M.I.T. Media Laboratory Vision and Modeling Group Technical Report No. 281
Appeared, Proc. Wkshp. Performance vs. Methodology in C.V., CVPR-94

Segmentation of Rigidly Moving Objects using Multiple Kalman Filters

Trevor Darrell, Ali Azarbayejani, and Alex P. Pentland

Perceptual Computing Group
The Media Laboratory
Massachusetts Institute of Technology
20 Ames Street
Cambridge MA, 02139

ABSTRACT

In this paper we describe a method for structure-from-motion recovery when there are multiple objects in the scene. We use our recently developed recursive estimation technique to recover shape and motion parameters given a set of features being tracked in the image. Support maps defined for these features limit the integration of information across space when there are multiple objects in a scene, due to occlusions or interleaved regions. Multiple hypothetical models are run concurrently, based on random initial groupings of the data. A minimum description length selection mechanism determines which 3-D structure/motion models and which groups of features constitute the best match to the data. We show results segmenting a synthetic sequence containing features on two rotating spheres, where there are no static cues available for segmentation, and on a real image sequence containing both camera and object motion.

1 Introduction

The problem of recovering shape information from a sequence of images containing moving objects is a central issue in computer vision. Many methods have been proposed for recovering the structure and motion parameters of a single rigidly moving object given a set of tracked features or reliable optical flow information [1, 3, 7, 11, 12, 14, 15, 16, 17, 18, 19]. In general these methods do not address the question of how to segment the motion information when there are multiple motions in the sequence (but see, for instance, [2]).

If there is only a single rigid motion in the scene, such as in the case of ego-motion and a static scene, or a single moving object with a uniform background, then all features can be integrated together. If, however, there are multiple moving objects and no obvious static segmentation cues, then some method is needed to segment the feature tracks into groups corresponding to each object. In this paper we provide a mechanism that can correctly segment the scene into objects while simultaneously recovering the motion and structure information for each object, given only a set of tracked features and no spatial segmentation cues.

Our method uses a multiple-model approach, testing several hypotheses about object motion and then selecting which subset of hypotheses best accounts for the data. To find whether an object exists in the scene with the hypothesized motion, we check which feature tracks are compatible with that motion. We then find the most parsimonious set of hypotheses, e.g. the smallest set which can account for all the feature tracks. Given that the rigid-body assumption is correct, after the selection stage each real object in the scene will have a separate structure from motion computation modeling shape and motion for that object.

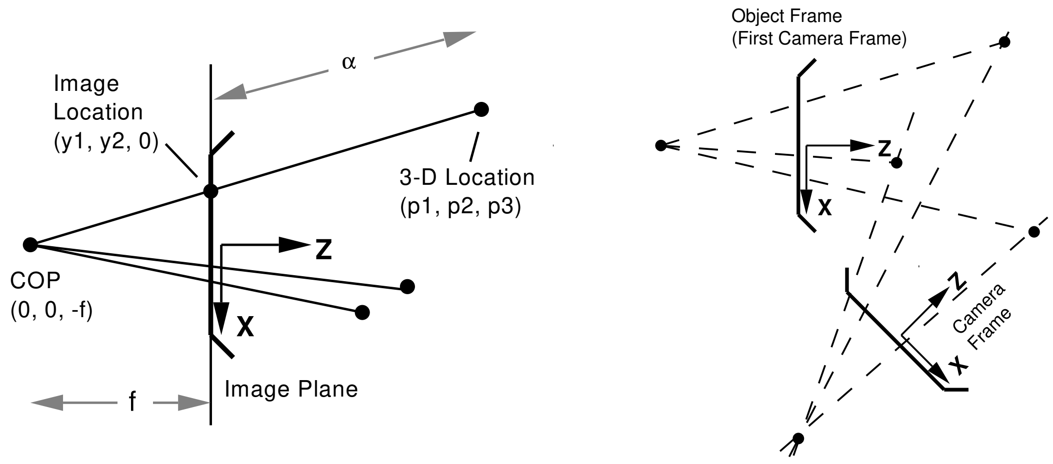

Figure 1: Imaging Geometry

The motion model we use is based on the recursive estimation approach of Azarbayejani and Pentland [1]. This method formulates the estimation problem as an instance of an Extended Kalman Filter (EKF) which solves for the optimal estimate of the relative orientation and pointwise structure of an object at each time step. This work provides for the optimal integration of information over time, under the assumption that a set of features constituting a coherent object is provided as input.

2 Recursive Structure From Motion

First we briefly review the recursive structure from motion model presented in [1]. This formulation embeds the constraints of the well-known relative orientation approach to structure from motion [8, 9] directly into an EKF measurement relationship. Here we are only considering the case of structure from motion based on explicitly tracked features. In our system features are selected and tracked as in reference [17].

For a single object, the formulation is as follows. Motion is defined as the 3-D translation and rotation of the camera (or object) with respect to the first frame in the sequence. Similarly, structure is represented as the 3-D locations of points seen in the first camera frame. However, since features are chosen from the image, there is actually only one unknown parameter for each feature point: the depth along the perspective ray established in the first camera frame.

Using a standard central projection imaging model (Figure 1), the 3-D position of a feature point $\mathbf{p}$ can thus be defined as the sum of two vectors, one pointing to the 2-D location of the feature in the image plane, $\mathbf{y}$, and one pointing along the perspective ray, weighted by a scale factor α.

$$\mathbf{p} = \begin{pmatrix} y_1 \\ y_2 \\ 0 \end{pmatrix} + \alpha_i \begin{pmatrix} y_1\beta \\ y_2\beta \\ 1 \end{pmatrix} \tag{1}$$

where β is the inverse principle distance, ($\beta = 1/f$). For each point $\mathbf{p}$, α is the only free parameter relating to its position that need be estimated.

The state vector for the EKF contains six motion parameters and N structure parameters.

$$\mathbf{x}(t) = \left(\mathbf{t}_{CO}^T(t), \rho^T(t), \alpha_1, \cdots, \alpha_N\right)^T \tag{2}$$

In the state vector, the rotational vector $\rho(t)$ contains three Euler angles which describe the (small) rotation between time $t - \Delta t$ and time t. After every time step, the small rotation is composed into the global rotation quaternion $\mathbf{q}_{CO}(t)$, which is maintained external to the EKF.

