A Feature-Based Approach for Determining Dense Long Range Correspondences

Josh Wills and Serge Belongie

University of California, San Diego
La Jolla, CA 92093 USA
{josh,sjb}@cs.ucsd.edu
http://vision.ucsd.edu

Abstract. Planar motion models can provide gross motion estimation and good segmentation for image pairs with large inter-frame disparity. However, as the disparity becomes larger, the resulting dense correspondences will become increasingly inaccurate for everything but purely planar objects. Flexible motion models, on the other hand, tend to overfit and thus make partitioning difficult. For this reason, to achieve dense optical flow for image sequences with large inter-frame disparity, we propose a two stage process in which a planar model is used to get an approximation for the segmentation and the gross motion, and then a spline is used to refine the fit. We present experimental results for dense optical flow estimation on image pairs with large inter-frame disparity that are beyond the scope of existing approaches.

1 Introduction

Layer-based motion segmentation based on differential optical flow [18,19] can provide good estimation of both the coherent groups in image sequence as well as the associated motion of each group. However, that work is only applicable to scenes where the inter-frame disparity is small. There are two major problems that arise as the disparity increases. The first is that if the disparity exceeds roughly 10-15% of the image size, then even coarse-to-fine optical flow will not be able to find the solution [7]. The second is that with large disparity the planar motion model associated with the layers (e.g. rigid, affine) likely becomes inaccurate for everything but purely planar objects.

In this paper our goal is to determine dense optical flow – by optical flow we are referring to dense correspondences and not the method of differential optical flow estimation – between image pairs with large inter-frame disparity. We propose a two-stage framework based on a planar motion model for capturing the gross motion followed by a regularized spline model for capturing finer scale variations. Our approach is related in spirit to the *deformotion* concept in [12], developed for the case of differential motion, which separates overall motion (a finite dimensional group action) from the more general deformation (a diffeomorphism).

T. Pajdla and J. Matas (Eds.): ECCV 2004, LNCS 3023, pp. 170–182, 2004.

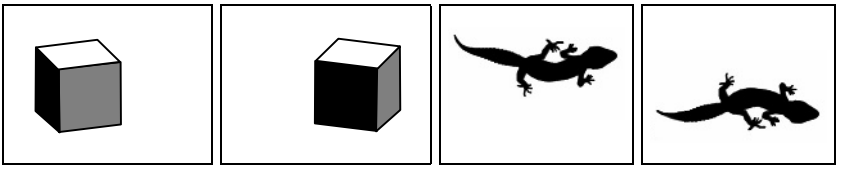

Fig. 1. Non-planarity vs. non-rigidity: The left image pair shows a non-planar object undergoing 3D rigid motion; the right pair shows an approximately planar object undergoing non-rigid motion. Both examples result in residual with respect to a 2D planar fit.

The types of image pairs that we wish to consider are illustrated in Figure 1. These have a significant component of planar motion but exhibit residual with respect to a planar fit because of either the non-planarity of the object (e.g. a cube) or the non-rigidity of the motion (e.g. a lizard). These are scenes for which the motion can be approximately described by a planar layer-based framework, i.e. scenes that have "shallow structure" [10].

It is important to remember that optical flow does not model the 3D motion of objects, but rather the changes in the image that result from this motion. Without the assumption of a rigid object, it is very difficult to estimate the 3D structure and motion of an object from observed change in the image, though there is existing work that attempts to do this [5,17]. For this reason, we choose to do all estimation in the image plane (i.e. we use 2D models), but we show that if the object is assumed to be rigid, the correspondences estimated can be used to recover the dense structure and 3D motion.

