Three dimensional transparent structure segmentation and multiple 3D motion estimation from monocular perspective image sequences

Stefano Soatto† and **Pietro Perona**†‡
† California Institute of Technology 116-81, Pasadena–CA 91125
‡ Università di Padova, Dipartimento di Elettronica, Padova–Italy
{soatto,perona}@caltech.edu

Technical Report **CIT-CNS 33/93** and **CIT-CDS 93-022**
California Institute of Technology
November 15,1993. Revised Feb. 8,1994

Abstract

A three dimensional scene can be segmented using different cues, such as boundaries, texture, motion, discontinuities of the optical flow, stereo, models for structure etc. . We investigate segmentation based upon one of these cues, namely three dimensional motion.

If the scene contains transparent objects, the two dimensional (local) cues are inconsistent, since neighboring points with similar optical flow can correspond to different objects.

We present a method for performing three dimensional motion-based segmentation of (possibly) transparent scenes together with recursive estimation of the motion of each independent rigid object from monocular perspective images.

To our knowledge this problem has never been addressed in the literature.

Our algorithm is based on a recently proposed method for rigid motion reconstruction and a validation test which allows us to initialize the scheme and detect outliers during the motion estimation procedure.

The scheme is tested on challenging real and synthetic image sequences. Segmentation is performed for the Ullmann's experiment of two transparent cylinders rotating about the same axis in opposite directions.

1 Introduction

Many cues may be used for scene segmentation, such as boundaries, texture, discontinuities of the optical flow, stereo, motion etc. . Ultimately a system for performing three dimensional scene segmentation ought to integrate all the information available by exploiting each cue.

There are two *motion cues* that might be used for scene segmentation: 2D motion on the image plane, where optical flow discontinuities are projections of scene depth and/or 3D motion discontinuities, or 3D motion itself. There are a number of assumptions as well: object rigidity, piecewise smoothness of the scene, object opaqueness (which, together with all previous assumptions translates into piecewise smoothness of the optical flow), existence of a "dominant motion". Accordingly, the motion-based segmentation algorithms may be classified into a number of categories. 2D optical flow region-based algorithms [3, 21, 5, 11], 3D region-based [8, 20], and *transparent* 3D motion [25, 4]. We call "transparent 3D motion" algorithms the ones which do not make use of regions-contiguity assumptions, and may therefore handle motion of transparent objects.

In some situations 3D motion may be the only available cue. Consider for example Ullmann's experiment [23, 22], in which we project onto a screen two coaxial transparent cylinders, rotating in opposite directions (see fig. 3 R). We are clearly able to perceive the existence of two independent motions; however, the local (2D) information is inconsistent. In fact in a neighborhood of each feature point on the image plane

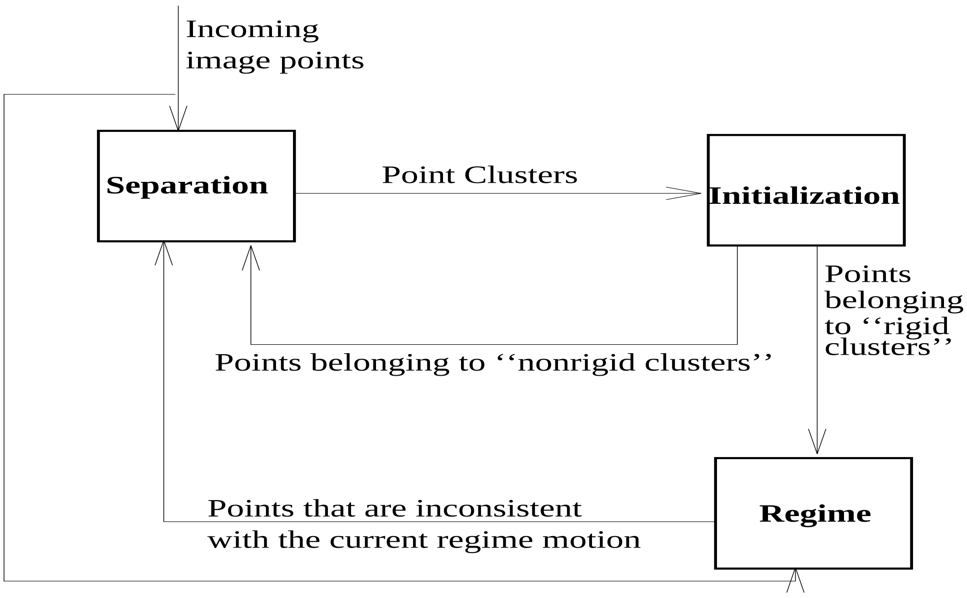


Figure 1: *Structure of the segmentation scheme*

there are points moving with the same 3D motion and similar 2D motion, same 3D motion and opposite 2D motion, opposite 3D motion and similar 2D motion, opposite 3D motion and opposite 2D motion.

