

A Variational Approach to Video Registration with Subspace Constraints

Ravi Garg · Anastasios Roussos · Lourdes Agapito

Received: 9 December 2011 / Accepted: 20 December 2012 / Published online: 2 April 2013
© The Author(s) 2013. This article is published with open access at Springerlink.com

Abstract This paper addresses the problem of non-rigid video registration, or the computation of optical flow from a reference frame to each of the subsequent images in a sequence, when the camera views deformable objects. We exploit the high correlation between 2D trajectories of different points on the same non-rigid surface by assuming that the displacement of any point throughout the sequence can be expressed in a compact way as a linear combination of a low-rank motion basis. This subspace constraint effectively acts as a trajectory regularization term 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 can be decoupled into the optimization of the brightness constancy and spatial regularization terms, leading to an efficient optimization scheme. Additionally, we propose a novel optimization scheme for the case of vector valued images, based on the dualization of the data term. This allows us to extend our approach to deal with colour images which results in significant improvements on the registration results. Finally, 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-frame optical flow algorithms for non-rigid surfaces. Our experiments show that our proposed approach outperforms 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. It has 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 consider long video sequences instead of a single pair of frames—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 (see Fig. 1).

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 ambiguity. In this paper we show that a multi-frame approach allows us to exploit temporal information, resolving these ambiguities and improving the overall quality of the optical flow. We make use of the strong 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 trajectory vector storing 2D positions of a point across time 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. Figure 2 depicts the lower dimensional trajectory subspace.

Subspace constraints have been used before both in the context of sparse point tracking (Irani 2002; Brand 2001;

R. Garg · A. Roussos · L. Agapito (✉)
Queen Mary University of London, Mile End Road,
London E1 4NS, UK
e-mail: lourdes@dcs.qmul.ac.uk

R. Garg
e-mail: rgarg@eecs.qmul.ac.uk

A. Roussos
e-mail: troussos@eecs.qmul.ac.uk

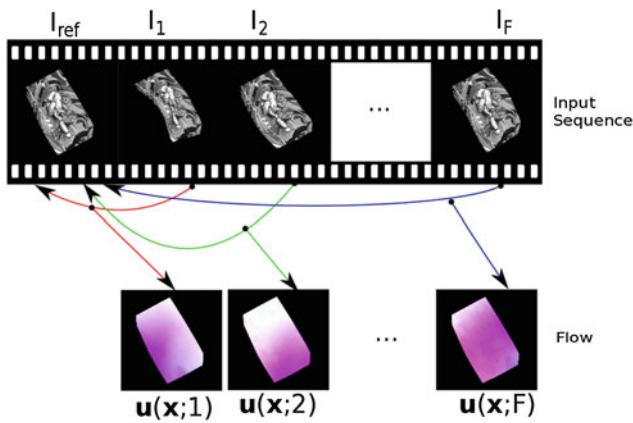


Fig. 1 Video registration is equivalent to the problem of estimating dense optical flow $u(x; n)$ between a reference frame I_{ref} and each of the subsequent frames I_n in a sequence. We propose a multi-frame optical flow algorithm that exploits temporal consistency by imposing subspace constraints on the 2D image trajectories

Torresani et al. 2001; Torresani and Bregler 2002) and optical flow (Irani 2002; Garg et al. 2010) in the rigid and non-rigid domains, to allow correspondences to be obtained in low textured areas. While Irani’s original rigid (Irani 2002) formulation along with its non-rigid extensions (Torresani et al. 2001; Brand 2001; Torresani and Bregler 2002) relied on minimizing the linearized brightness constraint without smoothness priors, Garg et al. (2010) extended the subspace constraints to the continuous domain in the non-rigid case using a variational approach. Nir et al. (2008) propose a variational approach to optical flow estimation based on a spatio-temporal model. However, all of the above approaches impose the subspace constraint as a hard constraint. Hard constraints are vulnerable to noise in the data 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 (Chambolle 2004; Chambolle and Pock 2011) in variational methods and optimize the energy defining a variant of the duality-based efficient numerical optimization scheme. We are also able to prove that our soft constraint is preferable to a hard constraint imposed via reparameterization. To do this we provide a formulation of the hard constraint and its optimization and we perform thorough experimental comparisons where we show that the results obtained via the soft constraint always outperform those obtained after reparameterization.

The paper is organized as follows. In Sect. 2 we describe related approaches and discuss the contributions of our work. Section 3 defines the trajectory subspace constraints that we use in our formulation. In Sect. 4 we describe the energy and provide a discussion on the design of our effective trajectory regularizer. Section 5 addresses the optimization of our proposed energy. This is followed by a description of the estimation of the motion basis in Sect. 6. In Sect. 7 we propose the extension of our algorithm to vector-valued images and Sect. 8 discusses implementation details. Finally Sect. 9 describes the alternative formulation of the subspace constraint as a hard constraint while Sect. 10 describes our experimental evaluation.

2 Related Work and Contribution

Variational methods formulate the optical flow or image alignment problems as the optimization of an energy functional in a continuous domain. Stemming from Horn and Schunck’s original approach (Horn and Schunck 1981), the energy incorporates a data term that accounts for the brightness constancy assumption 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

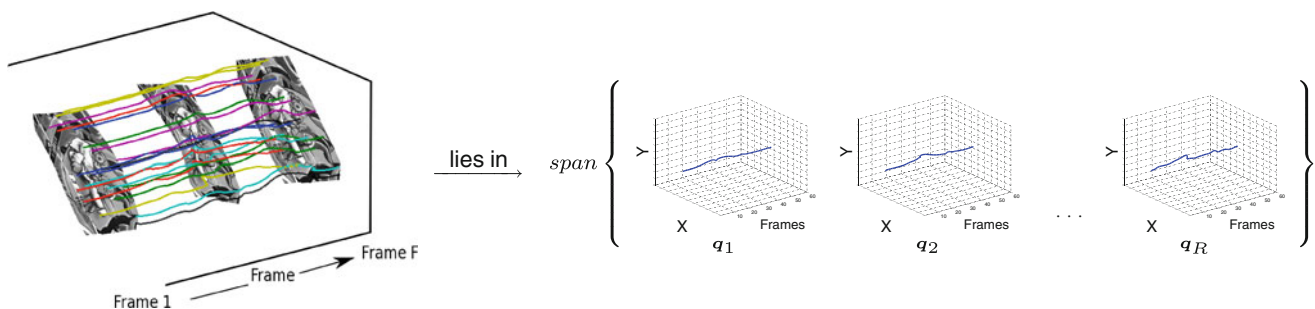


Fig. 2 The strong correlation between 2D trajectories of different points on the same non-rigid surface can be exploited to impose temporal coherence by modelling long term temporal coherence imposing

subspace constraints. These trajectories lie on a lower dimensional manifold which leads to a significant reduction in the dimensionality of the problem while implicitly imposing some form of temporal smoothness

or occlusions (Brox and Malik 2011; Brox et al. 2004); the addition of efficient regularization terms such as Total Variation (TV) (Zach et al. 2007; Wedel et al. 2008) or temporal smoothing terms (Weickert and Schnörr 2001b); and new optimization strategies that allow computation of highly accurate (Wedel et al. 2009) and real time optical flow (Zach et al. 2007) even in the presence of large displacements (Alvarez et al. 2000; Brox and Malik 2011; Steinbruecker et al. 2009).

One important recent advance in variational optical flow methods has been the development of the duality based efficient optimization of the so-called TV- L^1 formulation (Zach et al. 2007; Chambolle and Pock 2011) (which owes its name to the Total Variation that is used for regularization and the robust L^1 -norm that is used in the data fidelity term). An example of this class is the Improved TV- L^1 (ITV- L^1) method (Wedel et al. 2009), which yielded notable quantitative performance, by also carefully considering some practical aspects of the optical flow algorithm. Duplication of the optimization variable via a quadratic relaxation is used to decouple the linearized 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* independent, its optimization can be highly parallelized using graphics hardware (Zach et al. 2007). Following its success in optical flow computation, this optimization scheme has since been successfully applied to motion and disparity estimation (Pock et al. 2010) and real time dense 3D reconstruction (Newcombe et al. 2011; Stuehmer et al. 2010). In this work we adopt this efficient duality based TV- L^1 optimization scheme (Zach et al. 2007) and extend it to the case of multi-frame optical flow for video registration, by modelling long term temporal coherence imposing subspace constraints.

Despite being such a powerful cue most optical flow algorithms do not take advantage of temporal coherence and only work on pairs of images. Few previous attempts to multi-frame optical flow estimation exist in the literature (Weickert and Schnörr 2001b,a; Papadakis et al. 2007; Nir et al. 2008; Werlberger et al. 2009; Volz et al. 2011). Even in those cases, temporal smoothness constraints are only exploited over a very small number of frames (typically 1 or 2 frames either side of the current image) and not for an entire sequence. This is mostly due to the difficulty of providing an explicit model for longer term trajectories. In recent work Volz et al. (2011) report improvements in optical flow computation by imposing first and second order trajectory smoothness over 5 frames. We take this further and exploit temporal coherence through the entire video. Moreover,

while previous approaches incorporate explicit temporal smoothness regularization terms over a few frames, our subspace constraint acts as an implicit long term trajectory regularization term leading to temporally consistent optical flow.

Our approach is related to the recent work of Garg et al. (2010) in which dense multi-frame optical flow for non-rigid motion is computed under hard subspace constraints. Our approach departs in a number of ways. First, while Garg et al. (2010) imposes the subspace constraint via reparameterization of the optical flow, we use a soft constraint and optimize over two sets of closely coupled flows, one that lies on the low-rank manifold and one that does not. Secondly, our use of a robust penalizer for the data term allows us to have more resilience than Garg et al. (2010) against occlusions and appearance changes. Moreover, our use of a modified Total Variation regularizer instead of the non-robust L^2 -norm and quadratic regularizer used by Garg et al. (2010) allows to preserve object boundaries. Finally, by providing a generalization of the subspace constraint, we have extended the approach to deal with any orthonormal basis and not just the PCA basis. More recently Ricco and Tomasi (2012) also proposed the use of subspace constraints to model multi-frame optical flow with explicit reasoning for occlusions. However, their approach is restricted to hard subspace constraints with a known PCA basis which is computed from sparse feature tracking.

Non-rigid image registration, has recently seen substantial 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 parametric¹ approach of Pizarro and Bartoli (2010) 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 (2010) combine generative and discriminative approaches which results in lowering the total number of training samples.

2.1 Our contribution

In this paper we adopt a robust approach to non-rigid image alignment where instead of imposing the hard constraint that

¹ The parametric warp functions used in this work include Thin Plate Spline (TPS) and Free-Form Deformations (FFD) based on 2D Cubic B-Splines.

the optical flow must lie on the low-rank manifold (Garg et al. 2010), we penalize flow fields that lie outside it. Formulating the manifold constraint as a *soft constraint* using variational principles (Garg et al. 2011) leads to an energy with a quadratic term that allows us to adopt a decoupling scheme, related to the one described above (Zach et al. 2007; Chambolle and Pock 2011), for its efficient optimization. We propose a new *anisotropic trajectory regularization* term, parameterized in terms of the basis coefficients, instead of the full flow field. This results in an important dimensionality reduction in this term, which is usually the bottleneck of other quadratic relaxation duality based approaches (Zach et al. 2007; Chambolle and Pock 2011). Moreover, the optimization of our regularization step can be highly parallelized due to the independence of the orthonormal basis coefficients adding further advantages to previous approaches. Our approach can be seen as an extension of Zach et al. (2007) efficient TV- L^1 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. In practice, our approach is equivalent to Zach et al. (2007) in the degenerate case where the identity matrix is chosen as the motion basis.

We take advantage of the high level of parallelism inherent to our approach by developing a GPU implementation using the Nvidia CUDA framework. This parallel implementation vastly outperforms the equivalent Matlab code.

Additionally, we provide an extension of our multi-frame approach to the case of vector-valued images which allows us to use the information from all colour channels in image sequences, and further improve results. Our novel optimization scheme is based on the dualization of the linearized data term. Unlike Râket et al.'s previous attempt to extend TV- L^1 flow to vector valued images (Rakêt et al. 2011), our new algorithm is not restricted to the use of the L^1 -norm penaliser and instead allows the use of more general convex robust penalizers in the data 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 (Baker et al. 2011) (Middlebury) does not incorporate any such sequences. To facilitate the quantitative evaluation of multi-frame non-rigid registration and optical flow and to promote progress in this area, 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 different motion bases shows that our proposed approach improves on state of the art algorithms including large displacement (Brox and Malik 2011) and duality based (Zach et al. 2007) optical flow algorithms and the parametric dense non-rigid registration approach of Pizarro and Bartoli (2010).

3 Multi-frame Image Registration

Consider a video sequence of non-rigid objects 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, typically the first frame, our goal is to estimate the 2D trajectories of every point visible in the reference frame over the entire sequence, using a multi-frame approach (Fig. 1 illustrates our 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 Low-Rank Trajectory Space

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 lie near a low-dimensional linear subspace. This assumption is analogous to the non-rigid low-rank shape model, first proposed by Bregler et al. (2000), which states that the time varying 3D shape of a non-rigid object can be expressed as a linear combination of a low-rank shape basis. This rank constraint has been successfully exploited for 3D reconstruction by Non-Rigid Structure from Motion (NRSfM) algorithms (Torresani et al. 2008) where the matrix of 2D tracks is factorized into the product of two low-rank matrices: a motion matrix that describes the camera pose and time varying coefficients and a shape matrix that encodes the basis shapes.

The low-rank shape basis model of Bregler et al. (2000), Torresani et al. (2008) exploits the spatial properties of non-rigid motion, introducing rank constraints on the 3D location of the set of points (shape) at any given frame. Interestingly, the dual formulation of this model states that the rank constraint can be instead applied to the trajectories of each individual point, modelling them as a linear combination of *basis trajectories*. Therefore, the motion and shape matrices can exchange their roles as basis and coefficients and we can either interpret the 2D tracks as the projection of a linear combination of 3D basis shapes or as the linear combination of a 2D motion basis. This concept of non-rigid *trajectory basis* was first introduced in 2D by Torresani and Bregler (2002) who applied it to non-rigid 2D tracking as an extension of the rigid subspace constraints proposed by Irani (2002). Later Akhter et al. (2008, 2011) extended the *trajectory basis* to 3D to model non-rigid 3D trajectories using the Discrete Cosine Transform (DCT) basis.

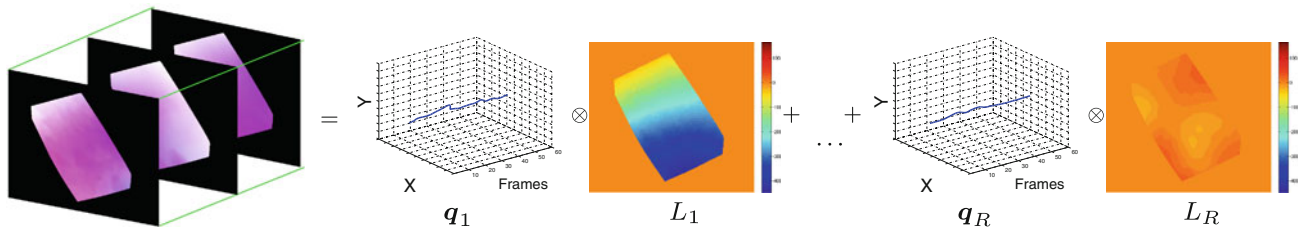


Fig. 3 The displacement of any point throughout the sequence can be expressed in a compact way as a linear combination of a low-rank trajectory basis. The basis vectors $q_i(n)$ encode the temporal information while the coefficient maps L_i describe the spatial distribution of the individual basis trajectories

3.2 Dense Trajectory Subspace Constraints

This paper extends the use of 2D trajectory subspace constraints to the case of estimating dense multi-frame optic flow using a variational approach.

More precisely, we assume that the input image sequence has F frames and the n_0 -th frame, $n_0 \in \{1, \dots, F\}$ has been chosen as the reference. We denote by $\Omega \subset \mathbb{R}^2$ the image domain and we define the function:

$$u(x; n) = \begin{bmatrix} u_1(x; n) \\ u_2(x; n) \end{bmatrix} : \Omega \times \{1, \dots, F\} \rightarrow \mathbb{R}^2 \quad (1)$$

that represents the point trajectories in the following way. For every visible point $x \in \Omega$ in the reference image, $u(x; \cdot) : \{1, \dots, F\} \rightarrow \mathbb{R}^2$ is its discrete-time 2D trajectory over all frames of the sequence. The coordinates of each trajectory $u(x; \cdot)$ are expressed with respect to the position of the point x at $n = n_0$, which means that $u(x; n_0) = 0$ and that the location of the same point in frame n is $x + u(x; n)$. We use the term *multi-frame optical flow* to describe u , since it corresponds to a multi-frame extension of the conventional optical flow: the latter is given by $u(x; 2)$ in the degenerate case where the sequence contains only $F = 2$ frames and the first one is considered as the reference ($n_0 = 1$).

