Proceedings of the 1999 IEEE
International Conference on Robotics & Automation
Detroit, Michigan • May 1999

Measurement error estimation for feature tracking

Kevin Nickels
knickels@trinity.edu
Dept. of Engineering Science
Trinity University
San Antonio TX 78212

Seth Hutchinson
seth@uiuc.edu
Dept. of Electrical and Computer Engineering
and The Beckman Institute
University of Illinois at Urbana-Champaign
Urbana IL 61801

Abstract

Performance estimation for feature tracking is a critical issue, if feature tracking results are to be used intelligently. In this paper, we derive quantitative measures for the spatial accuracy of a particular feature tracker. This method uses the results from the sum-of-squared-differences correlation measure commonly used for feature tracking to estimate the accuracy (in the image plane) of the feature tracking result. In this way, feature tracking results can be analyzed and exploited to a greater extent without placing undue confidence in inaccurate results or throwing out accurate results. We argue that this interpretation of results is more flexible and useful than simply using a confidence measure on tracking results to accept or reject features. For example, an extended Kalman filtering framework can assimilate these tracking results directly to monitor the uncertainty in the estimation process for the state of an articulated object.

1 Introduction

Estimating the effectiveness of feature tracking information is a very important topic in image processing today. In correspondence-based object tracking the results from several feature trackers, each tracking salient points or edges of an object, are combined to track a (possibly articulated) object [1]. Point tracking for the purpose of computing image flow requires a metric for the confidence in a motion estimate, so that estimates from regions of high confidence can be used to improve estimates in regions of low confidence [2]. Feature tracking confidence measures have also been used in a visual servo control to increase the robustness of the servo control [3].

We characterize the *accuracy* of a feature tracking result by the accuracy of the location computed by the feature tracking. This characterization leads to the evaluation of the *confidence* of a feature tracking result for a more general purpose than that of accepting or rejecting the feature for use in tracking. We analyze the uncertainty in the tracking process, so that we can keep track of the uncertainty in the estimation process being driven by the feature tracking.

We begin with a review of standard feature tracking methods. Following this, we describe our goals with respect to characterizing the spatial discrimination of features. Then we present a Gaussian approximation and describe how sufficient statistics can be used to characterize this approximation. Finally, we present some results from our implemented tracking system.

2 Correlation and feature templates

In correlation-based feature tracking, a *feature template* is used to detect a feature in an image. A feature template contains some representation of the feature and is compared against portions of an image to locate that feature in the image. This comparison utilizes a similarity metric to rate the similarity of the template and the image patch. The image region found to be the most similar to the template is usually taken to be the location of the feature.

The following sections discuss three areas crucial to correlation based tracking: the content of this template, the definition and use of a specific similarity metric for tracking, and the definition of confidence measures on the tracking results.

2.1 Template content

The content of the template is an important choice in feature tracking. If the template faithfully reproduces the actual appearance of the feature in the image, tracking will work well. However, if a template is oversimplified or does not match the appearance of a feature in the image due to unmodeled effects, feature tracking will perform poorly.

0-7803-5180-0-5/99 $10.00 © 1999 IEEE

A template can be generated from a canonical view of the feature, and template matching done in a search window centered about the predicted position of the image. Brunelli and Poggio give a good review of this technique in the context of facial feature tracking [4]. The main problem with this straightforward approach is that the simple template is a 2D entity, and the image patch may undergo transformations that the template cannot model, such as rotation.

A more complex algorithm that also works in certain situations is to use an image patch from the previous image, taken from the area around the last computed position of the feature in that image, for the template. Hager [5] uses this approach for visual servoing. Hager and Belhumeur [6] have also used previous tracking information to warp this image patch before use as a feature template, which increases the flexibility of this approach.