In this paper we present a method for segmenting a scene from a sequence of monocular images using only 3D motion cues. We make no use of spatial contiguity, and hence we are able to perform on transparent motions. The main assumption is that each object populating the scene is a rigid body. The scheme also estimates recursively all independent motions. To our knowledge transparent structure segmentation together with recursive estimation of multiple motions from monocular sequences has never been addressed in the literature.

We will first sketch an outline of the algorithm (section 2). It consists of a *separation* step, which composes clusters of points having high probability of belonging to the same rigid object, an *initialization* step in which a filter is assigned to each cluster, and then a *regime* phase, which is characterized by having a filter associated to each rigid object. During the regime phase the rigidity assumption is constantly checked and, if the object splits into more than one independent body, the points which are incompatible with the current motion are rejected and returned to the separation phase (see fig. 1). In the later sections the operation of each step is analyzed in detail. In section 3 we describe the essential filter, which is a recently proposed motion estimation scheme [18], and introduce an innovation-based self-validating test, the *predictive innovation test*. In section 4 the operation of the separation and initialization phases is explained and some open issues are discussed. Finally in the experimental section we study the behavior of the scheme on real and synthetic image sequences.

2 Outline of the segmentation method

The scheme which we propose consists of three "modes of operation" which are constantly active during the segmentation procedure. A supervision program is in charge of assigning to each feature point a mode of operation.

Separation Suppose we are at the initial time instant. We do not know how many objects are moving in the scene and which points belong to which objects. The separation step produces a set of clusters (one for each point) which have high probability of belonging to a single rigid motion.

Initialization The initialization mode takes the output of the separation step, namely a set of clusters of points, and runs a motion estimation algorithm (the essential estimator) in parallel for each cluster. After a settling time it gives either a convergence verdict, which promotes the cluster to the regime stage, or a divergence verdict, which causes the cluster to be assigned to the separation again.

Regime The clusters which are promoted from the initialization mode enter into "regime" mode. Each object is assigned to a filter (a motion estimator) which is in charge of estimating the rigid motion of the object and constantly checking for outliers (points whose motion is not consistent with the rigid interpretation). This is done using a very simple criterion which we call the "predictive innovation test".

3 *"Regime mode"*: essential filter and the innovation test

The literature proposes a variety of schemes for recursively estimating motion, for example [1, 17, 15, 7, 6, 9, 18], just to mention the most recent ones. All the schemes assume that all features in the scene move rigidly; they fail when this assumption is relaxed. In order to perform 3D motion-based segmentation, a motion estimation algorithm should produce, together with the motion estimates, also a measure of the "consistency" of each point with the current motion interpretation. Since we want to be able to reject and acquire points, we cannot include them in the state dynamics, otherwise we would have a variable number of states which is cumbersome and leads to convergence problems. On the other hand we want to measure the reliability of each point, so that they must be represented in the filter dynamics.

In this section we will briefly review a recently proposed algorithm for estimating rigid motion [18]. It is peculiar in that it does *not* include structure in the dynamics of the filter, allowing us to change the set of feature points at each step. The consistency of each point with the current motion, however, is represented *on line* via its *innovation* process. Other schemes include both motion and structure in the state dynamics [1], or include only the structure [17], and hence are not suitable for our purposes.

The essential filter

The essential filter [18] can be viewed as a recursive extension of the basic coplanarity constraint introduced by Longuet-Higgins [14] (see also [24, 16, 9])

$$\begin{cases} \mathbf{x}_i^T(t+1)\mathbf{Q}(t)\mathbf{x}_i(t) = 0 \\ \tilde{\mathbf{x}}_i(t) = \mathbf{x}_i(t) + n_i(t) \end{cases} \qquad \forall i = 1:N \tag{1}$$

where $\mathbf{x}_i(t)$ are the projective coordinates of each of the N visible points in the viewer's reference at time t, $\mathbf{Q} \doteq R(T\wedge)$, where (T, R) is the rigid motion undergone by the observer between time t and $t+1$ and $\tilde{\mathbf{x}}_i$ are the noisy measurements of the image plane coordinates. It is customary to assume $n_i \in \mathcal{N}(0, \sigma_i^2)$. The basic step of the essential estimator is of the form