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

$$u(x; n) = \sum_{i=1}^R q_i(n)L_i(x) + \epsilon(x; n), \quad (2)$$

which states that the trajectory $u(x; \cdot)$ of any point $x \in \Omega$ can be approximated as the linear combination of R basis trajectories $q_1(n), \dots, q_R(n) : \{1, \dots, F\} \rightarrow \mathbb{R}^2$ that are independent from the point location. We include a modeling error $\epsilon(x; n)$ which will allow us to impose the subspace constraint as a penalty term. Normally the values of $\epsilon(x; n)$ are relatively small, yet sufficient to improve the robustness of the multi-frame optical flow estimation.

Note that we consider that the chosen trajectory basis is *orthonormal*. We refer to the linear span of these basis

trajectories as a *trajectory subspace* and denote it by S_Q . The linear combination is controlled by coefficients $L_i(x)$ that depend on x , therefore we can interpret the collection of all the coefficients for all the points $x \in \Omega$ as a vector-valued image $L(x) \triangleq [L_1(x), \dots, L_R(x)]^T : \Omega \rightarrow \mathbb{R}^R$. Figure 3 illustrates the subspace constraint.

In many cases, effective choices for the model order (or rank) R correspond to values smaller than $2F$, which means that the above representation is compact and achieves a significant dimensionality reduction on the point trajectories.

We now re-write equation (2) in matrix notation, which will be useful in the subsequent presentation. Let $U(x)$ and $\mathcal{E}(x) : \Omega \rightarrow \mathbb{R}^{2F}$ be equivalent representations of the functions $u(x; n)$ and $\epsilon(x; n)$ 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 $q_1(n), \dots, q_R(n)$, after vectorizing them in the same way:

$$\underbrace{U(x)}_{2F \times 1} \triangleq \begin{bmatrix} u(x; 1) \\ \vdots \\ u(x; F) \end{bmatrix}, \quad \underbrace{Q}_{2F \times R} \triangleq \begin{bmatrix} q_1(1) & \dots & q_R(1) \\ \vdots & & \vdots \\ q_1(F) & \dots & q_R(F) \end{bmatrix},$$

$$\underbrace{\mathcal{E}(x)}_{2F \times 1} \triangleq \begin{bmatrix} \epsilon(x; 1) \\ \vdots \\ \epsilon(x; F) \end{bmatrix} \quad (3)$$

The subspace constraint (2) can now be written as follows:

$$U(x) = QL(x) + \mathcal{E}(x), \quad \forall x \in \Omega \quad (4)$$

3.3 Non-Rigid Video Registration from Multi-frame Optical Flow

Let $I(x; n) : \Omega \times \{1, \dots, F\} \rightarrow \mathbb{R}$ be the sequence of grayscale image frames, which are given either directly from the input frames or from the input frames after some pre-processing, such as structure-texture decomposition (Wedel et al. 2009).

In our formulation, the estimation of the multi-frame optical flow is equivalent to the simultaneous registration of all the frames with the reference frame n_0 : Recall that for every

frame n the coordinates $\mathbf{x} + \mathbf{u}(\mathbf{x}; n)$ yield the current location of any image point \mathbf{x} of the reference. Therefore, the image:

$$\mathbf{x} \rightarrow I(\mathbf{x} + \mathbf{u}(\mathbf{x}; n); n) \tag{5}$$

is the registered version of the image $I(\mathbf{x}; n)$ back to the reference $I(\mathbf{x}; n_0)$, or in other words it is the *warping* of the image $I(\mathbf{x}; n)$ to the image $I(\mathbf{x}; n_0)$. As it will be described later, we expect that the brightness differences between every registered image and the reference image to be small and therefore we use an appropriate brightness constancy term in our proposed energy.

4 Variational Multi-frame Optical Flow Estimation

In this section we show how dense motion estimation can be combined with the trajectory subspace constraints described in Sect. 3. In order to estimate the 2D trajectories of all the points, or equivalently simultaneously register all the frames with the reference frame n_0 , we propose the following energy:

$$E[\mathbf{u}(\mathbf{x}; n), \mathbf{L}(\mathbf{x})] = \alpha E_{data} + \beta E_{link} + E_{reg}, \tag{6}$$

where

$$E_{data} = \int_{\Omega} \sum_{n=1}^F |I(\mathbf{x} + \mathbf{u}(\mathbf{x}; n); n) - I(\mathbf{x}; n_0)| \, d\mathbf{x}, \tag{7}$$

$$E_{link} = \int_{\Omega} \sum_{n=1}^F \left| \mathbf{u}(\mathbf{x}; n) - \sum_{i=1}^R \mathbf{q}_i(n) L_i(\mathbf{x}) \right|^2 \, d\mathbf{x}, \tag{8}$$

$$E_{reg} = \int_{\Omega} \sum_{i=1}^R g(\mathbf{x}) |\nabla L_i(\mathbf{x})|_{\epsilon} \, d\mathbf{x}. \tag{9}$$

We minimize this energy jointly with respect to the point trajectories $\mathbf{u}(\mathbf{x}; n)$ and their components on the trajectory subspace that are determined by the linear model coefficients $\mathbf{L}(\mathbf{x})$. We also add the constraint that $\mathbf{u}(\mathbf{x}; n_0) = 0$, since this corresponds to the flow from the reference image frame to itself. The positive constants α and β weigh the balance between the terms of the energy. Also, $|\cdot|_{\epsilon}$ in (9) denotes the Huber norm of a vector and $g(\mathbf{x})$ is a space-varying weighting function (see Sect. 4 for more details).

Note that the functions $\mathbf{u}(\mathbf{x}; n)$ and $\mathbf{L}(\mathbf{x})$ determine two sets of trajectories that are relatively close to each other but not identical since the subspace constraint is imposed as a soft constraint. This improves the robustness of our method against overfitting to the image data in cases where the brightness constancy assumption fails. For this reason, we consider that the final output of our method are the trajectories $\mathcal{U}'(\mathbf{x}) = \mathbf{Q} \mathbf{L}(\mathbf{x})$ that lie on the trajectory subspace and are directly derived by the coefficients $\mathbf{L}(\mathbf{x})$.

4.1 Description of the Energy

In this section we provide more details about the properties of the proposed energy (6).

The **first term** (E_{data}) is a data attachment term that uses the robust \mathbf{L}^1 -norm and is a direct multi-frame extension of the brightness constancy term used by most optical flow methods, e.g. Zach et al. (2007). It is based on the assumption that the image brightness $I(\mathbf{x}; n_0)$ at every pixel \mathbf{x} of the reference frame is preserved at its new location, $\mathbf{x} + \mathbf{u}(\mathbf{x}; n)$, in every frame of the sequence. The use of an \mathbf{L}^1 -norm improves the robustness of the method since it allows deviations from this assumption, which might occur in real-world scenarios because of noise, illumination changes or occlusions of some points in some frames.

The **second term** (E_{link}) penalizes the difference between the two sets of trajectories $\mathbf{u}(\mathbf{x}; n)$ and $\mathbf{Q} \mathbf{L}(\mathbf{x})$ and acts as a coupling (linking) term between them. This term serves as a soft constraint that the trajectories $\mathbf{u}(\mathbf{x}; n)$ should be relatively close to the subspace spanned by the basis \mathbf{Q} . Concerning the weight β , the larger its value the more restrictive the subspace constraint becomes. Since the subspace of \mathbf{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.

An equivalent interpretation is that this term is derived from the constraint that the error $\boldsymbol{\epsilon}(\mathbf{x}; n)$ in (2) has a bounded \mathbf{L}^2 norm, i.e. $\int_{\Omega} \sum_{n=1}^F |\boldsymbol{\epsilon}(\mathbf{x}; n)|^2 \, d\mathbf{x} \leq C$, for some appropriate constant C . Then β corresponds to the Lagrange multiplier for this constraint.

The **third term** (E_{reg}) corresponds to the spatial regularization of the trajectory 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 neighbourhoods. Following Werlberger et al. (2009), Newcombe et al. (2011), we use the Huber norm over the gradient of each subspace coefficient $L_i(\mathbf{x})$, which is defined as:

$$|\nabla L_i(\mathbf{x})|_{\epsilon} = H_{\epsilon}(|\nabla L_i(\mathbf{x})|^2), \text{ with:} \tag{10}$$

$$H_{\epsilon}(s^2) = \begin{cases} \frac{s^2}{2\epsilon} & \text{if } s \leq \epsilon \\ s - \frac{\epsilon}{2} & \text{otherwise} \end{cases}$$

where ϵ is a relatively small constant. The Huber norm is a convex differentiable function that combines quadratic regularization in the interval $|\nabla L_i| \leq \epsilon$, with Total Variation regularization outside the interval. For small gradient magnitudes the Huber norm offers smooth solutions, whereas for larger magnitudes the discontinuity preserving properties of Total Variation are maintained. Following Alvarez et al. (1999), Wedel et al. (2009), Newcombe et al. (2011), we also

incorporate a space-varying weight $g(\mathbf{x})$ that depends on the reference image as follows:

$$g(\mathbf{x}) = \exp(-c_g |\nabla G_{\sigma_g}(\mathbf{x}) * I(\mathbf{x}; n_0)|^2) \tag{11}$$

where c_g is a constant and σ_g is the standard deviation of the 2D Gaussian $G(\mathbf{x})$ that convolves the reference image $I(\mathbf{x}; n_0)$. This weight encourages discontinuities in flow to coincide with edges of the reference image by reducing the regularisation strength near those edges. Further discussion on our proposed regularization term E_{reg} is provided in Sect. 4.

4.2 Connections to Previous Work

Interestingly, our adopted strategy of estimating two sets of trajectories, $\mathbf{u}(\mathbf{x}; n)$ and $\mathbf{Q} \mathbf{L}(\mathbf{x})$, resembles the techniques of quadratic relaxation and duplication of the optimization variable that have been previously used in the context of optical flow and depth map estimation (Zach et al. 2007; Pock et al. 2010; Stuehmer et al. 2010; Newcombe et al. 2011). Similarly, we benefit from the fact that the optimization problem can be decomposed into two parts, each of which is a convex energy² that can be solved efficiently and in a globally optimal manner. However, our formulation offers an additional advantage: the spatial regularization step, which is the bottleneck in these optimization schemes, is computationally much more efficient since it is applied to the coefficients $\mathbf{L}(\mathbf{x})$ that normally have smaller dimensionality than the flow $\mathbf{u}(\mathbf{x}; n)$.

Note that there is a degenerate case in which our proposed approach becomes equivalent to independently estimating the flow from the reference $I(\cdot; n_0)$ to each frame $I(\cdot; n)$ by applying $F - 1$ times the ITV- \mathbf{L}^1 optical flow algorithm (Wedel et al. 2009). This degenerate case occurs when:

- The motion basis is set to $\mathbf{Q} = \mathbf{I}_{2F}$, where \mathbf{I}_{2F} is the $2F \times 2F$ identity matrix, in which case $R = 2F$; and
- $c_g = 0$ and $\epsilon = 0$.

When $c_g = 0$ and $\epsilon = 0$, the terms $g(\mathbf{x}) |\nabla L_i(\mathbf{x})|_\epsilon$ become equivalent to $|\nabla L_i(\mathbf{x})|$ and therefore our regularization term E_{reg} is a summation of Total Variation terms. Furthermore, the choice $\mathbf{Q} = \mathbf{I}_{2F}$ converts the energy (6) into a summation of F decoupled energy terms J_n :

$$\begin{aligned} J_n = & \alpha \int_{\Omega} |I(\mathbf{x} + \mathbf{u}(\mathbf{x}; n); n) - I(\mathbf{x}; n_0)| \, d\mathbf{x} \\ & + \beta \int_{\Omega} \left| \mathbf{u}(\mathbf{x}; n) - \begin{bmatrix} L_{2n-1}(\mathbf{x}) \\ L_{2n}(\mathbf{x}) \end{bmatrix} \right|^2 \, d\mathbf{x} \\ & + \int_{\Omega} \sum_{i=2n-1}^{2n} |\nabla L_i(\mathbf{x})| \, d\mathbf{x} \end{aligned} \tag{12}$$

² After the linearization of the brightness constancy term

Each term J_n corresponds to a specific frame n and depends only on $\mathbf{u}(\mathbf{x}; n)$ and the two coefficients $L_{2n-1}(\mathbf{x})$ and $L_{2n}(\mathbf{x})$. These coefficients stacked together as a vector-valued function can be seen as the auxiliary variable of $\mathbf{u}(\mathbf{x}; n)$ so the energy term J_n is equivalent to the convex relaxation of the TV- \mathbf{L}^1 functional used in Wedel et al. (2009).

4.3 Effective Trajectory Regularization

In this section we provide further intuition into our choice of multi-frame optical flow regularization E_{reg} . The presentation of this section follows a constructive approach—we build our proposed regularizer from the simplest choice of regularization term in successive steps, each of which adds more complexity but improves its effectiveness. We start by revisiting common practices in the literature and conclude by proposing our novel *anisotropic trajectory regularization* term in the final step. Our goal is to regularize the multi-frame optical flow $\mathbf{U}'(\mathbf{x})$ that lies on the trajectory subspace. Note that $\mathbf{U}'(\mathbf{x})$ can be interpreted as a vector valued function with $2 \times F$ channels encoding the horizontal and vertical components of the optical flow at each frame as defined in equation (3).

Step 1. A simple choice would be to use homogeneous regularization of $\mathbf{U}'(\mathbf{x})$, which is a straightforward multiframe generalization of the model of Horn and Schunck (1981):

$$\begin{aligned} & \int_{\Omega} \sum_{n=1}^F |\nabla u'_1(\mathbf{x}; n)|^2 + |\nabla u'_2(\mathbf{x}; n)|^2 \, d\mathbf{x} \\ & = \int_{\Omega} \|D\mathbf{U}'(\mathbf{x})\|_F^2 \, d\mathbf{x} \end{aligned} \tag{13}$$

where $\|\cdot\|_F$ denotes the Frobenius norm of a matrix and $D\mathbf{U}'(\mathbf{x})$ is the Jacobian of $\mathbf{U}'(\mathbf{x})$ (each row contains the gradient of the corresponding channel of $\mathbf{U}'(\mathbf{x})$). However, this regularizer leads to oversmoothing on the motion boundaries since the quadratic term excessively penalizes large magnitudes of the gradients of $\mathbf{U}'(\mathbf{x})$, which correspond to motion discontinuities.

Step 2. A way to avoid this is by applying a *robust function* Ψ that penalizes outliers of the gradient less severely than the quadratic penalizer:

$$\int_{\Omega} \Psi \left(\|D\mathbf{U}'(\mathbf{x})\|_F^2 \right) \, d\mathbf{x} \tag{14}$$

This choice is used in Nir et al. (2008) and when only two frames are taken into account it is equivalent to the regularizers used in Schnörr (1994), Weickert (1998), Brox and Malik (2011) (*isotropic flow-driven regularization* in the terminol-

ogy of Weickert and Schnörr (2001a)). Some examples of the robust function Ψ include the following:

- $\Psi(s^2) = s$, in which case the regularizer is the *vectorial total variation* (Sapiro 1997) of the vector-valued function $\mathbf{U}'(\mathbf{x})$ that encodes the multi-frame optical flow.
- $\Psi(s^2) = H_\epsilon(s^2)$ or the Huber norm (10), which is the choice adopted in our approach.

The robust function Ψ in (14) penalizes outliers of the norm $\|D\mathbf{U}'(\mathbf{x})\|_F$ less strongly, therefore allows discontinuities to occur at $\mathbf{U}'(\mathbf{x})$. However, such outliers correspond only to the points \mathbf{x} where all the channels of $\mathbf{U}'(\mathbf{x})$ display sharp discontinuities. If for example only few channels of $\mathbf{U}'(\mathbf{x})$ have a high gradient at a point \mathbf{x} , then $\|D\mathbf{U}'(\mathbf{x})\|_F$ is not treated as an outlier, since it is still low (because of the sum of squares over all channels that is involved in this norm). This regularizer is thus much less tolerant to motion boundaries that occur at individual channels.

