Proceedings of the 2006 IEEE International Conference on Robotics and Automation
Orlando, Florida - May 2006

Adaptive Geometric Templates for Feature Matching

Harini Veeraraghavan Paul Schrater Nikolaos Papanikolopoulos
Department of Computer Science and Engineering
University of Minnesota
{harini,schrater,npapas}@cs.umn.edu

ABSTRACT

Robust motion recovery in tracking multiple targets using image features is affected by difficulties in obtaining good correspondences over long sequences. Difficulties are introduced by occlusions, scale changes, as well as disappearance of features with the rotation of targets. In this work, we describe an adaptive geometric template-based method for robust motion recovery from features. A geometric template consists of nodes containing salient features (e.g., corner features). The spatial configuration of the features is modeled using a spanning tree. This paper makes the following two contributions: (i) an adaptive geometric template to model the varying number of features on a target, and (ii) an iterative data association method for the features based on the uncertainties in the estimated template structure in conjunction with its individual features. We present experimental results for tracking multiple targets over long outdoor image sequences with multiple persistent occlusions. A comparison of the results of the data association method with a standard Mahalanobis distance gating applied to individual features is also presented.

KEYWORDS

Adaptive geometric templates, data association, variable dimension Kalman filtering, feature tracking.

I. INTRODUCTION

Image features provide a robust and simple way to recover target motion. However, point feature tracking methods generally suffer from poor feature correspondences owing to: (i) the similarity of the intensity information shared by features, (ii) the presence of clutter especially in uncontrolled, outdoor scenes, and (iii) the noise in the feature templates introduced from the image in general. This paper tries to address the following problem: Given a dynamic set of features on a target, can we obtain a robust data association for each feature using the information from other features on the target? Specifically, we try to address this problem by using an adaptive geometric template that models the spatial relation of features with respect to each other as a constraint for the measurements.

This paper makes two contributions: (i) an adaptive geometric template for representing rigid targets, and (ii) a robust solution for the data association problem for features using the constraints from the geometric template. Common methods for outlier detection for point features can be classified into one of the following two types based on (i) the motion of each feature independently of other features as in [1], [9], [14], and (ii) the relative configuration of features represented using a model, such as an affine motion model or the fundamental matrix. Examples of the latter include the research in [2], [6], [11], [12]. Reliable computation of the fundamental matrix requires a large number of features thereby, requiring computationally expensive matrix factorization in addition to making restrictive assumptions on tracking. Our method differs from all these methods in that we use the local configuration of features derived from a geometric template as shown in Fig. I. The dynamic template is flexible to addition or deletion of features, requiring no bootstrapping or large factorization. While superficially resembling adaptive mesh-based geometric models [10], no physical models are used. Instead, the link constraints are enforced based on the uncertainty in the estimation of an appearance model using a variable dimension Kalman filter. Data association for the features is obtained by minimizing the deformation of the estimated geometric appearance. The advantage of using an adaptive template for appearance is twofold. First, no prior knowledge of the target appearance is necessary, thereby, allowing it to be applied to the tracking of any generic rigid object. Second, this method provides a natural way to eliminate bad features by removing features moving differently from the remaining features on the model.

Fig. I. A geometric template imposed on a target. The template consists of a spanning tree connecting the individual features on the target.

This paper is arranged as follows: Relevant prior work is discussed in Section II. Section III presents the data association method, Section IV discusses the adaptive geometric template model. Details of the feature tracking method are discussed in Section V. Finally, experimental results are presented in Section VI and Section VII concludes the paper.

II. Previous Work

Most of the work on outlier detection for point image features consists of using a model describing the transformation of features in the image. Examples include, outlier detection using orthogonal regression by Torr and Murray [12], random sample consensus methods introduced by Fischler and Bolles [2], the X-84 criterion used by Tommasini *et al.* [11], as well as, subspace matching methods [6]. Harris [4] introduced rigid body constraints for outlier detection. Recently, Guo *et al.* [3] proposed an interesting linear combination representation-based approach for detecting feature outliers for affine motion images. A Support Vector Machine (SVM) regression method was used by Zhu *et al.* [16] for detecting outliers based on iterative pruning.