If object and scene modeling are part of the tracking framework, it is possible to create templates from this information. Lopez et al. [7] have a 3D registered texture of a face as part of their object model. Computer graphics techniques are used to render the relevant portion of the scene complete with sophisticated texture mapping to estimate the appearance of a feature in the image. This image patch is then used as a template in the feature tracking portion of the system. Our work uses 3D models for complex articulated objects, also in a graphics-based framework [8], to generate feature templates.

2.2 The SSD similarity metric

In correlation-based tracking, a similarity metric is used to compare the feature template described above to areas of the image to locate the feature in the image.

The standard sum-of-squared-differences (SSD) metric for grayscale images is defined as:

$$SSD(u,v) = \sum_{m,n \in N} [T(m,n) - I(u+m, v+n)]^2, \quad (1)$$

where T is the template image and I is the input image. The location (u,v) represents some location in the input image whose content is being compared to the content of the template. Papanikolopoulos [3] uses the SSD measure to generate tracking results that are then used for robotic visual servoing experiments. Anandan [2] and Singh and Allen [9] use this SSD metric for the computation of image flow.

Often, this measure is not computed for the entire input image, but only for some *search window* in the input image. Primarily for computational reasons, this restriction also serves as a focus of attention for the feature tracking algorithm. Singh and Allen [9] define a fixed size square search window surrounding the previous location of the feature. Kosaka and Kak [10] consider at length the shape and location of the search window. They model the scene and compute a spatial probability density function for the location of each feature, then search the image area corresponding to 85% of the probability mass. We use a constant-velocity model for an articulated object to predict 3D positions for relevant points on the object. Imaging models are then used to project these locations to points on the image plane. A fixed size rectangular search window centered at these locations is established in the input image. See [8] for more details.

3 Confidence, uncertainty estimation, and spatial uncertainty

It has been noted [2] that popular similarity measures often lead to some unreliable matches, particularly in image regions with little textural information. For this reason, it is often helpful to compute a *confidence* on the match found, as well as a location. This confidence measure typically gives information regarding the reliability of the match score. This scalar score often is used to estimate the reliability of the feature, i.e. for use in later tracking operations or to propagate image flow information from one portion of an image to another. Below, we will describe a matrix-valued covariance matrix that contains information both about the overall confidence in a feature measurement, and information about how accurate the measurement is in all directions.

Anandan [2] used the SSD matching scores of a template with a 5×5 image region to develop a match confidence measure based on the variation of the SSD values over the set of candidate matches. Anandan argued that if the variation of the SSD measure along a particular line in the search area surrounding the best match is small, then the component of the displacement along the direction of that line cannot be uniquely determined. Conversely, if there is significant variation along a given line in the search area, the displacement along this line is more likely correct.

Singh and Allen define a *response distribution* based on the SSD metric (1) as

$$\mathcal{RD}_c(u,v) = exp(-kSSD(u,v)), \quad (2)$$

where k is used as a normalization factor. The normalization factor k was chosen in [9] so that the maximum response was 0.95. Singh and Allen then argue that

each point in the search area is a candidate for the "true match." However, a point with a small response is less likely to be the true match than a point with a high response. Thus, the response distribution could be interpreted as a probability distribution on the true match location - the response at a point depicting the likelihood of the corresponding match being the true match. This interpretation of the response distribution allows the use of estimation-theoretic techniques.

Under the assumption of additive zero mean independent errors, a covariance matrix is associated with each location estimate:

$$\mathbf{P}_m = \begin{bmatrix} \widehat{\sigma_u^2} & \widehat{\rho\sigma_u\sigma_v} \\ \widehat{\rho\sigma_u\sigma_v} & \widehat{\sigma_v^2} \end{bmatrix} \quad (3)$$

$$\widehat{\sigma_u^2} = \sum_{u,v \in N} \mathcal{RD}(u,v)(u-u_m)^2/TP \quad (4)$$

$$\widehat{\sigma_v^2} = \sum_{u,v \in N} \mathcal{RD}(u,v)(v-v_m)^2/TP \quad (5)$$