$$\begin{bmatrix} \hat{T} \\ \hat{R} \end{bmatrix}(t+1) = \begin{bmatrix} \hat{T} \\ \hat{R} \end{bmatrix}(t) + L(t) \begin{bmatrix} \vdots \\ \tilde{\mathbf{x}}_i^T(t)\mathbf{Q}(\hat{T}, \hat{R})\tilde{\mathbf{x}}_i(t-1) \\ \vdots \end{bmatrix} \tag{2}$$

where L has the structure of the gain of an Extended Kalman Filter (EKF) [13, 12] whose states are the motion parameters. The quantities

$$\epsilon_i(t) \doteq \tilde{\mathbf{x}}_i^T(t)\mathbf{Q}(\hat{T}, \hat{R})\tilde{\mathbf{x}}_i(t-1) \qquad \forall i = 1:N \tag{3}$$

are the components of the *pseudo-innovation* vector, and measure how far each point is from the current motion interpretation $(\hat{T}, \hat{R})$. The essential filter also updates the variance of the motion estimation error through a discrete Riccati equation. Since the constraint (1) is *linear* in $\mathbf{Q}$, we use the (improper) notation ${\mathbf{x}'}_i^T\mathbf{Q}\mathbf{x}_i \doteq \chi_i(\mathbf{x}', \mathbf{x})\mathbf{Q} = 0$. Once N points are observed we can stack the measurements into a $N \times 9$ matrix χ and write $\chi\mathbf{Q} = 0$. We also use the shorthand $\hat{\mathbf{Q}}$ for $\mathbf{Q}(\hat{T}, \hat{R})$. The matrix $\mathbf{Q}$ belongs to the the so called "essential manifold" [18, 16].

Consistency with the rigidity assumption

Suppose at time t the filter is in steady state operation, and is estimating a rigid motion with some innovation norm (typically on the order of 10^{-2} to 10^{-4}). Suppose at time $t+1$ some points enter the scene which do not belong to that rigid motion. At time t the filter has produced the best prediction of motion at time $t+1$ given the measurements up to time t: $\hat{\mathbf{Q}}(t+1|t)$. We can therefore make a "prediction" of the innovation process $\hat{\epsilon}_i(t+1|t) \doteq \chi(\mathbf{x}', \mathbf{x})\hat{\mathbf{Q}}(t+1|t)$ and compare each component against the variance of the previous innovation: $\sigma^2_\epsilon(t)$.

In our implementation we reject at each time all the points which produce a residual error $\hat{\epsilon}_i(t+1|t)$ greater than one standard deviation of the innovation. Furthermore we can include into the filter any point which comes into the scene and produces a residual error within a standard deviation of the innovation. This allows dealing easily with occlusion, appearance of new feature points and splitting of rigid objects.

The above discussion relies on the standing assumption that the filter is in steady state operation, hence estimating the motion of a single moving object. What can we do at the initial time, when we have no clue of what the motions in the scenes are? We will show in the next sections how the innovation test can be exploited to initialize a filter for each moving object.

4 *"Separation"* and *"initialization"* modes

At the initial time instant we have a set of points for which we assume that the correspondence problem is solved. The first thing one is tempted to do is to run a filter until it converges to some "dominant" motion, rejecting progressively all the points which are not compatible, then assign the rejected points to a new filter, and so on, until all the points are assigned to a filter. However, the essential scheme is very sensitive to the presence of outliers (which is the key of the regime mode), and it does not converge if more than few points are inconsistent with a single rigid motion interpretation. Furthermore the innovation test can be done only when convergence is reached: if not, the norm of the innovation is large, which causes all the points to be rejected.

The separation mode is in charge of constructing a number of "clusters" of points which are likely to belong to the same rigid object, based only on 3-D motion (hence not exploiting local 2-D cues). The initialization phase runs a filter for each cluster and merges the clusters that have converged to similar motions.