Step 3. The above problem can be addressed by applying the penalizer Ψ independently to the squared norm of the gradient of each channel of $\mathbf{U}'(\mathbf{x})$:

$$\int_{\Omega} \sum_{n=1}^F \left\{ \Psi \left(|\nabla u'_1(\mathbf{x}; n)|^2 \right) + \Psi \left(|\nabla u'_2(\mathbf{x}; n)|^2 \right) \right\} d\mathbf{x} \quad (15)$$

This is a direct multi-frame extension of the regularizer used in Deriche et al. (1995), Kumar et al. (1996), Aubert et al. (1999), Zach et al. (2007), Wedel et al. (2009) for which efficient numerical implementations exist (Zach et al. 2007; Wedel et al. 2009). In this way, each channel of $\mathbf{U}'(\mathbf{x})$ can have different boundaries. However, this regularizer is on the other extreme of the regularizer of Step 2: where substantial correlation between the different channels exists, it is ineffective since it allows correlated trajectories to have different boundaries.

In addition, in contrast to the regularizers proposed in previous steps, it is not rotation invariant (Weickert and Schnörr 2001a).

Step 4. To avoid the aforementioned problems, we adopt our subspace model for the 2D trajectories $\mathbf{U}'(\mathbf{x}) = \mathbf{Q}\mathbf{L}(\mathbf{x})$ and rewrite the norm $\|D\mathbf{U}'(\mathbf{x})\|_F$ as a function of the coefficients:

$$\begin{aligned} \|D\mathbf{U}'(\mathbf{x})\|_F^2 &= \left| \frac{\partial \mathbf{U}'(\mathbf{x})}{\partial x_1} \right|^2 + \left| \frac{\partial \mathbf{U}'(\mathbf{x})}{\partial x_2} \right|^2 \\ &= \left| \mathbf{Q} \frac{\partial \mathbf{L}(\mathbf{x})}{\partial x_1} \right|^2 + \left| \mathbf{Q} \frac{\partial \mathbf{L}(\mathbf{x})}{\partial x_2} \right|^2 = \sum_{i=1}^R |\nabla L_i(\mathbf{x})|^2, \end{aligned} \quad (16)$$

where we have used the property of orthonormality of the basis \mathbf{Q} . Provided that the trajectory basis \mathbf{Q} has been chosen

appropriately, the coefficients $\mathbf{L}(\mathbf{x})$ are much less correlated than the channels of $\mathbf{U}'(\mathbf{x})$. We conclude that it is more effective to apply the robust function Ψ independently to the basis coefficients (instead of the flow fields) and we derive the regularizer:

$$\int_{\Omega} \sum_{i=1}^R \Psi \left(|\nabla L_i(\mathbf{x})|^2 \right) d\mathbf{x} \quad (17)$$

Furthermore, this regularizer leads to a much more efficient implementation for two main reasons. First, the resultant regularization is applied to the coefficients $\mathbf{L}(\mathbf{x})$, that typically have lower dimensionality than the flow $\mathbf{U}'(\mathbf{x})$. Second, this regularization is decoupled for each coefficient and can thus be highly parallelized. Note that the regularizer (15) derived in Step 3 can be considered as a special case of the above regularizer when the $2F \times 2F$ identity matrix is chosen as the basis \mathbf{Q} . However, in our work, we use two choices for \mathbf{Q} : DCT and PCA (derived from an initial flow). We now analyze each of these cases separately:

- When the basis matrix \mathbf{Q} has been estimated by applying PCA to some trajectory samples, the correlation between the coefficients can be considered negligible. Furthermore, in this case we regain the desirable property of rotation invariance, since the proposed regularizer (17) is consistent with the general design principle of Weickert and Schnörr (2001a) for *rotationally invariant anisotropic regularizers*. According to that principle³, given an appropriate decomposition of $\|D\mathbf{U}'(\mathbf{x})\|_F^2 = \sum_j \rho_j$ where ρ_j are rotationally invariant expressions, one should use the regularizer $\int_{\Omega} \sum_j \Psi(\rho_j)$, which is rotationally invariant and anisotropic. In our case, the expressions ρ_j correspond to the coefficients $L_i(\mathbf{x})$, which are indeed rotation invariant: If we assume that a rotation of the input frames causes the same rotation to be applied to the trajectory samples, then the basis trajectories will be equally rotated. Therefore, the coefficients $L_i(\mathbf{x})$ of a specific reference image point⁴ will remain invariant and the corresponding trajectory $\mathbf{U}'(\mathbf{x})$ will simply be rotated.
- In the case of the DCT basis, the above properties do not hold. However, the regularizer (17) with a DCT basis is much more effective than the regularizer (15), since the DCT frequency components of a trajectory are typically less correlated than its actual coordinates. This is due

³ In Weickert and Schnörr (2001a) this design principle is expressed for the classical optical flow case where the input is a single pair of frames, but here we present its straight-forward extension to the case of multiple frames.

⁴ By *specific reference point* we mean that we associate the new location (after rotation) of a point on the reference image with its original location.

to the fact that when the actual motions of the image points are compositions of different physical motions, these motions are expected to be much more localized in the frequency domain rather than in the time domain.

Step 5. Finally, it is reasonable to assume that the boundaries of all the motion components tend to be a subset of the edges on the reference image. Following Alvarez et al. (1999), Wedel et al. (2009), Newcombe et al. (2011), in order to prevent any smoothing along the motion boundaries our final regularizer E_{reg} is weighted by a space-varying function $g(\mathbf{x})$ that depends on the reference image as described in (11).

In our extensive experiments, we have empirically evaluated that the introduction of such a weighting improves the accuracy of the multiframe optical flow. This is in accordance with the experimental evidence reported in Wedel et al. (2009) for the classical optical flow.

5 Optimization of the Proposed Energy

In order to minimize the energy (6), we follow a coarse-to-fine technique with multiple warping iterations (Brox et al. 2004). In every warping iteration, we use an initialization $\mathbf{u}_0(\mathbf{x}; n)$ that comes from the previous iteration. We approximate the data term (7) by linearizing the image $I(\cdot; n)$ around $\mathbf{x} + \mathbf{u}_0(\mathbf{x}; n)$. After this approximation, the energy (6) becomes convex.

Following Zach et al. (2007), we implement the optimization of the energy (6) using an alternating approach. We decouple the data and regularization terms to decompose the optimization problem into two, each of which can be more easily solved. In this section we show how to adapt the method of Zach et al. (2007) to our problem, to take advantage of its computational efficiency and apply it to multiframe subspace-constrained optical flow. The key difference to Zach et al. (2007) 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.

We apply an alternating optimization, updating either $\mathbf{u}(\mathbf{x}; n)$ or $\mathbf{L}(\mathbf{x})$ in every iteration, as follows:

- Repeat until convergence:

Minimization Step 1: For $\mathbf{u}(\mathbf{x}; n)$ fixed, update $\mathbf{L}(\mathbf{x})$ by minimizing $E[\mathbf{u}(\mathbf{x}; n), \mathbf{L}(\mathbf{x})]$ w.r.t. $\mathbf{L}(\mathbf{x})$.

Minimization Step 2: For $\mathbf{L}(\mathbf{x})$ fixed, update $\mathbf{u}(\mathbf{x}; n)$ by minimizing $E[\mathbf{u}(\mathbf{x}; n), \mathbf{L}(\mathbf{x})]$ w.r.t. $\mathbf{u}(\mathbf{x}; n)$.

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

threshold. Since at every step the value of the energy does not increase and this value is bounded below by its global minimum, the above alternation is guaranteed to converge to a global minimum point.

5.1 Minimization Step 1

Since in this step we keep $\mathbf{u}(\mathbf{x}; n)$ fixed, we observe that only the last two terms, E_{link} and E_{reg} , of the energy (6) depend on $\mathbf{L}(\mathbf{x})$. Therefore we must minimize $\beta E_{link} + E_{reg}$ with respect to $\mathbf{L}(\mathbf{x})$. Using the matrix notation defined in (4), we can write the term E_{link} as:

$$\sum_{n=1}^F |\mathbf{u}(\mathbf{x}; n) - \sum_{i=1}^R \mathbf{q}_i(n) L_i(\mathbf{x})|^2 = |\mathbf{U}(\mathbf{x}) - \mathbf{Q}\mathbf{L}(\mathbf{x})|^2 \quad (18)$$

Let \mathbf{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 $[\mathbf{Q} \ \mathbf{Q}^\perp]$ is an orthonormal $2F \times 2F$ matrix, which means that its columns form a basis of \mathbb{R}^{2F} . Consequently, $\mathbf{U}(\mathbf{x})$ can be decomposed into two orthonormal vectors as

$$\mathbf{U}(\mathbf{x}) = \mathbf{Q}\mathbf{M}(\mathbf{x}) + \mathbf{Q}^\perp \mathbf{M}_{out}(\mathbf{x}) \quad (19)$$

where

$$\mathbf{M}(\mathbf{x}) \triangleq \mathbf{Q}^T \mathbf{U}(\mathbf{x}) \quad \text{and} \quad \mathbf{M}_{out}(\mathbf{x}) \triangleq (\mathbf{Q}^\perp)^T \mathbf{U}(\mathbf{x}) \quad (20)$$

are the coefficients that define the projections of $\mathbf{U}(\mathbf{x})$ onto the trajectory subspace \mathcal{S}_Q and its orthogonal complement. Equation (18) can now be further simplified:

$$\begin{aligned} & |\mathbf{U}(\mathbf{x}) - \mathbf{Q}\mathbf{L}(\mathbf{x})|^2 \\ &= \left| \mathbf{Q}^\perp \mathbf{M}_{out}(\mathbf{x}) + \mathbf{Q}(\mathbf{M}(\mathbf{x}) - \mathbf{L}(\mathbf{x})) \right|^2 \\ &= |\mathbf{M}_{out}(\mathbf{x})|^2 + |\mathbf{M}(\mathbf{x}) - \mathbf{L}(\mathbf{x})|^2, \end{aligned} \quad (21)$$

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

Thus, the minimization of Step 1 is equivalent to the minimization of:

$$\begin{aligned} & \beta \int_{\Omega} |\mathbf{M}(\mathbf{x}) - \mathbf{L}(\mathbf{x})|^2 d\mathbf{x} + \int_{\Omega} \sum_{i=1}^R g(\mathbf{x}) |\nabla L_i(\mathbf{x})|_{\epsilon} d\mathbf{x} \\ &= \sum_{i=1}^R \int_{\Omega} g(\mathbf{x}) |\nabla L_i(\mathbf{x})|_{\epsilon} + \beta (M_i(\mathbf{x}) - L_i(\mathbf{x}))^2 d\mathbf{x} \end{aligned} \quad (22)$$

where $M_i(\mathbf{x})$ is the i -th coordinate of $\mathbf{M}(\mathbf{x})$. We have finally obtained a new form of the energy that allows the trajectory model coefficients $L_i(\mathbf{x})$ to be decoupled. The minimization of each term in the above sum can be done independently and corresponds to a small modification of the TV- L^2 Rudin-Osher-Fatemi (ROF) model (Rudin et al. 1992) applied to each coefficient $L_i(\mathbf{x})$: This modification consists of incorporating an edge weighting $g(\mathbf{x})$ and replacing the L^2 norm $|\nabla L_i(\mathbf{x})|$ with the Huber norm $|\nabla L_i(\mathbf{x})|_\epsilon$. This modified ROF model has been recently studied in Newcombe et al. (2011) for the problem of depth estimation. The optimum $L_i(\mathbf{x})$ is actually a regularized version of $M_i(\mathbf{x})$ and the extent of this regularization increases as the weight β decreases.

The benefits of the computational efficiency of the above procedure are twofold. First, these independent minimizations can be parallelized. Second, several efficient algorithms exist to implement such regularization models. Appendix A describes the actual algorithm we used for the optimization of this energy, which is related to the method proposed in Newcombe et al. (2011).

5.2 Minimization Step 2

Keeping $\mathbf{L}(\mathbf{x})$ fixed, we observe that only the first two terms of the energy (6), E_{data} and E_{link} , depend on $\mathbf{u}(\mathbf{x}; n)$ and therefore we have to minimize with respect to $\mathbf{u}(\mathbf{x}; n)$ the following:

$$\alpha E_{data} + \beta E_{link} = \int_{\Omega} \sum_{n=1}^F \left\{ \alpha |I(\mathbf{x} + \mathbf{u}(\mathbf{x}; n); n) - I(\mathbf{x}; n_0)| + \beta |\mathbf{u}(\mathbf{x}; n) - \mathbf{u}'|^2 \right\} d\mathbf{x}, \tag{23}$$

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

We implement this point-wise minimization by applying the technique proposed in Zach et al. (2007) to every frame. More precisely, for every frame n and point \mathbf{x} the image $I(\cdot; n)$ is linearized around $\mathbf{x} + \mathbf{u}_0(\mathbf{x}; n)$, where $\mathbf{u}_0(\mathbf{x}; n)$ are the initializations of the trajectories $\mathbf{u}(\mathbf{x}; n)$. 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 Zach et al. (2007).

6 Derivation of the Trajectory Basis

Concerning the choice of 2D trajectory basis \mathbf{Q} , we consider orthonormal bases as it simplifies the analysis and calculations in our method (see Sect. 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:

$$\mathbf{q}_i(n) = \begin{cases} [w_i(n), 0]^T, & \text{if } i = 1, \dots, \frac{R}{2} \\ [0, w_{i-R/2}(n)]^T, & \text{if } i = \frac{R}{2} + 1, \dots, R \end{cases} \tag{24}$$

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. The obvious advantage of using a predefined basis is that it does not need to be estimated in advance.

An alternative is to estimate the basis by applying Principal Component Analysis (PCA) to some sample trajectories. Provided that it is possible to estimate a set of sample trajectories 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 a given rank R , in terms of representational power. In this work we consider two possibilities.

- (i) The sample trajectories could come from an initial estimate of optical flow. We have found that the flow obtained using the DCT basis provides a very good initial flow on which we then apply PCA to obtain an optimized basis.
- (ii) Alternatively, the sample trajectories could be a small subset of *reliable* point tracks, which we consider to be those where the texture of the image is strong in both spatial directions and can be selected using Shi and Tomasi's criterion (Shi and Tomasi 1994). However, this option is not resilient to outliers.

In practice, in our experimental evaluation section we show that the multi-frame optical flow obtained with the optimized PCA basis proposed in (i) provides the best results. It has the added advantage that, since we initialize the flow from our algorithm using the DCT basis, which is predefined and needs not be estimated, the entire process is automated and less affected by outliers.

7 Generalization to Sequences of Vector-Valued Images

The algorithm we have described so far assumes that the images in the sequence are grayscale. In this section we develop a generalization of our approach to the case of sequences of vector-valued images. We propose an optimization scheme that is based on the dualization of the data term of the energy.

The use of vector-valued images can significantly improve the accuracy of the estimated optical flow for various reasons. First of all, the vector-valued images can incorporate all the color channels of an image. The color cue in a video offers important additional information and resolves ambiguities that are present in the grayscale images. Furthermore, this generalization offers the potential for incorporating other powerful image cues as additional channels. For instance, the spatial derivatives of the color channels can be added to impose the gradient constancy assumption (Uras et al. 1988; Brox et al. 2004; Papenberg et al. 2006; Brox and Malik 2011) or even more complex features such as SIFT (Liu et al. 2011) features or others derived using a Field-of-Experts formulation (Sun et al. 2008), which can improve the robustness against illumination changes of the scene. Note that in our experimental evaluation we have only incorporated the color channels. To cope with illumination changes we have used structure-texture decomposition as a preprocessing step, which is an alternative way to gain robustness (Wedel et al. 2009).

7.1 Proposed Dual Formulation

Let us assume that the video frames that are used in our data term are vector-valued images with N_c channels:

$$I(\mathbf{x}; n) : \Omega \times \{1, \dots, F\} \rightarrow \mathbb{R}^{N_c} \tag{25}$$

To cope with this more general case, we only have to modify two elements of the formulation of our energy: (i) the data term E_{data} of the proposed energy (6) and (ii) the edge-weighting function of the regularization term $g(\mathbf{x})$ described in (11) that depends on the reference image.