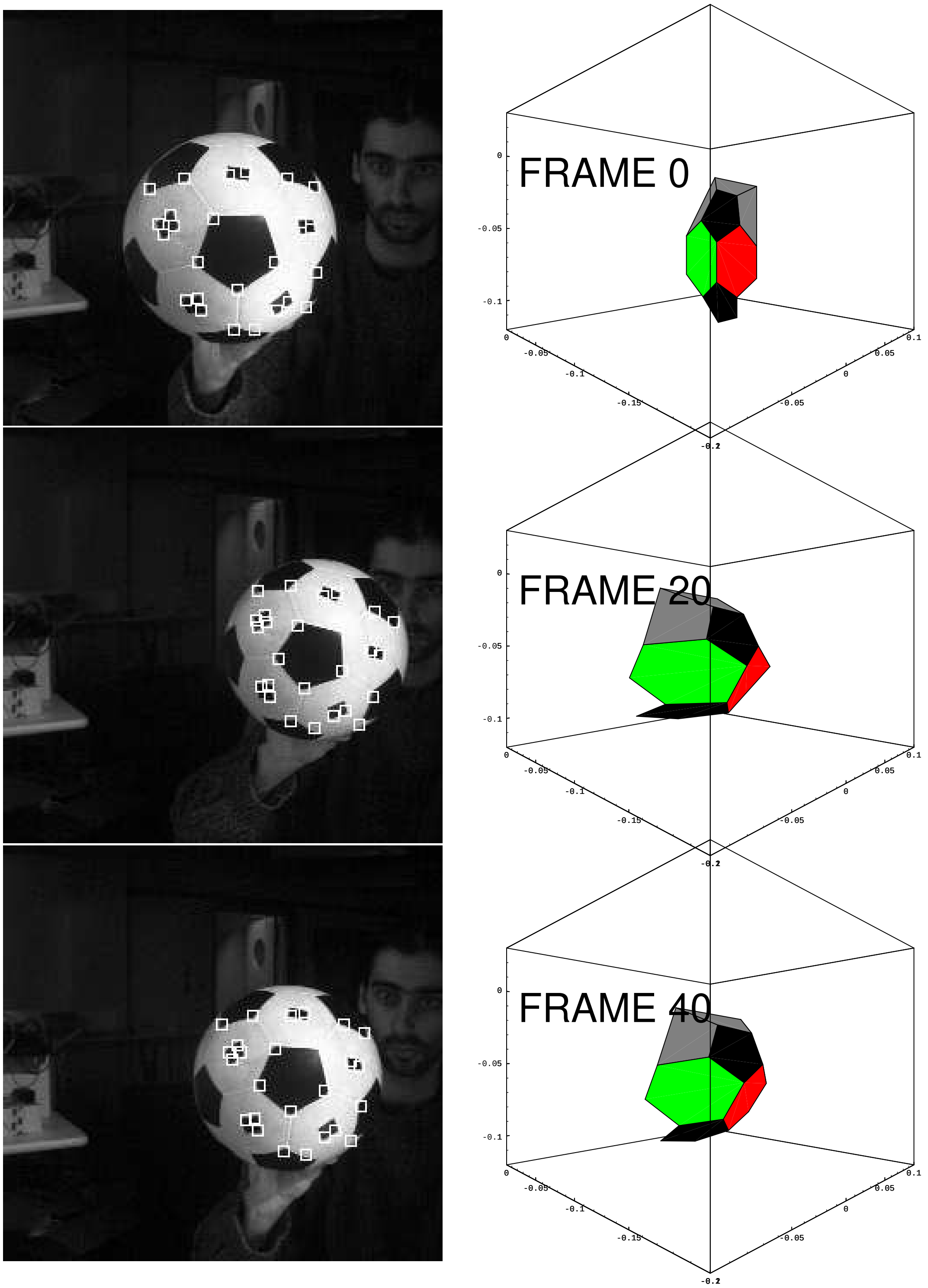

Figure 2: Image with (hand-selected) features and recovered 3-D structure

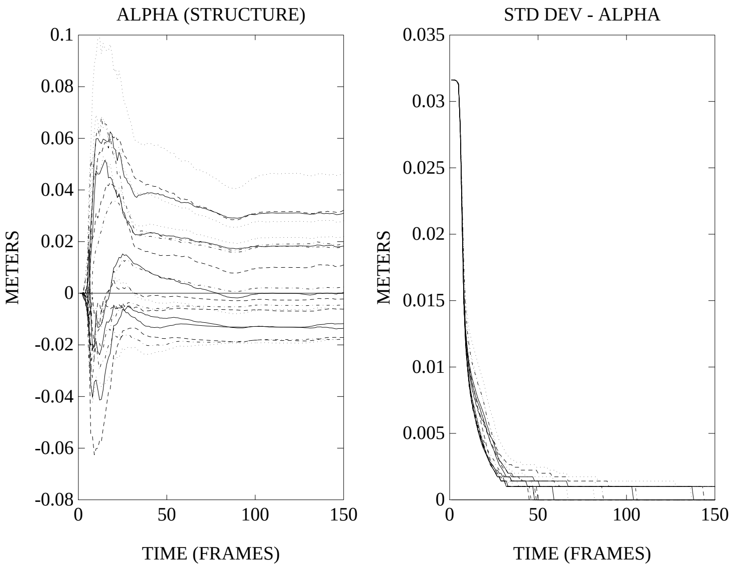

Figure 3: Convergence of shape parameters.

Appendix A presents the derivation of the EKF measurement and update equations, which compute the best linear least squares estimate, $\hat{\mathbf{x}}(t|t)$, and the error covariance, $\mathbf{P}(t|t)$. Figures 2,3 show the results from this method tracking a soccer ball, when appropriate features were selected by hand.

If the features from multiple objects are intermixed, then this method will not yield accurate estimates of structure or motion. We therefore need a method to isolate the features which belong to distinct objects. Our approach begins by testing multiple hypotheses about which points ought be grouped together.

3 Hypothesize and Test approach

When there are multiple rigid-body motions in a sequence, we need to consider multiple models in order to be able to accurately describe the image sequence. A separate structure from motion estimate should proceed for each object (or coherent motion) in the scene. We do not, in general, know which points belong to the same object, what the object motion is, nor even how many objects there are in the scene. We do have a model for motions which may be present, in the form of our structure from motion estimator.

Our approach is to use a Hypothesize, Test, and Select strategy. We obtain different hypotheses of possible rigid-body motions, test to see what, if any, support they have in the data, and then select the subset of hypotheses which comprise the most parsimonious description of the data. Figure 4 shows an overview of the model we propose. This work is an extension of the segmentation framework we developed based on minimal length description and robust estimation techniques [13, 4]. We have previously applied this formulation to the motion domain, developing a model for the perception of multiple motions in the case of structured optical flow fields with occlusion [5] or transparency [6].

For the task addressed in this paper – the computation of rigid-body structure from motion using tracked features – we generate hypotheses by taking random samples of sets of features, and using those to compute a motion estimate. These sets are constructed as follows: First, a feature is selected at random from the data and labeled F. The nearest feature to F is found, and with probability p is included in the set. The next nearest feature to F is then selected, and the process repeats until n features have been selected. If the farthest feature is reached the selection process restarts. Using this method we generate a sufficient number of such grouping hypotheses so that we have high confidence that there will be a homogeneous set of features for each object in the scene.