Adaptive meshes and geometric representation models such as [5], [7], [10], [15] are some examples of geometry-based methods for object recognition and tracking. However, the similarity of the proposed method with these methods ends with the use of local similarity of features to form the geometric appearance template. The nodes are variable, and the deformation of the links is guided by the uncertainty in the estimated appearance, in contrast to physics-based models which typically make use of a spring force system to guide the link deformations. The structure is inferred automatically without any need for bootstrapping with a template.

III. Geometric Template-based Data Association

Under the assumption that the tracked targets are rigid, the deformations in the geometric template will be restricted to slight changes in the spatial relations between the features on the target. Hence, the problem of data association consists of finding the set of measurements in a given frame that minimizes the extent of deformation in the template. This is the same as maximizing the joint likelihood of all the measurements given the appearance. This can be expressed as,

$$\Lambda(a) = p(Z|a) = p(x_1, x_2, \ldots, x_n|a) \tag{1}$$

where $\Lambda(a)$ is the likelihood of the appearance, and $Z = \{x_1, x_2, \ldots, x_n\}$ is the set of measurements for N features on the template, and a is the appearance. Since each feature measurement is obtained independent of one another, Equation (1) can be written as,

$$\begin{aligned} \Lambda(a) &= \sum_{i=1}^{N} \sum_{j=1, j\neq i}^{N} p(x_i|a)p(x_j|a) && (2)\\ a^{ML} &= argmax\Lambda(a). && (3) \end{aligned}$$

Thus, the maximum likelihood solution is obtained by the maximum of the appearance likelihood for all the feature measurements. The likelihood of a feature measurement is computed based on the appearance and the estimate of the feature motion as,

$$\begin{aligned} p(x_i, \hat{x}_i|a) &= p(x_i|\hat{x}_i)p(\hat{x}_i|a) && (4)\\ &= N(x_i; \hat{x}_i, \Sigma_i)N(\hat{x}_i; a, \Sigma_a) \end{aligned}$$

where x_i is the measurement for a given feature i, $\hat{x}_i$ and Σ_i are the estimated position and covariance associated with the feature i, respectively. Σ_a is the estimated covariance in the appearance a.

$$\begin{aligned} N(\hat{x}_i; a, \Sigma_a) &= \frac{1}{c_1} e^{-\frac{\sum_{j=1}^{K}(l_{ij}-\hat{l}_{ij})\Sigma_{ij}^{-1}(l_{ij}-\hat{l}_{ij})'}{2}}, && (5)\\ l_{ij} &= \hat{x}_i - \hat{x}_j \\ N(x_i; \hat{x}_i, \Sigma_i) &= \frac{1}{c_2} e^{-\frac{(x_i-\hat{x}_i)\Sigma_i^{-1}(x_i-\hat{x}_i)'}{2}} && (6) \end{aligned}$$

where c_1 and c_2 are the normalization constants. K corresponds to the number of links connected to a given feature in the template, $\hat{l}_{ij}$ is the predicted link length, with covariance Σ_{ij}, and l_{ij} is the link length computed from the position estimates of features i and j. Σ_i corresponds to the estimated covariance for the estimated position of feature i. As might be obvious to most readers, the terms in the exponent of both the equations, Equations (5) and (6), correspond to computing the Mahalanobis distance. Hence, the problem can be reduced to computing the set of measurements that minimizes the total deviation computed in the form of Mahalanobis distances. This can be summarized as,

