Robust Trajectory-Space TV-L1 Optical Flow for Non-rigid Sequences[*]

Ravi Garg, Anastasios Roussos, and Lourdes Agapito

Queen Mary University of London, Mile End Road, London E1 4NS, UK

Abstract. This paper deals with the problem of computing optical flow between each of the images in a sequence and a reference frame when the camera is viewing a non-rigid object. We exploit the high correlation between 2D trajectories of different points on the same non-rigid surface by assuming that the displacement sequence of any point can be expressed in a compact way as a linear combination of a low-rank motion basis. This subspace constraint effectively acts as a long term regularization leading to temporally consistent optical flow. We formulate it as a robust soft constraint within a variational framework by penalizing flow fields that lie outside the low-rank manifold. The resulting energy functional includes a quadratic relaxation term that allows to decouple the optimization of the brightness constancy and spatial regularization terms, leading to an efficient optimization scheme. We provide a new benchmark dataset, based on motion capture data of a flag waving in the wind, with dense ground truth optical flow for evaluation of multi-view optical flow of non-rigid surfaces. Our experiments, show that our proposed approach provides comparable or superior results to state of the art optical flow and dense non-rigid registration algorithms.

1 Introduction

Optical flow in the presence of non-rigid deformations is a challenging task and an important problem that continues to attract significant attention from the computer vision community given its wide ranging applications from medical imaging and video augmentation to non-rigid structure from motion. Given a template image of a non-rigid object and an input image of it after deforming, the task can be described as one of finding the displacement field (warp) that relates the input image back to the template. In this paper we are interested in the case where we deal with a long image sequence instead of a single pair of images – each of the images in the sequence must be aligned back to the reference frame. Our work concerns the estimation of the vector field of displacements that maps pixels in the reference frame to each image in the sequence.

Two significant difficulties arise. First, the image displacements between the reference frame and subsequent ones are large since we deal with long sequences. Secondly, as a consequence of the non-rigidity of the motion, multiple warps can explain the same pair of images causing ambiguities to arise. A multi-frame approach offers the advantage to exploit temporal information to resolve these ambiguities. In this paper we make

[*] This work is supported by the European Research Council under ERC Starting Grant agreement 204871-HUMANIS. We are grateful to D. Pizarro for providing results of their method [1] and tracks for the synthetic sequence.

use of the high correlation between 2D trajectories of different points on the same non-rigid surface. These trajectories lie on a lower dimensional subspace and we assume that the displacement field of any point can be expressed compactly as a linear combination of a low-rank motion basis. This leads to a significant reduction in the dimensionality of the problem while implicitly imposing some form of temporal smoothness. The flow field can be represented by the basis and a set of coefficients for each point in the template image. In contrast to previous multi-frame optical flow approaches that incorporate explicit temporal smoothness regularization [2] our subspace constraint implicitly acts as a long term smoothing term leading to temporally consistent optical flow.

Subspace constraints have been used before both in the context of sparse point tracking [3–5] and optical flow [3, 6] in the rigid and non-rigid domains, to allow correspondences to be obtained in low textured areas. While Irani's original rigid [3] and Torresani et al.'s non-rigid [5] formulations relied on minimizing the linearized brightness constraint in their discrete form, Garg et al. [6] extended the subspace constraints to the continuous domain in the non-rigid case using a variational approach. The common feature of all the above approaches is that the subspace constraint is imposed as a hard constraint. Hard constraints are vulnerable to noise in the model and can be avoided by substituting them with principled robust constraints. In this paper we extend the use of multi-frame temporal smoothness constraints within a variational framework by providing a more principled energy formulation with a robust soft constraint which leads to improved results. In practice, we penalize deviations of the optical flow trajectories from the low-rank subspace manifold, which acts as a temporal regularization term over long sequences. We then take advantage of recent developments [7, 8] in variational methods and optimize the energy defining a variant of the duality-based efficient numerical optimization scheme.

2 Related Work and Contribution

Variational methods formulate the optical flow or image alignment problems as the optimization of an energy functional in the continuous domain. Stemming from Horn and Schunck's original approach [9], the energy incorporates a data term that optimizes the brightness constancy constraint and a regularization term that allows to fill-in flow information in low textured areas. Variational methods have seen a huge surge in recent years due to the development of more sophisticated and robust data fidelity terms which are robust to changes in image brightness or occlusions [10, 11]; the addition of efficient regularization terms such as Total-Variation [12, 13] or temporal smoothing terms [2]; and new optimization strategies that allow computation of highly accurate [14] and real time optical flow [12] even in the presence of large displacements [15, 10, 16].

One of the most successful recent advances in variational methods has been the development of the duality based efficient numerical optimization scheme to solve the TV-L1 optical flow problem [12, 8]. Duplication of the optimization variable via a quadratic relaxation is used to decouple the data and regularization terms, decomposing the optimization problem into two, each of which is a convex energy that can be solved in a globally optimal manner. The minimization algorithm then alternates between solving for each of the two variables assuming the other one fixed. One of the key advantages of this decoupling scheme is that since the data term is *point-wise* its optimization can be highly parallelized using graphics hardware [12]. Following its success in optical flow computation, this optimization scheme has since been successfully applied to motion and disparity estimation [17] and real time dense 3D reconstruction [18].

Non-rigid image registration, is a long standing field that has recently seen important progress in its robust estimation in the case of severe deformations and large baselines both from keypoint-based and learning based approaches. Successful keypoint-based approaches to deformable image registration include the parametric approach of Pizarro and Bartoli [1] who propose a warp estimation algorithm that can cope with wide baseline and self-occlusions using a piecewise smoothness prior on the deforming surface. A direct approach that uses all the pixels in the image is used as a refinement step. Discriminative approaches, on the other hand, learn the mapping that predicts the deformation parameters given a distorted image but require a large number of training samples. In recent work, Tian and Narasimhan [19] propose to combine generative and discriminative approaches to reuse training samples far away from the test image which leads to the use of a significantly lower number of training samples.

Our contribution In this paper we adopt a robust approach to non-rigid image alignment where instead of imposing the hard constraint that the optical flow must lie on the low-rank manifold [6], we penalize flow fields that lie outside it. Formulating the manifold constraint as a *soft constraint* using variational principles leads to an energy with a quadratic relaxation term that allows us to adopt a decoupling scheme, similar to the one described above [12, 8], for its efficient optimization. Since our regularization term is parameterized in terms of the basis coefficients, instead of the full flow field, we achieve an important dimensionality reduction in this term, which is usually the bottleneck of other quadratic relaxation duality based approaches [12, 8]. Moreover, the optimization of this regularization step can be parallelized due to the independence of the orthogonal basis coefficients adding further advantages to the, already efficient, optimization scheme of Zach et al. [12]. Our approach can be seen as an extension of this efficient TV-L1 flow estimation algorithm to the case of multi-frame non-rigid optical flow, where the addition of subspace constraints acts as a temporal regularization term.

