computer Science*, pages 563–578. Springer, Berlin, 1992.

[22] O. Faugeras, Q.-T. Luong, and T. Papadopoulo. *The Geometry of Multiple Images*. MIT Press, Cambridge, MA, 2001.

[23] M. Fischler and R. Bolles. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24:381–385, 1981.

[24] A. W. Fitzgibbon. Simultaneous linear estimation of multiple view geometry and lens distortion. In *Proc. 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 125–132, Kauai, HI, June 2001. IEEE Computer Society Press.

[25] W. Förstner and E. Gülch. A fast operator for detection and precise location of distinct points, corners and centres of circular features. In *Proc. ISPRS Intercommission Conference on Fast Processing of Photogrammetric Data*, pages 281–305, Interlaken, Switzerland, June 1987.

[26] J.-M. Frahm and M. Pollefeys. RANSAC for (quasi-) degenerate data (QDEGSAC). In *Proc. 2006 IEEEComputer Society Conference on Computer Vision and Pattern Recognition*, pages 453–460, New York, NY, June 2006. IEEE Computer Society Press.

[27] G. H. Golub and C. M. Van Loan. *Matrix Computations*. The John Hopkins University Press, Baltimore, MD, 1989.

[28] P. Gwosdek, H. Zimmer, S. Grewenig, A. Bruhn, and J. Weickert. A highly efficient GPU implementation for variational optic flow based on the Euler-Lagrange framework. In *Proc. 2010 ECCV Workshop on Computer Vision with GPUs*, Heraklion, Greece, September 2010.

[29] K. J. Hanna. Direct multi-resolution estimation of ego-motion and structure from motion. In *Proc. Workshop on Visual Motion*, pages 156–162. IEEE Computer Society Press, October 1991.

[30] C. G. Harris and M. Stephens. A combined corner and edge detector. In *Proc. Fourth Alvey Vision Conference*, pages 147–152, Manchester, England, August 1988.

[31] R. Hartley. In defense of the eight-point algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(6):580–593, 1997.

[32] R. Hartley, R. Gupta, and T. Chang. Stereo from uncalibrated cameras. In *Proc. 1992 IEEE International Conference on Image Processing*, pages 761–764, Champaign, IL, June 1992. IEEE Computer Society Press.

[33] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision*. Cambridge University Press, Cambridge, UK, 2000.

[34] B. Horn and B. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[35] P. J. Huber. *Robust Statistics*. Wiley, New York, 1981.

[36] K Kanatani, Y. Shimizu, N. Ohta, M. J. Brooks, W. Chojnacki, and A. van den Hengel. Fundamental matrix from optical flow: optimal computation and reliability evaluation. *Journal of Electronic Imaging*, 9:194–202, April 2000.

[37] Y. H. Kim, A. M. Martinez, and A. C. Kak. Robust motion estimation under varying illumination. *Image and Vision Computing*, 23(4):365–375, April 2005.

[38] A. Klaus, M. Sormann, and K. Karner. Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure. In *Proc. 18th International Conference on Pattern Recognition, Part III*, volume 3, pages 15–18, Hong Kong, China, August 2006.

[39] V. Kolmogorov and R. Zabih. Multi-camera scene reconstruction via graph cuts. In A. Heyden, G. Sparr, M. Nielsen, and P. Johansen, editors, *Computer Vision – ECCV 2002, Part III*, volume 2352 of *Lecture Notes in Computer Science*, pages 82–96. Springer, Berlin, 2002.

[40] C. Lei, J. Selzer, and Y.-H. Yang. Region-tree based stereo using dynamic programming optimization. In *Proc. 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2378–2385, Washington, DC, June 2006. IEEE Computer Society Press.

[41] K. Levenberg. A method for the solution of certain non-linear problems in least squares. *The Quarterly of Applied Mathematics*, 2:164–168, July 1944.

[42] H. C. Longuet-Higgins. A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135, September 1981.

