Integration of motion fields through shape

Hui Ji and Cornelia Fermüller
Center for Automation Research, University of Maryland
College Park, MD 20742, USA
{jihu,fer}@cfar.umd.edu

Abstract

Structure from motion from single flow fields has been studied intensively, but the integration of information from multiple flow field has not received much attention. Here we address this problem by enforcing constraints on the shape (surface normals) of the scene in view, as opposed to constraints on the structure (depth). The advantage of integrating shape is two-fold. First, we do not need to estimate feature correspondences over multiple frames, but we only need to match patches. Second, the shape vectors in the different views are related only by rotation. This constraint on shape can be combined easily with motion estimation, thus formulating motion and structure estimation from multiple views as a practical constrained minimization problem using a rank=3 constraint. Based on this constraint, we develop a 3D motion technique, which locates through color and motion segmentation, planar patches in the scene, matches patches over multiple frames, and estimates the motion between multiple frames and the shape of the selected scene patches using the image gradients. Experiments evaluate the accuracy of the 3D motion estimation and demonstrate the motion and shape estimation of the technique by super-resolving an image sequence.

Keywords: 3D motion estimation, integration of motion fields, decoupling translation from rotation, shape and rotation

1 Introduction

Continuous approaches to structure from motion use as input motion fields extracted from video. Clearly, the information in one motion field is not rich enough to allow for accurate estimation of the 3D motion and structure. There are two issues. First, there is the ambiguity in the estimation of the motion parameters. For standard cameras with limited field of view, there is a confusion between translation and rotation [3, 5]. Thus, any motion algorithm, because of noise, can estimate the motion only within a range of the true solution. The second issue is the stability of structure estimation. An erroneous estimate of the motion parameters clearly will lead to errors in the estimation of structure. But even for the correct motion parameters, the estimation of structure, because of the small displacement between the cameras (small baseline), is very unreliable.

To obtain good motion and structure estimates requires to combine the information of consecutive motion fields, that is to *integrate motion fields*. One flow field, or in the abstraction two image frames, are constrained by the rigidity of 3D motion. Two or more motion fields are in addition constrained by the observed scene which remains constant. Existing approaches formulate this as a constraint enforcing, the structure, or depth of the scene to be the same. In order to make use of this constraint, we need to correspond image points (or lines) over multiple frames. But automatic correspondence usually is not possible. Drifting occurs, that is errors in correspondence accumulate till eventually the correspondence cannot be established anymore. Another problem is, that since structure in the coordinate system of one frame is related to structure in the coordinate system of another frame through both the translation and rotation, there do not exist simple constraints that would allow a robust estimation of structure and motion from multiple frames. We have the trilinear constraint resulting in 27 and the quadrilinear constraint resulting in 64 parameters whose estimations are very sensitive. Most commonly the photogrammetric techniques of bundle adjustment are used to refine 3D structure and viewing parameters, where one seeks to minimize some cost function of the feature reprojection errors.

In this paper we introduce an approach, which links multiple motion fields, not through depth or inverse depth values, but through shape. The shape of a scene patch is described by the surface normal of the patch, that is by two parameters. We model the scene as patch-wise planar. The surface normal of a planar patch in one view is related to the surface normal of the corresponding patch in another view only through rotation. That is, let r_1 and r_2 be the surface normals of a patch in the first and second frame, and let Ω be the rotation relating the two frames, then $r_2 = \Omega r_1$. This relationship can be formulated as a rank 3 constraint on a matrix containing the normal vectors of all patches which

are matched over many views. The advantage of this constraint is two-fold. First, we do not need to correspond image points over multiple frames, but only image patches corresponding to planar scene patches, which is much easier to accomplish. Second, the constraint can be easily combined with the estimation of motion and structure from individual flow fields, resulting in a simultaneous estimation of structure from motion from multiple flow fields as a practical constrained minimization, where the constraint is the rank constraint on the surface normals.