$$\widehat{\rho\sigma_u\sigma_v} = \sum_{u,v \in N} \mathcal{RD}(u,v)(u-u_m)(v-v_m)/TP \quad (6)$$

$$TP = \sum_{u,v \in N} \mathcal{RD}(u,v) \quad (7)$$

where u_m and v_m are the estimated locations, in the u and v directions, of the feature. The reciprocals of the eigenvalues of the covariance matrix are used as confidence measures associated with the estimate, along the directions given by the corresponding eigenvectors. To our knowledge, Singh and Allen are the first researchers treat the location of the best match as a random vector, and the (normalized) SSD surface is used to compute the spatial certainty of the estimate of this vector [9]. These confidence measures are used in the propagation of high confidence measurements for local image flow to regions with lower confidence measurements, such as caused by large homogeneous regions.

Our work develops a different normalization procedure for $\mathcal{RD}$ that is useful for the evaluation of isolated feature measurements from template images. As described in Section 4.2, we compute one covariance matrix and one location for each feature, and use this information in a model-based object tracking framework. We do not reject any tracking information, but weight each measurement on the basis of this covariance matrix, using as much information as possible from the feature tracking.

As the SSD measure is used to compare the template to areas of the image near the area generating the minimum SSD score, some measure of the *spatial discrimination* power of the template can be generated [2]. Spatial discrimination is defined as the ability to detect feature motion along a given direction in the image. This concept is quite similar to the confidence measures discussed in Section 3 that estimate the reliability of the location estimate. However, we interpret the confidences as spatial uncertainties in the returned location.

While conclusions about the efficacy of a given template for feature localization can be drawn from the fully computed SSDS, it is expensive both computationally and from a computer memory standpoint to maintain the complete surface for this purpose. In the next section, we derive a approximation for $\mathcal{RD}$ that is more useful.

4 A practical approximation for $\mathcal{RD}$

In order to maintain and use relevant information about the shape of the response distribution, we introduce a mathematical approximation to the distribution given in (2). By suppressing the off-peak response of the feature tracking result, this response distribution function converts the SSDS into an approximately Gaussian distribution that contains the feature tracking information we wish to maintain. Since many object tracking systems (including all Kalman filter-based systems) assume measurements are random vectors with Gaussian probability density functions, we explicitly model and approximate this density.

4.1 Uncertain feature measurements

The measurement vector $\mathbf{z}_k$ is interpreted as an uncertain location in the (u,v) plane, and modeled as a 2D Gaussian random vector. It is illustrative to analyze the behavior of the density function for this vector with respect to the spatial certainty of the feature tracking result as $\mathbf{R}_k$, the covariance matrix for the vector, changes. For example, if $\mathbf{R}_k = \sigma^2 I$, where σ^2 is the variance of the vector, the location is equally certain in each direction. The ellipses of equal probability on the density surface are circles. If $\sigma_u \neq \sigma_v$, where σ_u^2 and σ_v^2 are the variances in the u and v directions, the location is more certain in one direction (given by the minor axis of the ellipses of equal probability) than in the other direction (given by the major axis). As the length of the major axis approaches infinity, complete uncertainty on the location along this dimension is asserted. It is well known that the mean and covariance are sufficient statistics for a Gaussian random variable. Therefore, if this Gaussian density surface is sufficient to model the tracking behavior,

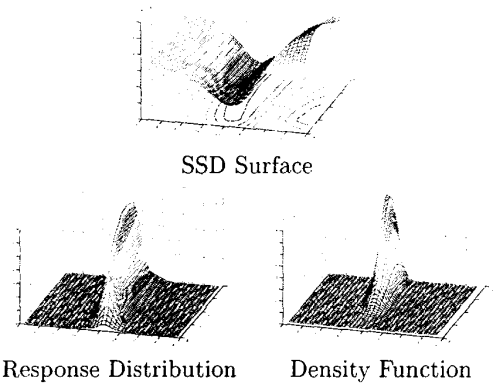


Figure 1: Approximation of response distribution by density function.

it is no surprise that the mean and covariance suffice to maintain this information. In the next section we explain how we estimate these quantities.