$$\begin{aligned} N_D &= \min\left(\sum_{i=1}^{N} r_f^i + r_l^i\right) && (7)\\ r_f^i &= [x_i - \hat{x}_{i-1}]\Sigma_{i-1}^{-1}[x_i - \hat{x}_{i-1}]' \\ r_l^i &= \sum_{j=1}^{k}[l_{ij} - \hat{l}_{ij}]\Sigma_{ij}^{-1}[l_{ij} - \hat{l}_{ij}]' \end{aligned}$$

where r_f^i and r_l^i are the residuals on node n^i based on the estimates of the node (or feature) motion $\hat{x}_{i-1}$, covariance Σ_{i-1} from frame $i-1$, and the predicted estimates of links $\hat{l}_{ij}$, and covariance Σ_{ij}. For any node, n_i, the total node residual when excluding the true measurement is obtained by replacing the estimated feature position in frame i with an estimate of the predicted position in the same frame. Hence, a true measurement will be discarded as an outlier when the residual resulting from its inclusion is more than the residual resulting from excluding the measurement. The algorithm for data association is described in Table I. As shown, each node is tested with two different hypotheses: (a) estimated position at time t resulting from inclusion of measurement for a feature i at time t, and (b) predicted position resulting from the Kalman filter's propagation step at time t.

IV. Geometric Template Adaptation

The geometric template consists of a tree structure whose nodes are comprised of the point features on the target and the links connecting the features. A minimum spanning tree is used to connect the features on a target. Hence, for a target consisting of N features, the template consists of N nodes with $N-1$ links.

The geometric template deforms with target motion owing to

```
Set Hypothesis of all features to estimated
compute total node residual NDmin

STEP I:
for each feature i <- 1 to N
  Hypothesis(i) <- alternate(Hypothesis(i))
  compute total node residual Ndnew
  if(Ndnew < Ndmin)
     Ndmin <- Ndnew
     alteredFeature <- i
  else
    Hypothesis(i) <- alternate(Hypothesis(i))
end
push(Stack, alteredFeature)
push(alteredList, alteredFeature)

STEP II:
while(Stack is not empty)
  alteredFeature <- pop(Stack)
  for each feature (j <- 1 to m
     connected to alteredFeature
     & feature j is not in alteredList)
     Hypothesis(j) <- alternate(Hypothesis(j))
     compute node residual Ndnew
     if(Ndnew < Ndmin)
       Ndmin <- Ndnew
       push(Stack, j)
       push(alteredList,j)
     else
       Hypothesis(j) <- alternate(Hypothesis(j))
  end
  alteredFeature <- pop(Stack)
end
```

TABLE I

ALGORITHM FOR COMPUTING DATA ASSOCIATION FOR FEATURES. ALTERNATE HYPOTHESIS FOR A NODE IS PREDICTED IF THE ORIGINAL HYPOTHESIS FOR THE NODE i IS ESTIMATED AND VICE-VERSA.

feature translation. The structure is altered by (i) new features added to the target, (ii) removal of untracked features, or when (iii) the template contains features that do not truly belong to the target. The last case results from the initialization of features that were truly outside the target, but were detected as part of the target owing to errors in the blob segmentation. The template is modified in each of the above cases either by: (i) addition or removal of links, or (ii) by replacing parts or the whole template with a new template.

In the case of (i), the template update consists of the addition of new links. A new feature is added to a node closest to the feature as shown in Fig. II. In the case of (ii), the template is updated by removing the links connected to the removed feature. In order to maintain the tree structure, the appearance is adapted by reconnecting the appropriate features. Fig. III illustrates examples for the removal of an internal node indicated by the shaded circle, and an external node (clear node). In this case, nodes are reconnected to preserve a tree structure without constructing a new minimum spanning tree so that the learned appearance model is not discarded. In the case of (iii), the features that do not truly belong to the target are identified as those that consistently lie outside the target's blob boundary, and are removed from the appearance model. As mentioned in the previous paragraph, in any of case (i), (ii), and (iii) only part of the template is modified leaving the rest intact (other than the case when the tree is completely reinitialized with a new template). Hence, the estimates for the parts of the template containing the old links are preserved while new links are initialized in the Kalman filter, using the covariance $\Sigma_{ij} = \sqrt{\Sigma_i \Sigma_j}$, where Σ_i and Σ_j are the covariances in the position of the features i and j. The subscripts for x and y coordinates are omitted for clarity.