This approach extends the capabilities of feature-based scene matching algorithms to include dense optical flow without the limits on allowable motion associated with techniques based on differential optical flow. Previously, feature-based approaches could handle image pairs with large disparity and multiple independently moving objects, while optical flow techniques could provide a dense set of pixel correspondences even for objects with non-rigid motion. However, neither type of approach could handle both simultaneously. Without the assumption of a rigid scene, existing feature-based methods cannot produce dense optical flow from the sparse correspondences, and in the presence of large disparity and multiple independently moving objects, differential optical flow (even coarse-to-fine) can break down. The strength of our approach is that dense optical flow can now be estimated for image pairs with large disparity, more than one independently moving object, and non-planar (including non-rigid) motion.

The structure of the paper is as follows. We will begin in Section 2 with an overview of related work. In Section 3, we detail the components of our approach. We discuss experimental results in Section 4. There is discussion in Section 5 and the paper concludes with Section 6.

2 Related Work

The work related to our approach comes from the areas of motion segmentation, optical flow and feature-based (sparse) matching. Several well known ap-

proaches to motion segmentation are based on dense optical flow estimation [1,3, 8]; in these approaches the optical flow field was assumed to be piecewise smooth to account for discontinuities due to occlusion and object boundaries. Wang & Adelson introduced the idea of decomposing the image sequence into multiple overlapping layers, where each layer represents an affine motion field [18]. However their work was based on differential optical flow, which places strict limits on the amount of motion between two frames.

In [19], Weiss uses regularized radial basis functions (RBFs) to estimate dense optical flow; Weiss' method is based on the assumption that while the motion will not be smooth across the entire image, the motion is smooth within each of the layers. Given the set of spatiotemporal derivatives, he used the EM algorithm to estimate the number of motions, the dense segmentation and the dense optical flow. This work along with other spline-based optical flow methods [13, 14] however, also assumes differential motion and therefore does not apply for the types of sequences that we are considering.

In [16], Torr et al. show that the trifocal tensor can be used to cluster groups of sparse correspondences that move coherently. This work addresses similar types of sequences to those of our work in that it is trying to capture more than simply a planar approximation of motion, but it does not provide dense assignment to motion layers or dense optical flow. The paper states that it is an initialization and that more work is needed to provide a dense segmentation, however the extension of dense stereo assignment to multiple independent motions is certainly non-trivial and there is yet to be a published solution. In addition, this approach is not applicable for objects with non-rigid motion, as the fundamental matrix and trifocal tensor apply only to rigid motion.

Our work builds on the motion segmentation found via planar motion models as in [20], where planar transformations are robustly estimated from point correspondences in a RANSAC framework. A dense assignment of pixels to transformation layers is then estimated using an MRF. We refine the planar estimation produced by [20] using a regularized spline fit. Szeliski & Shum [14] also use a spline basis for motion estimation, however their approach has the same limitations on the allowable motion as other coarse-to-fine methods.

3 Our Approach

Our goal in this paper is to determine the dense optical flow for pairs of images with large inter-frame disparity and in the presence of multiple independent motions. If the scene contains objects undergoing significant 3D motion or deformation, the optical flow cannot be described by any single low dimensional image plane transformation (e.g. an affine transformation or a homography). However, to keep the problem tractable we need a compact representation of these transformations; we propose the use of thin plate splines for this purpose. A single spline is not sufficient for representing multiple independent motions, especially when the motion vectors intersect [19]. Therefore we represent the optical flow between two frames as a set of disjoint splines. By disjoint we mean

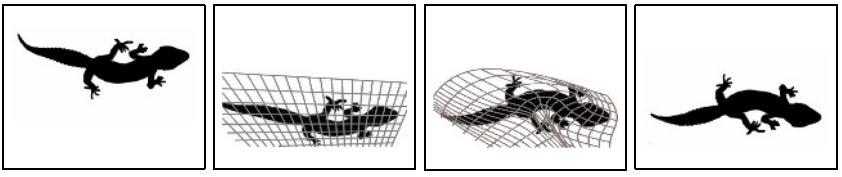

Fig. 2. Determining Long Range Optical Flow. The goal is to provide dense optical flow from the first frame (1), to the second (4). This is done via a planar fit (2) followed by a flexible fit (3).