4.2 Parameter estimation from the SSDS

This section describes a process for analyzing the SSDS to arrive at estimates for the mean and variance of a Gaussian random vector. The density function of this vector acts as an approximation to the response distribution $\mathcal{RD}$ (see (2)) for the purpose of tracking features.

Our computation of the normalization factor k in (2) differs from that of Singh and Allen [9]. We chose k such that

$$\sum_{u,v \in N} \mathcal{RD}(u,v) \approx 1. \tag{8}$$

This has the effect of suppressing the off-peak response of the feature detector, when compared with Singh and Allen's normalization. Since we are using correlation between synthetic templates and images, the off-peak response in our situation is more significant than for Singh and Allen. As shown in Figure 1, our normalization makes the response distribution approximate a Gaussian density function with the desired characteristics with respect to feature tracking.

The mode, or most probable value, of a random vector is located at the peak of the density function. We take the location of the minimum of the SSDS as our value for the mode of the vector,

$$\mathbf{z}_k = argmin_{u,v} SSD(u,v). \tag{9}$$

The variance of u (σ_u^2), the variance of v (σ_v^2), and the covariance between u and v ($\rho_{uv}\sigma_u\sigma_v$) can be estimated directly from the response distribution using Equations (2) and (3)-(7), yielding the desired covariance matrix,

$$\mathbf{R}_k = \begin{bmatrix} \widehat{\sigma_u^2} & \widehat{\rho_{uv}\sigma_u\sigma_v} \\ \widehat{\rho_{uv}\sigma_u\sigma_v} & \widehat{\sigma_v^2} \end{bmatrix}, \tag{10}$$

which, as described above, contains complete information about the *orientation* and *shape* of the error ellipsoids. Figure 1 illustrates this process for a vertical edge feature.

Of course, as we are only maintaining the mean and variance of the random vector, and not the complete SSDS, this is only an approximation to the complete information about local image structure given by the SSD. However, it does give an indication of both the absolute quality of the match and, in cases where edge features exist, the direction of the edge.

5 Results

5.1 Gripper feature

The feature illustrated in this section is the end-effector of a robot. Figure 2 shows the search region, tracking result, and measurement uncertainty estimates for two different cases. Note that since the SSD measure involves the image area surrounding a pixel, a border around the search region must be retained for each search region.

The results of feature tracking in normal circumstances are shown in Figure 2(a). The cross indicates the location of the minimum point of the SSD surface. The complete SSDS and the Gaussian approximation to this surface are shown in Figure 2(b)-(c), both indicating equal accuracy of the tracking result in all directions. Note the effect in (b) and (c) of our normalization procedure. Even though there is significant off-peak response, the relative certainty in the peak response with respect to the lower responses indicated a single proper match. This fact is evident from the final result shown in (c).

In (d)-(f), we present an illustration of the usefulness of the on-line estimation of template efficacy. This feature has the same template as in the previous case. However, a person has stepped between the camera and the feature, occluding the feature. The feature template thus does not match any portion of the search window well, as shown in (e). This mismatch causes larger values for the variances for this measurement, and (f) indicates high uncertainty of the feature location in all directions.

A 2D measurement, represented by the cross in Figure 2(a) and (d), and a 2×2 covariance measurement are the output of the feature tracking, and are used directly in the EKF framework described in [8].