The complete approach works as follows. We first segment planar patches in the scene and match the patches over the image sequence. The segmentation is based on color information and motion information (Sec. 5). Then the 3D motion parameters and the shape of the extracted scene patches are estimated using as input the image gradients (Sec. 4). Using as starting point the motion estimates from single flow fields, the constrained minimization is solved iteratively in a two-step optimization; in one step the surface normal are obtained, in the next step the motion parameters. Experiments evaluate the technique's 3D motion estimation and demonstrate its shape and motion estimation in the application of super-resolution imaging (Sec. 6). Finally, it is discussed how to use the 3D information together with image measurements to obtain better segmentation and scene models (Sec. 7).

Subspace constraints on planar patches in multiple views were previously introduced in [6] and [7]. The improvement of our method stems from the use of shape which makes the integration of frames computationally feasible. In particular, [6] defined a constraint on the relative homographies of planes, which however is hard to incorporate into the estimation. [7] presents a subspace constraint on image gradients, which assumes differential motion between any two frames. Since good reconstruction requires a large baseline and this assumption is valid only for short image sequences the improvement in structure estimation is hampered.

2 Preliminaries

Let $P = (X, Y, Z)$ denote a 3d scene point and $p = (x, y)$ denote its corresponding point in the image plane $Z = f$, where $(x, y) = \frac{f}{Z}(X, Y)$. Without loss of generality, we assume $f = 1$. Then the differential motion of the point p is expressed as:

$$\begin{pmatrix} \frac{dx}{dt} \\ \frac{dy}{dt} \end{pmatrix} = \frac{1}{Z} \begin{pmatrix} t_z x - t_x \\ t_z y - t_y \end{pmatrix} + \begin{pmatrix} xy\omega_x - (x^2+1)\omega_y + y\omega_z \\ (y^2+1)\omega_x - xy\omega_y - x\omega_z \end{pmatrix} \tag{1}$$

Let P be on the world plane $\alpha X + \beta Y + \gamma Z = 1$. The plane is described by the parameter vector $n = (\alpha, \beta, \gamma)$, which describes the depth and surface normal at P. Thus, the inverse depth at P amounts to $\frac{1}{Z} = \alpha x + \beta y + \gamma$. Substituting the expressions for the image motion and the plane into the the well known brightness consistency constraint:

$$\frac{\partial I}{\partial x} \cdot \frac{dx}{dt} + \frac{\partial I}{\partial y} \cdot \frac{dy}{dt} + \frac{\partial I}{\partial t} = 0,$$

we obtain the constraint

$$\begin{aligned} [-I_x t_x - I_y t_y + (I_x x + I_y y) t_z][x\alpha + y\beta + \gamma] & \\ +(I_x xy + I_y(y^2+1))\omega_x & \\ -(I_y xy + I_x(x^2+1))\omega_y + (I_x y - I_y x)\omega_z &= -I_t, \end{aligned}$$

which relates the image gradients to the motion and plane parameters. This equation is bilinear in the motion and plane parameters. There is a scaling ambiguity between translation and depth, and thus only the direction of vectors t and and n can be obtained. Let $p_i = (x_i, y_i), i = 1, \cdots, N$ denote the image points corresponding to a single world plane, from which we obtain an equation system that we write as

$$\begin{aligned} & f(t, \omega, n) = 0 = \\ & [t_x A_1 + t_y A_2 + t_z A_3] \begin{pmatrix} \alpha \\ \beta \\ \gamma \end{pmatrix} + B \begin{pmatrix} \omega_x \\ \omega_y \\ \omega_z \end{pmatrix} - b, \end{aligned} \tag{2}$$

where A_1, A_2, A_3, B are $N \times 3$ matrices and b is an $N \times 1$ vector, whose elements are described by the image point coordinates and the intensity gradients at points p_i.