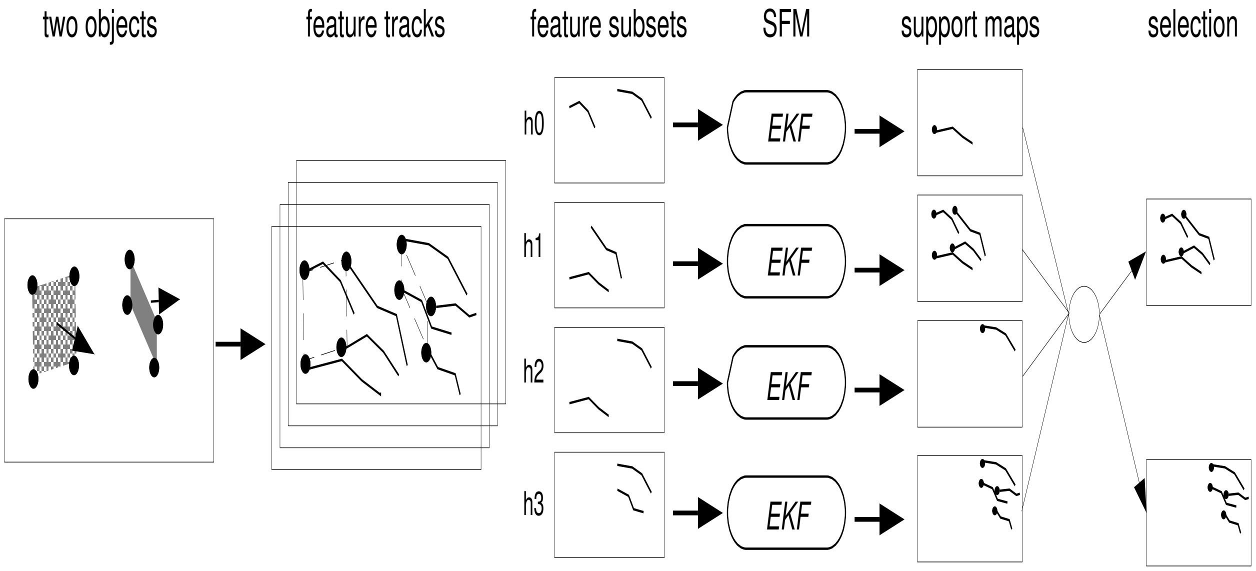

Figure 4: Block diagram of automatic rigid-body motion segmentation method

Figure 5 shows the construction of a synthetic sequence of feature tracks from two rotating spheres. For this example, we generated 200 hypotheses, taking random samples of 4 feature tracks ($n = 4$ and $p = 0.1$). For each hypothesis we compute an initial support map, s^*_{ij}, which is set to 1 if feature j is in the initial subset for hypothesis i, and 0 if not.

3.1 Re-computing SFM Estimates based on Limited Data

Given a support map, we need to make an estimate of structure and motion parameters based on only those features in the subset specified by the support map. In practice we can do this constructing a new filter with only the state variables corresponding to the set of supported features. However there is also a direct way to do this in the full EKF formulation, by adjusting the measurement noise covariance term **R**. (This approach also has the advantage that it can be more easily generalized to the case of non-boolean support.)

For each feature track, **R** specifies the confidence of the measurement, and weights the contribution of that feature to the estimation process through the Kalman gain matrix. When **R** is 0, there is implicitly no noise in the measurement, and the feature vector contributes fully to the estimation. If **R** is large, the contribution of the feature vector is minimized.

If we wish to specify that a feature track is an outlier and should not be included in the estimation, we can set its measurement noise model to be infinitely large, and it's contribution to the estimation will be infinitesimal. Thus, if we have a support map s which indicates which points ought to be supported in the estimate, we can simply set the measurement covariance to be the inverse support: $\mathbf{R} = 1$ if $s = 1$, $\mathbf{R} = \infty$ if $s = 0$, and run the full EKF structure-from-motion estimator. [1]

3.2 Updating support

After estimating motion and structure parameters for a hypothesized rigid-body motion h_i based on the initial support maps s^*, we compute an updated support map that indicates which feature tracks f_j, including those *not* in the original support, can be considered to be moving rigidly with the same motion parameters. We compute a new support map by thresholding a distance function which measures the

[1] ~~In this case it can be advantageous to use the~~ "information" form of the Kalman filter, since it is not adversely effected by numerical stability problems with infinite covariances.

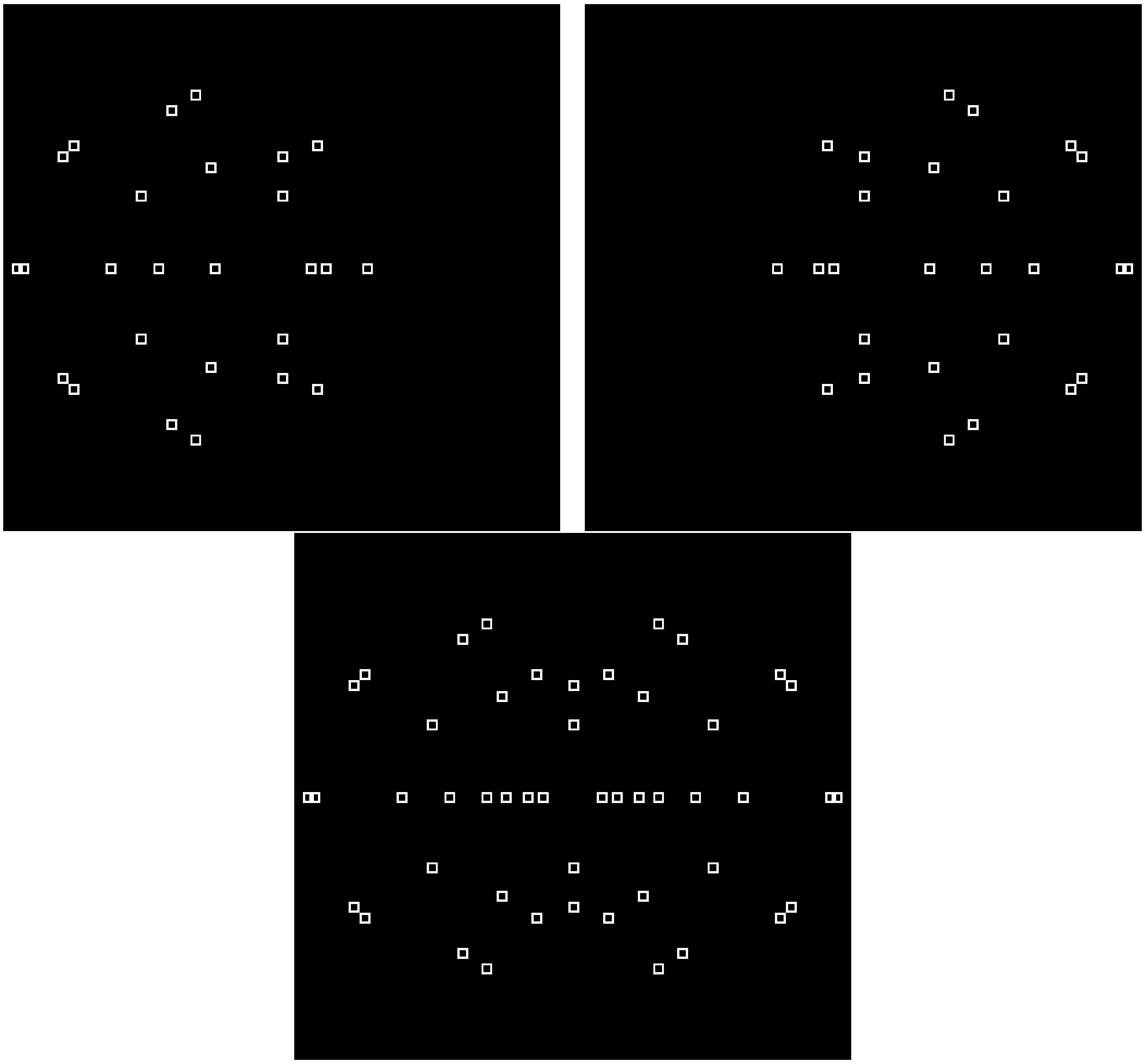

Figure 5: Construction of synthetic sequence with two rigid motions, and no static segmentation cues. Dots were placed on two spheres with different rotations; in (a) the sphere is rotating right, in (b) the sphere is rotating up. (c) shows the combined sequence (only the first frame is displayed)