[43] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[44] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proc. Seventh International Joint Conference on Artificial Intelligence*, pages 674–679, Vancouver, Canada, August 1981.

[45] Q.-T. Luong and O. D. Faugeras. The fundamental matrix: theory, algorithms, and stability analysis. *International Journal of Computer Vision*, 17(1):43–75, January 1996.

[46] M. Mainberger, A. Bruhn, and J. Weickert. Is dense optical flow useful to compute the fundamental matrix? In A. Campilho and M. Kamel, editors, *Image Analysis and Recognition*, volume 5112 of *Lecture Notes in Computer Science*, pages 630–639, Póvoa de Varzim, Portugal, 2008. Springer, Berlin.

[47] D. Marquardt. An algorithm for least-squares estimation of nonlinear parameters. *SIAM Journal on Applied Mathematics*, 11:431–441, 1963.

[48] D. Marr and E. Hildreth. Theory of edge detection. *Proceedings of the Royal Society of London, Series B*, 207:187–217, 1980.

[49] E. Mémin and P. Pérez. Dense estimation and object-based segmentation of the optical flow with robust techniques. *IEEE Transactions on Image Processing*, 7(5):703–719, May 1998.

[50] E. Mémin and P. Pérez. Hierarchical estimation and segmentation of dense motion fields. *International Journal of Computer Vision*, 46(2):129–155, 2002.

[51] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(10):1615–1630, October 2005.

[52] Y. Mileva, A. Bruhn, and J. Weickert. Illumination-robust variational optical flow with photometric invariants. In F.A. Hamprecht, C. Schnörr, and B. Jähne, editors, *Pattern Recognition*, volume 4713 of *Lecture Notes in Computer Science*, pages 152–162, Heidelberg, Germany, 2007. Springer, Berlin.

[53] H.-H. Nagel and W. Enkelmann. An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 8:565–593, 1986.

[54] T. Nir, A. M. Bruckstein, and R. Kimmel. Over-parameterized variational optical flow. *International Journal of Computer Vision*, 76(2):205–216, 2008.

[55] N. Ohta and K. Kanatani. Optimal structure from motion algorithm for optical flow. *IEICE Transactions on Information and Systems*, E78-D(12):1559–1566, December 1995.

[56] M. Proesmans, L. Van Gool, E. Pauwels, and A. Oosterlinck. Determination of optical flow and its discontinuities using non-linear diffusion. In J.-O. Eklundh, editor, *Computer Vision – ECCV '94*, volume 801 of *Lecture Notes in Computer Science*, pages 295–304. Springer, Berlin, 1994.

[57] R. Raguram, J. M. Frahm, and M. Pollefeys. A comparative analysis of RANSAC techniques leading to adaptive real-time random sample consensus. In D. Forsyth, P. Torr, and A. Zisserman, editors, *Computer Vision – ECCV 2008, Part II*, volume 5303 of *Lecture Notes in Computer Science*, pages 500–513. Springer, Berlin, 2008.

[58] S. Roth and M. Black. On the spatial statistices of optical flow. In *Proc. Tenth International Conference on Computer Vision*, volume 1, pages 42–49, Beijing, China, June 2005. IEEE Computer Society Press.

[59] P. J. Rousseeuw and A. M. Leroy. *Robust Regression and Outlier Detection*. Wiley, New York, 1987.

[60] J. Saragih and R. Goecke. Monocular and stereo methods for AAM learning from video. In *Proc. 2007 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Minneapolis, MN, June 2007. IEEE Computer Society Press.

[61] D. Scharstein and R. Szeliski. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1-3):7–42, 2002.

[62] D. Schlesinger, B. Flach, and A. Shekhovtsov. A higher order MRF-model for stereo-reconstruction. In C. E. Rasmussen, H. H. Bülthoff, M. A. Giese, and B. Schölkopf, editors, *Pattern Recognition*, volume 3175 of *Lecture Notes in Computer Science*, pages 440–446. Springer, Berlin, 2004.