2.1 Motion and shape estimation from individual flow fields

Our method uses as input 3D motion estimates obtained from the individual flow field. We could employ many of the algorithm for 3D motion estimation from the literature. We chose to use the algorithm in [1], which is based on the equations above and works as follows:

Consider a segmentation of the scene into P planar patches $V^1, V^2 \cdots V^P$. Combining equations (2) for all the patches we obtain an over-determined bilinear equation system of the form

$$f^p(t, \omega, n^p) = 0 \quad \text{for} \quad p = 1, \cdots, P, \tag{3}$$

whose solution provides estimates for the direction of translation, the rotation and the structure parameters for the individual patches. The algorithm in [1] solves this minimization as a search in the space of translational directions. For each candidate translation one can solve closed-form for the rotation and the plane parameters. The translational direction minimizing eq. (3) in the least squares sense provides the solution. Note the algorithm does not require optical flow, but only the image gradients, which define the so-called normal flow.

2.2 Ambiguities in 3D motion estimation from single flow fields

Several studies have addressed the noise sensitivity in structure from motion. In particular it has been shown that for standard cameras with small field of view there is a confusion between translation and rotation. Translation along the x-axis can be confused with rotation around the y-axis, and translation along the y-axis with rotation around the x-axis. Thus, in the presence of noise it is hard to distinguish these motions [3].

Since the estimated motion field always is noisy, we can obtain only a range of possible solutions for the motion parameters, among which the correct one lies. If we consider the 2d subspace of translational directions of this range and visualize it on the image (or on a sphere), we usually obtain an elongated region, to which we refer as the *motion valley* of solutions. Each translation direction in the motion valley, along with its best corresponding rotation and structure will agree with the observed noisy flow field. Figure 2 shows typical error functions (residuals of the minimization) plotted on the 2d spherical surface. The best solutions lie in the red area of the surface.

3 Rank constraint on shape parameters

Let the frames to be integrated be denoted as $I_1, I_2, \cdots, I_F$. The scene patches are denoted as V_f^p, where f indicates the frame index and p indicates the patch index. The translational and rotational velocities of the normal flow field at frame I_f are denoted as t_f and ω_f. One thing to be emphasized here is that the frames we integrate do not need to be consecutive, and actually there is no need to combine all consecutive frames; one may combine frames far apart. The reason is that because of temporal smoothing and because of the small baseline, views close by do not provide very different information. However, note, the motion parameters t_f and ω_f, are the differential velocities computed from the flow field at frame f between consecutive frames, and not the motions between the chosen frames.

The normal flow field at every frame I_f provides a bilinear equation system $\{f(t_f, \omega_f, n_f^p)\}$ for the estimation of motion and depth. In order to integrate the systems of the different I_f we formulate a constraint which enforces the shape, that is the orientation of scene patches to be the same. The shape representations in the different coordinate systems of the frames are easily related. The relative orientations of the different patches within a view are invariant across views. The absolute orientations of a patch in the different views are related by rotation only. This invariance can be expressed as a rank constraint on a matrix containing the surface normals.

Let n_f^p be the parameter vector of the inverse depth which describes the plane with index p in frame I_f. The normal vector r_f^p of this plane is obtained by normalizing the vector n_k^p, i.e $r_k^p = \frac{1}{||n_f^p||_2} n_f^p$. The normal vectors of a plane in two frames I_{f_1}, I_{f_2} are related by the rotation matrix Ω_{f_1,f_2} as

$$r_{f_1}^p = \Omega_{f_1,f_2} r_{f_2}^p \quad \text{for} \quad p = 1, \cdots, P,$$

which can be expressed in matrix form as:

$$\begin{aligned} R_{f_1} &= (r_{f_1}^1, r_{f_1}^2, \cdots, r_{f_1}^P) \\ &= (\Omega_{f_1,f_2} r_{f_2}^1, \Omega_{f_1,f_2} r_{f_2}^2, \cdots, \Omega_{f_1,f_2} r_{f_2}^P) \\ &= \Omega_{f_1,f_2} (r_{f_2}^1, r_{f_2}^2, \cdots, r_{f_2}^P) = \Omega_{f_1,f_2} R_{f_2}. \end{aligned}$$