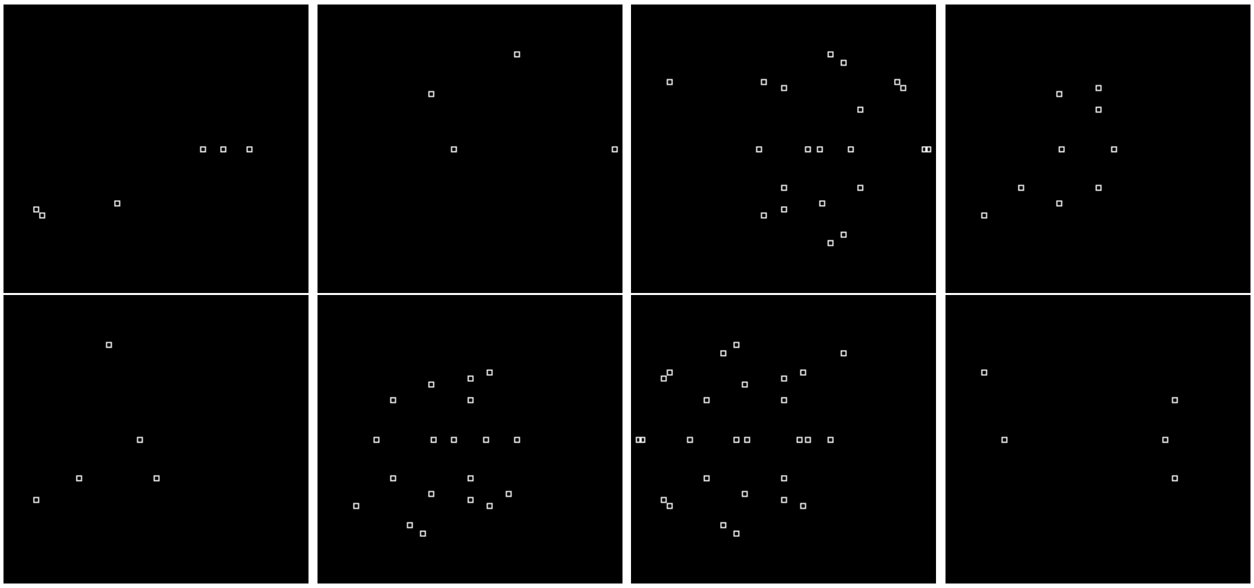

Figure 6: Eight (of 200) support maps for rigid-body motion hypotheses. Each hypotheses was based on computing a motion estimate using 4 randomly selected features from the sequnce. Most solutions are degenerate, but a few have "locked on" to a real object in the scene.

compatibility of a feature track and a known motion trace:

$$s_{ij} \;=\; \begin{pmatrix} 1 & \text{if } D(h_i, f_j) \leq \theta \\ 0 & \text{else} \end{pmatrix} \tag{3}$$

We define $D(h_i, f_j)$ to be the difference between the predicted feature locations for feature j given hypothesis i and the feature locations actually observed. This can be computed by estimating a structure parameter α given the motion parameters in h_i, and then generating the synthetic feature track predicted by those motion and structure parameters. We use the sum of squared differences of the velocities (instantaneous differences) between the predicted feature track and the observed feature track. We set $D(h_i, f_j) = r_j$, where

$$\mathbf{r} = (\nabla \mathbf{y}(t) - \nabla \mathbf{h}(\hat{\mathbf{x}}^{(i)}(T|T)))^2 \tag{4}$$

and $\hat{\mathbf{x}}^{(i)}(T|T)$ is the optimal estimate of $\mathbf{x}$ given the motion and structure parameters for hypothesis i, and ∇ is the gradient operator. We compute $\hat{\mathbf{x}}^{(i)}(T|T)$ by fixing the motion parameters (setting their initial covariances to 0) and running the SFM estimator. The function $\mathbf{h}()$ generates the predicted image locations for the features given the state vector $\mathbf{x}$. [2]

Figure 6 shows some of the support maps computed using this technique, for the rotating spheres example shown in Figure 5. Many of the initial hypotheses yield degenerate motion/structure estimates, and accordingly have few supported points, since their initial support will have come from heterogeneous or degenerate feature sets. However a few will exist that have initial support that covers only a single object (with known probability). These will have motion estimates that coarsely match the actual object motion, and will consequently have support maps that cover a majority of the object.

[2] Another method for defining $D(h_i, f_j)$ which has more generic application, and has yielded sucessful results in our method, is to add the feature f_j to the support of h_i, re-estimate the motion parameters $\mathbf{t}$ and ρ for h_i, and define D to be the sum of the squared difference between the new and old parameter estimates. This approach is simple to implement using the structure from motion method as a "black box", and has the advantage that it can be used with any feature-based structure from motion method, even if a forward model ($\mathbf{h}()$)is not available.

3.3 Selection

In this section we present the mechanism that selects the subset of hypotheses which best accounts for the overall visual signal. We follow the presentation adopted in [6] simplified to the case of pure occlusion (e.g., no transparent combination of two hypotheses at a single point).

Given a set of hypothesized rigid-body motions, $\mathcal{H}$:

$$\mathcal{H} = \{h_0, h_1, ..., h_M\} \tag{5}$$

where each h_i contains motion parameters $\mathbf{t}, \rho$ for each frame of the image sequence, we test to see which feature tracks f_j are compatible with which hypotheses. We use the support computation method described above, defining a support function s_{ij} which marks whether feature j is compatible with hypothesis i.

As in our previous formulations, the selection method proceeds by attempting to find the subset $\mathcal{L} \subset \mathcal{H}$ that best "covers" all the feature tracks using the smallest number of hypotheses possible. More formally, we optimize a criterion function that encourages features to be supported by exactly one hypothesis in the selected set.

We attempt to find the subset $\mathcal{L}$ which maximizes

$$E(\mathcal{L}) \;=\; \sum_{\{h|h\in\mathcal{L}\}} \sum_{j} s_{hj} \left(1 - \sum_{\{k|k\neq h, k\in\mathcal{L}\}} s_{kj}\right) - O||\mathcal{L}|| \tag{6}$$

where O is a overhead penalty that is incurred when an an additional hypothesis is added to the representation. This equation encourages hypotheses which support many features, subject to a strong inhibitory term that prevents more than one hypothesis from covering the same data, and a term that discourages hypothesis with too little support (e.g. less than O).