that the support of the splines are disjoint subsets of the image plane. The task of fitting a mixture of splines naturally decomposes into two subtasks: motion segmentation and spline fitting. Ideally we would like to do both of these tasks simultaneously, however these tasks have conflicting goals. The task of motion segmentation requires us to identify groups of pixels whose motion can be described by a smooth transformation. Smoothness implies that each motion segment has the the same gross motion, however except for the rare case in which the entire layer has exactly the same motion everywhere, there will be local variations. Hence the motion segmentation algorithm should be sensitive to inter-layer motion and insensitive to intra-layer variations. On the other hand, fitting a spline to each motion field requires attention to all the local variations. This is an example of different tradeoffs between bias and variance in the two stages of the algorithm. In the first stage we would like to exert a high bias and use models with a high amount of stiffness and insensitivity to local variations, whereas in the second stage we would like to use a more flexible model with a low bias.

The motion segmentation consists of a two stage RANSAC-based robust estimation procedure which operates on a sparse set of correspondences between the two frames. Any planar transformation can be used as the motion model in this stage; we use homographies in this paper. Once the dominant motions have been detected, a dense assignment is performed using a fast graph partitioning algorithm.

The output of the first stage, while sufficient to achieve a good segmentation is not sufficient to recover the optical flow accurately. However it serves two important purposes: firstly it provides an approximate segmentation of the sparse correspondences that allows for coherent groups to be processed separately. This is crucial for the second stage of the algorithm as a flexible model will likely find an unwieldy compromise between distinct moving groups as well as outliers. Secondly, since the assignment is dense, it is possible to find matches for points that were initially mismatched by limiting the correspondence search space to points in the same motion layer. The second stage then bootstraps off of these estimates of motion and layer support to iteratively fit a thin plate spline to account for non-planarity or non-rigidity in the motion. Figure 2 illustrates this process.

We now describe the two stages of the algorithm in detail.

3.1 Detecting Dominant Planar Motion

We begin by finding planar approximations of the motion in the scene as well as a dense assignment of pixels to motion layers. We use the motion segmentation algorithm of [20]. An example of this is shown in Figure 3

Example of Planar Fit and Segmentation. Figure 3 shows an example of the output from the planar fit and segmentation process. In this figure we show the two images, I and I', and the assignments for each pixel to a motion layer (one of the three detected motion fields). The columns represent the different motion fields and the rows represent the portions of each image that are assigned to a given motion layer. The motions are made explicit in that the pixel support from frame to frame is related exactly by a planar homography. Notice that the portions of the background and the dumpsters that were visible in both frames were segmented correctly, as was the man. The result of the spline fit for this example will be shown in Section 4.

Fig. 3. Notting Hill sequence. (Row 1) Original image pair of size 311×552, (Row 2) Pixels assigned to warp layers 1-3 in I, (Row 3) Pixels assigned to warp layers 1-3 in I'.

3.2 Refining the Fit with a Flexible Model

The flexible fit is an iterative process using regularized radial basis functions, in this case Thin Plate Spline (TPS). The spline interpolates the correspondences to result in a dense optical flow field. This process is run on a per-motion layer basis.

Feature Extraction and Matching. During the planar motion estimation stage, only a gross estimate of the motion is required so a sparse set of feature points will suffice. In the final fit however, we would like to use as many correspondences as possible to ensure a good fit. In addition, since the correspondence search space is reduced (i.e. matches are only considered between pixels assigned to corresponding motion layers), matching becomes somewhat simpler. For this reason, we use the Canny edge detector to find the set of edge points in each of the frames and estimate correspondences in the same manner as in [20].

Iterated TPS Fitting. Given the approximate planar homography and the set of correspondences between edge pixels, we would like to find the dense set of correspondences. If all of the correspondences were correct, we could jump

straight to a smoothed spline fit to obtain dense (interpolated) correspondences for the whole region. However, we must account for the fact that many of the correspondences are incorrect. As such, the purpose of the iterative matching is essentially to distinguish inliers from outliers, that is, we would like to identify sets of points that exhibit coherence in their correspondences.