We then combine the normal vectors in all the F frames into a matrix M, where

$$M = \begin{pmatrix} R_1 \\ R_2 \\ \vdots \\ R_F \end{pmatrix} = \begin{pmatrix} I_3 R_1 \\ \Omega_{2,1} R_1 \\ \vdots \\ \Omega_{F,1} R_1 \end{pmatrix} \tag{4}$$

$$= \begin{pmatrix} I & 0 & \cdots & 0 \\ 0 & \Omega_{2,1} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \Omega_{F,1} \end{pmatrix} \begin{pmatrix} R_1 \\ R_1 \\ \vdots \\ R_1 \end{pmatrix}. \tag{5}$$

The decomposition into a product, makes it clear that the $3F \times P$ matrix M is of rank 3. This constraint on the rank relates the bilinear equation systems $\{f(t_f, \omega_f, n_f^p)\}$ of the different motion fields.

4 Motion and shape estimation from multiple flow fields

We first reformulate the estimation of motion and structure on the basis of individual flow fields. Then we embed the shape estimation into this formulation to obtain a multi-frame motion estimation.

For every frame I_f there is a bilinear system $\{f_f^p(t_f, \omega_f, n_f^p) = 0\}$ which we rewrite as:

$$A_f^p(t_f) n_f^p = d_f^p(\omega_f), \tag{6}$$

with

$$A_f^p(t_f) = t_{x_f} A_{1_f}^p + t_{y_f} A_{2_f}^p + t_{z_f} A_{3_f}^p$$

and

$$d_f^p(\omega_f) = -B_f^p \omega_f + b_f^p.$$

Estimating motion and structure from the individual flow fields is formulated as a least squares optimization, that is

$$\min_{\omega_f, t_f, n_f^p} \sum_p ||A_f^p(t_f) n_f^p - d_f^p(\omega_f)||^2 = \quad (7)$$

$$\min_{\omega_f, t_f} \min_{n_f^p} \sum_p ||A_f^p(t_f) n_f^p - d_f^p(\omega_f)||^2. \quad (8)$$

We can address this minimization in two iterative steps as follows: Given initial values for t_f and ω_f,

Step 1: Solve for structure: Substitute the values t_f, ω_f into the system and solve the over-determined linear system for all n_f^p using linear least squares estimation.

Step 2: Solve for motion: Substitute the values n_f^p into the system, and solve for t_f and ω_f using least squares estimation.

Go back to Step 1 until the estimation converges.

4.1 Adding the shape constraint to the estimation

The shape is parameterized by the normal vectors r_f^p. Thus, we first need to transform the non-homogeneous system $A_f^p n_f^p = d_f^p$ for n_f^p into a homogeneous system $W_f^p r_f^p = 0$ with $||r_f^p|| = 1$. This is shown next.

Let H_f^p be the matrix, such that the vector d_f^p spans the null space of H_f^p, i.e. $H_f^p d_f^p = 0$. H_f^p amounts to

$$H_f^p = I - \frac{d_f^p d_f^{p\,t}}{d_f^{p\,t} d_f^p}.$$

Then multiplying H_f^p on the both sides of the equation $A_f^p n_f^p = d_f^p$ yields

$$(H_f^p A_f^p) n_f^p = H_f^p d_f^p = 0.$$

Let $W_f^p = H_f^p A_f^p$ and $r_f^p = \frac{n_f^p}{\|n_f^p\|}$ (the normalized version of n_f^p), and we arrive at an homogeneous constraint on the normal vectors r_f^p of the form:

$$W_f^p r_f^p = 0, \quad \text{for} \quad p = 1, \cdots, P \quad \text{and} \quad f = 1, \cdots, F.$$

Incorporating the shape constraint, the minimization becomes:

$$\min_{t_f, \omega_f} \min_{||r_f^p||=1} \sum_{p,f} ||W_f^p(t_f, \omega_f) r_f^p||^2,$$
$$\text{subject to} \quad rank(M(r_f^p)) = 3. \quad (9)$$

We adopt the two-step optimization to solve for motion and structure from multiple frames simultaneously. Step 2, that is, the estimation of t_f and ω_f remains the same; given n_f^p we minimize equation (8) using least squares. However, Step 1, the estimation of shape, becomes rather different and is described in the next section.