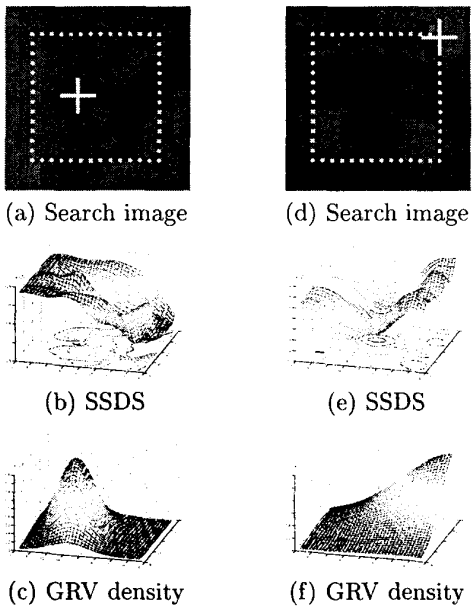


Figure 2: Tracking results for gripper feature (a)-(c) normal (d)-(f) externally occluded

5.2 Edge feature

This case illustrates the usefulness of the measurement uncertainty estimation for tracking features with poor spatial discrimination in one direction. An edge feature can be tracked well only in the direction orthogonal to the edge. This feature arises from a point on the edge of the robotic arm. Thus, the orientation of the edge in the feature depends on the configuration of the robot. As the configuration of the robot changes, the direction of the edge projected onto the image plane will change. In Figure 3 (a), the edge is in a diagonal orientation. The full SSDS shown in (b) has a ridge along this direction, indicating good match scores along the ridge. After normalization, the density function shown in (c) exhibits the same ridge, while suppressing the off-peak match scores on both sides of the ridge. Similarly, the edge in (d) is in a vertical orientation, so the ridges in (e) and (f) are in the vertical direction.

By maintaining this information, the system can exploit the feature tracking information to a greater extent: the result is neither endowed with inappropriate confidence due to the good accuracy in the direction orthogonal to the edge nor unduly devalued due to the poor accuracy in the direction along the edge.

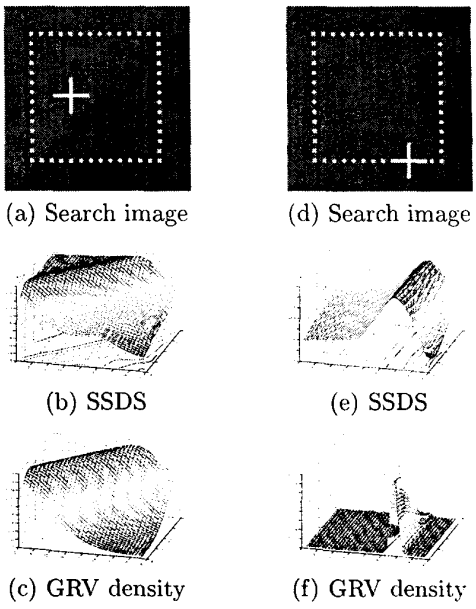


Figure 3: Tracking results for edge feature.

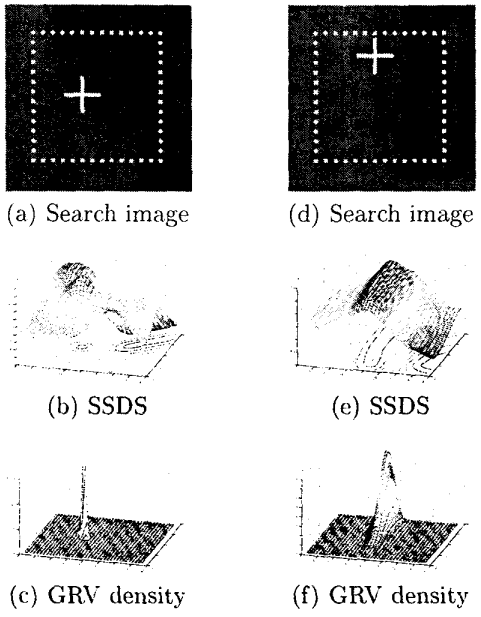
5.3 Degenerate point feature

In this section, we illustrate another aspect of the usefulness of on-line estimation of template efficacy. Since our object-tracking system works under widely varying configurations of the object, the appearance of features may change significantly during tracking. A single feature acceptance or rejectance decision will not suffice in this case. Figure 4 shows tracking results for a feature that undergoes such a change in appearance, a point of intersection of a black line on the edge of the robotic arm with the rear edge of the arm.