Separation of initial motions

Let us examine the structure of the innovation (or residual) ϵ. It is the image of χ via $\hat{\mathbf{Q}}$, considered as an element of the vector space $\mathbb{R}^9$. If all the N points which build up χ were part of a rigid body, and no noise was present, then $\hat{\mathbf{Q}}$ would span the null space of χ and the residual error would be zero. Suppose some point i is added which does not belong to the rigid motion, then the corresponding component of the residual error $\epsilon_i \doteq \chi_i\hat{\mathbf{Q}}$ is greater than zero and the point can be easily spotted. However, we do not know $\hat{\mathbf{Q}}$, and in fact there might be many objects moving, each with its corresponding motion $\hat{\mathbf{Q}}$. Now suppose two objects are undergoing independent and unknown motions. The matrix χ has now full rank [16]. Let us define the *"residual space"* as the span of χ. The intuition is that, if we pick up an arbitrary motion $\tilde{\mathbf{Q}}$, the errors $\chi\tilde{\mathbf{Q}}$ in the residual space corresponding to points which belong to the same motions are comparable. For example when $\tilde{\mathbf{Q}}$ is very close to the motion of one of the two objects, its points will produce a very small residual, while others will have larger errors. We want to explore experimentally the possibility of using a similar criterion for separating points based on their residual errors.

One could think of computing residuals with respect to an arbitrary motion set $< \mathbf{Q}_i >_{\{i=1:K\}}$ for grouping points which are associated by similar rigid motions. A question of *sufficient excitation* arises about the family of motions we choose [19]. If the family $< \mathbf{Q}_i >_{\{i=1:K\}}$ is chosen properly, points corresponding to different rigid motions will group into different clusters in the residual space. Some questions arise at this point: which family of motion vectors do we use? how do we perform the clustering if separation occurs?

Our choice for the family $< \mathbf{Q}_i >_{\{i=1:K\}}$ is the canonical basis of the motion space $\mathbb{R}^5$ lifted to the essential manifold [18]. This choice seems to be the most natural, though it might be far from the optimal. Another simple choice of sufficiently exciting motions are random vectors in $\mathbb{R}^5$ lifted to the essential manifold. We could also employ the canonical basis (or random vectors) in $\mathbb{R}^9$, although they may represent points which are not on the essential manifold.

Given any basis of K elements, for each measurement set χ we produce a matrix $\mathcal{E} \doteq [\epsilon_1, \epsilon_2, ...\epsilon_K]$. Then we cluster the points using a nearest neighbor criterion *in the residual space*. To do so, we produce a matrix $D = \{d_{i,j}\}$ measuring the distance of the error vectors corresponding to each couple of points: $d_{i,j} = \|\epsilon_i - \epsilon_j\|$. D is a $N \times N$ matrix, called the separation matrix. We mark for each point i (row) all the points j (columns) which have an error smaller than a threshold: $d_{i,j} \leq \gamma$. In our experiments we have used $\gamma = 3\ mean(\Delta)$, where Δ is the vector having as its elements i the minimum distance of the point i from the other points.

We have tested the separating power of this procedure on a variety of motions and points configurations. We have evaluated roughly as 0.3 the probability of having clusters which contain no spurious points and more than 40 % of the correct points. Therefore out of 100 clusters generated (one about each point), 30 contain at least 40 points which are moving with a coherent rigid motion. The essential filters initialized for such sets converge from an arbitrary initial conditions. Some instances are reported in the experimental section.

Initialization phase

The separation procedure has produced N clusters of points. For each of these clusters we start an essential filter. According to the estimates of the separation step, for 100 clusters, one about each point, 30 will have a set of at least 40 points all belonging to the same rigid motion. We initialize each filter with one step of the basic Longuet-Higgins algorithm [14].

After some settling time (20 steps) we evaluate the norm of the innovation process for each filter. We discard filters with high innovation norm (≥ 1), and we merge together points belonging to clusters which have produced motions whose difference is in the range of a standard deviation of the estimation error. At this point we have initialized the algorithm and we have one essential filter associated to each rigid cluster.

5 Experimental assessment

In this section we will show the results of some experiments on the operation of the segmentation scheme on real and synthetic image sequences. We will show each mode of operation separately: first the performance of the *separation* step is tested on a synthetic set of transparent clouds of points rotating about two orthogonal axes. The same is then repeated when the two clouds are rotating *about the same axis* in opposite directions (Ullmann's experiment [23, 22]). Then the *initialization mode* is tested on typical sets of points of the rotating clouds. We show the convergence of a filter associated to a cluster containing no spurious points and the divergence of a filter attached to a cluster with 20 % of spurious points. We then show the behavior of the *regime* phase when a rigid object attached to a filter splits into two objects which move with independent motions. Finally the same experiments are performed on a sequence obtained from the "rocket" scene (see fig. 8) by mirroring the motion of some points.

Throughout the experiments we have used initial information about the scale factor (norm of initial translation or distance from the centroid) and then propagated it through the estimation procedure. In the synthetic sequences the images are generated by a simulation program which adds gaussian noise to the image plane measurements with 1 pixel std, according to the performance of the most common feature tracking and optical flow schemes [2].