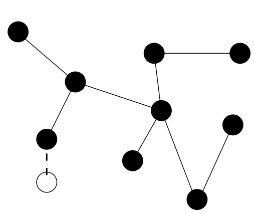


Fig. II. Feature addition. The new feature, indicated by light circle, is added to the closest node in the template.

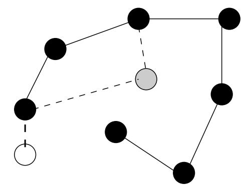


Fig. III. Feature removal. The features to be removed are indicated by lightly shaded and clear circles. As shown, removal can be simply performed by deleting the node and its associated link for a external node like the clear circle, while removal of an internal node such as the lightly shaded circle, requires addition of new links among the remaining features.

The geometric template is estimated using a variable state dimension Kalman filter. The state of the filter consists of the spatial distances between the features forming the links in the template. Since, only portions of the filter state corresponding to the newly added or removed links in the template are altered, the filter is merely augmented or diminished rather than initializing a new filter on every appearance update. A zeroth order motion model is used to model the spatial constraint of features. The filter state and transition matrices are expressed as,

$$X = \begin{bmatrix} L_1 \\ L_2 \\ \vdots \\ L_{N-1} \end{bmatrix}, \quad A = \begin{bmatrix} 1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix},$$

$$H = \begin{bmatrix} 0 & -1 & 0 & \dots & 1 \\ 0 & 1 & \dots & -1 & 0 \\ 1 & \dots & -1 & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & 1 & -1 \end{bmatrix}$$

where X corresponds to the state of the filter and $L_1, \dots, L_{N-1}$ are the link lengths or spatial distance between features, A is the state transition matrix, and H is the observation transition matrix. Note that the dimensionality of H is $N \times N$ since it transforms the individual feature positions to the links. The measurement error for the links is expressed as a product of the error covariances of the features forming the link.

V. Feature Initialization and Tracking

An adaptive background segmentation method as in [13] is used to automatically detect and track the targets of interest in the scene. Target tracking consists of using both the blob's position measurement and the features' velocity measurement. Using the blob region as the region of interest, features are initialized using the standard Harris' corner operators. As the target moves in the scene, newly detected features replace those lost due to rotations, scale changes, and occlusions. The Sum-of-Squared Differences metric as proposed in [8] is used to eliminate any bad or unreliable features. The feature dynamics are modeled as a first-order motion model using a linear Kalman filter, while the target dynamics are also modeled by a first-order motion model albeit with an extended Kalman filter.

Feature correspondences between consecutive frames is ob-

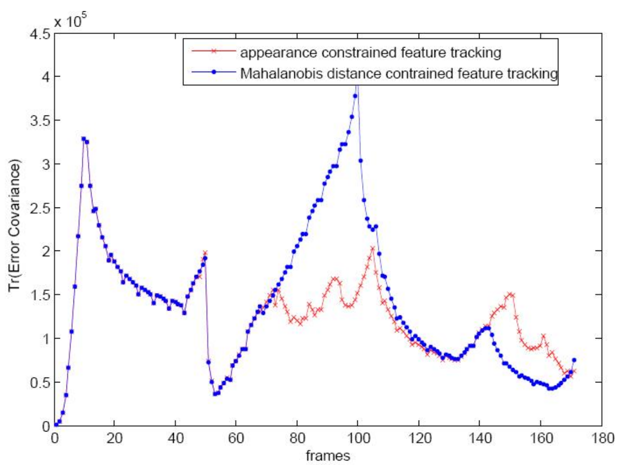


Fig. IV. Tracking error for feature tracking using the Mahalanobis distance, and feature tracking using the geometric template.

tained through Sum-of-Squared Differences (SSD) matching. The measurements obtained from the set of features are then checked for validity using the geometric template-based data association as discussed in Section III.