One of the assumptions that we make about the scenes we wish to consider is that the motion of the scene can be approximated by a set of planar layers. Therefore a good initial set of inliers are those correspondences that are roughly approximated by the estimated homography. From this set, we use TPS regression with increasingly tighter inlier thresholds to identify the final set of inliers, for which a final fit is used to interpolate the dense optical flow. We now briefly describe this process.

The Thin Plate Spline is the Radial Basis Function (RBF) that minimizes the following bending energy or integral bending norm [4],

$$I_f = \iint_{\mathbb{R}^2} (f_{xx}^2 + 2f_{xy}^2 + f_{yy}^2) dxdy$$

where $f = f(x, y)$ represents the x or y component of the transformation for the pixel at position (x, y). In our approach we use a regularized version of TPS fitting in which μ controls the tradeoff between data and smoothing in the cost functional

$$H[f] = \sum_i (v_i - f(x_i, y_i))^2 + \mu I_f$$

where v_i represents the target of the transformation and $f(x_i, y_i)$ is the mapped value for the point at location (x_i, y_i). Since each point gets mapped to a new 2D position, we require two TPS transformations, one for the x-coordinates and another for the y-coordinates. We solve for this transformation as in [9].

We estimate the TPS mapping from the points in the first frame to those in the second where μ_t is the regularization factor for iteration t. The fit is estimated using the set of correspondences that are deemed inliers for the current transformation, where τ_t is the threshold for the t^{th} iteration. After the transformation is estimated, it is applied to the entire edge set and the set of correspondences is again processed for inliers, using the new locations of the points for error computation. This means that some correspondences that were outliers before may be pulled into the set of inliers and vice versa. The iteration continues on this new set of inliers where $\tau_{t+1} \leq \tau_t$ and $\mu_{t+1} \leq \mu_t$. We have found that three iterations of this TPS regression with incrementally decreasing regularization and corresponding outlier thresholds suffices for a large set of real world examples. Additional iterations produced no change in the estimated set of inlier correspondences.

I. Estimate planar motion
1. Find correspondences between I and I'
2. Robustly estimate the motion fields
3. Densely assign pixels to motion layers

II. Refine the fit with a flexible model
4. Match edge pixels between I and I'
5. For t=1:3
6. Fit all correspondences within τ_t using TPS regularized by μ_t
7. Apply TPS to set of correspondences

Note: $(\tau_{t+1} \leq \tau_t, \mu_{t+1} \leq \mu_t)$

Fig. 4. Algorithm Summary

This simultaneous tightening of the pruning threshold and annealing of the regularization factor aid in differentiating between residual due to localization error or mismatching and residual due to the non-planarity of the object in motion. When the pruning threshold is loose, it is likely that there will be some incorrect correspondences that will pass the threshold. This means that the spline should be stiff enough to avoid the adverse effect of these mismatches. However, as the mapping converges we place higher confidence in the set of correspondences passing the tighter thresholds. This process is similar in spirit to iterative deformable shape matching methods [2,6].

4 Experimental Results

We now illustrate our algorithm, which is summarized in Figure 4, on several pairs of images containing objects undergoing significant, non-planar motion. Since the motion is large, displaying the optical flow as a vector field will result in a very confusing figure. Because of this, we show the quality of the optical flow in other ways, including (1) examining the image and corresponding reconstruction error that result from the application of the estimated transform to the original image (we refer to this transformed image as $\mathcal{T}(I)$), (2) showing intermediate views (as in [11]), or by (3) showing the 3D reconstruction induced by the set of dense correspondences. Examples are presented that exhibit either non-planarity, non-rigidity or a combination of the two. We show that our algorithm is capable of providing optical flow for pairs of images that are beyond the scope of existing algorithms. We performed all of the experiments on grayscale images using the same parameters[1].