5.1 Separation

Transparent objects rotating about orthogonal axes

Two clouds of points in the same 3D region undergo a rotational motion about two orthogonal axes. An example of an optical flow generated by this sequence is shown in fig. 2 (L). As it can be seen the two clusters can be separated quite easily based on the direction of the 2D flow. However, neighboring points moving with the same 3D motion can have opposite 2D velocity. In fig. 2 (R) is shown the matrix D described

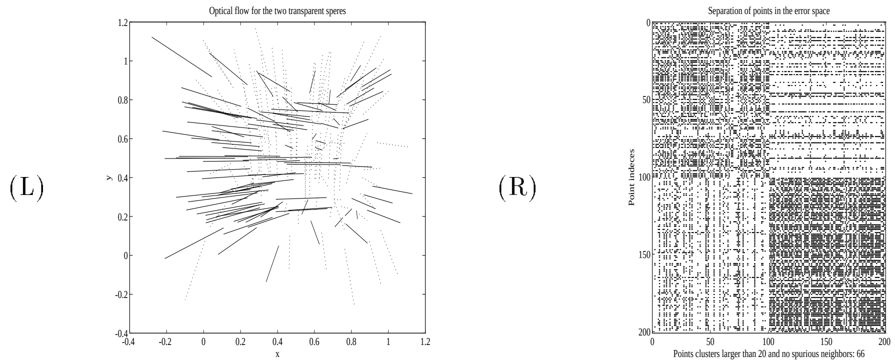


Figure 2: *(L) Optical flow generated by two clouds of points rotating about two orthogonal axes. Points belonging to one cloud are plotted with dotted lines, while the other cloud is plotted in solid lines. (R) Separation matrix. For each point (row) we mark a dot on each other point (column) for which the difference of the residuals ($d_{i,j}$) is smaller than a threshold. The points belonging to one object are ordered from row 1 to row 100, while points of the second object are labeled from 101 to 200. Ideally we would like to see two black diagonal blocks, meaning that each cluster contains all and only the points moving coherently. This does not happen in the experiments; however, the number of clusters having no spurious neighbors and collecting more than 20 points are 66 out of 200 (circa 30%).*

in section 4 (the separation matrix). Points satisfying the neighboring criterion in the residual space are marked as dots. In this example points from 1 to 100 belong to one objectn, and from 101 to 200 belong to the object rotating about the orthogonal axis. Hence in an ideal situation we expect a symmetric, block diagonal structure with zeros on the off-diagonal blocks. Instead, the number of clusters having no spurious neighbors and collecting more than 20 points are 66 out of 200 (circa 30%). Hence for 200 filters which run independently in the initialization phase, at least 66 will converge to a rigid motion. In fig. 3 (L) we show an image plane view of the selected points for the cluster No. 66. It can be seen that the selected points are mixed with other points which belong to the orthogonal motion.

Transparent objects rotating about the same axis with different directions

The same experiment described in the previous section is repeated when the two clouds of points are rotating about the same axis in opposite directions (see figure 3). Psychophysical experiments showed that this is a difficult task for humans; 3D motion is the only available cue.

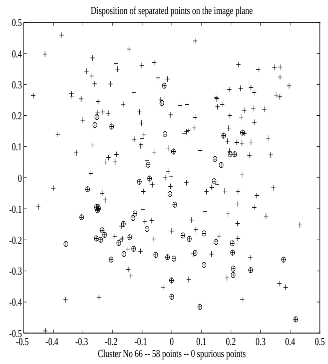
The image plane view is reported in fig. 4 (L), and the corresponding separation matrix D in fig. 4 (R). The number of clusters collecting no spurious neighbors is smaller than in the previous experiment. However the number of pure clusters with more than 20 points is still 12, which corresponds to 5% circa of the original feature set. Filters initiated with one of the 12 pure (rigid) clusters converge to the proper motion allowing the scheme to be initialized correctly.

5.2 Initialization

In this section we show a prototype of a converging cluster (fig. 5 R) and a diverging one (fig. 5 L). Motion is represented using six components (three of translation and three of rotational velocity); ground truth is shown in dotted lined.