Currently, there are no benchmark datasets for the evaluation of optical flow that include long sequences of non-rigid deformations. In particular, the most popular one [20] (Middlebury) does not incorporate any such sequences. In order to facilitate quantitative evaluation of multi-frame non-rigid registration and optical flow and to promote progress in this area, in this paper we provide a new dataset based on motion capture data of a flag waving in the wind, with dense ground truth optical flow. Our quantitative evaluation on this dataset using three different motion bases (Principal Components Analysis (PCA), Discrete Cosine Transform (DCT) and Uniform Cubic B-Splines) shows that our proposed approach improves or has equivalent performance to state of the art large displacement [10] and duality based [12] optical flow algorithms and a parametric dense non-rigid registration approach [1].

3 Multi-frame Image Alignment

Consider an image sequence of a non-rigid object moving and deforming in 3D. In the classical optical flow problem, one seeks to estimate the vector field of image point displacements independently for each pair of consecutive frames. In this paper, we adopt the following multi-frame reformulation of the problem. Taking one frame as the reference template, usually the first frame, our goal is then to estimate the 2D trajectories of every point visible in the reference frame over the entire sequence, using a multi-frame approach. The use of temporal information in this way allows us to predict the location of points not visible in a particular frame making us robust to self-occlusions or external occlusions by other objects.

3.1 Subspace Trajectory Model

In order to solve the multi-frame optical flow problem, we make use of the fact that the 2D image trajectories of points on an object are highly correlated, even when the object is deforming. We model this property by assuming that the trajectories are near a low-dimensional subspace. This is induced by the non-rigid low-rank shape model, first proposed by Bregler et al. [21], which states that the time varying 3D shape of a non-rigid object can be expressed as as a linear combination of a low-rank shape basis. This assumption has been successfully exploited for 3D reconstruction by Non-Rigid Structure from Motion (NRSfM) algorithms [22] and non-rigid 2D tracking [5].

More precisely, assume that the input image sequence has F frames and the n_0-th frame, $n_0 \in \{1, \ldots, F\}$ has been chosen as the reference. If $\Omega \subset \mathbb{R}^2$ denotes the image domain, we define the function $\boldsymbol{u} : \Omega \times \{1, \ldots, F\} \to \mathbb{R}^2$ that represents the point trajectories in the following way. For every visible point $\boldsymbol{x} \in \Omega$ in the reference image, $\boldsymbol{u}(\boldsymbol{x}; \cdot) : \{1, \ldots, F\} \to \mathbb{R}^2$ is its discrete-time 2D trajectory over all frames of the sequence. The coordinates of each trajectory $\boldsymbol{u}(\boldsymbol{x}; \cdot)$ are expressed with respect to the position of the point $\boldsymbol{x}$ at $n = n_0$, which means that $\boldsymbol{u}(\boldsymbol{x}; n_0) = 0$ and that the location of the same point in frame n is $\boldsymbol{x} + \boldsymbol{u}(\boldsymbol{x}; n)$.

Mathematically, the linear subspace constraint on the 2D trajectories $\boldsymbol{u}(\boldsymbol{x}; n)$ can be expressed in the following way. For all $\boldsymbol{x} \in \Omega$ and $n \in \{1, \ldots, F\}$:

$$\boldsymbol{u}(\boldsymbol{x}; n) = \sum_{i=1}^{R} \boldsymbol{q}_i(n) L_i(\boldsymbol{x}) \; + \; \boldsymbol{\varepsilon}(\boldsymbol{x}; n) \, , \tag{1}$$

which states that the trajectory $\boldsymbol{u}(\boldsymbol{x}; \cdot)$ of any point $\boldsymbol{x} \in \Omega$ can be approximated as the linear combination of R basis trajectories $\boldsymbol{q}_1(n), \ldots, \boldsymbol{q}_R(n) : \{1, \ldots, F\} \to \mathbb{R}^2$ that are independent from the point location. We include a modeling error term $\boldsymbol{\varepsilon}(\boldsymbol{x}; n)$ which will allow us to impose the subspace constraint as a penalty term. We refer to the subspace where any such combination lies, i.e. the linear span of the basis trajectories, as a *trajectory subspace* and we denote it by $\mathcal{S}_Q$. The linear combination is controlled by coefficients $L_i(\boldsymbol{x})$ that depend on $\boldsymbol{x}$, therefore we can interpret the collection of all the coefficients for all the points $\boldsymbol{x} \in \Omega$ as a vector-valued image $\boldsymbol{L}(\boldsymbol{x}) \triangleq [L_1(\boldsymbol{x}), \ldots, L_R(\boldsymbol{x})]^T : \Omega \to \mathbb{R}^R$. Effective choices for the model order (or rank) R usually correspond to values much smaller than $2F$, which means that the above representation is very compact and achieves a dramatic dimensionality reduction on the point trajectories. Normally the values of $\boldsymbol{\varepsilon}(\boldsymbol{x}; n)$ are relatively small, yet sufficient to improve the robustness of the multi-frame optical flow estimation.

We now re-write equation (1) in matrix notation, which will be useful in the subsequent presentation. Let $\boldsymbol{\mathcal{U}}(\boldsymbol{x})$ and $\boldsymbol{\mathcal{E}}(\boldsymbol{x}) : \Omega \to \mathbb{R}^{2F}$ be equivalent representations of the functions $\boldsymbol{u}(\boldsymbol{x}; n)$ and $\boldsymbol{\varepsilon}(\boldsymbol{x}; n)$ that are derived by vectorizing the dependence on the discrete time n and let Q be the trajectory basis matrix whose columns contain the basis elements $\boldsymbol{q}_1(n), \dots, \boldsymbol{q}_R(n)$, after vectorizing them in the same way:

$$\underbrace{\boldsymbol{\mathcal{U}}}_{2F\times 1}(\boldsymbol{x}) \triangleq \begin{bmatrix} \boldsymbol{u}(\boldsymbol{x};1) \\ \vdots \\ \boldsymbol{u}(\boldsymbol{x};F) \end{bmatrix}, \quad \underbrace{\boldsymbol{\mathcal{E}}}_{2F\times 1}(\boldsymbol{x}) \triangleq \begin{bmatrix} \boldsymbol{\varepsilon}(\boldsymbol{x};1) \\ \vdots \\ \boldsymbol{\varepsilon}(\boldsymbol{x};F) \end{bmatrix}, \quad \underbrace{\mathrm{Q}}_{2F\times R} \triangleq \begin{bmatrix} \boldsymbol{q}_1(1) & \cdots & \boldsymbol{q}_R(1) \\ \vdots & & \vdots \\ \boldsymbol{q}_1(F) & \cdots & \boldsymbol{q}_R(F) \end{bmatrix}$$