The original definition of the function $g(\mathbf{x})$ is based on the term $|\nabla G_{\sigma_g}(\mathbf{x}) * I(\mathbf{x}; n_0)|^2$ used as a simple edge-strength predictor. For vector-valued images, we use a common and natural extension of this predictor (Blomgren and Chan 1998; Tschumperlé and Deriche 2005) by adding the contributions of the different image channels. We thus generalize the edge-weighting function as follows:

$$g(\mathbf{x}) = \exp\left(-c_g \sum_{i=1}^{N_c} |\nabla G_{\sigma_g}(\mathbf{x}) * I_i(\mathbf{x}; n_0)|^2\right) \tag{26}$$

Concerning the data term E_{data} , we also make a further generalization by applying a generic robust function Φ to the image differences:

$$\Phi : \mathbb{R}^{N_c} \rightarrow \mathbb{R}. \tag{27}$$

Our generalized data term becomes:

$$E_{data} = \int_{\Omega} \sum_{n=1}^F \Phi(I(\mathbf{x} + \mathbf{u}; n) - I(\mathbf{x}; n_0)) \, d\mathbf{x} \tag{28}$$

Since only the data term is affected by the extension to vector-valued images, the optimization of our proposed energy (6) only requires a modification of the minimization of $\alpha E_{data} + \beta E_{link}$ with respect to $\mathbf{u}(\mathbf{x}; n)$ (Step 2 in Sect. 5). Similarly to the case of grayscale images, this minimization is independent for every spatio-temporal point $(\mathbf{x}; n)$. But the point-wise energy that must be minimized with respect to \mathbf{u} is now the following:

$$E^{aux}(\mathbf{u}) = \alpha \Phi(I(\mathbf{x} + \mathbf{u}; n) - I(\mathbf{x}; n_0)) + \beta |\mathbf{u} - \mathbf{u}'|^2$$

For every point \mathbf{x} in every frame n each channel of $I(\cdot; n)$ is linearized around $\mathbf{x} + \mathbf{u}_0(\mathbf{x}; n)$, where $\mathbf{u}_0(\mathbf{x}; n)$ are the initializations of the trajectories $\mathbf{u}(\mathbf{x}; n)$. With this approximation, E^{aux} can be written as:

$$E^{aux}(\mathbf{u}) = \alpha \Phi(\mathbf{A}\mathbf{u} + \mathbf{b}) + \beta |\mathbf{u} - \mathbf{u}'|^2 \tag{29}$$

where $\mathbf{b} = I(\mathbf{x} + \mathbf{u}_0; n) - I(\mathbf{x}; n_0) - \mathbf{A}\mathbf{u}_0$ and $\mathbf{A} = \frac{\partial I(\mathbf{x} + \mathbf{u}_0; n)}{\partial \mathbf{x}}$ is the $N_c \times 2$ (spatial) Jacobian of the n -th frame $I(\cdot; n)$, evaluated at $\mathbf{x} + \mathbf{u}_0$.

Assuming that the function Φ is proper convex and lower semi-continuous, we dualise it by using its convex bi-conjugate (Rockafellar 1997; Chambolle and Pock 2011):

$$\Phi(s) = \sup_{\mathcal{I}} \{\langle s, \mathcal{I} \rangle - \Phi^*(\mathcal{I})\} \tag{30}$$

where, $\Phi^*(\mathcal{I})$ is the Legendre-Fenchel transform of $\Phi(s)$ and \mathcal{I} is the dual variable to s . We can now rewrite the energy E^{aux} (29) as:

$$E^{aux}(\mathbf{u}) = \alpha \max_{\mathcal{I}} \{\langle \mathbf{A}\mathbf{u} + \mathbf{b}, \mathcal{I} \rangle - \Phi^*(\mathcal{I})\} + \beta |\mathbf{u} - \mathbf{u}'|^2 \tag{31}$$

Based on the above expression, we propose to minimise E^{aux} by solving the following saddle point problem:

$$\min_{\mathbf{u}} \max_{\mathcal{I}} E^{sp}(\mathbf{u}, \mathcal{I}), \tag{32}$$

⁵ Note that, for the sake of clarity in our presentation, the generic robust function Φ defined here differs from the robust function Ψ that we used in Sect. 4: Φ is applied directly to the vectorial differences whereas Ψ is applied to their squared norms. The two definitions are linked by: $\Phi(\mathbf{v}) = \Psi(|\mathbf{v}|^2)$

where

$$E^{sp}(\mathbf{u}, \mathcal{I}) \triangleq \alpha \left(\langle \mathbf{A}\mathbf{u} + \mathbf{b}, \mathcal{I} \rangle - \Phi^*(\mathcal{I}) \right) + \beta \|\mathbf{u} - \mathbf{u}'\|^2 \tag{33}$$

Given a specific choice for the robust function Φ , one can derive efficient algorithms to solve the saddle point problem (32), using a similar framework as in Esser et al. (2010), Chambolle and Pock (2011), Pock and Chambolle (2011). In Appendix B we provide such algorithms for two special cases of Φ of particular interest:

- $\Phi(\mathbf{v}) = |\mathbf{v}|$, which leads to L^1 - norm of the image differences in E_{data} (28). This is the choice that we use in our experiments on colour images.
- $\Phi(\mathbf{v}) = H_\epsilon(|\mathbf{v}|^2)$, which corresponds to the Huber norm (10).

Note that Rakêt et al. (2011) recently proposed an extension of the TV- L^1 algorithm for vector-valued images. Their method corresponds to the choice $\Phi(\mathbf{v}) = |\mathbf{v}|$ and uses a step of projection onto an elliptic ball. The formulation that we propose in this section can be seen as an alternative to the aforementioned work. The advantage of our approach is that it allows the use of more general robust functions Φ .

8 Implementation Details

In this section we provide details about the implementation of the numerical optimization schemes for our grayscale and vector-valued multi-frame subspace optical flow algorithms.

We used a similar numerical optimisation scheme and pre-processing of images⁶ to the one proposed in Wedel et al. (2009) to minimise the energy (6), 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 parameters, the default values proposed in Wedel et al. (2009) for the ITV- L^1 algorithm were found to give the best results for ITV- L^1 and our method on the benchmark sequence (5 warp iterations, 20 alternation iterations and the weights α and β were set to 30 and 2). The same settings were used in all our experiments on real sequences. Note that when we ran the colour version of our algorithm we downweighed the value of α by a factor of $\frac{1}{\sqrt{3}}$ to account for the three colour channels. Regarding the parameters of the space varying weight of the regularization term $g(\mathbf{x})$ defined in (11), we used the following values: $\sigma_g = 1$ pixel, $c_g = 0.8$ and $\epsilon = 0.1$.

⁶ Note that we have normalized the image intensity values to lie between 0 and 1.

Since our algorithm can be efficiently parallelized on standard graphics hardware we have developed a GPU implementation using the CUDA framework. We run our algorithm on an NVIDIA GTX-580 GPU card hosted on a dual-core CPU. We obtain an average speedup of $\times 50$ with respect to our CPU Matlab implementation which runs on a 4 quad-core server with 192Gb of memory.

9 Reparameterization of the Optical Flow: Hard Subspace Constraint

In the special case where the error $\epsilon(\mathbf{x}; n)$ in (2) is close to zero everywhere in the image, or equivalently when $\beta \rightarrow \infty$ in (6), our soft constraint becomes a hard constraint and the optical flow $\mathbf{u}(\mathbf{x}; n)$ can be reparameterized as:

$$\sum_{i=1}^R \mathbf{q}_i(n) L_i(\mathbf{x}) \tag{34}$$

where the coefficients of the motion basis $L_i(\mathbf{x})$ are the unknown variables. In this case the energy for vector valued images with N_c channels can be rewritten as:

$$E_h = \int_{\Omega} \sum_{n=1}^F |I(\mathbf{x} + Q_n L(\mathbf{x}); n) - I(\mathbf{x}; n_0)| \, d\mathbf{x} + \int_{\Omega} \sum_{i=1}^R g(\mathbf{x}) |\nabla L_i(\mathbf{x})|_{\epsilon} \, d\mathbf{x} \tag{35}$$

where Q_n is the $2 \times R$ matrix $[\mathbf{q}_1(n) \cdots \mathbf{q}_R(n)]$, i.e. two rows of the basis matrix Q which correspond to frame n . Appendix C describes a primal-dual optimization algorithm to minimize this energy obtained via reparameterization of the flow.

A valid question at this point would be: how does this hard subspace constraint compare with respect to our proposed soft constraint? In Sect. 3 we argued that a soft constraint would provide increased robustness. For this reason, in Sect. 10 we have conducted a thorough experimental comparison between the two approaches which in fact reveals that it is indeed beneficial to allow deviations from the subspace constraint. Our robust soft constraint consistently outperforms imposing a hard constraint via reparameterization of the optical flow.

10 Experimental Results

In this section we evaluate our method and compare its performance with state of the art optical flow (Brox and Malik 2011; Zach et al. 2007) and image registration (Pizarro and Bartoli 2010) algorithms. We show quantitative comparative results

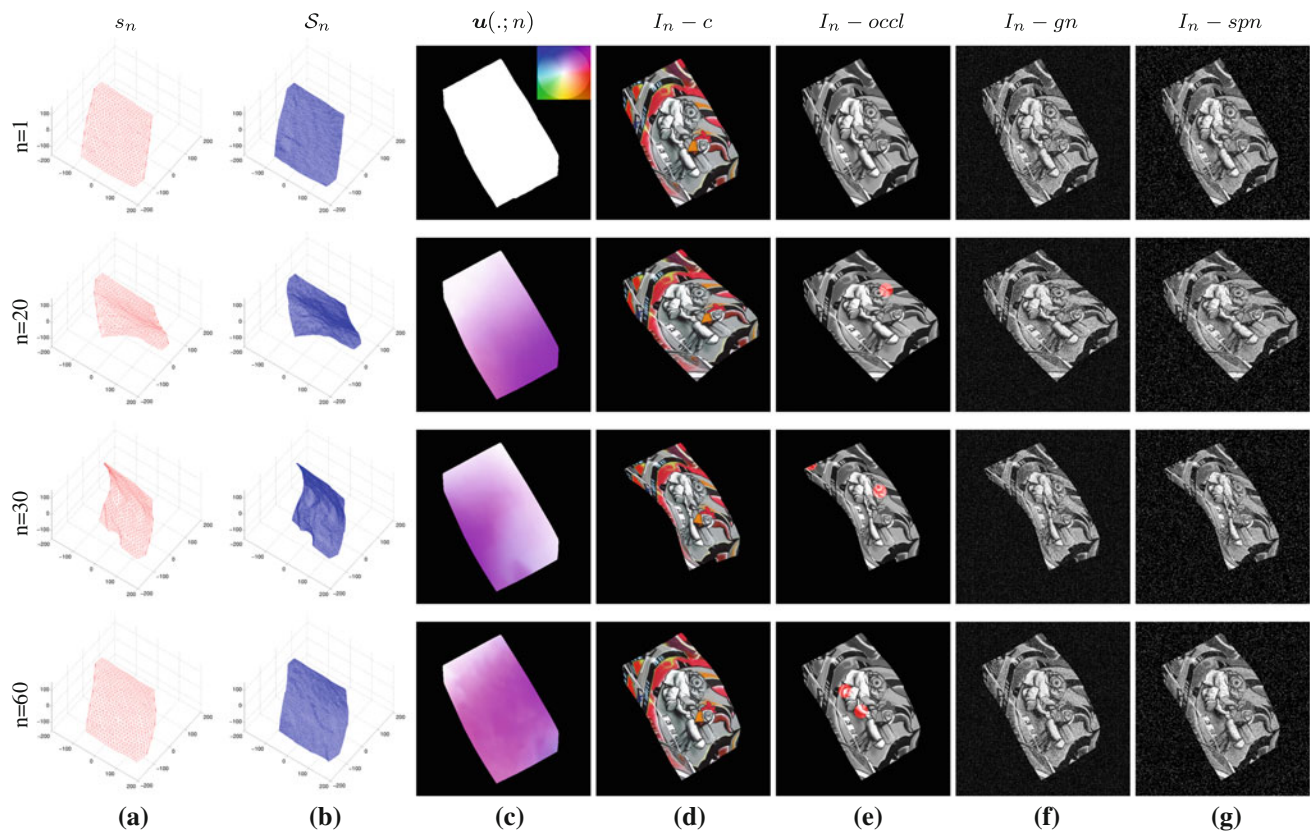


Fig. 4 Rendering process for ground truth optical flow sequence of a non-rigid object for different images in each row. (a) Sparse surface s_n representing MOCAP data (White et al. 2007), (b) Dense surfaces S_n constructed using thin plate spline interpolation, (c) Ground truth optical flow $u(x; n)$ visualized with the color coding that is shown at $n = 1$, ((d)) Color sequence $I_n - c$ rendered from S_n using texture

mapping of a graffiti image, (e) Grayscale version $I_n - occl$ of the same sequence with superimposed red disks indicate regions where intensities are replaced by black in the case of synthetic occlusions, (f) Grayscale sequence $I_n - gn$ with synthetic gaussian noise, (g) Grayscale sequence $I_n - spn$ with synthetic salt and pepper noise (Color figure online)

on our new benchmark ground truth optical flow dataset and qualitative results on real-world sequences⁷.

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 such as Weickert and Schnörr (2001b). However, due to the lack of publicly available implementations we chose to compare with LDOF (Large Displacement Optical Flow) (Brox and Malik 2011), one of the best performing 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 against the duality-based ITV- L^1 (Improved TV-

L^1) algorithm (Wedel et al. 2009), which we use as a baseline since our method can be seen as its generalization to the case of multi-frame non-rigid optical flow via robust trajectory subspace constraints (see Sect. 4). 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 (Pizarro and Bartoli 2010).

Note that all these algorithms can only be used on grayscale images.

10.1 Construction of a Ground Truth Benchmark Dataset

For the purpose of quantitative evaluation of multi-frame non-rigid optical flow we have generated a new benchmark sequence with ground truth optical flow data. 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 White et al. (2007) to capture the real

⁷ 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

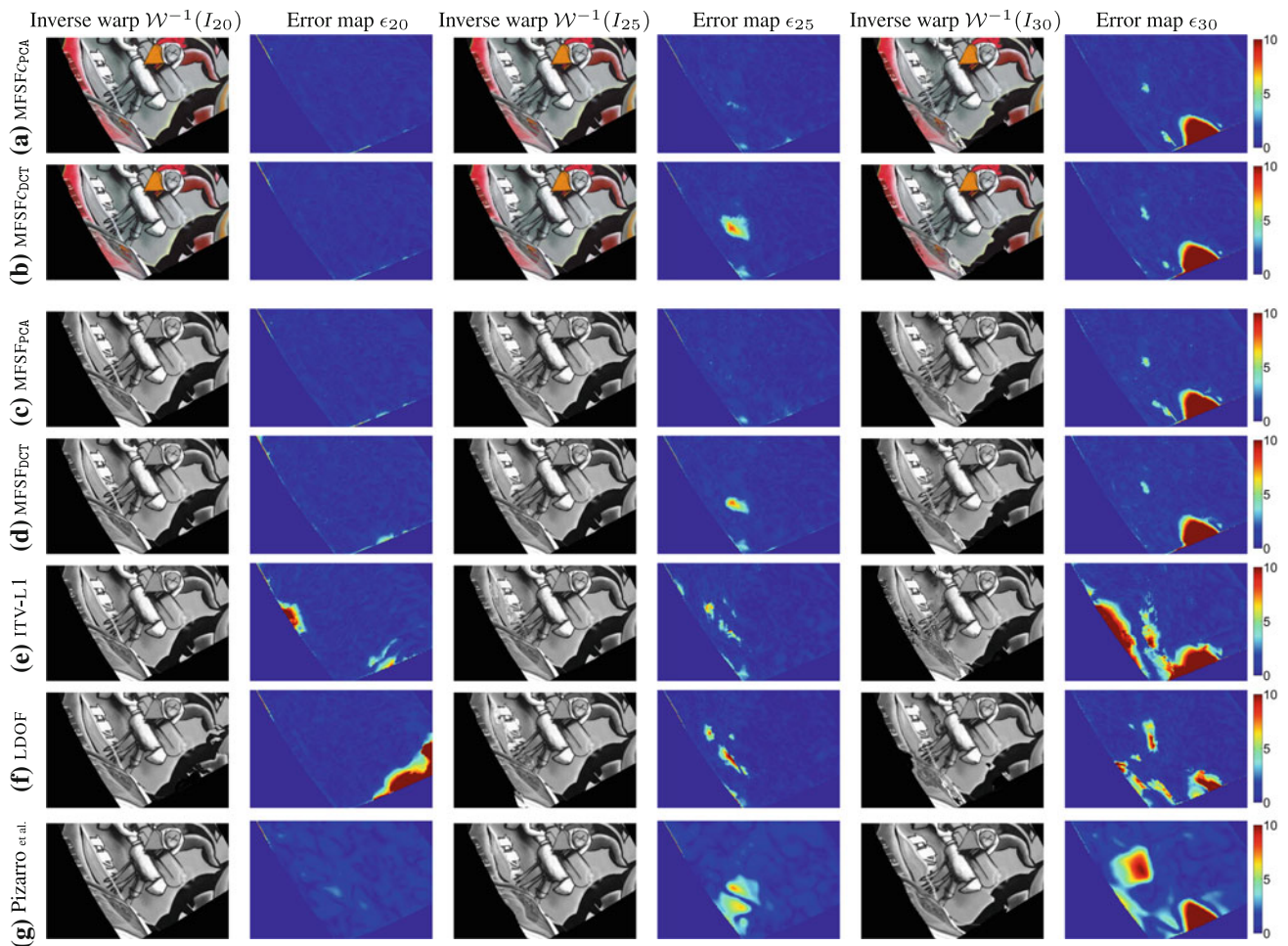


Fig. 5 Inverse warps $\mathcal{W}^{-1}(I_i)$ and error maps $\epsilon_i(\mathbf{x})$ for frames ($i = 20, 25, 30$) of the original benchmark sequence. Each row shows results for different methods. **(a–b)** Multi-frame subspace flow on color images: **(a)** MFSF_{PCA}, **(b)** MFSF_{DCT}. **(c–d)** Multi-frame subspace flow

on grayscale images: **(c)** MFSF_{PCA}, **(d)** MFSF_{DCT}. Against **(e)** ITV-L1, **(f)** LDOF [Brox and Malik \(2011\)](#), **(g)** [Pizarro and Bartoli \(2010\)](#)

deformations of a waving flag in 3D. This sparse data is interpolated to create a continuous dense 3D surface using the motion capture markers as the control points for smooth Spline interpolation. Figure 4 shows four frames of the (a) sparse and (b) dense interpolated 3D flag surface. 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 frames of size 500×500 pixels. We provide both grayscale and colour sequences. 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 sequences by adding (i) Gaussian noise, of standard deviation 0.2 relative to the range of image intensities, (ii) salt & pepper (S&P) noise of density 10% and (iii) synthetic

occlusions generated by superimposing some black circles of radius 20 pixels moving in linear orbits. Figure 4 shows four frames of the original colour sequence, the ground truth optical flow and the equivalent frames of the grayscale sequence with: synthetic occlusions, Gaussian noise and salt & pepper noise.