VI. Results and Discussions

A. Experimental Setup

The objective of the experiments was to test the efficacy of using the geometric template-based data association for obtaining robust tracking. All experiments were performed on fairly crowded outdoor image sequences. In order to evaluate the performance of the proposed data association scheme, standard Mahalanobis distance gating-based data association applied to each feature individually was used as a benchmark.

The input to the target tracker consists of the velocity measurements obtained from the features and position measurements obtained from standard motion segmented blobs. However, the measurements from features were weighted higher compared to the position measurements to reduce the influence of the position measurements on tracking.

Fig. IV and Fig. V show the tracking errors for two different

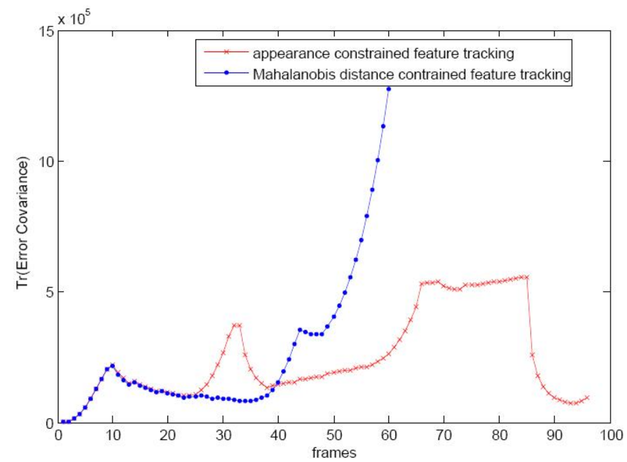


Fig. V. Tracking errors for feature tracking using the Mahalanobis distance and feature tracking using the geometric template. The increase in error for the Mahalanobis distance constrained feature tracker after frame 40 is the result of tracker switching during an occlusion.

targets for the following two cases: (a) feature measurements constrained using Mahalanobis distance-based gating applied to individual features, and (b) feature measurements constrained using the geometric template. The plots show the trace of the error covariance in the estimated target state (position and velocity in the scene coordinates) with time. As shown in Fig. IV, the error covariance for the case where the feature measurements are constrained using the geometric template produces the lowest errors even during occlusions. Occlusions occurred during the frames (65 to 110). Similarly, in Fig. V, the increase in the error covariance after frame 40 for the Mahalanobis distance-gated features is the result of tracker switching to another target, whereas, the geometric appearance constrained tracker produces consistent results without losing the target.

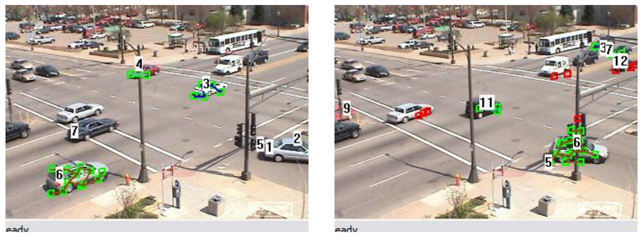
Fig. VI and Fig. VII show examples of tracking in two different outdoor sequences using geometric template constrained tracking.

B. Discussion of Results

As indicated by the results, geometric template constrained tracking helps attain a consistent tracking as well as minimizes the covariance of the estimated trajectory in comparison to the Mahalanobis distance constrained tracking. In comparison to

(a) Frame 246 (b) Frame 321

Fig. VI. Outdoor tracking sequence I.