[63] C. Schnörr. Segmentation of visual motion by minimizing convex non-quadratic functionals. In *Proc. Twelfth International Conference on Pattern Recognition*, volume A, pages 661–663, Jerusalem, Israel, October 1994. IEEE Computer Society Press.

[64] S. Seitz, B. Curless, J. Diebel, D. Scharstein, and R. Szeliski. A comparison and evaluation of multi-view stereo reconstruction algorithms. In *Proc. 2006 IEEE Conference on Computer Vision and Pattern Recognition*, pages I: 519–528, New York, NY, June 2006. IEEE Computer Society Press.

[65] Y. Sheikh, A. Hakeem, and M. Shah. On the direct estimation of the fundamental matrix. In *Proc. 2007 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Minneapolis, MN, June 2007. IEEE Computer Society Press.

[66] J. Shi and C. Tomasi. Good features to track. In *Proc. 1994 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 593–600, Seattle, WA, June 1994. IEEE Computer Society Press.

[67] N. Slesareva, A. Bruhn, and J. Weickert. Optic flow goes stereo: a variational approach for estimating discontinuity-preserving dense dispartiy maps. In W. Kropatsch, R. Sablatnig, and A. Hanbury, editors, *Pattern Recognition*, volume 3663 of *Lecture Notes in Computer Science*, pages 33–40. Springer, Berlin, 2005.

[68] F. Steinbrücker, T. Pock, and D. Cremers. Advanced data terms for variational optic flow estimation. In M.A. Magnor, B. Rosenhahn, and H. Theisel, editors, *Proceedings of the Vision, Modeling, and Visualization Workshop (VMV)*, pages 155–164. DNB, November 2009.

[69] F. Steinbrücker, T. Pock, and D. Cremers. Large displacement optical flow computation without warping. In *Proc. Twelfth International Conference on Computer Vision*, Kyoto, October 2009. IEEE Computer Society Press.

[70] C. V. Stewart. Robust parameter estimation in computer vision. *SIAM Review*, 41(3):513–537, 1999.

[71] C. Strecha, R. Fransens, and L. Van Gool. A probabilistic approach to large displacement optical flow and occlusion detection. In D. Comaniciu, K. Kanatani, R. Mester, and D. Suter, editors, *Statistical Methods in Video Processing*, volume 3247 of *Lecture Notes in Computer Science*, pages 71–82, Berlin, 2004. Springer.

[72] C. Strecha, T. Tuytelaars, and L. Van Gool. Dense matching of multiple wide-baseline views. In *Proc. Ninth International Conference on Computer Vision*, volume 2, pages 1194–1201. IEEE Computer Society Press, October 2003.

[73] C. Strecha, W. von Hansen, L. J. Van Gool, P. Fua, and U. Thoennessen. On benchmarking camera calibration and multi-view stereo for high resolution imagery. In *Proc. 2008 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Anchorage, AK, June 2008. IEEE Computer Society Press.

[74] D. Sun, S. Roth, J. P. Lewis, and M. J. Black. Learning optical flow. In D. Forsyth, P. Torr, and A. Zisserman, editors, *Computer Vision – ECCV 2008, Part III*, volume 5304 of *Lecture Notes in Computer Science*, pages 83–97. Springer, Berlin, 2008.

[75] C. Tomasi and T. Kanade. Detection and tracking of point features. Technical Report CMU-CS-91-132, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, June 1991.

[76] P. H. S. Torr and D. W. Murray. The development and comparison of robust methods for estimating the fundamental matrix. *International Journal of Computer Vision*, 24(3):271–300, 1997.

[77] R. Tsai and T. Huang. Uniqueness and estimation of thee-dimensional motion parameters of rigid objects with curved surfaces. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(1):13–26, 1984.