10.2 Quantitative Results on Benchmark Sequence

We tested our *Multi-Frame Subspace Flow* algorithm for grayscale (MFSF) and colour images (MFSF C) using the three different proposed motion basis: PCA, DCT and Cubic B-Spline (Figs. 5, 6). In Table 1, we provide a quantitative comparison of the performance of the different versions of our algorithm, against the state of the art methods listed above, using the four different versions of the rendered flag sequence as input. We report the *root mean square*

(RMS) of the endpoint error, i.e. the amplitude of the difference between the ground truth and estimated flow $\mathbf{u}(\mathbf{x}; n)$. These measures are computed over all the frames and for all the foreground pixels. Note that the results obtained with the Spline basis were omitted since they were almost equivalent to those obtained with the DCT basis, as Fig. 7a reveals.

First we compare the performance of our original algorithm for grayscale images (MFSF) with ITV-L¹ (Wedel et al. 2009), LDOF Brox and Malik (2011) and Pizarro and Bartoli (2010), since these algorithms can only be used on grayscale images. We report results for our algorithm using the full rank ($R = 2F$) DCT basis (MFSF_{DCT}) and a full rank PCA basis (MFSF_{PCA}). Note that the PCA basis was estimated using as input the flow obtained after running our algorithm with the DCT basis (MFSF_{DCT}). We also ran our algorithm using the identity matrix as the basis (MFSF_{I_{2F}}) to show the degradation of the results when subspace constraints are not applied to compute the multi-frame optical flow.

Table 1 shows that our proposed algorithms (MFSF_{PCA}) and (MFSF_{DCT}) rank top amongst the grayscale algorithms, outperforming all other methods and yielding the lowest RMS errors on all the sequences: original, occlusions, Gaussian noise and salt & pepper noise. The best results are obtained using the PCA basis.

Moreover, the top two rows of Table 1 show that using the novel extension of our algorithm to colour images (MFSF_C) described in Sect. 7 improves significantly the results in all versions of the sequence. Once more, the results obtained using a full rank PCA basis (MFSF_{CPCA}) outperform those obtained with the DCT basis (MFSF_{C_{DCT}}).

Regarding the choice of parameters, as we described in Sect. 8 the default values proposed in Wedel et al. (2009) for the ITV-L¹ algorithm were also found to give best results on our grayscale algorithm (MFSF).⁸

However, we found that these parameters needed some tuning on the noisy and occluded versions of our benchmark sequence. A lower value of the data term weight $\alpha = 18$ was found to provide best results. Additionally, on the noisy sequences, the weight of the quadratic term was lowered to $\beta = 0.4$. These modified values were used on MFSF_{PCA}, MFSF_{DCT} and MFSF_{I_{2F}}.

Figure 5 shows a visual comparison of the results on the benchmark sequence reported in Table 1. We show a closeup of the reverse warped images $\mathcal{W}^{-1}(I_i)$ of three frames in the sequence ($i = 20, 25, 30$) which should look identical to the template frame; and the error in the flow estimation ϵ_i for

the same frames, expressed in pixels, encoded as a heatmap. Notice the significant improvements that our proposed algorithms for colour images (MFSF_{CPCA}, MFSF_{C_{DCT}}) show with respect to their grayscale counterparts (MFSF_{PCA}, MFSF_{DCT}). Overall, all our approaches outperform state of the art methods: ITV-L¹ optical flow (Wedel et al. 2009); LDOF (Brox and Malik 2011) and Pizarro and Bartoli's registration algorithm (Pizarro and Bartoli 2010).

Figure 7 shows results of the experiments on the benchmark sequence with synthetic occlusions. The error maps ϵ_i for images ($i = 20, 25, 30, 60$) encoded as heatmaps are shown for all the variants of our grayscale (MFSF_{PCA}, MFSF_{DCT}) and colour (MFSF_{CPCA}, MFSF_{C_{DCT}}) algorithms as well as ITV-L¹ (Wedel et al. 2009), LDOF (Brox and Malik 2011) and Pizarro and Bartoli (2010). We notice the same behaviour as in the experiments without occlusions—the error maps obtained with our algorithms show a superior performance with respect to state of the art approaches. Amongst our proposed approaches, one can observe significant improvements of the colour versions over their grayscale equivalents.

Figure 6a shows a graph of the RMS error 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 6c establishes the same fact in the case of noisy images and sequences with occlusions. Figure 6b 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 Fig. 6a–c we used ground truth tracks to compute the PCA basis to remove the bias from tracking.

10.3 Experimental Comparison of Soft Versus Hard Subspace Constraint

In this section we use the synthetic grayscale flag sequence to conduct an experimental comparison of the optical flow obtained using our proposed soft subspace constraint with that obtained imposing the hard constraint described in Sect. 9. The energy associated with the hard constraint (59)

⁸ Note that, as we discussed in Sect. 4, MFSF_{I_{2F}} and ITV-L¹ (Wedel et al. (2009)) are equivalent algorithms and should therefore provide the same results. The difference in the numerical results is due to two factors: (i) in MFSF_{I_{2F}} $c_g = 0$ and $\epsilon \neq 0$ (ii) the ITV-L¹ algorithm was run with its default parameters and MFSF_{I_{2F}} with the tuned parameters described above.

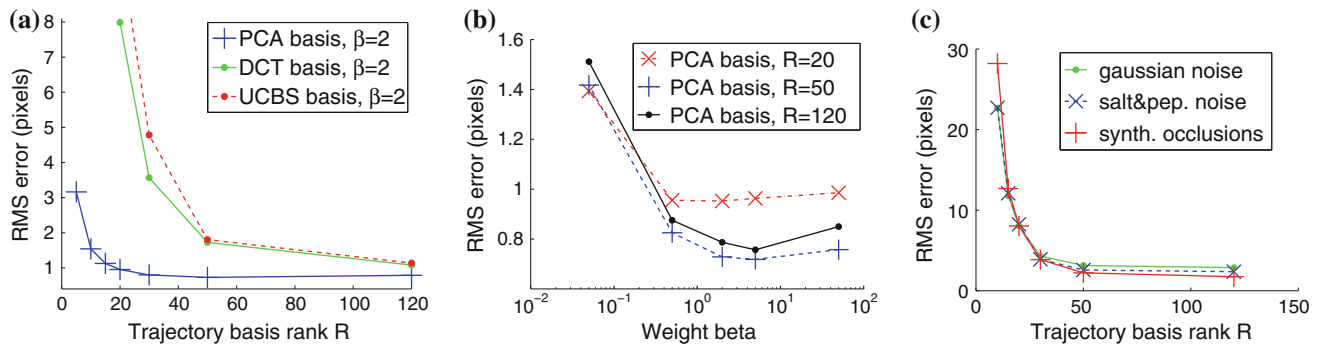


Fig. 6 (a) RMS flow error vs increasing values of the rank of the different trajectory bases (PCA, DCT, UCBS). The graph shows that the PCA motion basis provides best results and that our algorithm does not overfit when the rank of the basis is overestimated. (b) RMS flow error vs increasing values of the weight of the subspace constraint β . (c) RMS flow error for increasing value of the rank of the PCA basis on the different variants of the benchmark sequence (occlusions, Gaussian noise, salt & pepper noise). All experiments are for our grayscale multi-frame subspace flow algorithm MFSF

Table 1 RMS endpoint errors in pixels on the benchmark sequences of our proposed method for colour (MFSF_C) and grayscale (MFSF) images using different motion basis (PCA, DCT and I_{2F})

Image type	Method	Version of input sequence:			
		Original	Occlusions	Gauss. noise	S&P noise
Color	MFSF _{PCA}	0.69	0.80	1.25	1.01
	MFSF _{DCT}	0.80	1.00	1.52	1.17
	MFSF _{PCA}	0.75	0.85	1.52	1.18
	MFSF _{DCT}	0.89	1.12	1.84	1.38
Grayscale	MFSF _{I_{2F}}	1.13	1.43	1.83	1.60
	ITV-L ¹ (Wedel et al. 2009)	1.43	1.89	2.61	2.34
	LDOF (Brox and Malik 2011)	1.71	2.01	4.35	5.05
	Pizarro and Bartoli (2010)	1.24	1.27	1.94	1.79

We compare the different versions of our grayscale algorithm (MFSF) against state of the art optical flow (ITV-L¹ (Wedel et al. 2009), LDOF (Brox and Malik 2011)) and non-rigid registration (Pizarro and Bartoli 2010) methods
 Numbers in bold highlight best performing color/grayscale algorithm

can be obtained by removing the quadratic term E_{link} from our energy (6) and reparameterizing the optical flow in terms of the trajectory coefficients.

We use the primal-dual algorithm described in Appendix C to minimise the energy obtained via reparameterization (59) with 200 iterations per warp to ensure convergence. We observed that 200 iterations were enough for the convergence of the cost function to a reasonable tolerance (which we consider to be when the change in cost per iteration is <1000th of the total change).

Our energy (6) based on the soft subspace constraint, is minimized using our optimization scheme described in Sect.5. To establish a fair comparison, we used 20 denoising iterations for the regularization step and 20 alternation iterations between the minimisation of Step 1 and Step 2 to ensure convergence.

Table 2 reports the RMS endpoint error, measured in pixels, of the flow obtained with the soft (S) and hard (H) constraints using 3 different basis:

1. Low rank ($R = 75$) PCA basis obtained from sparse tracking using Pizarro and Bartoli (2010).
2. Full rank PCA basis obtained from ground truth optical flow.
3. Full rank DCT basis.

The comparative results in Table 2 show that the optical flow obtained with our soft constraint consistently outperforms the flow obtained after reparameterization (hard constraint) in all three experiments on all the different sequences (original, noisy and with occlusions). This is particularly the case in the presence of Gaussian noise when the endpoint

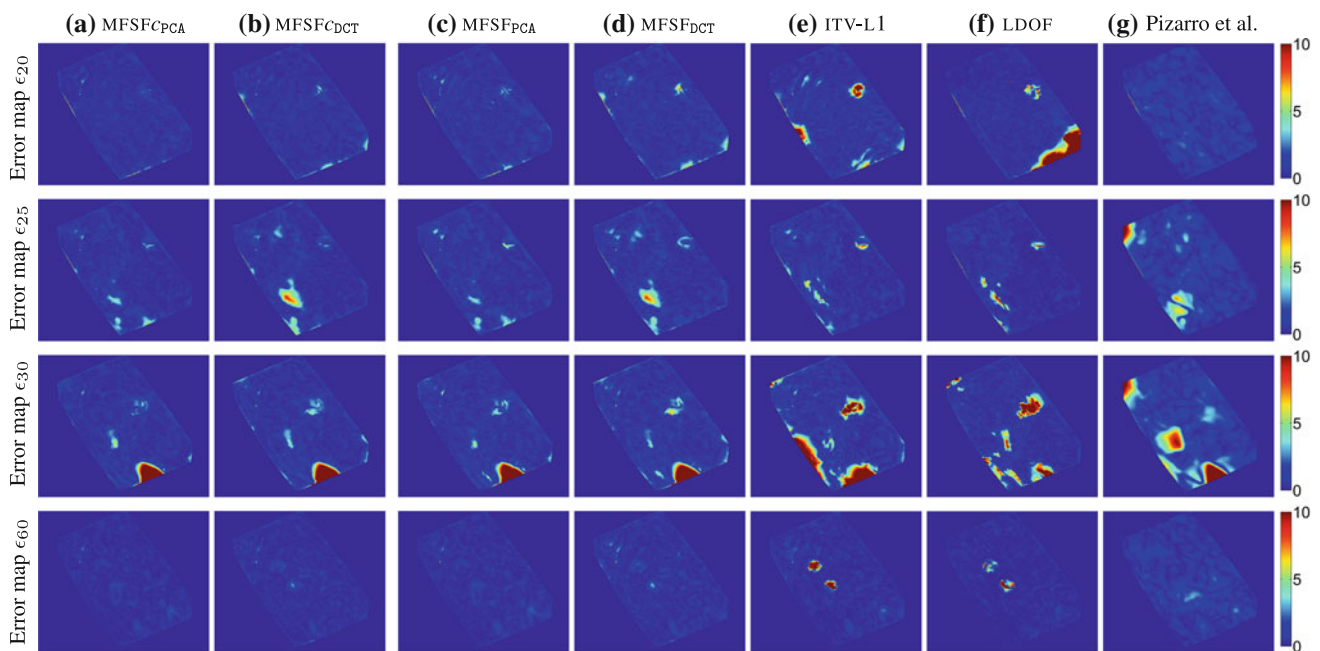


Fig. 7 Flow error maps $\epsilon_i(\mathbf{x})$ on the benchmark sequence with synthetic occlusions for frames ($i = 20, 25, 30, 60$). Each column shows results for different methods and errors are displayed as heatmaps. **(a–b)** Multi-frame subspace flow on color images: **(a)** MFSF_{CPCA}, **(b)** MFSF_{CDCT}. **(c–d)** Multi-frame subspace flow on grayscale images: **(c)**

MFSF_{PCA}, **(d)** MFSF_{DCT}. **(e)** ITV- L¹ [Wedel et al. \(2009\)](#). **(f)** LDOF [Brox and Malik \(2011\)](#) **(g)** [Pizarro and Bartoli \(2010\)](#). It is easy to see from the error maps for frames 20 or 25 that the colour versions of our algorithm **(a)** MFSF_{CPCA} and **(b)** MFSF_{CDCT} improve substantially on their grayscale counterparts **(c)** MFSF_{PCA} and **(d)** MFSF_{DCT}

Table 2 RMS endpoint error in pixels for the optical flow obtained with the hard (H) versus soft (S) constraints

Basis	Rank	Constraint	Version of input sequence:			
			Original	Occl.	Gauss. noise	S&P noise
Sparse PCA	75	SOFT (S)	0.90	1.01	1.80	1.46
		HARD (H)	0.98	1.05	2.22	1.60
GT PCA	120	SOFT (S)	0.69	0.76	1.43	1.07
		HARD (H)	0.70	0.77	1.65	1.08
DCT	120	SOFT (S)	0.89	1.12	1.83	1.38
		Hard (H)	1.09	1.28	2.00	1.42

We carry out 3 experiments using: (top) a low-rank sparse PCA basis (using tracks given by [Pizarro and Bartoli \(2010\)](#)); (middle) a full rank ground truth PCA basis (computed using the ground truth optical flow); and (bottom) a full rank DCT basis. The algorithms were tested on all the different types of sequence (original, noisy and with occlusions)

errors differ most. However, this is to be expected since our soft constraint allows some deviations from the subspace manifold.

In the first experiment we used a low rank PCA basis estimated from sparse tracking (obtained using [Pizarro and Bartoli's](#) matching algorithm ([Pizarro and Bartoli 2010](#))) to test the case of an inaccurate basis. This is the case when it is most clearly beneficial to allow deviations from the subspace manifold. This is naturally reflected on significantly higher endpoint errors on the flow computed with the hard constraint compared with that computed with our soft constraint.

It is also interesting to observe that even in the case when we used the full rank PCA basis computed from the ground truth flow the soft constraint performs marginally better than the hard constraint. In the sequence with Gaussian noise it provides a more clear benefit. Finally, the third experiment with a full rank DCT basis also shows that it is beneficial to use a soft constraint in all the different image sequences.