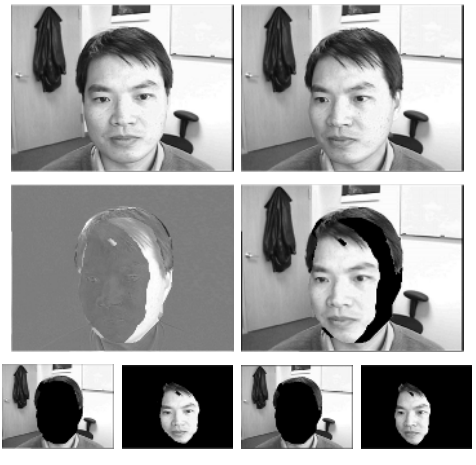

Fig. 5. Face Sequence. (1) The two input images, I and I' of size 240×320. (2) The difference image is show first where grey regions indicate zero error regions and the reconstruction, $\mathcal{T}(I)$ is second. (3) The initial segmentation found via planar motion.

4.1 Face Sequence

The first example is shown in Figures 5 and 6. The top row of Figure 5 shows the two input frames, I and I', in which a man moves his head to the left in

[1] $k = 2, \lambda = .285, \tau_p = 15, \mu_1 = 50, \mu_2 = 20, \mu_3 = 1, \tau_1 = 15, \tau_2 = 10, \tau_3 = 5$. Here, k, λ, and τ_p refer to parameters in [20].

front of a static scene (the nose moves more than 10% of the image width). The second row shows first the difference image between $\mathcal{T}(I)$ and I' where error values are on the interval [-1,1] and gray regions indicate areas of zero error. This image is followed by $\mathcal{T}(I)$; this image has two estimated transformations, one for the face and another for the background. Notice that error in the overlap of the faces is very small, which means that according to reconstruction error, the estimated transformation successfully fits the relation between the two frames. This transformation is non-trivial as seen in the change in the nose and lips as well as a shift in gaze seen in the eyes, however all of this is captured by the estimated optical flow. The final row in Figure 5 shows the segmentation and planar approximation from [20], where the planar transformation is made explicit as the regions' pixel supports are related exactly by a planar homography.

Dense correspondences allow for the estimation of intermediate views via interpolation as in [11]. Figure 6 shows the two original views of the segment associated with the face as well as a synthesized intermediate view that is realistic in appearance. The second row of this figure shows an estimation of relative depth that comes from the disparity along the rectified horizontal axis. Notice the shape of the nose and lips as well as the relation of the eyes to the nose and forehead. It is important to remember that no information specific to human faces was provided to the algorithm for this optical flow estimation.

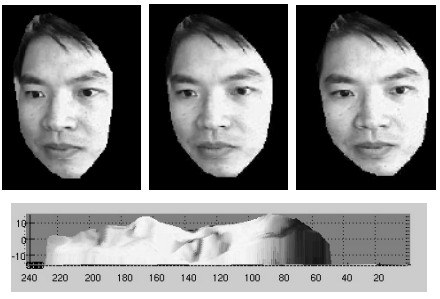

Fig. 6. Face Sequence – Interpolated views. (1) Original frame I', synthesized intermediate frame, original frame I, (2) A surface approximation from computed dense correspondences.

4.2 Notting Hill Sequence

The next example shows how the spline can also refine what is already a close approximation via planar models. Figure 7 shows a close up of the planar error image, the reconstruction error and finally the warped grid for the scene that was shown in Figure 3. The planar approximation was not able to capture the 3D nature of the clothing and the non-rigid motion of the head with respect to the torso, however the spline fit captures these things accurately.