4.2 Estimating the shape parameters

We are give the t_f's and ω_f's and wish to estimate the r_f^ps. We can pick any frames for integration, for example the first, tenth and twentieth frame of the sequence. In the general case, we will not concatenate the ω_f's to obtain a rotation between the selected frames, but re-estimate the rotation matrices.

Recall from Section 3 that $R_f = \Omega_{f,1} R_1$. Therefore

$$R_f^t R_f - R_1^t R_1 = R_1^t (\Omega_{f,1}^t \Omega_{f,1} - I) R_1 = 0.$$

In words, $R_f^t R_f$, which is the matrix encoding the relative orientations between the different patches, does not depend on the frame number. Using this we can rewrite the minimization (9) as:

$$\min \sum_{j,k} ||W_f^p r_f^p||^2 \quad \text{subject to} \quad ||r_f^p|| = 1, R_f^t R_f = R_1^t R_1. \quad (10)$$

This is a well defined least square minimization with quadratic constraints for which standard algorithms exist. We used the Mukai-Polak version of the Augmented Lagrangian method (ALS) ([8]) which guarantees super-linear convergence.

After having obtained the r_f^p from equation (10), we need to estimate the length of the n_f^p, in order to solve subsequently for translation and rotation. This is done using the first equation in $A_f^p n_f^p = d_f^p$.

4.2.1 Consecutive frames

When integrating a few consecutive frames (as opposed to significantly separated frames) $\Omega_{f,1}$ is well approximated by concatenating the differential motions ω_f. That is, $\Omega_{f,1} = (I + [\omega_f]_\times)\Omega_{f-1,1}$. For this case, the minimization is much simpler. Since $\Omega_{f,1}$ is known, the problem (9) becomes finding the least squares solution of an homogeneous system, that is:

$$\min_{||r_1^p||=1, r_f^p = \Omega_{f,1} r_1^p} \sum_{p,f} ||W_f^p r_f^p||^2 =$$
$$\min_{||r_1^p||=1} \sum_p (\sum_f ||W_f^p \Omega_{f,1} r_1^p||^2). \quad (11)$$

This minimization is solved using SVD decomposition (i.e r_1^p is the singular vector corresponding to the smallest singular value of $W_f^p \Omega_f$.)

5 Patch Segmentation and Matching

For our purposes the image segmentation does not need to give an "object-related" division of the scene, but merely

needs to locate some planar patches which are suitable for egomotion estimation and which can be matched over the image sequence. Thus oversegmentation is not a problem.

We perform segmentation and matching using a graph-based approach, which consists of the the following three components:

1. An edge enforced color segmentation in the individual frames, which provides an over-segmentation.

2. Matching of the color patches in the different frames of the sequence. The geometric transformation of the patches between the frames is modeled by the transformation $T(x, y)$, involving 2D translation, 2D rotation and scaling. The matching is carried out by phase correlation.

3. Taking a bottom-up approach to segmentation, neighboring color patches are merged using a graph-based algorithm.

5.1 Detailed algorithm and Motivation

Segmentation based on color usually works best for weakly textured regions, while segmentation based on flow information works best for well textured regions, which makes these two cues complement each other.

The procedure starts with an edge-enforced color segmentation which leads to an over-segmentation, and most likely many of the components will be too small for egomotion and shape estimation. However, color segmentation usually does not break smooth regions in space. There exist many color segmentation techniques which can achieve such a segmentation. We used the algorithm in [2].

Next the patches need to be merged on the basis of flow information. This is done not on the frames individually, but by linking the patches between the frames first, and then merging them using the flow information in all frames. Specifically, we model the movement of each color patch by the transform T and find the best matching color patch in every frame. Then we perform merging on the linked patches using a graph based approach. We adapted the algorithm in [4] (which was used for color matching) by using a measurement function based on flow measurements. That is, let $V = \{V_f, f = 1, \cdots, F\}$ be the patches matched from frame I_1 to I_F. The measurement function $R(V)$ is defined as

$$R(V) = \sum_f \frac{1}{k_f} ||(S(V_f))||_2,$$

where k_f encodes how good the matching of the patch in frame f is (the residual of the least squares fit to the transform T), and $S(V_f)$ encodes how well the flow of patch V_f fits the planar motion model (the residual of the least squares fit of the flow to the planar motion model).