In conclusion, the optical flow obtained using the subspace constraint as a soft constraint consistently outperforms the flow obtained by reparameterization when both algorithms were ran until convergence. The benefits of the soft constraint

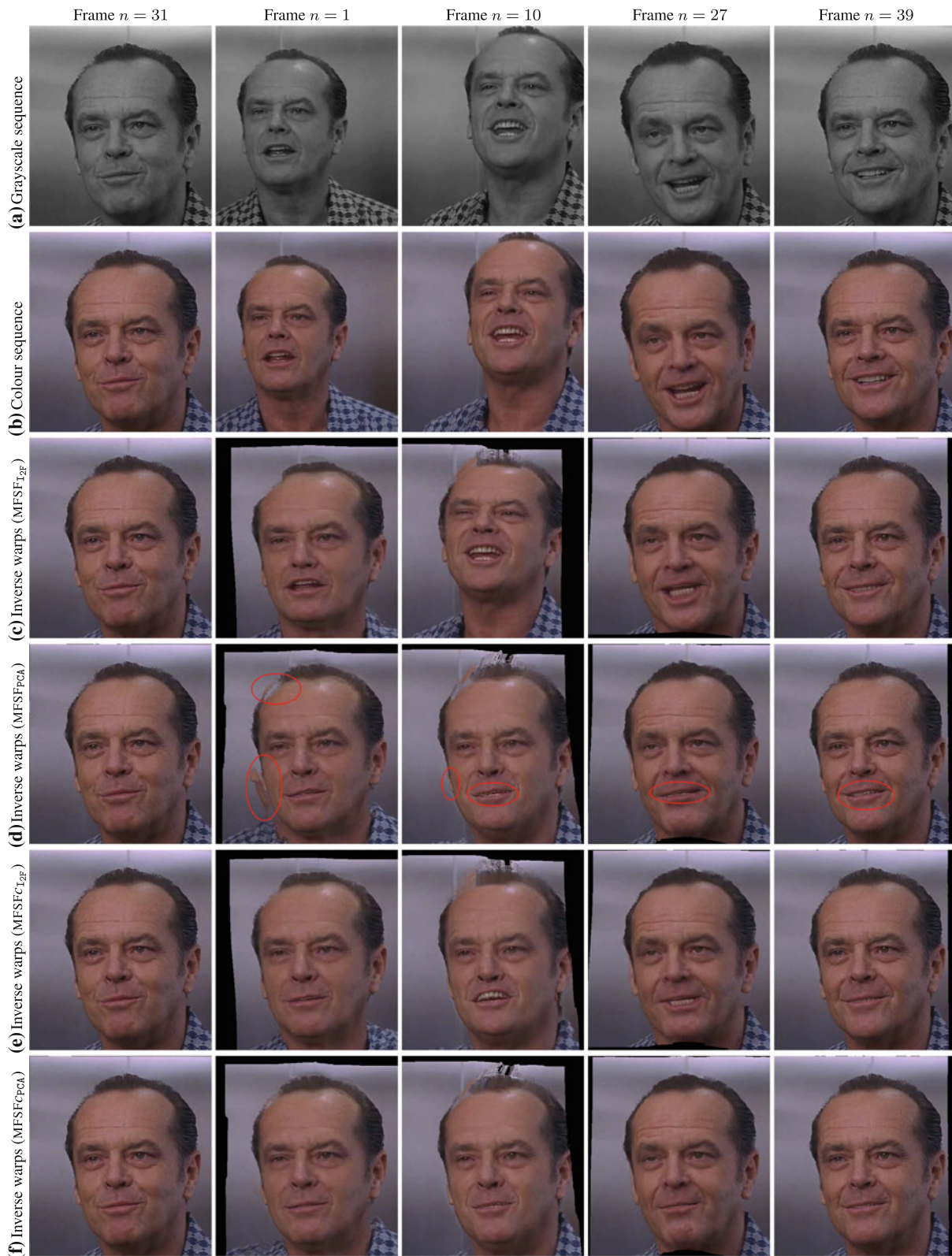


Fig. 8 Results on the Actor sequence: (a–b) Some frames of the grayscale and colour input sequences. This is a challenging sequence with large displacements and strong deformations. Frame 31 I_{31} is used as the reference frame. (c–d) Inverse warp images $\mathcal{W}^{-1}(I_n)$ comparing two versions of our grayscale algorithm: c without subspace constraints

(MFSF_{I_{2F}}) and (d) with subspace constraints (MFSF_{I_{2F}}). (e–f) Inverse warp images $\mathcal{W}^{-1}(I_n)$ comparing two versions of our colour algorithm: (e) without subspace constraints (MFSF_{I_{2F}}) and (f) with subspace constraints (MFSF_{PCA})



Fig. 9 Results on the grayscale Actor sequence: *Top row (a)* shows some frames of the original grayscale sequence. *Middle (b)* and *bottom (c)* rows compare the optical flow results obtained with two of our proposed grayscale algorithms: (c) with subspace constraints (MFSF_{PCA}) and (b) without subspace constraints (MFSF_{I_{2F}}). The flow is visualized

with a grid superimposed on the images to reveal the optical flow in a sparse subset of points. Points on the mouth are shown in yellow to highlight the results on the area with strongest deformations (Color figure online)



Fig. 10 Results on the colour Actor sequence: *Top row (a)* shows some frames of the original colour sequence. *Middle (b)* and *bottom (c)* rows compare the optical flow results obtained with two of our proposed colour algorithms: (c) with subspace constraints (MFSFc_{PCA}) and (b)

without subspace constraints (MFSFc_{I_{2F}}). The flow is visualized with a grid superimposed on the images to reveal the optical flow in a sparse subset of points. Points on the mouth are shown in yellow to highlight the results on the area with strongest deformations (Color figure online)



Fig. 11 Results on the Actress sequence: **(a)** Some frames of the original grayscale sequence. **(b)** Inverse warp images obtained with our best performing grayscale method using subspace constraints ($MFSF_{PCA}$). **(c)** Original images augmented with some texture to simulate a tattoo

are stronger when dealing with noisy images and in the case of an inaccurate motion basis which is to be expected.

10.4 Experiments on Real Sequences

In this section we provide details about the experiments we have carried out on four video sequences which display large displacements and strong deformations.

10.4.1 Actor sequence

This challenging sequence is a 39 frame long clip from a well known film, acquired at 25 frames per second with images of size 500×550 pixels. The top two rows of Fig. 8 show 5 frames of this sequence in grayscale and colour. Note that frame 31 was used as the reference frame⁹. The bottom four rows in Fig. 8 show comparative results of the inverse warp images (using the computed optical flow to warp the current image back to the reference frame) estimated using the following different versions of our algorithm: $MFSF_{I_{2F}}$, $MFSF_{PCA}$, $MFSFC_{I_{2F}}$, $MFSFC_{PCA}$. The first two methods work on grayscale images and use the identity matrix and PCA basis as the motion basis respectively while the last two are

their equivalent colour versions. Comparing the results of $MFSF_{I_{2F}}$ and $MFSF_{PCA}$ (or $MFSFC_{I_{2F}}$ and $MFSFC_{PCA}$) allows us to show the advantages of using subspace constraints (PCA basis) versus not using a temporal model for the trajectories (I_{2F} basis). We use a full rank PCA basis obtained after applying principal components analysis to an initial flow estimated with our algorithm using the DCT basis.

The advantages of using subspace constraints are clear. For instance, notice that for grayscale images $MFSF_{I_{2F}}$ failed completely to warp frame 10 while $MFSF_{PCA}$ provides an accurate inverse warp image for the same frame and consistently superior results throughout the sequence. It is also clear that making use of all three colour channels using the extension of our algorithm to vector valued images provides substantial improvements. Both $MFSFC_{I_{2F}}$ and $MFSFC_{PCA}$ outperform their grayscale equivalents. In row (d) of Fig. 8 we have highlighted in red areas where the flow has clearly failed on the grayscale $MFSF_{PCA}$ algorithm but have been correctly warped in its colour version $MFSFC_{PCA}$.

Notice also that $MFSFC_{I_{2F}}$ copes with the large displacements in frame 10 much better than $MFSF_{I_{2F}}$. However, just using colour without subspace constraints is not enough to estimate accurate flow. Comparing the bottom two rows of Fig. 8 reveals that using subspace constraints significantly improves results also in the case of colour. In conclusion, the best overall results are obtained with $MFSFC_{PCA}$, our colour algorithm with subspace constraints using the PCA basis.

⁹ We choose the reference frame to be one in which the points we are interested in tracking are all visible and also to reduce the maximum displacements.

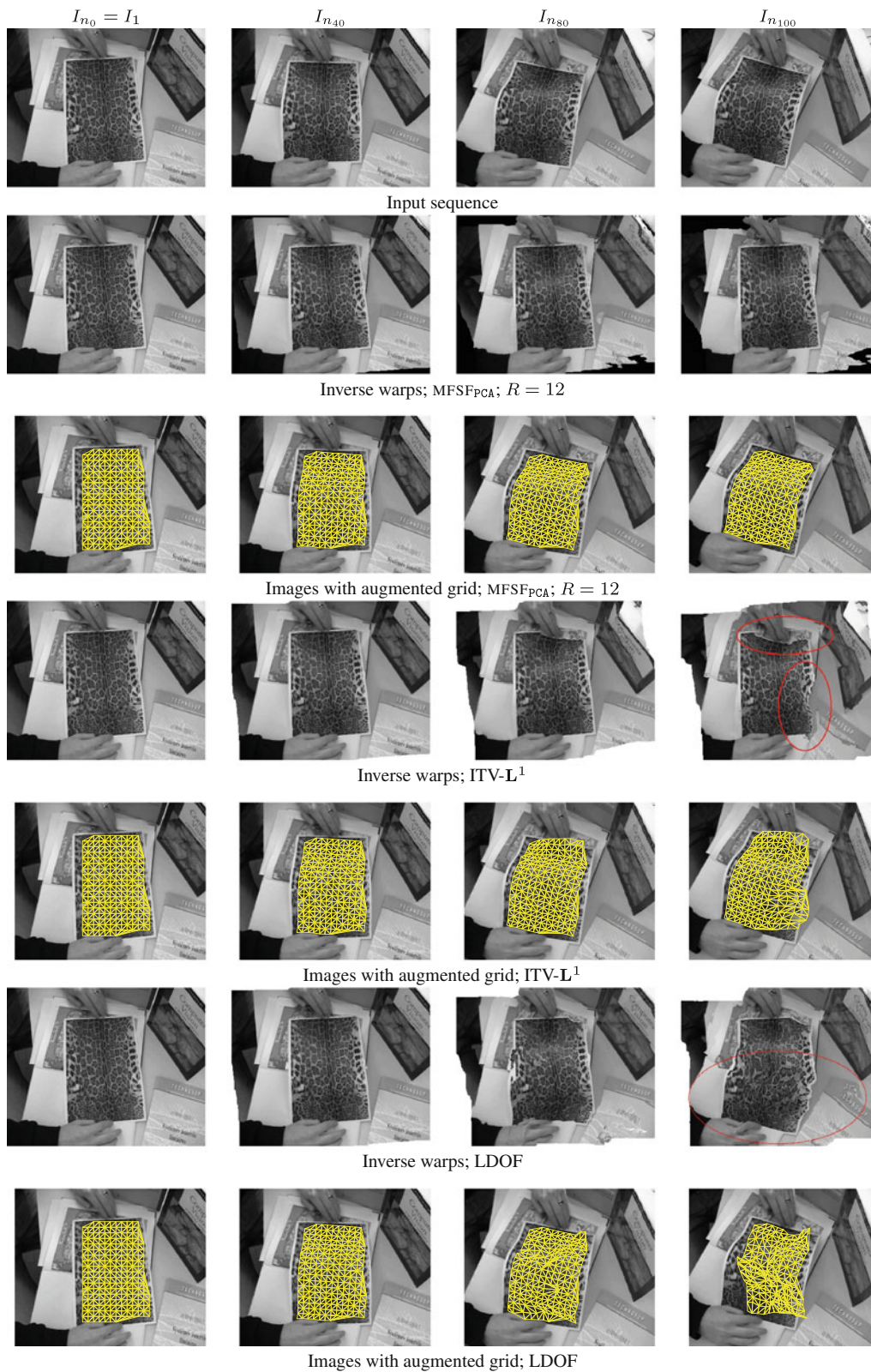


Fig. 12 Results on Paper Bending-1 grayscale sequence: Comparative results of the optical flow estimated with our best performing grayscale algorithm (MFSF_{PCA}) against state of the art optical flow methods (ITV-L¹ (Wedel et al. 2009), LDOF (Brox and Malik 2011)). We

show two visualizations of the optical flow estimated with the three methods in alternate rows: (i) the inverse warped images and (ii) a grid superimposed on the images to reveal the optical flow in a sparse subset of points. *Top row* shows some frames of the original sequence

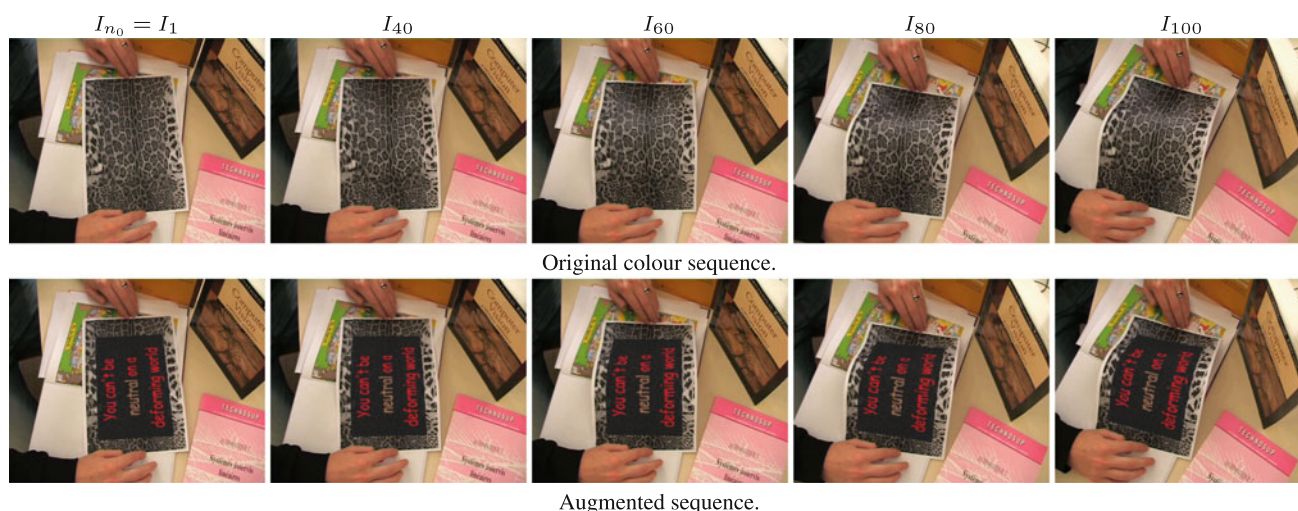


Fig. 13 Results on Paper Bending-1 colour sequence: The *top* row shows some frames of the original colour sequence. The *bottom* row displays the same sequence augmented with some new texture. The optical

flow obtained with our best performing colour algorithm MFSF_{PCA} was used to re-texture the original sequence

Figures 9 and 10 support our claims by showing a grid superimposed on the images to reveal the optical flow in a sparse subset of points. The points on the mouth are highlighted in yellow since that is where most of the deformations occur. Once more, Fig. 9 reveals that the quality of the flow computed using trajectory regularization constraints on grayscale images (MFSF_{PCA}) is far better than that obtained without using subspace constraints ($\text{MFSF}_{\text{T}_2\text{F}}$). Notice the complete failure of $\text{MFSF}_{\text{T}_2\text{F}}$ on frame 10. Similar conclusions can be drawn from the results on the colour images shown in Fig. 10. Notice the improvements particularly on the lips.

10.4.2 Actress sequence

This 72 frame long clip from the same film shows a close-up of an actress opening the mouth widely. The resolution of the images was 640×360 pixels. This sequence is similarly challenging to the previous one with very large displacements and deformations. In this case we only ran our best performing method on grayscale images MFSF_{PCA} with subspace constraints using a PCA basis of rank $R = 100$. Figure 11 shows the original sequence (top row); the inverse warp images estimated from the optical flow (middle row) and the original images augmented with some texture (bottom row) to simulate a tattoo.