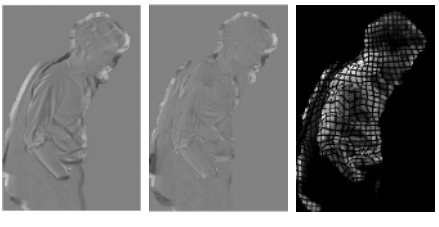

Fig. 7. Notting Hill. Detail of the spline fit for a layer from Figure 3, difference image for the planar fit, difference image for the spline fit, grid transformation.

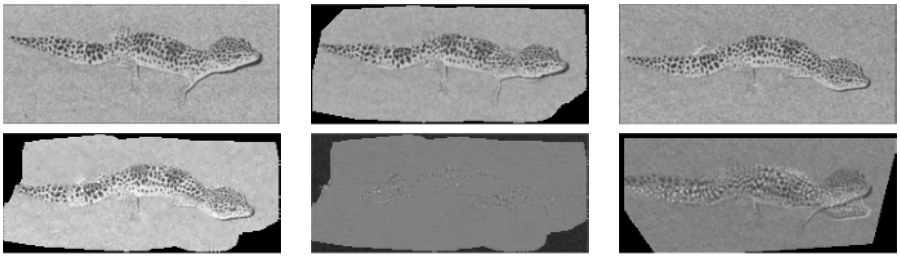

Fig. 8. Gecko Sequence. (1) Original frame I of size 102×236, synthesized intermediate view, original frame I'. (2) $\mathcal{T}(I)$, Difference image between the above image and I' (gray is zero error), Difference image for the planar fit.

4.3 Gecko Sequence

The second example, shown in Figure 8, displays a combination of a non-planar object (a gecko lizard), undergoing non-rigid motion. While this is a single object sequence, it shows the flexibility of our method to handle complicated motions. In Figure 8(1), the two original frames are shown as well as a synthesized intermediate view (here, intermediate refers to time rather than viewing direction since we are dealing with non-rigid motion) . The synthesized image is a reasonable guess at what the scene would look like midway between the two input frames. Figure 8(2) shows $\mathcal{T}(I)$ as well as the reconstruction error for the spline fit ($\mathcal{T}(I) - I'$), and the error incurred with the planar fit. We see in the second row of Figure 8(2) that the tail, back and head of the gecko are aligned very well and those areas have negligible error. When we compare the reconstruction error to the error induced by a planar fit, we see that the motion of the gecko is not well approximated by a rigid plane. Here, there is also some 3D motion present in that the head of the lizard changes in both direction and elevation. This is captured by the estimated optical flow.

4.4 Rubik's Cube

The next example shows a scene with rigid motion of a non-planar object. Figure 9 displays a Rubik's cube and user's manual switching places as the cube rotates in 3D. Below the frames, we see the segmentation that is a result of the planar approximation. It is important to remember that the background in this scene has no distinguishing marks so there is nothing to say that pieces of the background didn't actually move with the objects.

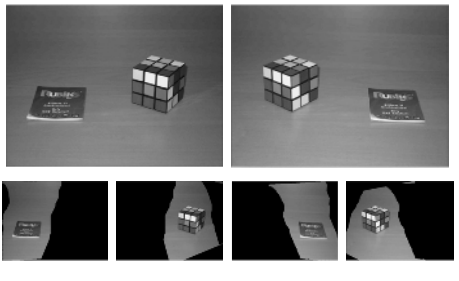

Fig. 9. Rubik's Cube. (1) Original image pair of size 300×400, (2) assignments of each image to layers 1 and 2.

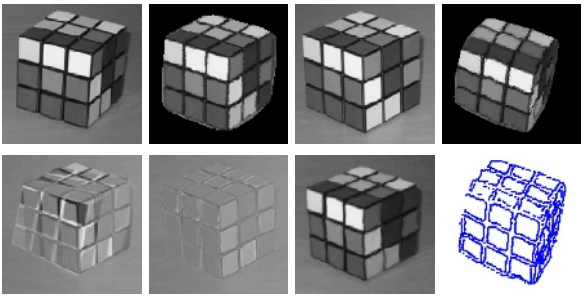

Fig. 10. Rubik's Cube – Detail. (1) Original frame I, synthesized intermediate frame, original frame I', A synthesized novel view, (2) difference image for the planar fit, difference image for the spline fit, $\mathcal{T}(I)$, the estimated structure shown for the edge points of I. We used dense 3D structure to produce the novel view.