Our approach uses flow from multiple frames. Thus usually it does not use segments which are non-planar or don't track well (for example patches close to occlusions). It may sometimes leave only a small fraction of patches in the image. However, this does not matter for the purpose of egomotion estimation.

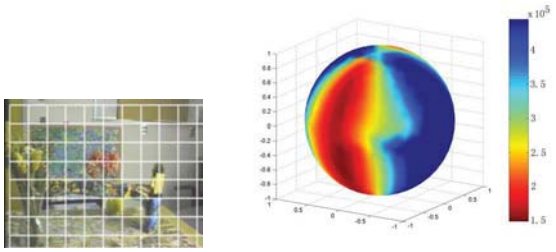
6 Experiments

6.1 3D motion estimation

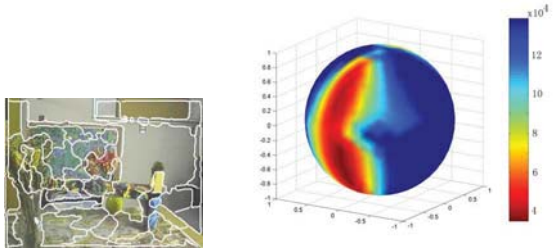
Figure 1: Key frame for the sequence "Office"

The motion for the color image sequence "office" (see Fig. 1a) (taken by a hand-held camera) is a translation mostly along the x- and y-axes and a rotation. For this motion, because of the camera's small field of view, the ambiguity between translation and rotation makes the estimation very difficult. Fig. 2 compares the effects of three different segmentations on the motion estimation from single flow fields. For every direction of translation we estimated the corresponding rotation, as the one which minimizes $\sum_p ||f^p(t, \omega, n^p)||^2$ and show the residual on the sphere of translational directions As can be seen, although the residual value decreases with better segmentation, the valleys are qualitatively very similar, demonstrating that the ambiguity in motion estimation cannot be avoided. Next we integrated multiple flow fields using the algorithm described in the paper. (We used three frames separated by a significant baseline.) Fig. 3 shows that the ambiguity has been reduced significantly, and it also also shows that our segmentation performs similar to the manual segmentation. The table below compares the estimation of the translational directions, $(\frac{t_x}{t_z}, \frac{t_y}{t_z})$, where we show the absolute smallest values found in the valley.

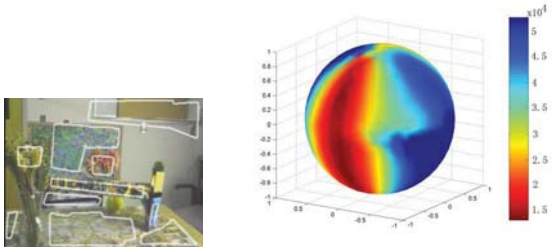
	Single frame	Multiple frame
Brute force Part.	(3.09, 0.00)	
Automatic Part.	(3.14, 0.02)	(2.15, -0.02)
Manual Part.	(2.30, 0.01)	(2.07, -0.01)



(a)Brute-force partition

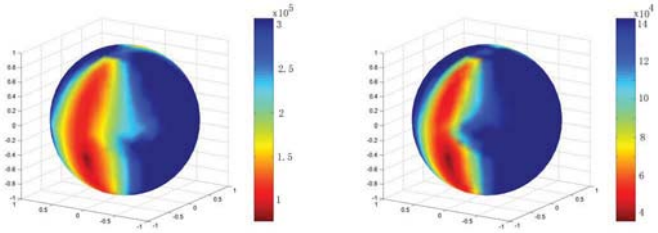


(b)Automatic segmentation by the method in the paper



(c)Manual segmentation

Figure 2: Residual sphere from egomotion estimation for one single frame in the sequence "office". The smallest values (shown in red) denote the possible candidate solutions, to which we refer to as the "motion valley".

