

# Maximum-Likelihood Image Matching

Clark F. Olson, *Member, IEEE*

**Abstract**—Image matching applications such as tracking and stereo commonly use the sum-of-squared-difference (SSD) measure to determine the best match. However, this measure is sensitive to outliers and is not robust to template variations. Alternative measures have also been proposed that are more robust to these issues. We improve upon these using a probabilistic formulation for image matching in terms of maximum-likelihood estimation that can be used for both edge template matching and gray-level image matching. This formulation generalizes previous edge matching methods based on distance transforms. We apply the techniques to stereo matching and feature tracking. Uncertainty estimation techniques allow feature selection to be performed by choosing features that minimize the localization uncertainty.

**Index Terms**—Image matching, tracking, stereo, maximum-likelihood estimation.

## 1 INTRODUCTION

IMAGE matching is a common tool in many applications, including object recognition, stereo matching, and feature tracking. Some applications use a template composed of edge pixels, which are robust to variations in lighting and viewpoint, but others operate directly on the gray-levels present in the image. Many applications of gray-level image matching use the sum-of-squared-differences (SSD) measure to determine the best match. However, this measure is sensitive to outliers and is not robust to variations in the template, such as those that occur at occluding boundaries in the image. This method has been improved through the use of robust measures to compare image pixel intensities [1], [2]. However, these methods still compare only the overlapping pixels at some relative position between the images.

We describe a new measure for image matching. Rather than using a nonrobust measure, such as the SSD, or even a robust measure comparing overlapping pixels, we generalize previous work on matching edge images using a probabilistic framework. This method considers the distance from each pixel in the template (at some position with respect to the image) to the closest matching pixel in the image. It is not restricted to overlapping pixels. The distance encompasses both the image distance and the intensity variation. This technique is robust to small changes in viewpoint, illumination, and other variations in the template and is applicable to both edge templates and gray-level templates.

Our method is general with respect to the set of pose parameters allowed. We formulate the method using two-dimensional edge and intensity templates with the pose space restricted to translations in the plane in order to simplify the presentation. However, the techniques can be adapted to more complex transformations, as we have demonstrated by applying these techniques to object recognition under similarity transformations of the object template [3].

We first describe a new image matching measure that is based upon maximum-likelihood estimation of the template position using the distance values from the template pixels (edges or

intensities) to the image pixels. When a particular likelihood function is introduced for the distance from each template feature to the closest matching image feature, this formulation yields a conventional matching method. We use a different likelihood function to generate a new image matching measure that more accurately models errors and outliers in the data.

This probabilistic formulation for image matching yields several benefits. Perhaps most importantly, we can consider an arbitrary likelihood function for the matching error between edge or intensity features. In contrast to Hausdorff distance-based measures for edge matching [4], [5], this allows us to eliminate the sharp distinction between matched and unmatched template features. In contrast to previous image matching applications, we allow matches between pixels that do not directly overlap. An efficient search strategy is described that generalizes previous work for detecting the template position that optimizes a variation on the Hausdorff distance.

We apply these techniques to real images in two application areas. The first application is stereo matching, where we obtain dense, subpixel disparities through the use of these techniques. Next, we show that uncertainty estimation techniques can be adapted to optimal feature selection for tracking features in an image sequence and we compare the performance of the method to image matching using the SSD, SAD (sum-of-absolute-differences), and Lorentzian function for distance comparisons.

## 2 MAXIMUM-LIKELIHOOD MATCHING

We use a maximum-likelihood matching formulation that has been previously applied to robot localization [6]. This section reviews this method and discusses a novel application of it to gray-level image matching.

### 2.1 Maximum-Likelihood Measure

To formalize the problem, let us say that we have a set of model (or template) features,  $M = \{\mu_1, \dots, \mu_m\}$ , and a set of image features,  $I = \{\nu_1, \dots, \nu_n\}$ . Both  $M$  and  $I$  can be considered to be discrete sets of points at the locations of the occupied pixels in the image or template. For now, the elements of  $M$  and  $I$  will be vectors of the  $x$  and  $y$  image coordinates of edge pixels.

We let  $t$  be a random variable describing the position of the template in the image. This position can be thought of as a function that maps the template features into the image. The position could be a simple translation of the template with respect to the image, but it also may encompass more complex transformations, such as similarity transformations or affine transformations. Here, we make an implicit assumption that exactly one instance of the template appears in the image. Note that, if we are seeking templates that may appear more than once in an image or not at all, we must only set some threshold on the likelihood function, as is usually done in matching applications. The template positions that surpass the threshold are reported as the possible positions of the template.

In order to formulate the problem in terms of maximum-likelihood estimation of the template position, we must have some set of measurements that are a function of the position of the template. Similar to methods based on the Hausdorff distance [4], we use the distance from each template pixel (at the position specified by  $t$ ) to the closest occupied pixel in the image as our set of measurements. We denote these distances  $D_1(t), \dots, D_m(t)$ . In general, these distances can be found quickly for any  $t$ , if we precompute the distance transform of the image [7], [8], [9].

The joint probability density function (PDF) for the distances, given  $t$ , can be approximated as the product of each individual PDF, if the distance measurements are independent. We have

• The author is with Computing and Software Systems, University of Washington, Bothell, 18115 Campus Way NE, Box 358534, Bothell, WA 98011-8246. E-mail: cfolson@u.washington.edu.

Manuscript received 31 Oct. 2000; revised 10 July 2001; accepted 6 Nov. 2001.

Recommended for acceptance by D. Fleet.