Figure 10 shows $\mathcal{T}(I)$, the result of the spline fit applied to this same scene. The first row shows a detail of the two original views of the Rubik's cube as well as a synthesized intermediate view. Notice that the rotation in 3D is accurately captured and demonstrated in this intermediate view. The second row shows the reconstruction errors, first for the planar fit and then for the spline fit, followed by $\mathcal{T}(I)$. Notice how accurate the correspondence is since the spline applied to the first image is almost identical to the second frame.

Correspondences between portions of two frames that are assumed to be projections of rigid objects in motion allow for the recovery of the structure of the object, at least up to a projective transformation. In [15], the authors show a sparse point-set from a novel viewpoint and compare it to a real image from the same viewpoint to show the accuracy of the structure. Figure 10 shows a similar result, however since our correspondences are dense, we can actually render the novel view that validates our structure estimation. The novel viewpoint is well above the observed viewpoints, yet the rendering as well as the displayed structure is fairly accurate. Note that only the set of points that were identified as edges in I are shown; this is not the result of simple edge detection on the rendered view. We use this display convention because the entire point-set is too dense to allow the perception of structure from a printed image. However, the rendered image shows that our estimated structure was very dense. It is important to note that the only assumption that we made about the object is that it is a rigid, piecewise smooth object. To achieve similar results from sparse correspondences would require additional object knowledge, namely that the object in question is a cube and has planar faces. It is also important to point out that this is not a standard stereo pair since the scene contains multiple objects undergoing independent motion.

5 Discussion

Since splines form a family of universal approximators over $\mathbb{R}^2$ and can represent any 2D transformation to any desired degree of accuracy, it raises the question

as to why one needs to use two different motion models in the two stages of the algorithm. If one were to use the affine transform as the dominant motion model, splines with an infinite or very large degree of regularization can indeed be used in its place. However, in the case where the dominant planar motion is not captured by an affine transform and we need to use a homography, it is not practical to use a spline. This is so because the set of homographies over any connected region of $\mathbb{R}^2$ are unbounded, and can in principal require a spline with an unbounded number of knots to represent an arbitrary homography. So while a homography can be estimated using a set of four correspondences, the corresponding spline approximation can, in principle, require an arbitrarily large number of control points. This poses a serious problem for robust estimation procedures like RANSAC since the probability of hitting the correct model decreases exponentially with increasing degrees of freedom. Many previous approaches for capturing long range motion are based on the fundamental matrix. However, since the fundamental matrix maps points to lines, translations in a single direction with varying velocity and sign are completely indistinguishable, as pointed out, e.g. by [16]. This type of motion is observed frequently in motion sequences. The trifocal tensor does not have this problem; however, like the fundamental matrix, it is only applicable for scenes with rigid motion and there is not yet a published solution for dense stereo correspondence in the presence of multiple motions.

6 Conclusion

In this paper, we have presented a new method for determining long range optical flow. We have shown that dense optical flow can now be estimated for image pairs with large disparity, multiple independently moving objects, and non-planar (including non-rigid) motion. Our approach is a two-stage framework based on a planar motion model for capturing the gross motion of the group followed by regularized spline fitting for capturing finer scale variations.

Our approach is intentionally generic in that it requires no object knowledge. However, in many cases, information about the types of objects in question could be used. The partitioning and initial estimation of gross motion may benefit from the use of articulated/object models. While a general solution using known models would require a solution to object recognition, incorporating object knowledge and models in specific domains will be the subject of future research.