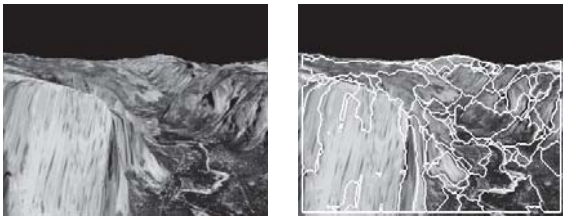


(a)Automatic segmentation (b)Manual segmentation

Figure 3: Residual for multi-frame motion estimation in the sequence "office"

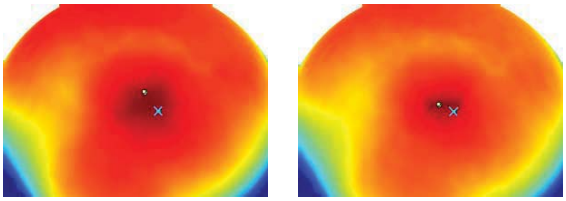
For the Yosemite sequence (Fig. 4), because of the large field of view and translation mostly along the z-axis, the ambiguity is not significant. We still achieved an improvement on this short sequence (5 frames only). (see Fig 5). The estimated directions of translations were as follows:

Single frame	Multiple frames	Ground truth
(0.00,0.149)	(0.00,0.158)	(0.00, 0.170)



(a) Key frame (b) Automatic segmentation

Figure 4: Yosemite sequence.



(a) Single frame (b) Multiple frames

Figure 5: Magnified local part of residual surface , with "." denoting the estimation and "x" denoting the ground truth.

6.2 Super-resolution imaging

We demonstrate the 3-D motion and shape estimation of our technique by super-resolving an image sequence containing documents, and then warping the document planes to a fronto-parallel view. A prerequisite for superresolution reconstruction from video is an accurate estimate of image motion. Usually the image motion is estimated with an affine motion model. We computed the image motion as the optical flow of a 3D plane using the 3D motion estimate obtained from the algorithm. Five frames are combined in a super-resolved frame (the middle frame), and the interpolation is obtained with the POCS (projection onto convex sets) technique [9], which models the point spread function; we used a simple Gaussian model. Referring to Figure 6, we compared the effect of estimating image motion using our method against modeling the flow with an affine transform. The improvement of our technique is clearly observable, most noticeable on the left patch where our technique makes the letters clearly readable (Fig. 6f).