The subspace constraint (1) can now be written as follows:

$$\boldsymbol{\mathcal{U}}(\boldsymbol{x}) = \mathrm{Q}\,\boldsymbol{L}(\boldsymbol{x}) \;+\; \boldsymbol{\mathcal{E}}(\boldsymbol{x})\,,\ \forall \boldsymbol{x} \in \Omega \tag{2}$$

3.2 Choice of Basis

Concerning the choice of 2D trajectory basis $\{\boldsymbol{q}_1(n), \dots, \boldsymbol{q}_R(n)\}$, we consider orthonormal bases as it simplifies the analysis and calculations in our method (see Section 4). Of course this assumption is not restrictive, since for any basis an orthonormal one can be found that will span the same subspace. We now describe several effective choices of trajectory basis that we have used in our formulation.

Predefined bases for single-valued discrete-time signals with F samples can be used to model separately each coordinate of the 2D trajectories. Assuming that the rank R is an even number, this single-valued basis should have $R/2$ elements $w_1(n), \dots, w_{R/2}(n)$ and the trajectory basis would be given by:

$$\boldsymbol{q}_i(n) = \begin{cases} [w_i(n), 0]^T, & \text{if } i = 1, \dots, R/2 \\ [0, w_{i-R/2}(n)]^T, & \text{if } i = R/2+1, \dots, R \end{cases} \tag{3}$$

Provided that the object moves and deforms smoothly, effective choices for the basis $\{w_i(n)\}$ are (*i*) the first $\frac{R}{2}$ low-frequency basis elements of the 1D Discrete Cosine Transform (DCT) or (*ii*) a sampling of the basis elements of the Uniform Cubic B-Splines of rank $R/2$ over the sequence's time window, followed by orthonormalization of the yielded basis. An alternative is to compute the basis by applying Principal Component Analysis (PCA) to a small subset of *reliable* point tracks. *Reliable* tracks are those where the texture of the image is strong in both spatial directions and could be selected using Shi and Tomasi's criterium [23]. Provided that it is possible to estimate a set of *reliable* tracks that adequately represent the trajectories of the points over the whole object, the choice of the PCA basis is optimum for the linear model of given rank R, in terms of representational power.

4 Variational multi-frame optical flow estimation

In this section we aim to combine dense motion estimation with the trajectory subspace constraints described in Section 3.1 following variational principles. If $I(\boldsymbol{x};n) : \Omega \times \{1, \dots, F\} \to \mathbb{R}$ denotes the input image sequence and n_0 is the index of the reference frame, then we propose to minimize the following energy:

$$\begin{aligned} E\big[\boldsymbol{u}(\boldsymbol{x};n)\,,\,\boldsymbol{L}(\boldsymbol{x})\big] &= \alpha \int_\Omega \sum_{n=1}^{F} \left| I\left(\boldsymbol{x} + \boldsymbol{u}(\boldsymbol{x};n)\ ;\ n\right) - I(\boldsymbol{x};n_0)\right| \mathrm{d}\boldsymbol{x} \\ +\beta \int_\Omega \sum_{n=1}^{F} &\Big\| \boldsymbol{u}(\boldsymbol{x};n) - \sum_{i=1}^{R} \boldsymbol{q}_i(n) L_i(\boldsymbol{x}) \Big\|^2 \mathrm{d}\boldsymbol{x} \;+\; \int_\Omega \sum_{i=1}^{R} \|\nabla L_i(\boldsymbol{x})\|\ \mathrm{d}\boldsymbol{x} \end{aligned} \tag{4}$$

jointly with respect to the point trajectories $\boldsymbol{u}(\boldsymbol{x};n)$ and their components on the trajectory subspace that are determined by the linear model coefficients $\boldsymbol{L}(\boldsymbol{x})$. The positive constants α and β weigh the balance between the terms of the energy. Note that the functions $\boldsymbol{u}(\boldsymbol{x};n)$ and $\boldsymbol{L}(\boldsymbol{x})$ determine two sets of trajectories that are relatively close to each other but not exactly the same since the subspace constraint is imposed as a soft constraint (i.e. the model error $\boldsymbol{\varepsilon}$ in equation (1) is not zero). Since we regard the linear trajectory model as an approximation, we consider that the final output of our method are the trajectories $\boldsymbol{u}(\boldsymbol{x};n)$.

The **first term** in the above energy is a data attachment term that uses the robust $\mathcal{L}_1$-norm and is a direct multi-frame extension of the brightness constancy term used by most optical flow methods, e.g. [12]. It is based on the assumption that the image brightness $I(\boldsymbol{x};n_0)$ at every pixel $\boldsymbol{x}$ of the reference frame is preserved at its new location, $\boldsymbol{x}+\boldsymbol{u}(\boldsymbol{x};n)$, in every frame of the sequence. The use of an $\mathcal{L}_1$-norm improves the robustness of the method since it accounts for deviations from this assumption, which might occur in real-world scenarios because of occlusions of some points in some frames. The **second term** of the energy (4) penalizes trajectories $\boldsymbol{u}(\boldsymbol{x};n)$ that do not lie on the trajectory subspace $Q\boldsymbol{L}(\boldsymbol{x})$. In fact, this term corresponds to the energy of the trajectory model error $\boldsymbol{\varepsilon}$ (c.f. equation (1)) and serves as a soft constraint that the trajectories $\boldsymbol{u}(\boldsymbol{x};n)$ should be relatively close to the subspace spanned by the basis Q. Concerning the weight β, the larger its value the more restrictive the subspace constraint becomes. We normally use a relatively high value for this weight. Since the subspace of Q is low-dimensional, this constraint operates also as a temporal regularization that is able to perform temporal filling-in in cases of occlusions or other distortions. Note that, unlike [12], we do not need to introduce an auxiliary variable since this quadratic term allows us to decouple the data term and the regularizer directly. The **third term** of (4) corresponds to Total Variation - based spatial regularization of the trajectory model coefficients. This term penalizes spatial oscillations of each coefficient caused by image noise or other distortions but not strong discontinuities that are desirable in the borders of each object. In addition, this term allows to fill in textural information into flat regions from their neighborhoods.