Acknowledgments. We would like to thank Sameer Agarwal, Charless Fowlkes and Ben Ochoa for helpful discussions. The images in Figures 5 and 6 are used courtesy of Dr. Philip Torr. This work was partially supported under the auspices of the U.S. Department of Energy by the Lawrence Livermore National Laboratory under contract No. W-7405-ENG-48 and by an NSF IGERT Grant (Vision and Learning in Humans and Machines, #DGE-0333451).

References

1. S. Ayer and H. Sawhney. Layered representation of motion video using robust maximum-likelihood estimation of mixture models and mdl encoding. In *ICCV 95*, pages 777–784, 1995.
2. S. Belongie, J. Malik, and J. Puzicha. Shape matching and object recognition using shape contexts. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 24(4):509–522, April 2002.
3. M. Black and A. Jepson. Estimating optical flow in segmented images using variable-order parametric models with local deformations. *T-PAMI*, 18:972–986, 1996.
4. F. L. Bookstein. Principal warps: thin-plate splines and decomposition of deformations. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 11(6):567–585, June 1989.
5. M. Brand. Morphable 3D models from video. In *IEEE Computer Vision and Pattern Recognition*, December 2001.
6. H. Chui and A. Rangarajan. A new algorithm for non-rigid point matching. In *Proc. IEEE Conf. Comput. Vision and Pattern Recognition*, pages 44–51, June 2000.
7. M. Irani and P. Anandan. All about direct methods. In W. Triggs, A. Zisserman, and R. Szeliski, editors, *Vision Algorithms: Theory and Practice*. Springer-Verlag, 1999.
8. J.-M. Odobez and P. Bouthemy. Direct incremental model-based image motion segmentation for video analysis. *Signal Processing*, 66(2):143–155, 1998.
9. M. J. D. Powell. A thin plate spline method for mapping curves into curves in two dimensions. In *Computational Techniques and Applications (CTAC95)*, Melbourne, Australia, 1995.
10. H. S. Sawhney and A. R. Hanson. Trackability as a cue for potential obstacle identification and 3D description. *International Journal of Computer Vision*, 11(3):237–265, 1993.
11. S. M. Seitz and C. R. Dyer. View morphing. In *SIGGRAPH*, pages 21–30, 1996.
12. S. Soatto and A. J. Yezzi. DEFORMOTION: Deforming motion, shape average and the joint registration and segmentation of images. In *European Conference on Computer Vision*, pages 32–47. Springer, 2002. Copenhagen.
13. R. Szeliski and J. Coughlan. Hierarchical spline-based image registration. In *IEEE Conference on Computer Vision Pattern Recognition*, pages 194–201, Seattle, Washington, 1994.
14. R. Szeliski and H.-Y. Shum. Motion estimation with quadtree splines. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(12):1199–1210, 1996.
15. C. Tomasi and T. Kanade. Factoring image sequences into shape and motion. In *Proc. IEEE Workshop on Visual Motion*. IEEE, 1991.
16. P. H. S. Torr, A. Zisserman, and D. W. Murray. Motion clustering using the trilinear constraint over three views. In R. Mohr and C. Wu, editors, *Europe-China Workshop on Geometrical Modelling and Invariants for Computer Vision*, pages 118–125. Springer–Verlag, 1995.
17. L. Torresani, D. Yang, G. Alexander, and C. Bregler. Tracking and modelling non-rigid objects with rank constraints. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 493–500, Kauai, Hawaii, 2001.
18. J. Wang and E. H. Adelson. Layered representation for motion analysis. In *Proc. Conf. Computer Vision and Pattern Recognition*, pages 361–366, 1993.

19. Y. Weiss. Smoothness in layers: Motion segmentation using nonparametric mixture estimation. In *Proc. IEEE Conf. Comput. Vision and Pattern Recognition*, pages 520–526, 1997.
20. J. Wills, S. Agarwal, and S. Belongie. What went where. In *Proc. IEEE Conf. Comput. Vision and Pattern Recognition*, June 2003.