5.3 Regime: a motion splitting experiment

In this section we show an experiment of a splitting object: one of the clouds of points is rotating and a regime filter is tracking its motion. After 25 frames the cloud breaks into two sets of points: one keeps

(L) 

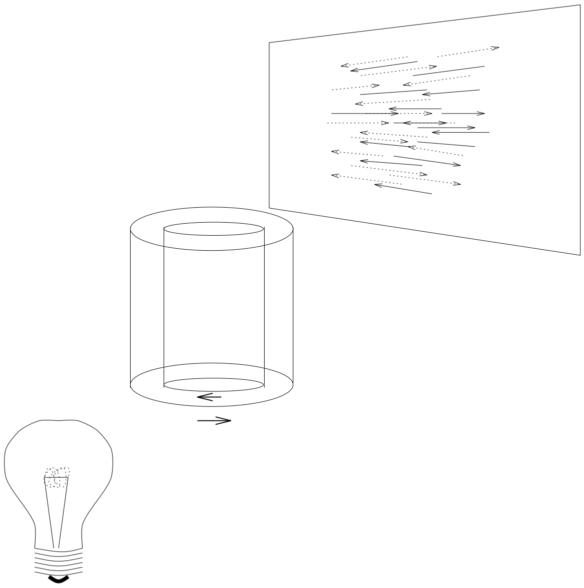
(R) 

Figure 3: *(L) Disposition of selected points (circled) on the image plane. It can be seen that the selected points are mixed with other points which belong to the orthogonal motion.(R) Illustration of the Ullmann experiment. Two transparent cylinders rotate about the same axis and in opposite directions. The only cue for segmentation is three dimensional motion.*

on rotating with the same motion, while the other starts rotating about an orthogonal axis. All the points which belong to the split cloud are rejected by the filter. Since all of them belong to the same rigid motion, the new filter initialized with the rejected points converges rapidly to the motion of the new split cloud. In fig. 6 (L) we show the motion for the cluster which continues after the splitting, and in fig. 6 (R) we show the motion estimates for the split cloud.

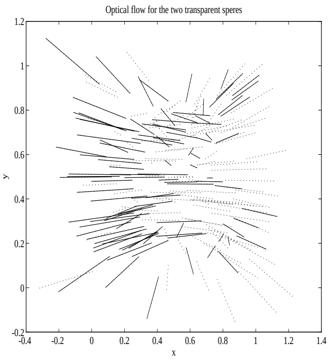
6 Experiment with a real image sequence: the mirror-rocket scene

In this section we show the performance of the segmentation and motion estimation scheme on a real image sequence which is obtained by mirror-imaging the rocket scene (fig. 8). In fig. 7 (L) the optical flow is plotted for one frame of the 11 available (10 correspondences). In fig. 7 (L) the separation matrix is plotted. Among the clusters that collect no spurious points there are 10 out of 22 with more than 5 points. In this case we have to perform a more accurate initialization. A Horn basic step [10] on 5 successive correspondences suffices for the purpose. In fig. 8 (left) we can see the selected points for one of the clusters, and finally in fig. 8 (center) the motion components for one of the clusters as estimated by the essential estimator. The iteration is run twice on the same data to allow the scheme to converge (there is a transient of about 20 steps, while the sequence is only 10 frames long).

7 Conclusions

We have presented a method for performing three dimensional transparent structure segmentation and multiple motion estimation. It is based on a motion estimation scheme [18] integrated with a self-validation test, called the predictive innovation test.

Experiments with challenging real and synthetic image sequences have been performed, and the algorithm

(L) 

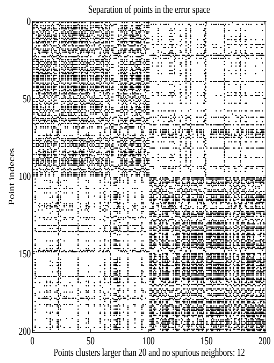
(R) 

Figure 4: *Optical flow generated by the Ullmann experiments. Two clouds are rotated about the same axis in opposite directions. Observe that in this case no region-based algorithm could work and 3D "transparent" motion is the only available cue. (R) Separation matrix. The number of pure clusters with more than 20 points is 12, which corresponds to 5% circa of the original feature set.*

is able to segment the classical Ullmann's scene of two transparent clouds rotating about the same axis in opposite directions.

To our knowledge no previous work in the literature addressed the problem of transparent structure segmentation and multiple motion estimation from monocular sequences. There are a number of open issues, like the choice of a sufficiently exciting set (or a basis) of the motion space for clustering (separation phase), different methods to perform the unsupervised grouping, evaluation on the separating power of the basis we currently use.