10.4.3 Paper bending-1 sequence

Figure 12 shows results on a sequence of textured paper bending smoothly (Bartoli et al. 2008); a challenging sequence due to its length (100 frames) and the large camera rotation. We show results comparing our best performing grayscale

algorithm (MFSF_{PCA}) against state of the art optical flow methods (ITV- \mathbf{L}^1 (Wedel et al. 2009), LDOF (Brox and Malik 2011)). For completeness in our experimental evaluation, in this case we computed the motion basis by applying PCA to KLT tracks (Lucas and Kanade 1981) keeping the first 12 components. We ran the LDOF and ITV- \mathbf{L}^1 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, especially after some frames (see e.g. the artifacts annotated by the red ellipses in the results of LDOF and ITV- \mathbf{L}^1). We show an alternative visualization of the same results with a grid superimposed on the images to reveal the optical flow in a sparse subset of points. This visualization helps to highlight the superiority of the optical flow estimated with our algorithm (MFSF_{PCA}) with respect to others.

In Fig. 13 we show results on the colour version of this sequence, subsampled taking every fifth frame to give a 25 frame long sequence. In this case, we augment the images with new texture using the optical flow results given by our colour multi-frame subspace algorithm using a PCA basis (MFSF_{PCA}). In this case we use a full rank PCA basis obtained after applying principal components analysis to an initial flow estimated with our algorithm using the DCT basis (MFSF_{DCT}).

10.4.4 Paper bending-2 sequence

Figure 14 shows a 71 frame long grayscale sequence introduced in Varol et al. (2009) of a paper being bent backwards

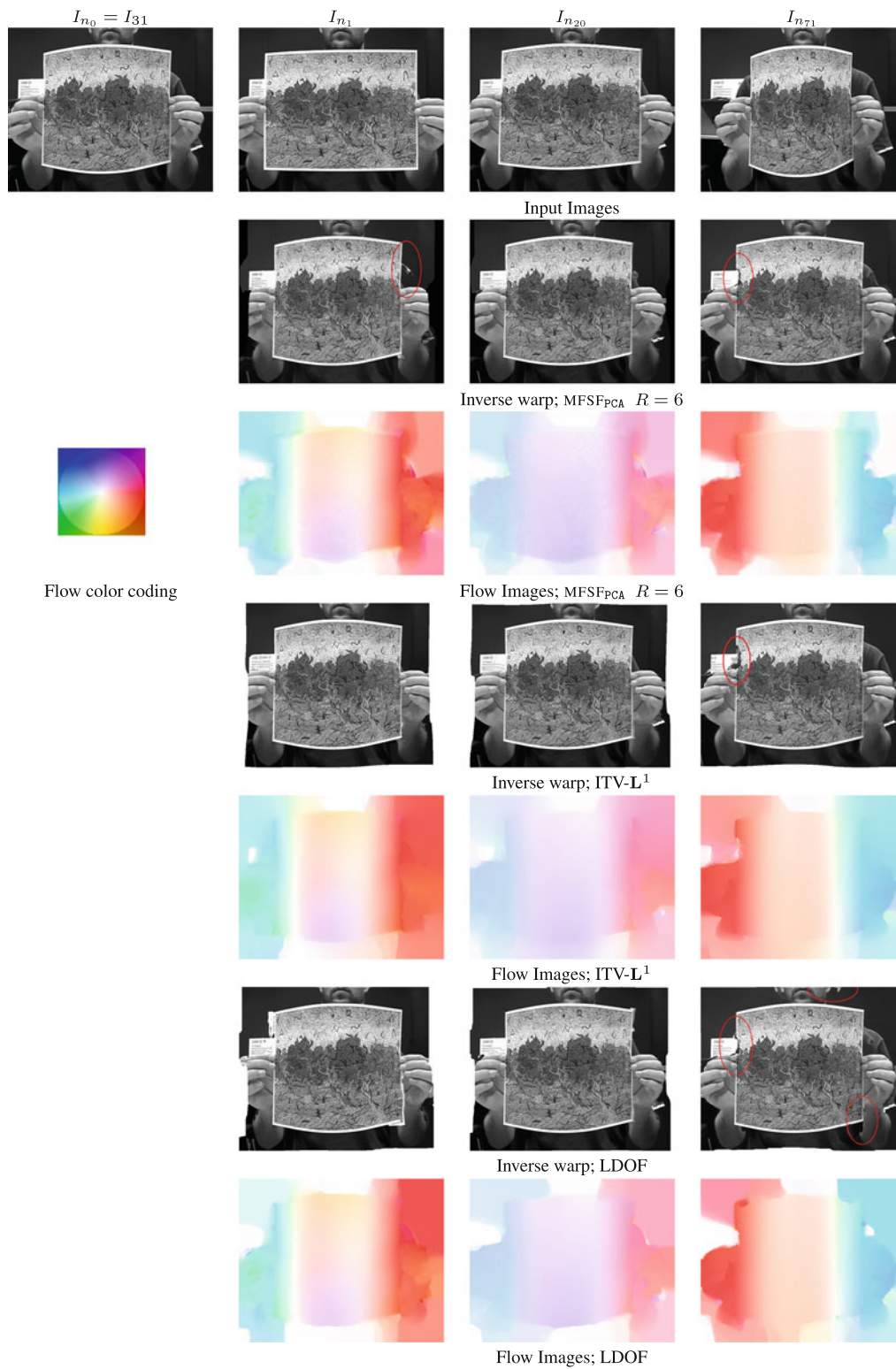


Fig. 14 Results on the Paper bending-2 sequence: *Top row* shows some images of this grayscale sequence. The 30th frame is used as the reference. Next rows show inverse warp images and colour coded optical flow comparing our best performing grayscale algorithm (MFSF_{PCA})

using a very low rank PCA decomposition ($R = 6$) against state of the art optical flow methods (ITV-L¹ (Wedel et al. 2009), LDOF (Brox and Malik 2011))

which is widely used for 3D reconstruction in non-rigid structure from motion (NRSfM). Our method used a PCA basis of rank $R = 6$ obtained from KLT tracks. The 30th frame is used as the reference. Once more, we compare results of our algorithm (MFSF_{PCA}) against the same state of the art approaches as in previous experiments. The inverse warped images and the colour coded optical flow in Fig. 14 reveal that despite having used a very low rank PCA motion basis, our results outperform LDOF and provide more accurate flow boundaries than ITV-L¹.

11 Conclusions

We have provided a new formulation for the computation of multi-frame optical flow exploiting the high correlation between 2D trajectories of points in a long sequence by assuming that these lie close to a low dimensional subspace. Our main contribution is to formulate the manifold constraint as a *soft constraint* which, using variational principles, leads to a *robust* energy that can be efficiently optimized. We propose a new *anisotropic trajectory regularization* term that acts on the coefficients of the trajectory basis. We take advantage of the high level of parallelism inherent to our approach by developing a GPU implementation using the Nvidia CUDA framework. We also provide an extension of our approach to the case of vector-valued images which allows us to exploit all three colour channels and gain substantial improvements in the accuracy of the estimated optical flow. We also provide a new benchmark dataset, with ground truth optical flow. Our experimental results on the benchmark dataset and on real video footage reveal that using subspace constraints significantly improves results. Our approach outperforms state of the art optical flow and non-rigid registration algorithms.

Acknowledgments This work is supported by the European Research Council under ERC Starting Grant agreement 204871-HUMANIS. We thank T. Collins for his texture mapping code and D. Pizarro for providing results of their method (Pizarro and Bartoli 2010) and tracks for the synthetic sequence. We also thank A. Handa and L. Pizarro for fruitful discussions.

Open Access This article is distributed under the terms of the Creative Commons Attribution License which permits any use, distribution, and reproduction in any medium, provided the original author(s) and the source are credited.

Appendix A: Primal Dual Algorithm for Denoising

This appendix describes the optimization of the energy minimized in Step 1 of our algorithm as defined in (22):

$$E_d = \int_{\Omega} g(x) |\nabla L(x)|_{\epsilon} + \beta(M(x) - L(x))^2 dx \quad (36)$$

which corresponds to a small modification of the TV-L² Rudin-Osher-Fatemi (ROF) model (Rudin et al. 1992), as described in Sect. 5.1. Note that as the trajectory model coefficients $L_i(x)$ in (22) are decoupled for each i , in the following derivation we have dropped the subscript for simplicity.

The first step in the optimization is the dualisation of the weighted Huber functional $g(x)H_{\epsilon}(|\nabla L(x)|^2)$ of the above energy with respect to the gradient $\nabla L(x)$ using its Legendre-Fenchel transform (Rockafellar 1997). After spatial discretization, the minimisation of (36) is equivalent to the following saddle point problem:

$$\min_L \max_{\mathcal{L}} \sum_{x \in X} \left\{ \langle \nabla L(x), \mathcal{L}(x) \rangle - \delta \left(\frac{\mathcal{L}(x)}{g(x)} \right) - \epsilon \frac{|\mathcal{L}(x)|^2}{2g(x)} + \beta(M(x) - L(x))^2 \right\}. \quad (37)$$

where X is the set of image grid points, ∇ denotes the discrete gradient operator as defined in Chambolle and Pock (2011), $\mathcal{L}(x) \in \mathbb{R}^2$ are the dual variables for every (x) , and $\delta(\mathcal{L})$ is the indicator function of the unit ball:

$$\delta(\mathcal{L}) \triangleq \begin{cases} 0 & \text{if } |\mathcal{L}| \leq 1 \\ \infty & \text{otherwise} \end{cases} \quad (38)$$

The problem (37) can be considered as a special case of the following general form of primal-dual problems that are studied in Chambolle and Pock (2011):

$$\min_p \max_q \langle Kp, q \rangle - F^*(q) + G(p). \quad (39)$$

In the case of (37), the norm of the linear operator $K = \nabla$ is bounded by $\sqrt{8}$. Also, both G and F^* are uniformly convex with convexity parameters 2β and ϵ respectively.

Therefore, we solve (37) by applying Algorithm 3 of Chambolle and Pock (2011). The steps of the algorithm can be written as follows :

- Initialize $\mathcal{L}^0 = \mathbf{0}$, $L^0(x) = \bar{L}^0(x) = M(x)$
- Iterate for $k = 0, 1, 2, \dots$ until a convergence criterion is satisfied:

$$\mathcal{L}^{k+1}(x) = g(x) \Pi \left(\frac{\mathcal{L}^k(x) + \tau \nabla \bar{L}^k(x)}{g(x) + \tau \epsilon} \right) \quad (40)$$

$$L^{k+1}(x) = \frac{1}{1+2\sigma\beta} \left(2\sigma\beta M(x) + L^k(x) + \sigma \operatorname{div}(\mathcal{L}^{k+1}(x)) \right) \quad (41)$$

$$\bar{L}^{k+1}(x) = 2L^{k+1}(x) - L^k(x) \quad (42)$$

where $\operatorname{div}(\cdot)$ is the discrete divergence operator and the operator $\Pi(s)$ projects a vector s onto the unit ball as:

$$\Pi(s) = \frac{s}{\max(1, |s|)} \quad (43)$$

We choose the following values for the steps σ, τ , that guarantees the convergence:

$$\sigma = \sqrt{\frac{\epsilon}{16\beta}}, \tau = \sqrt{\frac{\beta}{4\epsilon}} \tag{44}$$

Appendix B: Primal Dual Algorithm for Robust Vector-Valued Image Matching

This appendix provides the details of the algorithm to optimise the saddle point problem (32) for vector-valued images using Euclidean norm and Huber penalisers.

Euclidean Norm Penaliser

This case corresponds to $\Phi(\mathbf{v}) = |\mathbf{v}|$ and is a straightforward extension of the absolute value of image differences that we used for E_{data} in (7) for grayscale images. After dualisation, (32) can be written as:

$$\min_u \max_{\mathcal{I}} \left\{ \langle \alpha \mathbf{A} \mathbf{u}, \mathcal{I} \rangle + \alpha \langle \mathbf{b}, \mathcal{I} \rangle - \delta(\mathcal{I}) + \beta |\mathbf{u} - \mathbf{u}'|^2 \right\} \tag{45}$$

This problem is also a special case of the general saddle point problem (39) with the linear operator $K = \alpha \mathbf{A}$. Since the function $\beta |\mathbf{u} - \mathbf{u}'|^2$ is uniformly convex with convexity parameter 2β , we apply Algorithm 2 of Chambolle and Pock (2011) and derive following optimisation algorithm:

- Choose $\sigma_0 = \tau_0 = \frac{1}{\alpha B_A}$
- Initialize \mathbf{u}^0 from the previous alternation iteration.
- Initialize $\bar{\mathbf{u}}^0 = \mathbf{u}^0, \mathcal{I}^k = \mathbf{0}$.
- Iterate for $k = 0, 1, 2, \dots$ until a convergence criterion is satisfied:

$$\mathcal{I}^{k+1} = \Pi(\mathcal{I}^k + \alpha \tau_k (\mathbf{A} \bar{\mathbf{u}}^k + \mathbf{b})) \tag{46}$$

$$\mathbf{u}^{k+1} = \frac{1}{1 + 2\sigma_k \beta} \left(2\sigma_k \beta \mathbf{u}' + \mathbf{u}^k - \sigma_k \mathbf{A}^T \mathcal{I}^{k+1} \right), \tag{47}$$

$$\theta_k = \frac{1}{\sqrt{1 + 4\beta \sigma_k}}, \quad \sigma_{k+1} = \theta_k \sigma_k, \quad \tau_{k+1} = \frac{\tau_k}{\theta_k} \tag{48}$$

$$\bar{\mathbf{u}}^{k+1} = \mathbf{u}^{k+1} + \theta_k (\mathbf{u}^{k+1} - \mathbf{u}^k) \tag{49}$$

where B_A can be any upper bound on the norm of \mathbf{A} . Although the saddle point problem is minimised separately for each spatio-temporal point of the video and \mathbf{A} is spatially varying, for simplicity we choose a common upper bound on the linear operator for all the points. It can be shown that L_A as defined below is a valid upper bound.

$$B_A = \sqrt{\max_n \sum_{i=1}^{N_c} \left(\max_x \left| \frac{\partial \mathbf{I}_i(\mathbf{x}; n)}{\partial x} \right|^2 + \max_x \left| \frac{\partial \mathbf{I}_i(\mathbf{x}; n)}{\partial y} \right|^2 \right)} \tag{50}$$

where (x, y) are the horizontal and vertical coordinate axes of the image plane.

Huber Penaliser

When the robust function used in the data term of the energy for vector-valued images is the Huber norm: $\Phi(\mathbf{v}) = H_\epsilon(|\mathbf{v}|^2)$, the saddle point problem (32) can be written as:

$$\min_u \max_{\mathcal{I}} \left\{ \langle \mathbf{A} \mathbf{u}, \mathcal{I} \rangle + \langle \mathbf{b}, \mathcal{I} \rangle - \frac{\epsilon}{2\alpha} |\mathcal{I}|^2 - \delta\left(\frac{\mathcal{I}}{\alpha}\right) + \beta |\mathbf{u} - \mathbf{u}'|^2 \right\} \tag{51}$$

This problem is again of the form (39) with the linear operator $K = \mathbf{A}$. The corresponding G and F^* functions are both uniformly convex with parameters 2β and $\frac{\epsilon}{\alpha}$. We thus solve (51) using Algorithm 3 of Chambolle and Pock (2011) and derive the following optimisation algorithm:

- Initialize \mathbf{u}^0 from the previous alternation iteration.
- Initialize $\bar{\mathbf{u}}^0 = \mathbf{u}^0, \mathcal{I}^k = \mathbf{0}$.
- Iterate for $k = 0, 1, 2, 3, \dots$ until a convergence criterion is satisfied:

$$\mathcal{I}^{k+1} = \alpha \Pi \left(\frac{\mathcal{I}^k + \tau (\mathbf{A} \bar{\mathbf{u}}^k + \mathbf{b})}{\alpha + \tau \epsilon} \right) \tag{52}$$

$$\mathbf{u}^{k+1} = \frac{1}{1 + 2\sigma \beta} \left(2\sigma \beta \mathbf{u}' + \mathbf{u}^k - \sigma \mathbf{A}^T \mathcal{I}^{k+1} \right), \tag{53}$$

$$\bar{\mathbf{u}}^{k+1} = 2\mathbf{u}^{k+1} - \mathbf{u}^k \tag{54}$$

We choose the following step-sizes which ensure the convergence of our algorithm:

$$\sigma = \frac{1}{B_A} \sqrt{\frac{\epsilon}{2\beta\alpha}}, \tau = \frac{1}{B_A} \sqrt{\frac{2\beta\alpha}{\epsilon}} \tag{55}$$

where B_A is, again, any upper bound on the operator norm of \mathbf{A} . As in the case of Euclidean norm penalisation, we choose B_A as defined in (50).