Our approach is related to the recent work of Garg et al. [6] in which dense multi-frame optical flow for non-rigid motion is computed imposing hard subspace constraints. Our approach departs in a number of ways. First, while [6] imposes the subspace constraint via re-parameterization of the optical flow, we use a soft constraint and do not optimize directly on the low-rank manifold but impose that the flow should lie close to it. Secondly, the use of the $\mathcal{L}_1$-norm for the data term and a Total Variation regularizer instead of the non-robust $\mathcal{L}_2$-norm and quadratic regularizer used by [6] allow us to deal with occlusions and appearance changes and to preserve object boundaries. Finally, by providing a generalization of the subspace constraint, we have extended the approach to deal with any orthogonal basis and not just the PCA basis [6].

5 Optimization of the Proposed Energy

As we described in the previous section, the energy in (4) is related to the TV-L1 formulation of the optical flow problem described in [12], therefore we follow a similar alternating approach to solve the optimization problem. We decouple the data and regularization terms to decompose the optimization problem into two, each of which can be solved in a globally optimal manner. The key difference is that we do not solve for pairwise optical flow but instead we optimize over all the frames of the sequence while imposing the trajectory subspace constraint as a soft constraint. In this section we show how to adapt the method of [12] to our problem, to take advantage of its computational efficiency and apply it to multi-frame subspace-constrained optical flow. Assuming an initialization $\boldsymbol{u}_0(\boldsymbol{x};n)$ is available for $\boldsymbol{u}(\boldsymbol{x};n)$, we apply an alternating optimization, updating either $\boldsymbol{u}(\boldsymbol{x};n)$ or $\boldsymbol{L}(\boldsymbol{x})$ in every iteration, as follows:

– Repeat until convergence:

Step 1. For $\boldsymbol{u}(\boldsymbol{x};n)$ fixed, update $\boldsymbol{L}(\boldsymbol{x})$ by minimizing $E\big[\boldsymbol{u}(\boldsymbol{x};n)\,,\,\boldsymbol{L}(\boldsymbol{x})\big]$ wrt $\boldsymbol{L}(\boldsymbol{x})$.
Step 2. For $\boldsymbol{L}(\boldsymbol{x})$ fixed, update $\boldsymbol{u}(\boldsymbol{x};n)$ by minimizing $E\big[\boldsymbol{u}(\boldsymbol{x};n)\,,\,\boldsymbol{L}(\boldsymbol{x})\big]$ wrt $\boldsymbol{u}(\boldsymbol{x};n)$.

Convergence is declared if the relative update of $\boldsymbol{L}(\boldsymbol{x})$ and $\boldsymbol{u}(\boldsymbol{x};n)$ is negligible according to some appropriate distance threshold.

5.1 Minimization of Step 1

Since in this step we keep $\boldsymbol{u}(\boldsymbol{x};n)$ fixed, we can observe that only the last two terms of the energy (4) depend on $\boldsymbol{L}(\boldsymbol{x})$. Regarding the second term, using the matrix notation defined in (2), we can write this penalty term as:

$$\sum_{n=1}^{F}\Big\|\boldsymbol{u}(\boldsymbol{x};n)-\sum_{i=1}^{R}\boldsymbol{q}_i(n)L_i(\boldsymbol{x})\Big\|^2=\|\boldsymbol{\mathcal{E}}(\boldsymbol{x})\|^2=\|\boldsymbol{\mathcal{U}}(\boldsymbol{x})-\mathrm{Q}\,\boldsymbol{L}(\boldsymbol{x})\|^2 \qquad (5)$$

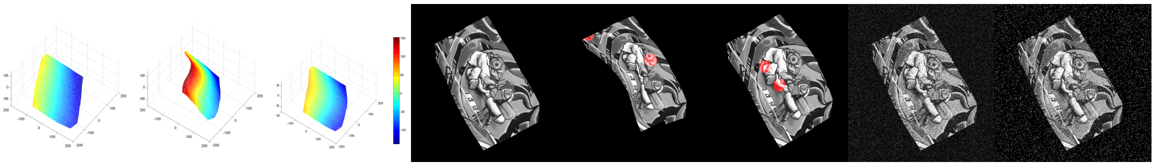
Let $\mathrm{Q}^\perp$ be an $2F\times(2F-R)$ matrix whose columns form an orthonormal basis of the orthogonal complement of the trajectory subspace $\mathcal{S}_Q$. Then the block matrix $[\mathrm{Q}\ \mathrm{Q}^\perp]$ is an orthonormal $2F\times 2F$ matrix, which means that its columns form a basis of $\mathbb{R}^{2F}$. Consequently, $\boldsymbol{\mathcal{U}}(\boldsymbol{x})$ can be decomposed into two orthogonal vectors as $\boldsymbol{\mathcal{U}}(\boldsymbol{x})=\mathrm{Q}\boldsymbol{\mathcal{U}}_{in}(\boldsymbol{x})+\mathrm{Q}^\perp\boldsymbol{\mathcal{U}}_{out}(\boldsymbol{x})$ where $\boldsymbol{\mathcal{U}}_{in}(\boldsymbol{x})\triangleq\mathrm{Q}^T\boldsymbol{\mathcal{U}}(\boldsymbol{x})$ and $\boldsymbol{\mathcal{U}}_{out}(\boldsymbol{x})\triangleq(\mathrm{Q}^\perp)^T\boldsymbol{\mathcal{U}}(\boldsymbol{x})$ are the coefficients that define the projections of $\boldsymbol{\mathcal{U}}(\boldsymbol{x})$ onto the trajectory subspace $\mathcal{S}_Q$ and its orthogonal complement. Equation (5) can now be further simplified:

$$\|\boldsymbol{\mathcal{E}}(\boldsymbol{x})\|^2=\big\|\mathrm{Q}^\perp\boldsymbol{\mathcal{U}}_{out}(\boldsymbol{x})+\mathrm{Q}\,(\boldsymbol{\mathcal{U}}_{in}(\boldsymbol{x})-\boldsymbol{L}(\boldsymbol{x}))\big\|^2=\|\boldsymbol{\mathcal{U}}_{out}(\boldsymbol{x})\|^2+\|\boldsymbol{\mathcal{U}}_{in}(\boldsymbol{x})-\boldsymbol{L}(\boldsymbol{x})\|^2\ ,$$

due to the orthonormality of the columns of Q and $\mathrm{Q}^\perp$ (which makes the corresponding transforms isometric) and Pythagoras' theorem. The component $\|\boldsymbol{\mathcal{U}}_{out}(\boldsymbol{x})\|^2$ is constant with respect to $\boldsymbol{L}(\boldsymbol{x})$; therefore it can be neglected from the current minimization. In other words, with $\boldsymbol{\mathcal{U}}$ being fixed and $\mathrm{Q}\,\boldsymbol{L}$ lying on the linear subspace $\mathcal{S}_Q$, penalizing the distance between $\mathrm{Q}\,\boldsymbol{L}$ and $\boldsymbol{\mathcal{U}}$ is equivalent to penalizing the distance between $\mathrm{Q}\,\boldsymbol{L}$ and the projection of $\boldsymbol{\mathcal{U}}$ onto $\mathcal{S}_Q$. To conclude, the minimization of step 1 is equivalent to the minimization of:

$$\beta\int_\Omega\|\boldsymbol{\mathcal{U}}_{in}(\boldsymbol{x})-\boldsymbol{L}(\boldsymbol{x})\|^2+\int_\Omega\sum_{i=1}^{R}\|\nabla L_i(\boldsymbol{x})\|=\sum_{i=1}^{R}\int_\Omega\big\{\|\nabla L_i(\boldsymbol{x})\|+\beta\big(\mathcal{U}_{in}^{(i)}(\boldsymbol{x})-L_i(\boldsymbol{x})\big)^2\big\}$$

where $\mathcal{U}_{in}^{(i)}(\boldsymbol{x})$ is the i-th coordinate of $\boldsymbol{\mathcal{U}}_{in}(\boldsymbol{x})$. We have finally obtained a new form of the energy that offers a decoupling between the trajectory model coefficients $L_i(\boldsymbol{x})$. The



(a) $\mathcal{S}_1$ (b) $\mathcal{S}_{30}$ (c) $\mathcal{S}_{60}$ (d) I_1 (e) I_{30} (f) I_{60} (g) gaus. noise (h) salt-pep. noise

Fig. 1. Rendering process for ground truth optical flow sequence of a non-rigid object. **(a-c)**: dense surfaces $\mathcal{S}_n$, constructed using thin plate spline interpolation of sparse MOCAP data [25]. **(d-f)**: rendered image sequence I_n using texture mapping of a graffiti image. Superimposed red circles indicate regions where intensities have been replaced by black to simulate synthetic occlusions. **(g-h)**: sample reference images for (g) Gaussian and (h) Salt and paper noise.

minimization of each term in the above sum can be done independently and corresponds to the Total Variation - based denoising model of Rudin,Osher and Fatemi (ROF) [24] applied to each coefficient $L_i(\boldsymbol{x})$. The optimum $L_i(\boldsymbol{x})$ is actually a regularized version of $\mathcal{U}_{in}^{(i)}(\boldsymbol{x})$ and the extent of this regularization increases as the weight β decreases.

The benefits in the computational efficiency of the above procedure are twofold. First, these independent minimizations can be parallelized. Second, there exist several efficient algorithms for implementing the ROF model. We have used the method of [7], which uses a dual formulation of the minimization and proposes a globally convergent scheme (c.f. [7] for details). Note that this method has been also used by [12] for the problem of optical flow, but under its classical formulation of finding the frame-by-frame displacements.

5.2 Minimization of Step 2

Keeping $\boldsymbol{L}(\boldsymbol{x})$ fixed, we observe that only the first two terms of the energy (4) depend on $\boldsymbol{u}(\boldsymbol{x};n)$ and furthermore these terms can be written in the following way:

$$\int_\Omega \sum_{n=1}^{F} \left\{ \alpha \left| I\left(\boldsymbol{x}+\boldsymbol{u}(\boldsymbol{x};n)\ ;\ n\right) - I(\boldsymbol{x};n_0)\right| + \beta \left\|\boldsymbol{u}(\boldsymbol{x};n)-\boldsymbol{u}'\right\|^2 \right\} \mathrm{d}\boldsymbol{x}\ , \qquad (6)$$

where $\boldsymbol{u}' = \sum_{i=1}^{R} \boldsymbol{q}_i(n) L_i(\boldsymbol{x})$. This quantity depends only on the value of $\boldsymbol{u}$ on the specific point $\boldsymbol{x}$ and the discrete time n (and not on the derivatives of $\boldsymbol{u}$). Therefore the variational minimization of Step 2 boils down to the minimization of a bivariate function of the value of $\boldsymbol{u}$ for every spatiotemporal point $(\boldsymbol{x};n)$ independently.

We implement this pointwise minimization by applying the technique proposed in [12] to every frame. More precisely, for every frame n and point $\boldsymbol{x}$ the image $I(\cdot;n)$ is linearized around $\boldsymbol{x}+\boldsymbol{u}_0(\boldsymbol{x};n)$, where $\boldsymbol{u}_0(\boldsymbol{x};n)$ are the initializations of the trajectories $\boldsymbol{u}(\boldsymbol{x};n)$. The function to be minimized at every point will then have the simple form of a summation of a quadratic term with the absolute value of a linear term. The minimum can be easily found analytically using the thresholding scheme reported in [12].

5.3 Implementation Details

The above image linearizations are effective only if the initialization $\boldsymbol{u}_0(\boldsymbol{x};n)$ is relatively close to the actual solution $\boldsymbol{u}(\boldsymbol{x};n)$. To ensure the linearisation assumptions hold

	RMS endpoint error (pix)				99^{th} percentile of endpoint error (pix)			
Version of input:	Original	Occlusions	Gaus.noise	S&P noise	Original	Occlusions	Gaus.noise	S&P noise
Ours, PCA basis	**0.98**	1.33	2.28	1.84	**3.08**	**4.92**	8.33	**7.09**
Ours, DCT basis	1.06	1.72	2.78	2.29	6.70	5.18	**7.92**	8.53
Pizarro et al. [1]	1.24	**1.27**	**1.94**	**1.79**	4.88	5.05	8.67	8.54
ITV-L1 [26]	1.43	1.89	2.61	2.34	6.28	9.44	9.70	9.98
LDOF [10]	1.71	2.01	4.35	5.05	3.72	6.63	18.15	20.35

Table 1. Measures of endpoint errors for different methods on the benchmark sequences.

in the case of large optic flow we use coarse-to-fine techniques with multiple warping iterations.

We used a similar numerical optimisation scheme and preprocessing of images to the one proposed in [26] to minimise the energy (4), i.e. we use the structure-texture decomposition to make our input robust to illumination artifacts due to shadows and shading reflections. We also used blended versions of the image gradients and a median filter to reject flow outliers. Concerning the choice of the parameters of the algorithm, we used the same values for both ITV-L1 [26] and our method, i.e. 5 warp iterations, 20 alternation iterations and the weights α and β were set to 30 and 2.