To maximize this numerically, we need to express $E(\mathcal{L})$ in a differentiable form. First, we represent $\mathcal{L}$ by enumerating those elements of $\mathcal{H}$ that are in $\mathcal{L}$ using a vector $\mathbf{a}$. The sign of each a_i indicates whether a hypothesis $h_i \in \mathcal{H}$ is an element of the subset $\mathcal{L} \subset \mathcal{H}$: a positive value of a_i indicated $h_i \in \mathcal{L}$, and a negative value means $h_i \notin \mathcal{L}$.

We thus rewrite Eq. 6 as

$$E(\mathbf{a}) \;=\; \sum_{\mathbf{i}} \sigma(\mathbf{a_i}) \sum_{\mathbf{j}} \mathbf{s_{ij}} \left(\mathbf{1} - \sum_{\mathbf{k\neq i}} \sigma(\mathbf{a_k})\mathbf{s_{kj}}\right) - \mathbf{O} \sum_{\mathbf{i}} \sigma(\mathbf{a_i}) \tag{7}$$

where $\sigma(a_i)$ is function which transforms the a_i values into a multiplicative weight between 0 and 1. Ideally, $\sigma()$ would be the unit step function centered at the origin. However, this hard non-linearity makes it very difficult to maximize $E(\mathbf{a})$ numerically. Instead, we use the "softer" sigmoid function,

$$\sigma(x) = \frac{1}{1 + e^{-x}} \tag{8}$$

For the classes of problems described in this paper, we have found that straight-forward gradient ascent techniques can find good solutions to the segmentation problem. We use

$$\frac{da_i}{dt} = \frac{1}{K} \frac{dE(\mathbf{a})}{d\sigma(a_i)} =$$

$$= \frac{1}{K} \sum_{j} s_{ij} \left(1 - \sum_{k\neq i} s_{kj}\sigma(a_k)\right) - O \tag{9}$$

where K is an integration constant. This can be expressed in matrix terms as a Hopfield-Tank network

$$\nabla \mathbf{a} \;=\; \mathbf{U} - \mathbf{T}\mathbf{A} \tag{10}$$

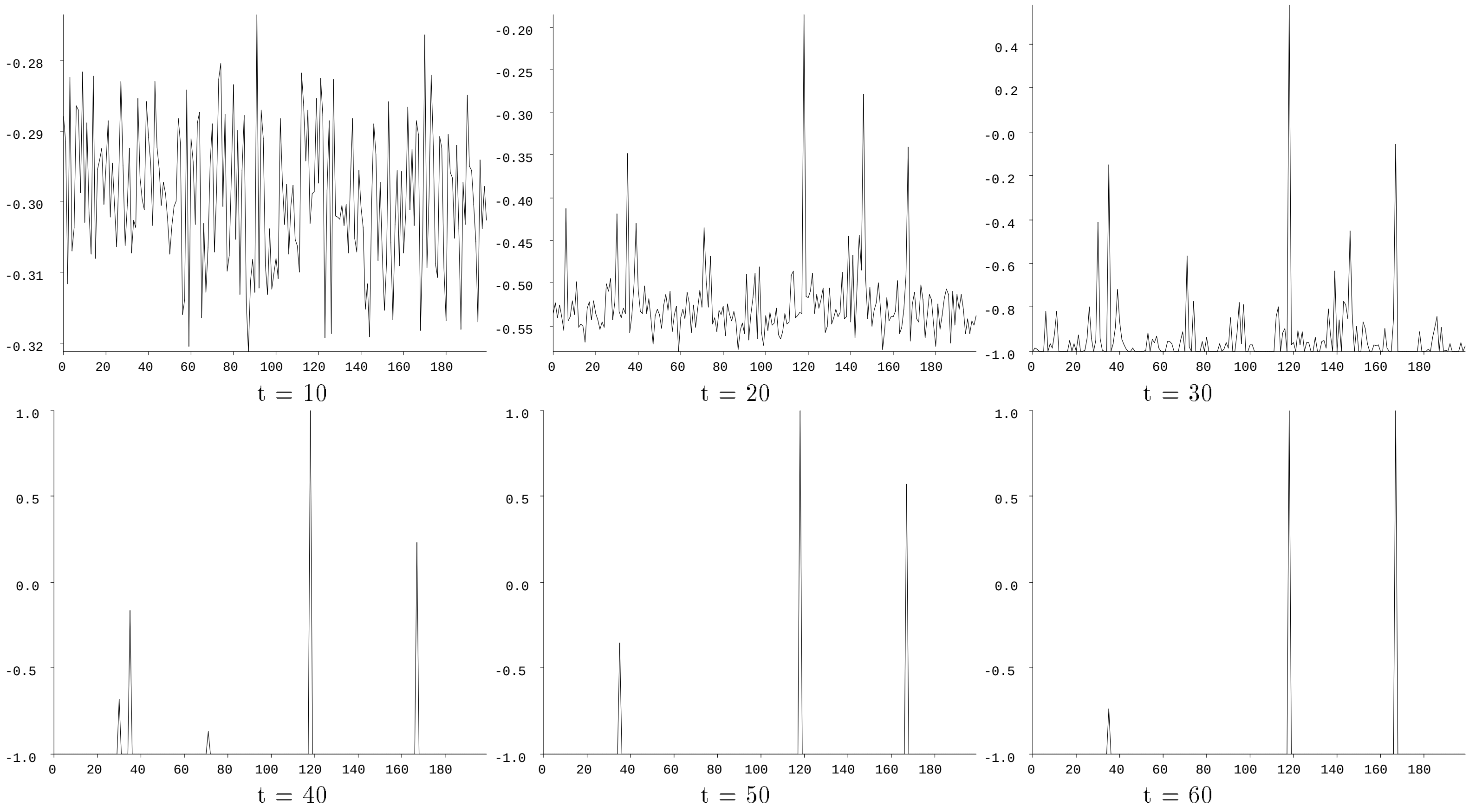

Figure 7: Convergence of selection net for synthetic rotation example

where $A_i = \sigma(a_i)$ and

$$U_i \;=\; \sum_j s_{ij}$$

and

$$T_{ij} \;=\; \sum_j s_{ij} s_{kj} \qquad i \neq j$$

except on the diagonal, where $T_{ii} = 0$. Since $\mathbf{T}$ is symmetric, the network will converge to a stable solution which is a local maxima of $E(\mathbf{a})$ [10].

Figure 7 shows the convergence of the selection network for the example presented in Figure 5. There are 200 hypotheses, and thus a selection vector **a** with 200 elements. Initially they are set to zero, and after 60 time steps they have converged to select two hypotheses which best account for all the data. The two hypotheses correspond to the two original motions used to synthetically create the sequence; their support maps are shown in Figure 8.