Appendix C: Optimization of the Hard Subspace Constraint

This appendix describes the optimization of the energy

$$E_h = \int_{\Omega} \sum_{n=1}^F |I(\mathbf{x} + \mathcal{Q}_n \mathbf{L}(\mathbf{x}); n) - I(\mathbf{x}; n_0)| \, d\mathbf{x} + \int_{\Omega} \sum_{i=1}^R g(\mathbf{x}) |\nabla L_i(\mathbf{x})|_{\epsilon} \, d\mathbf{x} \tag{56}$$

which corresponds to the case when the subspace constraint is imposed as a hard constraint and the 2D flow $\mathbf{u}(\mathbf{x}; n)$ can be

reparameterized as $\sum_{i=1}^R \mathbf{q}_i(n)L_i(\mathbf{x})$. First, each image channel of $\mathbf{I}(\cdot; n)$ is linearised around $Q_n L_0(\mathbf{x})$, using an initial estimate $L_0(\mathbf{x})$. Under this approximation the data term can be written as:

$$E_{data} = \int_{\Omega} \sum_{n=1}^F |\mathbf{C}(\mathbf{x}; n)\mathbf{L}(\mathbf{x}) + \mathbf{d}(\mathbf{x}; n)| \, d\mathbf{x} \tag{57}$$

where, for every spatio-temporal point $(\mathbf{x}; n)$,

$$\mathbf{C}(\mathbf{x}; n) = \frac{\partial \mathbf{I}(\mathbf{x} + Q_n L_0(\mathbf{x}); n)}{\partial \mathbf{x}} Q_n \tag{58}$$

is the $N_c \times R$ Jacobian matrix and $\mathbf{d}(\mathbf{x}; n) = \mathbf{I}(\mathbf{x} + Q_n L_0(\mathbf{x}); n) - \mathbf{I}(\mathbf{x}; n_0) - \mathbf{C}(\mathbf{x}; n)L_0(\mathbf{x})$ is a N_c dimensional vector.

Thus, the following minimization problem must be solved:

$$\min_{L(\mathbf{x})} \int_{\Omega} \left\{ \alpha \sum_{n=1}^F |\rho(L(\mathbf{x}); n)| + \sum_{i=1}^R g(\mathbf{x}) |\nabla L_i(\mathbf{x})|_{\epsilon} \right\} d\mathbf{x}. \tag{59}$$

where $\rho(L(\mathbf{x}); n) = \mathbf{C}(\mathbf{x}; n)\mathbf{L}(\mathbf{x}) + \mathbf{d}(\mathbf{x}; n)$ is the linearised color constancy. After dualisation of the data and regularisation terms and spatial discretization, the minimisation (59) is equivalent to the following saddle point problem:

$$\min_L \max_{\mathcal{I}, \mathcal{L}} \sum_{\mathbf{x} \in X} \left\{ \alpha \sum_{n=1}^F (\langle \rho(L(\mathbf{x}); n), \mathcal{I}(\mathbf{x}; n) \rangle - \delta(\mathcal{I}(\mathbf{x}; n))) + \sum_{i=1}^R \left(\langle \nabla L_i(\mathbf{x}), \mathcal{L}_i(\mathbf{x}) \rangle - \delta \left(\frac{\mathcal{L}_i(\mathbf{x})}{g(\mathbf{x})} \right) - \epsilon \frac{|\mathcal{L}_i(\mathbf{x})|^2}{2g(\mathbf{x})} \right) \right\} \tag{60}$$

where $\mathcal{I}(\mathbf{x}; n) \in \mathbb{R}^{N_c}$ and $\mathcal{L}_i(\mathbf{x}) \in \mathbb{R}^2$ are the dual variables for every $(\mathbf{x}; n)$ and $(\mathbf{x}; i)$ respectively.

The energy (60) can be considered as a special case of the general form of primal-dual problem (39) where the linear operator K is the $(N_c F + 2R)N_p \times RN_p$ dimensional matrix:

$$K = \begin{bmatrix} \alpha \tilde{\mathbf{C}} \\ \nabla \end{bmatrix}; \quad \tilde{\mathbf{C}} = \begin{bmatrix} \mathbf{C}(\mathbf{x}_1; 1) \\ \vdots \\ \mathbf{C}(\mathbf{x}_1; F) & & \\ & \ddots & \\ & & \mathbf{C}(\mathbf{x}_{N_p}; 1) \\ & & \vdots \\ & & \mathbf{C}(\mathbf{x}_{N_p}; F) \end{bmatrix} \tag{61}$$

where $\mathbf{x}_1, \dots, \mathbf{x}_{N_p}$ are the image grid points and $n \in \{1, \dots, F\}$.

Thus, we solve (52) by applying Algorithm 1 of Chambolle and Pock (2011). In this case, the steps of this algorithm can be written as follows :

- Initialize $L^0(\mathbf{x}) = \bar{L}^0(\mathbf{x}) = L_0(\mathbf{x})$
- Initialize $\mathcal{I}^0(\mathbf{x}; n) = \mathcal{L}^0(\mathbf{x}; n) = \mathbf{0}$
- Iterate for $k = 0, 1, 2, \dots$ until a convergence criterion is satisfied:

$$\mathcal{I}^{k+1}(\mathbf{x}; n) = \Pi \left(\mathcal{I}^k(\mathbf{x}; n) + \tau \alpha \rho(\bar{L}^k(\mathbf{x}); n) \right) \tag{62}$$

$$\mathcal{L}_i^{k+1}(\mathbf{x}) = g(\mathbf{x}) \Pi \left(\frac{\mathcal{L}_i^k(\mathbf{x}) + \tau \nabla \bar{L}_i^k(\mathbf{x})}{g(\mathbf{x}) + \tau \epsilon} \right) \tag{63}$$

$$L_i^{k+1}(\mathbf{x}) = L_i^k(\mathbf{x}) - \sigma \alpha \sum_{n=1}^F C(\mathbf{x}; n)^T \mathcal{I}^{k+1}(\mathbf{x}; n) + \sigma \operatorname{div}(\mathcal{L}_i^{k+1}(\mathbf{x})) \tag{64}$$

$$\bar{L}^{k+1}(\mathbf{x}) = 2L^{k+1}(\mathbf{x}) - L^k(\mathbf{x}) \tag{65}$$

We use the following step-sizes, which guarantee the convergence of this algorithm too:

$$\sigma = \tau = \frac{1}{B_K} \tag{66}$$

B_K is the following upper bound on the operator norm of K (61):

$$B_K = \sqrt{8 + \alpha^2 B_A^2} \tag{67}$$

where B_A is given by (50).

References

Akhter, I., Sheikh, Y., Khan, S., & Kanade, T. (2008). Nonrigid structure from motion in trajectory space. In *Neural Information Processing Systems*, pp. 41–48.

Akhter, I., Sheikh, Y., Khan, S., & Kanade, T. (2011). Trajectory space: A dual representation for nonrigid structure from motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(7), 1442–1456.

Alvarez L., Esclarín J., Lefébure M., Sánchez J. (1999). A PDE model for computing the optical flow. In *Proceedings of the XVI Congreso de Ecuaciones Diferenciales y Aplicaciones* (pp. 1349–1356). Las Palmas de Gran Canaria, Spain, Sept. 1999.

Alvarez, L., Weickert, J., & Sánchez, J. (Aug. 2000). Reliable estimation of dense optical flow fields with large displacements. *International Journal of Computer Vision*, 39(1), 41–56.

Aubert, G., Deriche, R., & Kornprobst, P. (1999). Computing optical flow via variational techniques. *SIAM Journal on Applied Mathematics*, 60(1), 156–182.

Baker, S., Scharstein, D., Lewis, J., Roth, S., Black, M., & Szeliski, R. (2011). A database and evaluation methodology for optical flow. *International Journal of Computer Vision*, 92, 1–31.

Bartoli, A., Gay-Bellile, V., Castellani, U., Peyras, J., Olsen, S. I., & Sayd, P. (2008). Coarse-to-fine low-rank structure-from-motion. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8.

Blomgren, P., & Chan, T. (1998). Color TV: Total variation methods for restoration of vector-valued images. *IEEE Transactions on Image Processing*, 7(3):304–309, Special issue on partial differential equations and geometry-driven diffusion in image processing and analysis.

Brand, M. (2001). Morphable models from video. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 456–463.

- Bregler, C., Hertzmann, A., & Biermann, H. (2000). Recovering non-rigid 3D shape from image streams. In *IEEE Conference on Computer Vision and Pattern Recognition*, Vol. 2, pp. 690–696.
- Brox, T., Bruhn, A., Papenberg, N., & Weickert, J. (2004). High accuracy optical flow estimation based on a theory for warping. In T. Pajdla & J. Matas (Eds.), *European Conference on Computer Vision—ECCV 2004* (pp. 25–36). Part IV, volume 3024 of LNCS Berlin: Springer.
- Brox, T., & Malik, J. (2011). Large displacement optical flow: Descriptor matching in variational motion estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(3), 500–513.
- Chambolle, A. (2004). An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20, 89–97.
- Chambolle, A., & Pock, T. (2011). A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1), 120–145.
- Deriche, R., Kornprobst, P., & Aubert, G. (1995). Optical-flow estimation while preserving its discontinuities: A variational approach. In *Proceedings of the Second Asian Conference on Computer Vision* (Vol. 2, pp. 290–295). Singapore, Dec. 1995.
- Esser, E., Zhang, X., & Chan, T. F. (2010). A general framework for a class of first order primal-dual algorithms for convex optimization in imaging science. *SIAM Journal on Imaging Sciences*, 3(4), 1015–1046.
- Garg, R., Pizarro, L., Rueckert, D., & Agapito, L. (2010). Dense multi-frame optic flow for non-rigid objects using subspace constraints. In *Asian Conference on Computer Vision*, pp. 460–473.
- Garg, R., Roussos, A., & Agapito, L. (2011). Robust trajectory-space TV-L1 optical flow for non-rigid sequences. In *8th International Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR)*, pp. 300–314.
- Horn, B., & Schunck, B. (1981). Determining optical flow. *Artificial Intelligence*, 17, 185–203.
- Irani, M. (2002). Multi-frame correspondence estimation using subspace constraints. *International Journal of Computer Vision*, 48(3), 173–194.
- Kumar, A., Tannenbaum, A. R., & Balas, G. J. (Apr. 1996). Optic flow: A curve evolution approach. *IEEE Transactions on Image Processing*, 5(4), 598–610.
- Liu, C., Yuen, J., & Torralba, A. (2011). SIFT flow: Dense correspondence across scenes and its applications. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(5), 978–994.
- Lucas, B., & Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. In *International Joint Conference on Artificial Intelligence*.
- Newcombe, R., Lovegrove, S., & Davison, A. (2011). DTAM: Dense tracking and mapping in real-time. In *International Conference on Computer Vision*, pp. 2320–2327.
- Nir, T., Bruckstein, A. M., & Kimmel, R. (February 2008). Over-parameterized variational optical flow. *International Journal of Computer Vision*, 76, 205–216.
- Papadakis, N., Corpetti, T., & Mémin, E. (2007). Dynamically consistent optical flow estimation. In *ICCV* (pp. 1–7). Rio de Janeiro, Brazil, October 2007.
- Papenberg, N., Bruhn, A., Brox, T., Didas, S., & Weickert, J. (Apr. 2006). Highly accurate optic flow computation with theoretically justified warping. *International Journal of Computer Vision*, 67(2), 141–158.
- Pizarro, D., & Bartoli, A. (2010). Feature-based deformable surface detection with self-occlusion reasoning. In *International symposium on 3D data processing, visualization and transmission, 3DPVT'10*.
- Pock, T., & Chambolle, A. (2011). Diagonal preconditioning for first order primal-dual algorithms in convex optimization. In *International Conference on Computer Vision*, pp. 1762–1769.
- Pock, T., Cremers, D., Bischof, H., & Chambolle, A. (2010). Global solutions of variational models with convex regularization. *SIAM Journal on Imaging Sciences*, 3(4), 1122–1145.
- Rakêt, L. L., Roholm, L., Nielsen, M., & Lauze, F. (2011). TV-L1 optical flow for vector valued images. In *8th International Conference on Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR)*, pp. 329–343.
- Ricco, S., & Tomasi, C. (2012). Dense lagrangian motion estimation with occlusions. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1800–1807.
- Rockafellar, R. T. (1997). Convex analysis. Princeton Landmarks in Mathematics. Princeton University Press, Princeton, NJ, 1997. Reprint of the 1970 original, Princeton Paperbacks.
- Rudin, L., Osher, S., & Fatemi, E. (1992). Nonlinear total variation based noise removal algorithms. *Physica D*, 60, 259–268.
- Sapiro, G. (1997). Color snakes. *Computer Vision and Image Understanding*, 68(2), 247–253.
- Schnörr, C. (1994). Segmentation of visual motion by minimizing convex non-quadratic functionals. In *Proceedings of the twelfth international conference on pattern recognition* (Vol. A, pp. 661–663). Jerusalem, Israel, Oct. 1994. IEEE Computer Society Press.
- Shi, J., & Tomasi, C. (1994). Good features to track. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 593–600.
- Steinbruecker, F., Pock, T., & Cremers, D. (2009). Large displacement optical flow computation without warping. In *International Conference on Computer Vision*, pp. 1609–1614.
- Stuehmer, J., Gumhold, S., & Cremers, D. (2010). Real-time dense geometry from a handheld camera. In *Pattern recognition (Proc. DAGM)* (pp. 11–20), September 2010.
- Sun, D., Roth, S., Lewis, J. P., & Black, M. (2008). Learning optical flow. In *European Conference on Computer Vision*, pp. 83–97.
- Tian, Y., & Narasimhan, S. (2010). A globally optimal data-driven approach for image distortion estimation. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1277–1284.
- Torresani, L., & Bregler, C. (2002). Space-time tracking. In *European Conference on Computer Vision*, pp. 801–812.
- Torresani, L., Hertzmann, A., & Bregler, C. (2008). Non-rigid structure-from-motion: Estimating shape and motion with hierarchical priors. *PAMI*, 30(5), 878–892.
- Torresani, L., Yang, D., Alexander, E., & Bregler, C. (2001). Tracking and modeling non-rigid objects with rank constraints. In *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 493–500.
- Tschumperlé, D., & Deriche, R. (2005). Vector-valued image regularization with PDE's: A common framework for different applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(4), 506–517.
- Uras, S., Girosi, F., Verri, A., & Torre, V. (1988). A computational approach to motion perception. *Biological Cybernetics*, 60, 79–87.
- Varol, A., Salzmann, M., Tola, E., & Fua, P. (2009). Template-free monocular reconstruction of deformable surfaces. In *International Conference on Computer Vision*, pp. 1811–1818.
- Volz, S., Bruhn, A., Valgaerts, L., & Zimmer, H. (2011). Modeling temporal coherence for optical flow. In *International Conference on Computer Vision*, pp. 1116–1123.
- Wedel, A., Cremers, D., Pock, T., & Bischof, H. (2009). Structure- and motion-adaptive regularization for high accuracy optic flow. In *International Conference on Computer Vision*, pp. 1663–1668.
- Wedel, A., Pock, T., Braun, J., Franke, U., & Cremers, D. (2008). Duality TV-L1 flow with fundamental matrix prior. In *Image and Vision Computing New Zealand*, pp. 1–6.
- Wedel, A., Pock, T., Zach, C., Bischof, H., & Cremers, D. (2009). An improved algorithm for TV-L1 optical flow. In *Statistical and geometrical approaches to visual motion analysis, LNCS* (pp. 23–45). Springer, Berlin.
- Weickert, J. (1998). On discontinuity-preserving optic flow. In S. Orphanoudakis, P. Trahanias, J. Crowley, & N. Katevas (Eds.), *Pro-*

- ceedings of the computer vision and mobile robotics workshop* (pp. 115–122). Santorini, Greece, Sept 1998.
- Weickert, J., & Schnörr, C. (Dec. 2001). A theoretical framework for convex regularizers in PDE-based computation of image motion. *International Journal of Computer Vision*, 45(3), 245–264.
- Weickert, J., & Schnörr, C. (May 2001). Variational optic flow computation with a spatio-temporal smoothness constraint. *Journal of Mathematical Imaging and Vision*, 14(3), 245–255.
- Werlberger, M., Trobin, W., Pock, T., Wedel, A., Cremers, D., & Bischof, H. (2009). Anisotropic Huber-L1 optical flow. In *British Machine Vision Conference*, Vol. 34, pp. 1–11.
- White, R., Crane, K., Forsyth, D. (2007). Capturing and animating occluded cloth. In *ACM Transactions on Graphics*.
- Zach, C., Pock, T., & Bischof, H. (2007). A duality based approach for realtime TV-L1 optical flow. In *Pattern recognition (Proc. DAGM)*, pp. 214–223.