[78] L. Valgaerts, A. Bruhn, and J. Weickert. A variational model for the joint recovery of the fundamental matrix and the optical flow. In G. Rigoll, editor, *Pattern Recognition*, volume 3663 of *Lecture Notes in Computer Science*, pages 314–324. Springer, Berlin, June 2008.

[79] L. Valgaerts, A. Bruhn, H. Zimmer, J. Weickert, C. Stoll, and C. Theobalt. Joint estimation of motion, structure and geometry from stereo sequences. In K. Daniilidis, P. Maragos, and N. Paragios, editors, *Computer Vision – ECCV 2010*, volume 6314 of *Lecture Notes in Computer Science*, pages 568–581. Springer, Berlin, 2010.

[80] T. Viéville and O. Faugeras. Motion analysis with a camera with unknown, and possibly varying intrinsic parameters. In *Proc. Fifth International Conference on Computer Vision*, pages 750–756, Cambridge, MA, June 1995. IEEE Computer Society Press.

[81] H. Wang and M. Brady. A practical solution to corner detection. In *Proc. 1994 IEEE International Conference on Image Processing*, volume 1, pages 919–923, Austin, TX, November 1994. IEEE Computer Society Press.

[82] A. Wedel, D. Cremers, T. Pock, and H. Bischof. Structure- and motion-adaptive regularization for high accuracy optic flow. In *Proc. Twelfth International Conference on Computer Vision*, Kyoto, October 2009. IEEE Computer Society Press.

[83] A. Wedel, T. Pock, J. Braun, U. Franke, and D. Cremers. Duality TV-$L1$ flow with fundamental matrix prior. In *Proc. Image and Vision Computing New Zealand*, Auckland, New Zealand, November 2008. IEEE Computer Society Press.

[84] J. Weickert and C. Schnörr. A theoretical framework for convex regularizers in PDE-based computation of image motion. *International Journal of Computer Vision*, 45(3):245–264, December 2001.

[85] J. Weng, N. Ahuja, and T. Huang. Optimal motion and structure estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(9):864–884, 1993.

[86] J. Weng, T. Huang, and N. Ahuja. Motion and structure from two perspective views: algorithms, error analysis and error estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(5):451–476, 1989.

[87] M. Werlberger, T. Pock, and H. Bischof. Motion estimation with non-local total variation regularization. In *Proc. 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pages 2464–2471, San Francisco, CA, June 2010. IEEE Computer Society Press.

[88] C. Zach, T. Pock, and H. Bischof. A duality based approach for realtime TV-L^1 optical flow. In F.A. Hamprecht and B. Jähne C. Schnörr, editors, *Pattern Recognition*, volume 4713 of *Lecture Notes in Computer Science*, pages 214–223, Berlin, 2007. Springer.

[89] C. Zach, T. Pock, and H. Bischof. A globally optimal algorithm for robust TV-L^1 range image integration. In *Proc. Ninth International Conference on Computer Vision*, Rio de Janeiro, Brazil, October 2007. IEEE Computer Society Press.

[90] B. Zeisl, P. F. Georgel, F. Schweiger, E. Steinbach, and N. Navab. Estimation of location uncertainty for scale invariant feature points. In *Proc. 2009 British Machine Vision Conference*, London, England, September 2009.

[91] Z. Zhang. Determining the epipolar geometry and its uncertainty: A review. *International Journal of Computer Vision*, 27(2):161–195, 1998.

[92] H. Zimmer, A. Bruhn, and J. Weickert. Optic flow in harmony. *International Journal of Computer Vision*, 93(3):368–388, April 2011.

[93] H. Zimmer, A. Bruhn, J. Weickert, L. Valgaerts, A. Salgado, B. Rosenhahn, and H.-P. Seidel. Complementary optic flow. In D. Cremers, Y. Boykov, A. Blake, and F. R. Schmidt, editors, *Energy Minimization Methods in Computer Vision and Pattern Recognition – EMMCVPR*, volume 5681 of *Lecture Notes in Computer Science*, pages 207–220. Springer, Berlin, 2009.