6 Experimental results

In this section we evaluate our method and compare its performance with state of the art optical flow [10, 12] and image registration [1] algorithms. We show quantitative comparative results on our new benchmark ground truth optical flow dataset and qualitative results on real-world sequences. Furthermore, we analyse the sensitivity of our algorithm to some of its parameters, such as the choice of trajectory basis and regularization weight. Since our algorithm computes multi-frame optical flow and incorporates an implicit temporal regularization term, it would have been natural to compare its performance with a spatiotemporal optical flow formulation [2]. However, due to the lack of publicly available implementations we chose to compare with LDOF (Large Displacement Optical Flow) [10], one of the best performing current optical flow algorithms, that can deal with large displacements by integrating rich feature descriptors into a variational optic flow approach to compute dense flow. We also compare with the duality based TV-L1 algorithm [12] since our method can be seen as its extension to the case of multi-frame non-rigid optical flow via robust trajectory subspace constraints. To be more exact, we compare with the *Improved TV-L1* (ITV-L1) algorithm [26] since we use a similar numerical optimization scheme and preprocessing steps (see Section 5.3). In both cases, we register each frame in the sequence independently with the reference frame. We also compare with Pizarro and Bartoli's state of the art keypoint-based non-rigid registration algorithm [1]. Additionally, we show comparative results with Garg et al. [6] which support our claim that imposing the subspace constraint as a soft instead of a hard constraint results in improved performance and higher resilience to noise[1].

[1] Videos of the results as well as our benchmark dataset can be found on the following URL: `http//www.eecs.qmul.ac.uk/~lourdes/subspace_flow`

6.1 Construction of a ground truth benchmark dataset

For the purpose of quantitative evaluation of multi-frame non-rigid optical flow and to promote progress in this area we generated a benchmark sequence with ground truth. To the best of our knowledge, this is one of the first attempts to generate a long image sequence of a deformable object with dense ground truth 2D trajectories. We use sparse motion capture (MOCAP) data from [25] to capture the real deformations of a waving flag in 3D. We interpolated this sparse data to have a continuous dense 3D surface using the motion capture markers as the control points for smooth Spline interpolation. This dense 3D surface is then projected synthetically onto the image plane using an orthographic camera. We use texture mapping to associate some texture to the surface while rendering 60 images of size 500x500 pixels. The advantage of this new sequence is that, since it is based on MOCAP data, it captures the complex natural deformations of a real non-rigid object while allowing us to have access to dense ground truth optical flow. We have also used three degraded versions of the original rendered sequence by adding (a) gaussian noise, of standard deviation 0.2 relative to the range of image intensities, (b) salt & peper noise of density 10% and (c) synthetic occlusions generated by superimposing some black circles of radius 20 pixels moving in linear orbits. Figure 1 shows the interpolated 3D flag surface and some of the frames of the 60 frame long sequence.

6.2 Quantitative Results on Benchmark Sequence

We tested our algorithm using the three different proposed motion basis: PCA, DCT and Cubic B-Spline. Similarly to Garg et al. [6] we compute the PCA basis from sparse tracked features. For the experiments on the benchmark sequence we used the tracks provided by the feature matching algorithm of Pizarro and Bartoli [1] where a robust method based on local surface smoothness is used to discard outliers from an initial set of SIFT feature matches. Temporal cubic spline interpolation is then used to fill in the missing data in each track independently for the computation of the PCA basis.

In Table 1, the error measures of various methods are compared using the different versions of the rendered flag sequence as inputs. Note that the results obtained with the Spline basis were omitted since they were almost equivalent those obtained with the DCT basis, as Figure 3(a) reveals. We observe that our proposed method yields the best RMS measure in the case of the original sequence and outperforms ITV-L1 and LDOF methods in all other cases. Also, in the case of data with synthetic noise it performs comparably to the best performing method, Pizarro et al. [1]. In the case of external occlusions, the method of Pizarro et al. [1] yields the best RMS error. As far as the percentile measures are concerned, the best measures are in all the cases yielded by the two versions of the proposed method. Furthermore, we can observe from the error maps of Figs. 2 and 3 that in the case of self-occlusions the situation seems reversed and our proposed method yields a more accurate result.

Figure 2 shows a comparison of the results on the Flag sequence of our algorithm using a PCA basis of rank $R = 75$ and a full rank DCT basis $R = 120$; ITV-L1 optical flow [26]; LDOF [10] and Pizarro and Bartoli's registration algorithm [1]. We show a closeup of the reverse warped images of 3 frames in the sequence (20 30 60) which should be identical to the template frame; and the error in flow estimation, expressed in pixels, encoded as a heatmap. Our method, both using PCA and DCT, gives lower

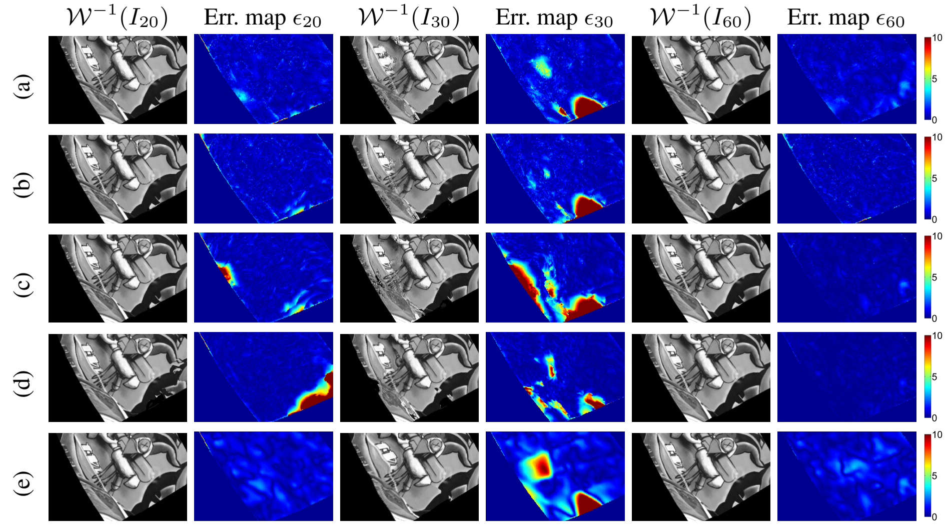


Fig. 2. Inverse warps $\mathcal{W}^{-1}(I_n)$ and error maps $\epsilon_n(\boldsymbol{x})$ for some frames of the flag sequence. **(a)** Proposed method: PCA basis, **(b)** DCT basis. **(c)** ITV-L1 [26]. **(d)** LDOF [10].**(e)** Pizarro et al.[1].

errors on these frames than the state of the art methods we compare with. Figure 3(c-g) shows a similar comparison in the presence of synthetic occlusions and it is evident that our method and [1] perform much better than others in occluded regions as they model the flow of a non rigid surface in the reference template.