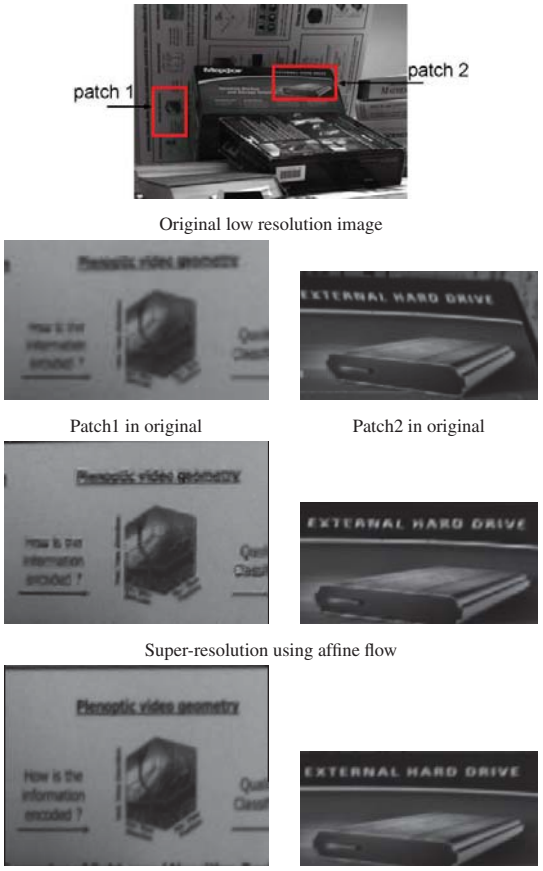


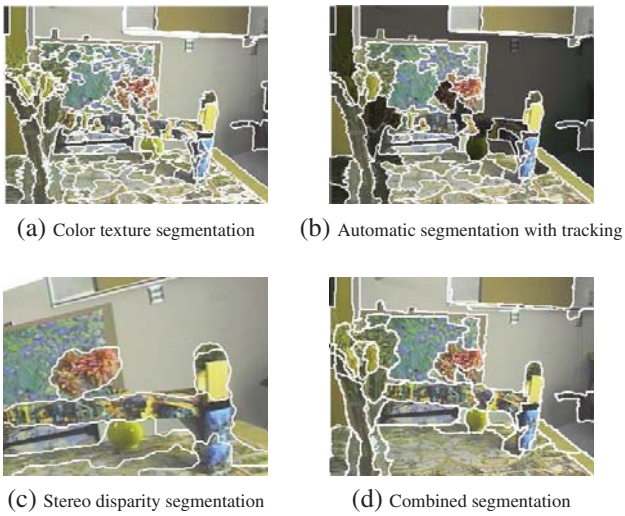
Figure 6: Super-resolution imaging and rewarping to fronto-parallel view: A comparison of the affine flow model against flow computed on the basis of 3D motion estimate.

7 Structure from motion in feedback

We argue that we cannot obtain good models using only local measurements (image motion or correspondence) in a bottom up approach, as they not allow good structure estimates and accurate localization of the discontinuities. For this, we need processes that involve larger spatial areas. But to employ such processes we need models of the scene. In other words, there need to be feed-back loops.

Our algorithm provides us with very good 3D motion estimates over multiple frames. In the sequel we can obtain depth from image motion and we can fit shape models to the segmented patches. Using this information we can now go back to the images to better segment and estimate structure.

First experimental results for segmentation are shown. Figure 7a shows the color segmentation and Figure 7b the final segmentation of the algorithm, where the patches which have been discarded (because of non-planarity or lack of gradients) are dimmed. Using the 3D motion estimate we rectified two frames significantly separated, and computed the depth map using a graph-cut stereo algorithm (Fig. 7c). The figure demonstrates that our motion estimate lets us reduce the structure from motion problem to a stereo problem. Then we inserted the boundaries obtained from stereo into the segmentation of our algorithm and merged areas for which the flow gave continuous depth values (Fig. 7d). This gives a segmentation based on motion, stereo and color.



(a) Color texture segmentation (b) Automatic segmentation with tracking

(c) Stereo disparity segmentation (d) Combined segmentation

Figure 7: Segmentations

References

[1] T. Brodský, C. Fermüller, and Y. Aloimonos. Structure from motion: Beyond the epipolar constraint. *International Journal of Computer Vision*, 37:231–258, 2000.

[2] D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619, 2002.

[3] K. Daniilidis and H.-H. Nagel. Analytical results on error sensitivity of motion estimation from two views. *Image and Vision Computing*, 8:297–303, 1990.

[4] P. Felzenszwalb and D. Huttenlocher. Efficient graph-based image segmentation. *International Journal of Computer Vision*, 59(2):167–181, 2004.

[5] D. J. Heeger and A. D. Jepson. Subspace methods for recovering rigid motion I: Algorithm and implementation. *International Journal of Computer Vision*, 7:95–117, 1992.

[6] L.Zelnik-Manor and M. Irani. Multi-view subspace constraints on homographies. In *Proc. International Conference on Computer Vision*, 1999.

[7] L.Zelnik-Manor and M. Irani. Multi-frame estimation of planar motion. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(10):1105–1116, 2000.

[8] E. Polak. *Optimization: algorithm and consistent approximation*. Springer, 1996.

[9] H. Stark and P. Oskoui. High-resolution image recovery from image-plane arrays using convex projections. *Optical Society of America A*, 6:1715–1726, 1989.