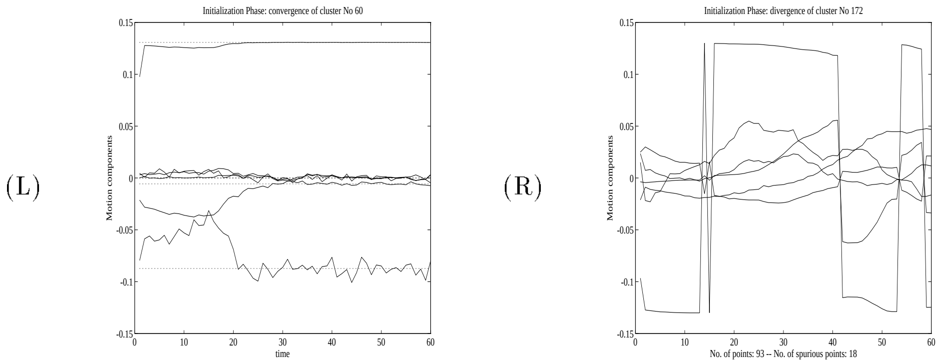


Figure 5: *Initialization phase: convergence (L) or divergence (R) of clusters of points. The motion coordinates (three for rotation and three for translation) are plotted in solid lines as estimated in the initialization phase. The behavior of a typical converging cluster and a typical diverging one is plotted. Ground truth is in dotted lines. Note that 20 steps are sufficient for deciding whether a filter has converged or not. Also note that the diverging cluster has 18 spurious points out or 93, i.e. circa 20%, which is sufficient not to reach convergence on the "dominant motion".*

References

[1] A. Azarbayejani, B. Horowitz, and A. Pentland. Recursive estimation of structure and motion using relative orientation constraints. *Proc. CVPR*, New York, 1993.

[2] J. Barron, D. Fleet, and S. Beauchemin. Performance of optical flow techniques. RPL-TR 9107, Queen's University Kingston, Ontario, Robotics and perception laboratory, 1992. Also in Proc. CVPR 1992, pp 236-242.

[3] M. Black. Combining intensity and motion for incremental segmentation and tracking over a long image sequence. *Proc. of the European Conf. on Comp. Vision*, 1992.

[4] Terrance E. Boult and Lisa G. Brown. Factorization-based segmentation of motions. In *Proceedings of the IEEE Workshop on Visual Motion*, pages 179–186, October 1991.

[5] P. Bouthemin and E. Francois. Motion segmentation and qualitative scene analysis from an image sequence. *Int. J. Comp. Vision IJCV*, 1993.

[6] T. Broida and R. Chellappa. Estimating the kinematics and structure of a rigid object from a sequence of monocular frames. *IEEE Trans. Pattern Anal. Mach. Intell.*, 1991.

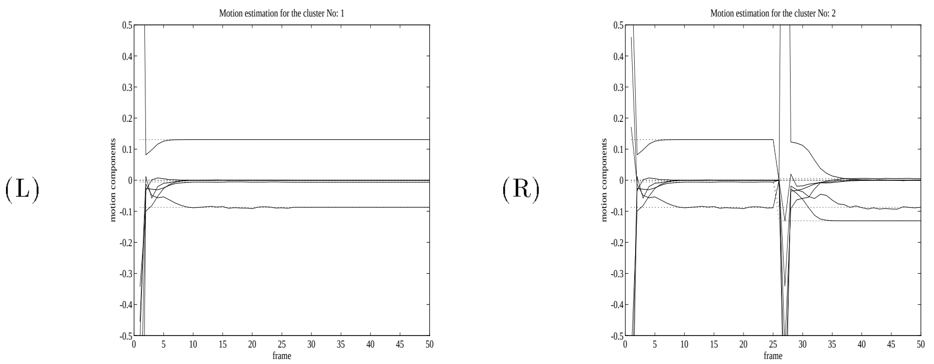
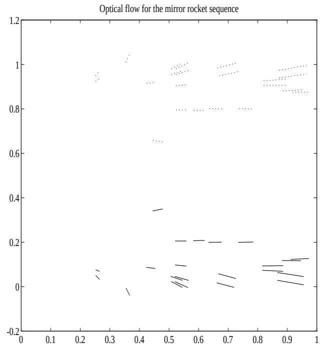


Figure 6: *Motion estimates for the splitting experiment: cluster of points with continuous motion (L) and split cluster (R). Filter estimates (solid) vs. ground truth (dotted)*

(L) 

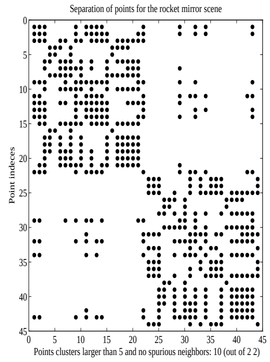
(R) 