Figure 3(a) shows a graph of the *root mean square* (RMS) error (measured in pixels) over all the frames of the optical flow estimated using the 3 different bases for different values of the rank and of the weight β associated with the soft constraint. For a reasonably large value of β all the basis can be used with a significant reduction in the rank. The optimization also appears not to overfit when the dimensionality of the subspace is overly high. Figure 3(b) explores the effect of varying the value of the weight β on the accuracy of the optical flow. While low values of β cause numerical instability (data and regularization terms become completely decoupled) high values of β, on the other hand, lead to slow convergence and errors since the point-wise search is not allowed to leave the manifold, simulating a hard constraint. Another interesting observation is that our proposed method with a PCA basis of rank R=50, yields a better performance than with a full rank PCA basis R=120. This reflects the fact that the temporal regularization due to the low dimensional subspace is often beneficial. Note that to analyze the sensitivity of our algorithm to its parameters in Figure 3(a-b) we used ground truth tracks to compute the PCA basis to remove the bias from tracking.

6.3 Experiments on Real Sequences

Figure 4 presents comparative results of optical flow methods in two real sequences of textured paper bending smoothly. The **first input sequence** is particularly challenging because of its length (100 frames) and large rotation of the camera. The trajectory basis of the proposed method is derived by applying PCA on KLT tracks [27] and keeping

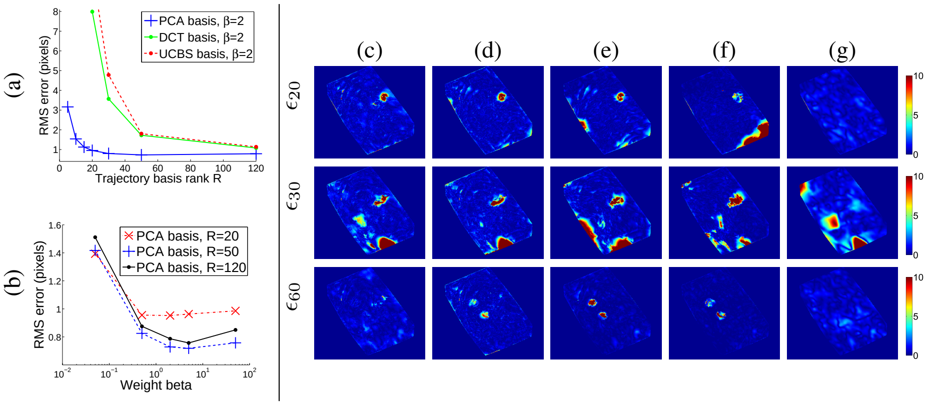
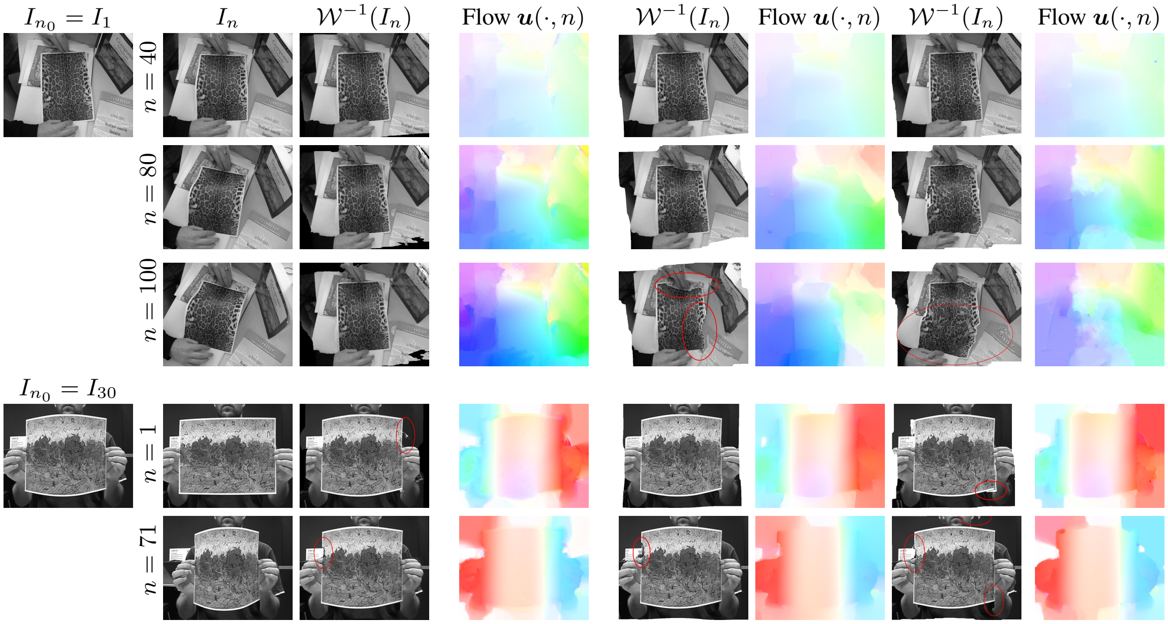


Fig. 3. (a-b) RMS flow error for proposed method in flag sequence varying basic parameters. **(c-f)** Flow error maps $\epsilon_n(\boldsymbol{x})$ for flag sequence with *synthetic occlusions*: **(c)** Proposed method: PCA basis, **(d)** DCT basis. **(e)** ITV-L1 [26]. **(f)** LDOF [10]. **(g)** Pizarro et al. [1].



(a) (b) (c) Proposed method, PCA basis (d) ITV-L1 [26] (e) LDOF [10], direct registr.

Fig. 4. Multi-frame optical flow for different methods, on 2 paper bending sequences (with 100 and 71 frames respectively). **(a)** Reference frames and **(b)** frames from the input sequences. **(c-e)** Flow-based inverse warps $\mathcal{W}^{-1}(I_n)$ in the reference frames and color-coded flow fields $\boldsymbol{u}(\cdot, n)$.

only the first 10 components. Note that our method achieves similar results with the DCT basis of rank R=14. We run the LDOF and ITV-L1 algorithms using a multi-resolution scaling factor of 0.95, whereas for our algorithm the value 0.75 was sufficient (pointing to faster convergence). Comparing the warped images $\mathcal{W}^{-1}(I_n)$, we observe that our method yields a significant improvement on the accuracy of the optical flow,