3.4 Example with real imagery

Figure 9 shows a real image sequence with two motions (egomotion and the motion of the soccer ball). Features were tracked on this sequence using the (human assisted) method desribed in [1]. Hypotheses were constructed by taking 500 random selections of 4 points, and setting the intial support map for each hypothesis to be the randomly selected points. Motion estimates were computed for each of these hypotheses, based only on the supported points. Support maps were then recomputed as described above, to see which other points in the scene agreed with the recovered motion. Figure 10 shows 8 of

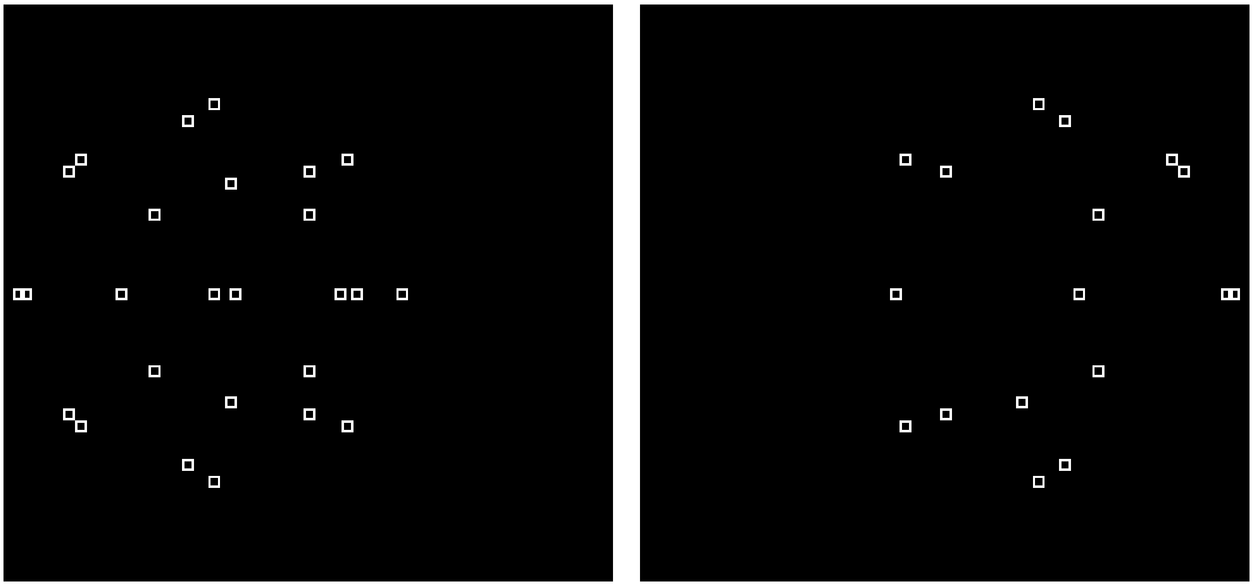

Figure 8: Selected Hypotheses for synthetic rotation example

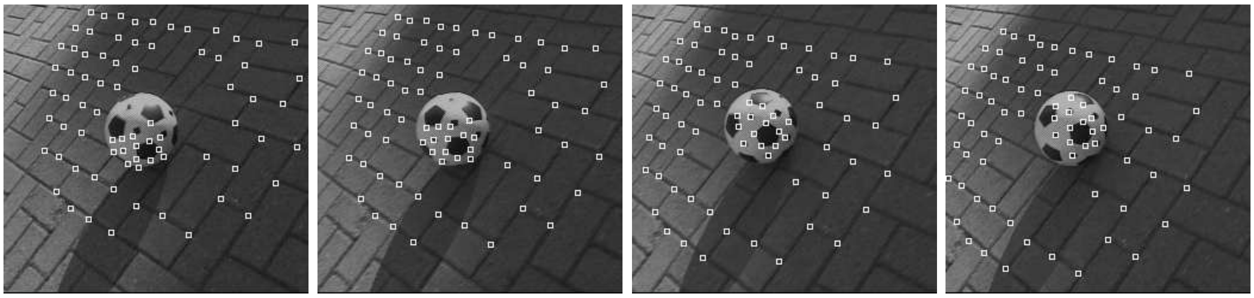

Figure 9: Intensity images with tracked features from a image sequence with two motions: a rotating soccer ball and a moving ground plane due to camera motion.

these support maps. The selection method was run with $O = 4$, indicating that at least four points were required to maintain a hypothesis. The convergence of the selection net is shown in Figure 12; of the 500 hypotheses, two corresponding to the actual motions in the scene are selected.

4 Conclusion

When features are tracked on a multitude of different objects, methods for structure from motion must be applied that can simultaneously estimate how many objects there are, which features belong to which objects, and of course, what the motion and shape parameters are for each object in the scene. We have presented a method which can perform these estimation steps automatically.

Our method is based on a Hypothesize, Test, and Select paradigm, which derives hypotheses about rigid motions, tests their support in the data, and selects the best description of the data using the available hypotheses. This "multiple-models" approach employs multiple structure-from-motion estimators, one for each hypothesized rigid motion in the scene. When the selection stage has completed, we are left with one Kalman filter SFM model for each object in the scene, which can then compute the optimal estimate based on only the subset of the data corresponding to the object it represents.

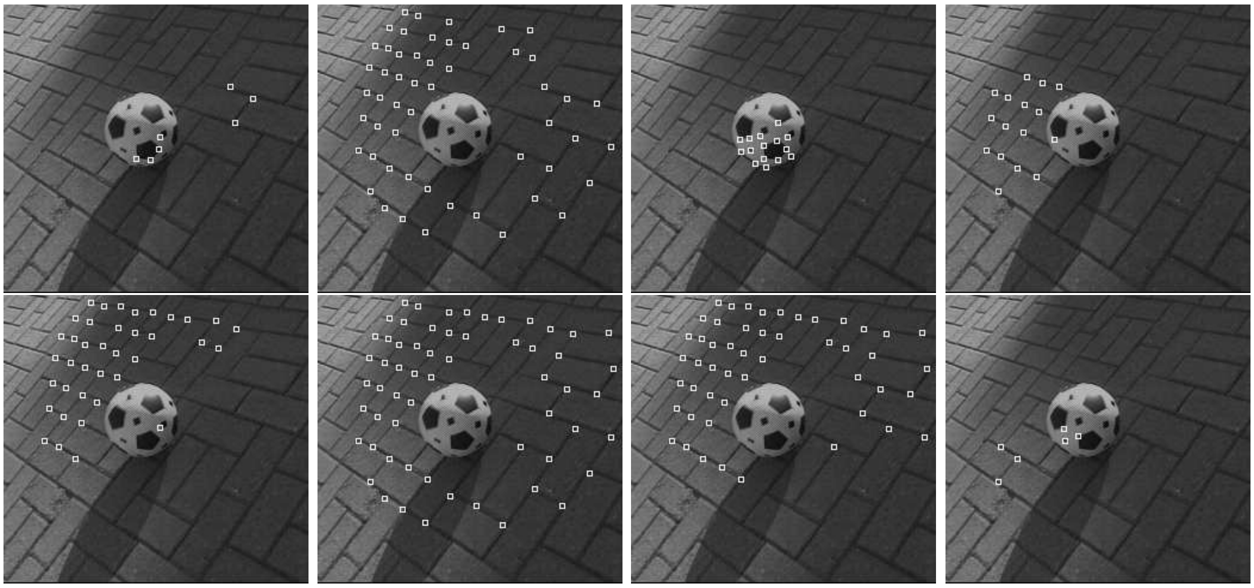

Figure 10: Eight support maps for different rigid-body motion hypotheses for the rolling ball sequence. 500 such support maps were generated in total, by randomly selecting features from the sequence, making a motion estimate with them, and then thresholding the residual error of the predicted feature locations for all other features. Since the random selection process will usually generate hypotheses which intermingle features from more than one object most solutions are degenerate, but a few are based on homogeneous initial features and have "locked on" to a real object in the scene.