Figure 7: *(L) Optical flow for the mirrored rocket sequence. Flow corresponding to the mirror image is plotted in dotted arrows. (R) Separation matrix for the rocket scene. Among the clusters that collect no spurious points there are 10 with more than 5 points*

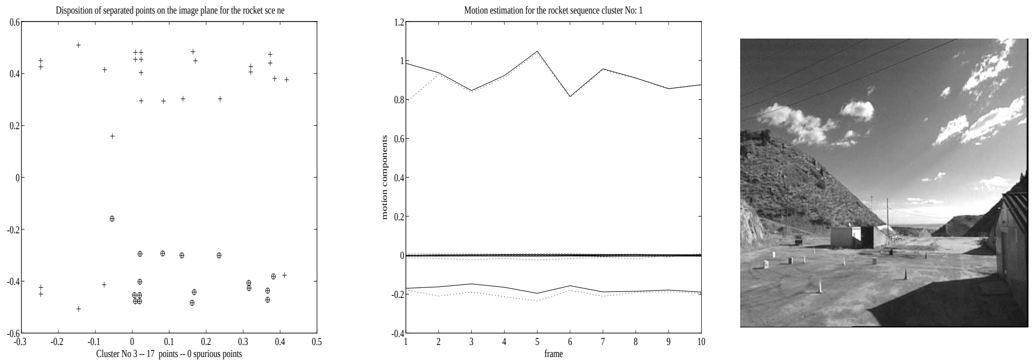


Figure 8: *(Left)Selected points for the rocket experiments: no spurious points are selected. (Center) Motion estimates for the rocket sequence. Ground truth is plotted in dotted lines. The iteration is run twice on the same data to allow the scheme to converge. (Right) One image of the rocket scene.*

[7] T. Broida and R. Chellappa. Estimation of object motion parameters from noisy images. *IEEE trans. PAMI*, Jan. 1986.

[8] C. Debrunner and N. Ahuja. Motion and structure factorization and segmentation of long multiple motion image sequences. *Proc. of the European Conf. on Comp. Vision*, 1992.

[9] O. Faugeras. *Three dimensional vision, a geometric viewpoint.* MIT Press, 1993.

[10] B.K.P. Horn. Relative orientation revisited. *J. Opt. Soc. Am. A*, 8(19):1630–1638, 1991.

[11] M. Irani, B. Rousso, and S. Peleg. Detecting and tracking of multiple moving objects using temporal integration. *Proc. of the European Conf. on Comp. Vision*, 1992.

[12] A.H. Jazwinski. *Stochastic Processes and Filtering Theory.* Academic Press, 1970.

[13] R.E. Kalman. A new approach to linear filtering and prediction problems. *Trans. of the ASME-Journal of basic engineering.*, 35-45, 1960.

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

[15] L. Matthies, R. Szelisky, and T. Kanade. Kalman filter-based algorithms for estimating depth from image sequences. *Int. J. of computer vision*, 1989.

[16] S. Maybank. *Theory of reconstruction from image motion*, volume 28 of *Information Sciences*. Springer-Verlag, 1992.

[17] J. Oliensis and J. Inigo-Thomas. Recursive multi-frame structure from motion incorporating motion error. *Proc. DARPA Image Understanding Workshop*, 1992.

[18] S. Soatto, R. Frezza, and P. Perona. Motion estimation on the essential manifold. *Proc. of the ECCV 94 – To appear in "Lecture Notes in Computer Sciences", Springer Verlag*, May 1994.

[19] T. Soderstorm and P. Stoica. *System Identification*. Prentice Hall, 1989.

[20] W. Thompson, P. Lechleider, and E. Stuck. Detecting moving objects using the rigidity constraint. *IEEE Trans. Pattern Anal. Mach. Intell.*, 1993.

[21] P. Torr and D. Murray. Statistical detection of independent movement from a moving camera. *Image and Vision Computing*, 1993.

[22] Shimon Ullmann. The interpretation of structure from motion. *Proc. R. Soc. London 203*, 1979.

[23] Shimon Ullmann. *The Interpretation ov Visual Motion*. MIT Press, 1979.

[24] J. Weng, T.S. Huang, and N. Ahuja. Motion and structure from line correspondences: closed-form solution, uniqueness and optimization. *IEEE Trans. Pattern Anal. Mach. Intell.*, 14(3):318–336, 1992.

[25] Z. Zhang and O. Faugeras. Three dimensional motion computation and object segmentation in a long sequence of stereo frames. *Int. J. of Computer Vision IJCV*, 1992.