(a) Frame 1300 (b) Frame 1464

Fig. VII. Outdoor tracking sequence II.

methods such as [6], [11], [12] that compute a fundamental matrix through optimization, the proposed method does not require a lot of features to obtain a robust appearance estimate. The accuracy of the model and data association improves with consistent tracking of the features over time. Further, no prior knowledge of the template is required. Thus, in comparison to adaptive mesh-based methods, no explicit physical models are required. Robust data association can be obtained as long as stable features on the target can be attained. Clearly, a more stable appearance model containing a large number of features would provide a much better estimate for feature measurements. Note that the data association using this method reduces to data association utilizing the standard Mahalanobis distance constraint for individual features on a template having a large uncertainty.

VII. Conclusions

This paper presented an adaptive geometric template for modeling the appearance of targets. The adaptive templates are automatically initialized from the image features detected using the motion blobs representing the target as region of interest. This paper showed how the geometric template can be used as an additional constraint for obtaining good data associations for image features on the targets for a multiple target tracking application. We also showed that using the geometric template-based constraint helps obtain a more robust tracking solution in applications where a standard data association such as a Mahalanobis distance-based method applied to each feature individually fails.

VIII. Acknowledgments

This work has been supported in part by the National Science Foundation through grant #IIS-0219863, the Minnesota Department of Transportation, and the ITS Institute. The authors would also like to thank the reviewers for their comments.

References

[1] Y. Bar-Shalom and T. E. Fortmann. *Tracking and Data Association*. Academic Press, 1987.

[2] M. A. Fischler and R. C. Bolles. Random sample consensus: a paradigm for model fitting and application to image analysis and automated cartography. In *Communication Association and Computing Machine*, volume 24, pages 381–395, 1981.

[3] G. Guo, C.R. Dyer, and Z. Zhang. Linear combination representation for outlier detection in motion tracking. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, volume 2, pages 274–281, June 2005.

[4] C. Harris. *Active Vision*, chapter Tracking with Rigid Models, pages 59–73. MIT Press, 1992.

[5] W. C. Huang and D. B. Goldgof. Adaptive-size meshes for rigid and nonrigid shape analysis and synthesis. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 15(6):611–616, 1993.

[6] D.Q. Huynh and A. Heyden. Outlier detection in video sequences under affine projection. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, volume 1, pages 695–701, 2001.

[7] D. Metaxas and D. Terzopoulos. Shape and nonrigid motion estimation through physics-based synthesis. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 15(6):580–591, 1993.

[8] K. Nickels and S. Hutchinson. Estimating uncertainty in SSD-based feature tracking. *Image and Vision Computing*, 20(1):47–58, January 2002.

[9] C.F. Olson. Maximum-likelihood template matching. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, volume 2, pages 52–57, 2000.

[10] D. Terzopoulos and M. Vasilescu. Sampling and reconstruction with adaptive meshes. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, pages 70–75, June 1991.

[11] T. Tommasini, A. Fusiello, E. Trucco, and V. Roberto. Making good features track better. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, page 178. IEEE Computer Society, 1998.

[12] P. H. S. Torr and D.W. Murray. Outlier detection and motion segmentation. In P. S. Schenker, editor, *Sensor Fusion VI*, volume 2059, pages 432–443. SPIE, 1993.

[13] H. Veeraraghavan, O. Masoud, and N.P. Papanikolopoulos. Computer vision algorithms for intersection monitoring. *IEEE Trans. on Intelligent Transportation Systems*, 4(2):78–89, June 2003.

[14] Y.S. Yao and R. Chellappa. Tracking a dynamic set of feature points. *IEEE Trans. on Image Processing*, 4(10):1382–1395, October 1995.

[15] S.C. Zhu and A. L. Yuille. Forms: A flexible object recognition and modeling system. *Intl. Journal of Computer Vision*, 20(3):187–212, December 1996.

[16] W. Zhu, S. Wang, R. S. Lin, and S. E. Levinson. Tracking of object with svm regression. In *Proc. IEEE Conf. on Computer Vision and Pattern Recognition*, volume 2, pages 240–245, 2001.