Appendix A. Extended Kalman Filter

The following presentation of the EKF is adapted from [1]. Given the state vector described by Eq. 2, the linear dynamics model is

$$\mathbf{x}(t+\Delta t) = \mathbf{\Phi}(\Delta t)\mathbf{x}(t) + \xi(t) \tag{11}$$

where

$$\mathbf{\Phi}(\Delta t) = \begin{pmatrix} I & 0 & I\Delta t & 0 & 0 \\ 0 & I & 0 & I\Delta t & 0 \\ 0 & 0 & I & 0 & 0 \\ 0 & 0 & 0 & I & 0 \\ 0 & 0 & 0 & 0 & I \end{pmatrix} \tag{12}$$

and $\xi(t)$ is an error term, modeled as multi-variate Gaussian distributed white noise.

The observation at each time step is a set of N 2D measurements which is a nonlinear function of the state vector:

$$\mathbf{y}(t) = \begin{pmatrix} \mathbf{y}_1(t) \\ \vdots \\ \mathbf{y}_N(t) \end{pmatrix} = \mathbf{h}(\mathbf{x}(t)) + \eta(t) \tag{13}$$

The imaging geometry outlined in Section 2 (see [1] for details) define the nonlinear function $\mathbf{h}(\mathbf{x}(t))$, with an uncertainty term $\eta(t)$, modeled as Gaussian distributed white noise.

The notation $\hat{\mathbf{x}}(t_1|t_2)$ represents the linear least squares estimate (LLSE) of $\mathbf{x}(t_1)$ based upon $\mathbf{y}(\tau), \tau \leq t_2$. The notation

$$\mathbf{P}(t_1|t_2) \triangleq E(\mathbf{x}(t_1) - \hat{\mathbf{x}}(t_1|t_2))(\mathbf{x}(t_1) - \hat{\mathbf{x}}(t_1|t_2))^T \tag{14}$$

represents the error covariance of the LLSE.