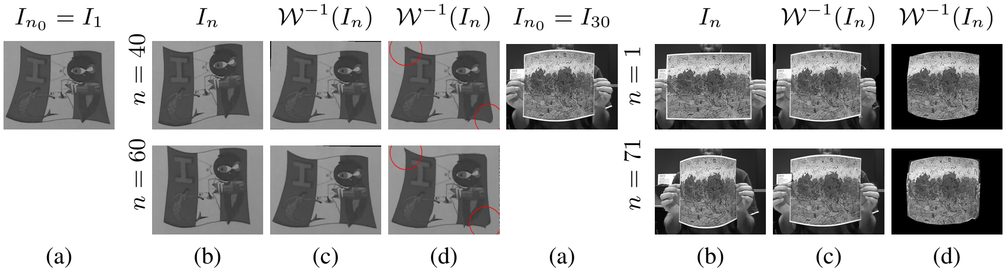


Fig. 5. Multi-frame optical flow results on a T-shirt and a paper-bending sequences. **(a)** Reference frames and **(b)** representative frames of the input sequences. **(c-d)** Inverse warps $\mathcal{W}^{-1}(I_n)$ for different methods: **(c)** Proposed method, PCA basis. **(d)** Garg et al. [6].

especially after some frames (see e.g. the artifacts annotated by the red ellipses in the results of LDOF and ITV-L1). The **second input sequence** in Fig. 4 is widely used in the structure from motion literature and contains 71 frames. We ran our method using a PCA basis on KLT tracks and choosing rank R=6. In this sequence we used the 30th frame as the reference. We observe that our method yields an accurate result and suffers from less artifacts than others.

In Fig.5, we show results on 2 input sequences to compare our new approach against Garg et al. [6]. The first sequence captures a T-shirt deforming as it is stretched from the bottom two corners and contains 60 frames. The second sequence is the same as in Fig. 4 (bottom). For the method of [6], we tested different values for the basis rank R and we kept the best value for each sequence, which turned out to be R=3 for the T-shirt and R=8 for the paper bending sequence. For our method, the choice of rank is less crucial and we selected R=8 for the T-shirt and R=6 for the paper bending sequence. We observe that both methods output a plausible result for the T-shirt sequence. However, [6] cannot reliably estimate the optical flow in the corners that are marked with red circles, whereas our proposed method can. On the paper bending sequence, we observe that our method performs significantly better than [6]. We believe that these improvements can be attributed to our use of robust soft subspace constraints and robust Total Variation and L1 data terms.

7 Conclusions

We have provided a new formulation for the computation of optical flow of a non-rigid surface exploiting the high correlation in a long sequence between 2D trajectories of points by assuming that these lie close to a low dimensional subspace. Our contribution is to formulate the manifold constraint as a *soft constraint* which, using variational principles, leads to a *robust* energy with a quadratic relaxation term that allows its efficient optimization. We also provide a new benchmark dataset, with ground truth optical flow. Our proposed approach improves or has equivalent performance to state of the art optical flow algorithms and a non-rigid registration approach.

References

1. Pizarro, D., Bartoli, A.: Feature-based deformable surface detection with self-occlusion reasoning. In: International Symposium on 3D Data Processing, Visualization and Trans-

mission, 3DPVT'10. (2010)

2. Weickert, J., Schnörr, C.: Variational optic flow computation with a spatio-temporal smoothness constraint. JMIV **14** (2001) 245–255
3. Irani, M.: Multi-frame correspondence estimation using subspace constraints. IJCV (2002)
4. Torresani, L., Yang, D., Alexander, E., Bregler, C.: Tracking and modeling non-rigid objects with rank constraints. In: CVPR. (2001)
5. Torresani, L., Bregler, C.: Space-time tracking. In: ECCV. (2002)
6. Garg, R., Pizarro, L., Rueckert, D., Agapito, L.: Dense multi-frame optic flow for non-rigid objects using subspace constraints. In: ACCV. (2010)
7. Chambolle, A.: An algorithm for total variation minimization and applications. JMIV **20** (2004) 89–97
8. Chambolle, A., Pock, T.: A first-order primal-dual algorithm for convex problems with applications to imaging. JMIV (2011)
9. Horn, B., Schunck, B.: Determining optical flow. Artificial Intelligence **17** (1981) 185–203
10. Brox, T., Malik, J.: Large displacement optical flow: Descriptor matching in variational motion estimation. TPAMI (2010)
11. Brox, T., Bruhn, A., Papenberg, N., Weickert, J.: High accuracy optical flow estimation based on a theory for warping. In: ECCV. (2004)
12. Zach, C., Pock, T., Bischof, H.: A duality based approach for realtime tv-l1 optical flow. In: Pattern Recognition (Proc. DAGM). (2007) 214–223
13. Wedel, A., Pock, T., Braun, J., Franke, U., Cremers, D.: Duality tv-l1 flow with fundamental matrix prior. In: Image and Vision Computing New Zealand. (2008)
14. Wedel, A., Cremers, D., Pock, T., Bischof, H.: Structure- and motion-adaptive regularization for high accuracy optic flow. In: ICCV. (2009)
15. Alvarez, L., Weickert, J., Sánchez, J.: Reliable estimation of dense optical flow fields with large displacements. IJCV **39** (2000) 41–56
16. Steinbruecker, F., Pock, T., Cremers, D.: Large displacement optical flow computation without warping. In: ICCV. (2009)
17. Pock, T., Cremers, D., Bischof, H., Chambolle, A.: Global solutions of variational models with convex regularization. SIAM Journal on Imaging Sciences (2010)
18. Stuehmer, J., Gumhold, S., Cremers, D.: Real-time dense geometry from a handheld camera. In: Pattern Recognition (Proc. DAGM). (2010) 11–20
19. Tian, Y., Narasimhan, S.: A globally optimal data-driven approach for image distortion estimation. In: CVPR. (2010)
20. Baker, S., Scharstein, D., Lewis, J., Roth, S., Black, M., Szeliski, R.: A database and evaluation methodology for optical flow. IJCV **92** (2011) 1–31
21. Bregler, C., Hertzmann, A., Biermann, H.: Recovering non-rigid 3D shape from image streams. In: CVPR. (2000)
22. Torresani, L., Hertzmann, A., Bregler., C.: Non-rigid structure-from-motion: Estimating shape and motion with hierarchical priors. PAMI **30** (2008)
23. Shi, J., Tomasi, C.: Good features to track. CVPR (1994)
24. Rudin, L., Osher, S., Fatemi, E.: Nonlinear total variation based noise removal algorithms. Physica D **60** (1992) 259–268
25. White, R., Crane, K., Forsyth, D.: Capturing and animating occluded cloth. In: ACM Trans. on Graphics. (2007)
26. Wedel, A., Pock, T., Zach, C., Bischof, H., Cremers, D.: An improved algorithm for tv-l1 optical flow. In: Statistical and Geometrical Approaches to Visual Motion Analysis. LNCS. (2009) 23–45
27. Lucas, B., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: IJCAI81. (1981)