This feature is a point of high texture in both directions when the arm is approximately parallel to the image plane, as shown in (a), and acts like a point feature. This feature location can be found with high accuracy in all directions, shown in (c).

However, this feature acts like an edge feature in other configurations, such as the configuration shown in (d), where the arm is pointing roughly toward the camera. In this configuration, the feature appears as a vertical edge feature. The location of the feature in this case can be found with high accuracy in only the horizontal direction, as seen in (e) and (f).

Again, the maintenance of the covariance matrix instead of a single confidence measure makes this suboptimal tracking result not only tolerable, but useful.



(a) Search image (d) Search image

(b) SSDS (e) SSDS

(c) GRV density (f) GRV density

Figure 4: Tracking results for a point feature (a)-(c) nondegenerate and (d)-(f) degenerate (acting as edge feature)

6 Conclusions

The method presented uses the SSDS, a common intermediate result in correlation-based feature tracking, to compute quantitative estimates for the spatial accuracy of the feature tracking result. This estimate consists of a covariance matrix for a Gaussian random vector. Analysis of this matrix yields information about the directions (if any) in which the template is discriminating the feature from the image background, and provides a quantitative measure of confidence in each direction.

The feature tracking results, combined with this matrix, yields a composite measure that is useful when analyzing the tracking results. An example of the use of this matrix in model-based object tracking can be found in [8]. Analysis of the results can detect templates that do not discriminate effectively in any direction. By associating spatial confidence measures with feature tracking results, those results can be more fully exploited: the fact that some directions may have high confidence does not lead us to accept the entire measurement, and the fact that some directions may have low confidence does not lead us to disregard useful data.

References

[1] D. Lowe, "Robust model-based motion tracking through the integration of search and estimation," *International Journal of Computer Vision*, vol. 8, pp. 113–122, Aug. 1992.

[2] P. Anandan, "A computational framework and an algorithm for the measurement of visual motion," *International Journal of Computer Vision*, vol. 2, pp. 283–310, Jan. 1989.

[3] N. P. Papanikolopoulos, "Selection of features and evaluation of visual measurements during robotic visual servoing tasks," *Journal of Intelligent and Robotic Systems*, vol. 13, pp. 279–304, July 1995.

[4] R. Brunelli and T. Poggio, "Face recognition: Features versus templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, pp. 1042–1052, Oct. 1993.

[5] G. Hager, "Real-time feature tracking and projective invariance as a basis for hand-eye coordination," in *Proceedings IEEE Computer Society Conference on Computer Vision Pattern Recognition*, pp. 533–539, 1994.

[6] G. Hager and P. Belhumeur, "Real-time tracking of image regions with changes in geometry and illumination," in *Proceedings IEEE Computer Society Conference on Computer Vision Pattern Recognition*, pp. 403–410, 1996.

[7] R. Lopez, A. Colmenarez, and T. S. Huang, "Vision-based head and facial feature tracking," in *Advanced Displays and Interactive Displays Federated Laboratory Consortium, Annual Symposium*, Advanced Displays and Interactive Displays Federated Laboratory Consortium, Jan. 1997.

[8] K. M. Nickels, *Model Based Articulated Object Tracking*. PhD thesis, University of Illinois, Urbana, IL, 61821, Oct. 1998.

[9] A. Singh and P. Allen, "Image flow computation: An estimation-theoretic framework and a unified perspective," *Computer Vision Graphics and Image Processing: Image Understanding*, vol. 56, pp. 152–177, Sept. 1992.

[10] A. Kosaka and A. C. Kak, "Fast vision-guided robot navigation using model-based reasoning and prediction of uncertainties," *Computer Vision and Image Understanding*, vol. 56, pp. 271–329, Nov. 1992.