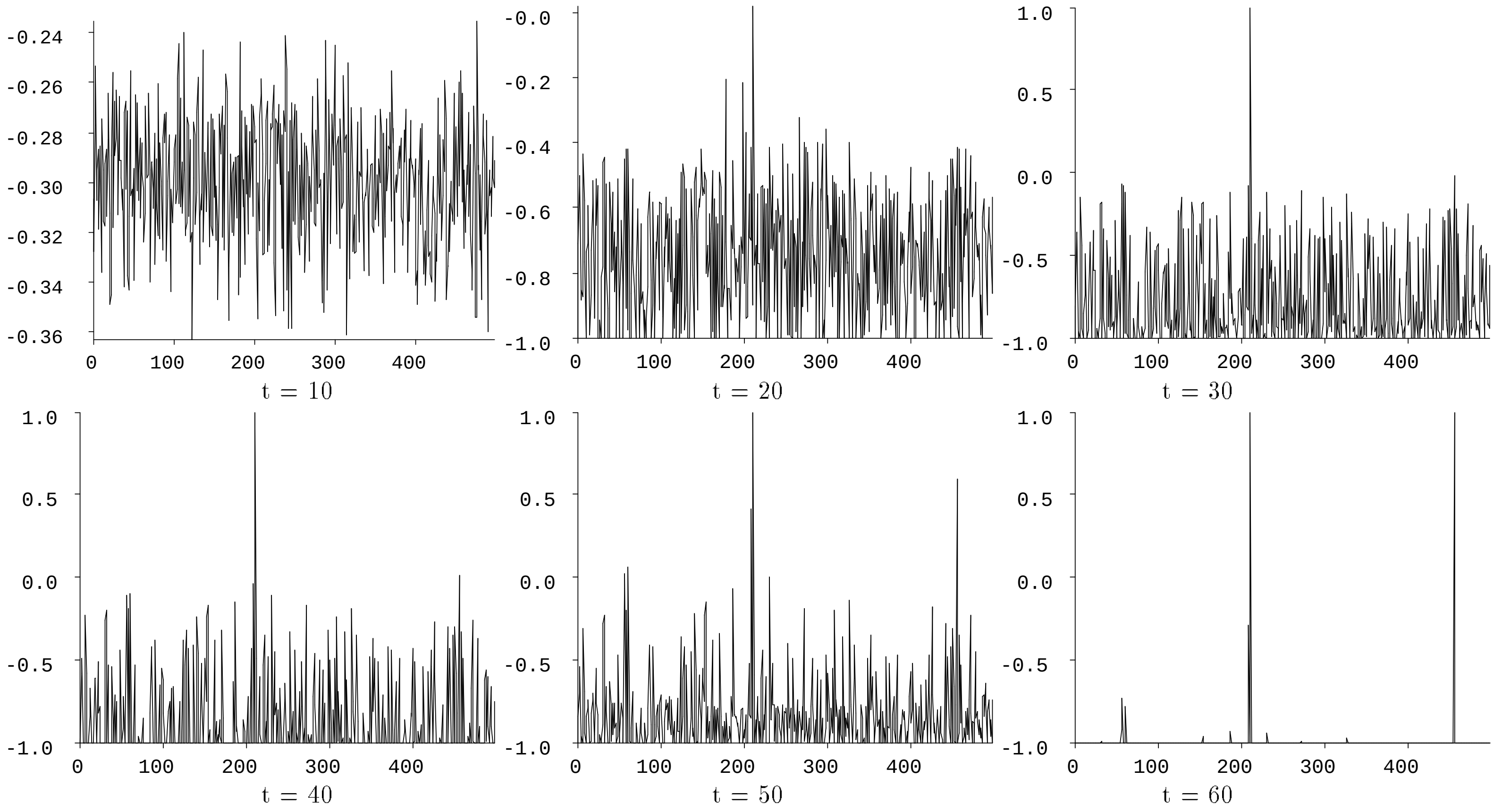

Figure 11: Convergence of selection net for soccer ball example

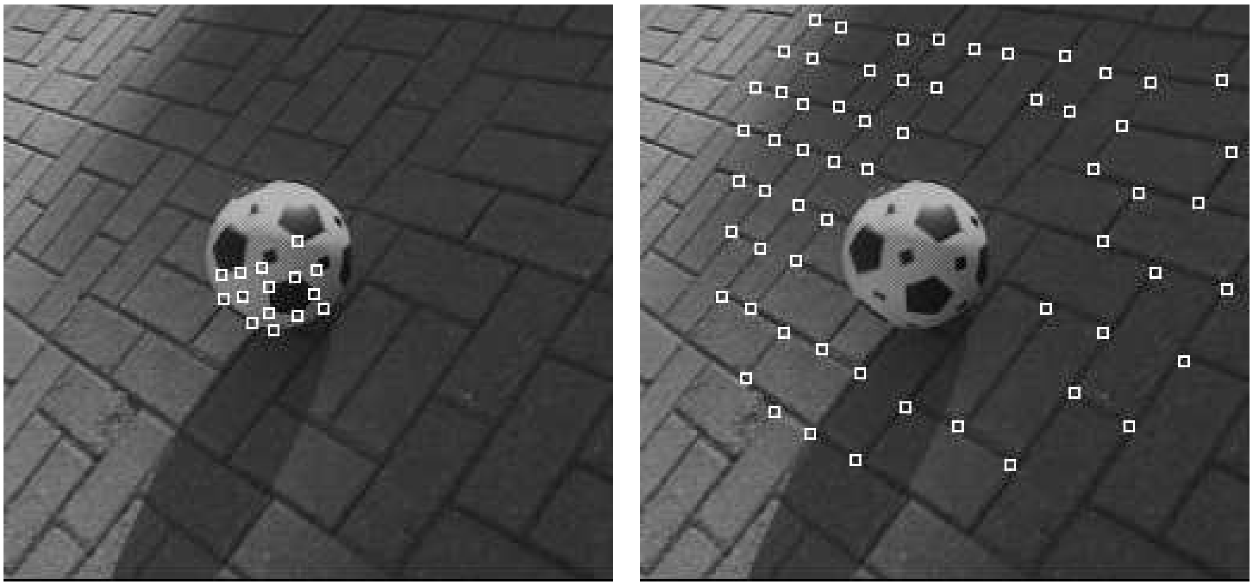

Figure 12: Selected Hypotheses for soccer ball example

The EKF update equations are

$$\hat{\mathbf{x}}(t|t) = \mathbf{\Phi}(\Delta t)\hat{\mathbf{x}}(t|t-1) + \mathbf{K}(t)\left(\mathbf{y}(t) - \mathbf{h}(\hat{\mathbf{x}}(t|t-1))\right) \tag{15}$$

$$\mathbf{P}(t|t) = \left(\mathbf{I} - \mathbf{K}(t)\mathbf{C}(t)\right)\mathbf{P}(t|t-1) \tag{16}$$

where the *Kalman gain* is defined as

$$\mathbf{K}(t) \triangleq \mathbf{P}(t|t-1)\mathbf{C}^T(t)(\mathbf{C}(t)\mathbf{P}(t|t-1)\mathbf{C}^T(t) + \mathbf{R}(t))^{-1} \tag{17}$$

and

$$\mathbf{C}(t) \triangleq \left[\frac{\partial \mathbf{h}}{\partial \mathbf{x}}\right]_{\mathbf{x}=\hat{\mathbf{x}}(t|t-1)} \tag{18}$$

$$\mathbf{R}(t) = E\eta(t)\eta^T(t), \quad E\eta(t) = \mathbf{0}. \tag{19}$$

The prediction step is

$$\hat{\mathbf{x}}(t+1|t) = \mathbf{\Phi}(\Delta t)\hat{\mathbf{x}}(t|t) \tag{20}$$

$$\mathbf{P}(t+1|t) = \mathbf{\Phi}(\Delta t)\mathbf{P}(t|t)\mathbf{\Phi}(\Delta t)^T + \mathbf{Q}(t) \tag{21}$$

where

$$\mathbf{Q}(t) = E\xi(t)\xi^T(t), \quad E\xi(t) = \mathbf{0}. \tag{22}$$

References

[1] A. J. Azarbayejani, B. Horowitz, A. P. Pentland, Recursive Estimation of Structure and Motion using the Relative Orientation Constraint In *Proceedings IEEE Conference on Computer Vision and Pattern Recognition*, New York, June 1993.

[2] T. E. Boult and L. G. Brown, Factorization-Based Segmentation of Motions *Proc. IEEE Workshop on Visual Motion*, Princeton, October 1991.

[3] Ted J. Broida, S. Chandrashekhar, and Rama Chellappa. Recursive estimation of 3-d motion from a monocular image sequence. *IEEE Trans. Aerosp. Electron. Syst.*, 26(4):639–656, July 1990.

[4] T. Darrell and A. P. Pentland. Segmentation by minimal description. In *Proceedings Thrid Intl. Conference on Computer Vision*, Osaka, Japan, December 1990.

[5] T. Darrell and A. P. Pentland. Robust estimation of a multi-layer motion representation. In *Proceedings IEEE Workshop on Visual Motion*, Princeton, October 1991.

[6] T. Darrell and E.P. Simoncelli Separation of Transparent Motion into Layers using Velocity-tuned Mechanisms. Vision and Modeling Group TR-244, 1993.

[7] Ernst Dieter Dickmanns and Volker Graefe. Dynamic monocular machine vision. *Machine Vision and Applications*, 1:223–240, 1988.

[8] B. K. P. Horn. *Robot Vision.* MIT Press, 1986.

[9] B. K. P. Horn. Relative orientation. *International Journal of Computer Vision*, 4(1):59–78, January 1990.

[10] J. J. Hopfield, and D. W. Tank, Neural computation of decisions in Optimization Problems. *Biological Cybernetics,* Vol. 52, pp. 141-152. 1985.

[11] Ratnam V. Raja Kumar, Arun Tirumalai, and Ramesh C. Jain. A non-linear optimization algorithm for the estimation of structure and motion parameters. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 136–143, June 1989. (San Diego, CA.).

[12] J. Oliensis and J. Inigo Thomas. Incorporating motion error in multi-frame structure from motion. In *IEEE Workshop on Visual Motion*, pages 8–13, Los Alamitos, CA, October 1991. IEEE Computer Society, IEEE Computer Society Press. (Nassau Inn, Princeton, NJ.).

[13] A. P. Pentland. Part Segmentation for Object Recognition. *Neural Computation*, 1:82-91, 1989.

[14] S. Soatto, P. Perona, R. Fraezza, and G. Picci. Recursive motion and structure estimation with complete error characterization. In *1993 IEEE Conference on Computer Vision and Pattern Recognition*, pages 428–433, Los Alamitos, CA, June 1993. IEEE Computer Society, IEEE Computer Society Press. (New York).

[15] Minas Spetsakis and Yiannis Aloimonos. A multi-frame approach to visual motion perception. *International Journal of Computer Vision*, 6(3):245–255, August 1991.

[16] Richard Szeliski and Sing Bing Kang. Recovering 3d shape and motion from image streams using non-linear least squares. In *1993 IEEE Conference on Computer Vision and Pattern Recognition*, pages 752–753, Los Alamitos, CA, June 1993. IEEE Computer Society, IEEE Computer Society Press. (New York).

[17] C. Tomasi and T. Kanade. Shape and motion from image streams under orthography: a factorization method. *International Journal of Computer Vision*, 9(2):137–154, November 1992.

[18] Juyang Weng, Narendra Ahuja, and Thomas S. Huang. Optimal motion and structure estimation. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 15(9):864–884, September 1993.

[19] G-S. Young and Rama Chellappa. 3-d motion estimation using a sequence of noisy stereo images: Models, estimation and uniqueness. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 12(8):735–759, January 1990.