For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 113085.

found that this yields accurate results since the correlation between the distances falls off quickly as the points become farther apart.

$$p(D_1(t), \dots, D_m(t) | t) = \prod_{i=1}^m p(D_i(t)), \quad (1)$$

where  $p(D_i(t))$  is the probability density function (PDF) of  $D_i(t)$  evaluated at the template position  $t$ . In order to find the most likely  $t$ , we find the position that maximizes (1).

It is often easier to work with the logarithm of (1) since this involves addition, rather than multiplication, and yields a measure that preserves the ordering of the template positions.

$$\ln L(D_1(t), \dots, D_m(t) | t) = \sum_{i=1}^m \ln p(D_i(t)). \quad (2)$$

We seek the position  $t$  that maximizes this likelihood function and this results in a maximum likelihood estimate for the template position.

## 2.2 Estimating the PDF

While it is only formally correct to use a probability density in (1) or (2), any function  $p(d)$  will yield a matching criterion for image templates in the above formulation. We note that a measure equivalent to the Hausdorff fraction [5], [10] can be obtained through the use of:

$$\ln p(d) = \begin{cases} k_1 + k_2, & \text{if } d \leq \delta, \\ k_1, & \text{otherwise.} \end{cases} \quad (3)$$

Each feature in the model template contributes one of two values, depending on the distance to the closest image feature (as in the Hausdorff fraction). If there is support for the template feature in the image at this position (i.e., an image feature lies within  $\delta$  of it), then some constant value is assigned to  $p(d)$  (a uniform likelihood in the distance to nearest feature). Otherwise, some smaller constant likelihood is assigned to  $p(d)$  (also uniform, but less likely). The precise values of  $k_1$  and  $k_2$  are unimportant in this case (they do not change the ranking of template positions) as long as  $k_2 > 0$ . Note, however, that this function is not a probability density function since it does not integrate to unity.

Previous formulations for edge template matching using the Hausdorff fraction do not accurately model the uncertainties in the matching process (since (3) is not accurate). The use of a function that closely models the uncertainties results in a measure that is able to detect templates with fewer false positives and allows accurate uncertainty estimation to be performed. We use a function based on the principle that the density can be modeled as the sum of two terms (one for inliers and one for outliers):

$$p(d) = \alpha p_1(d) + (1 - \alpha) p_2(d). \quad (4)$$

The first term describes the error density when the edge pixel is an inlier (in the sense that the edge pixel also appears in the image), where  $\alpha$  is the probability that the pixel is an inlier. In this case,  $d$  is the error in the location of the pixel. Usually, we can model the probability density as normal in the distance to the closest edge pixel. Assuming  $\sigma_x = \sigma_y = \sigma$  we have:

$$p_1(d) = \frac{1}{2\pi\sigma^2} e^{-(d_x^2 + d_y^2)/2\sigma^2} = \frac{1}{2\pi\sigma^2} e^{-d^2/2\sigma^2}. \quad (5)$$

Note that this is a bivariate probability density in  $d_x$  and  $d_y$ , rather than a univariate probability density in

$$d = \sqrt{d_x^2 + d_y^2},$$

which would imply different assumptions about the error density. Formally, we should think of  $d$  as a 2-vector of the  $x$  and  $y$  distances to the closest edge pixel in the image. However, to compute the probability density function, it will only be necessary to know the magnitude of  $d$ . Thus, the orientation of the error vector is irrelevant (as long as the errors in each direction are independent.) The second term in (4) describes the probability density of the distances when the edge is an outlier. In this case, the template pixel does not appear in the image for some reason (such as occlusion). In theory, this term should also decrease as  $d$  increases since even outliers are usually not far from some edge pixel in the image. However, this allows pathological cases to have an undue effect on the likelihood for a particular template position. In practice, we have found that using the expected probability density for a random outlier yields excellent results:

$$p_2(d) = p_{exp}. \quad (6)$$

Note that, like (3), (6) is not a probability density function. This is unavoidable in a robust measure. Any true probability density function must become arbitrarily close to zero for large values of  $d$  (owing to the limited density that can be assigned over the range of the function). This means that any true PDF will yield a nonrobust measure when used in our maximum-likelihood method since a single outlier can cause the likelihood function to become arbitrarily close to zero (and, thus, the logarithm becomes an arbitrarily large negative number). We do not wish this to occur and have found that using the expected probability density is a useful alternative. Note that, while our probability density function is not derived from a generative model, this does not cause a problem in practice.

## 2.3 General Search Strategy

In order to determine the template position that maximizes the likelihood function developed above, we use a search strategy that generalizes previous methods for matching with the Hausdorff distance [11], [12]. This method has been previously described in some detail [6], so we give only an overview here. The basic idea of these methods is to use a multiresolution search that examines a hierarchical cell decomposition of the space of possible template positions. This method divides the space of template positions into rectilinear cells and determines which cells could contain a position satisfying the acceptance criterion. The cells that pass the test are divided into subcells, which are examined recursively, while the rest are pruned. Note that each template position can be evaluated quickly, if we precompute the distance transform of the image, allowing each distance to be determined using a single memory access.

In addition to determining the most likely template position, we can estimate the uncertainty in the localization in terms of both the variance of the estimated positions and the probability that a qualitative failure has occurred. The uncertainty is estimated by measuring the rate at which the likelihood function falls off from the peak. This is performed by fitting a normal distribution to a neighborhood of values in the likelihood function around the location of the maximum. The probability of failure is estimated by comparing the sum of the likelihood scores under the selected peak with the remainder of the pose space.

## 2.4 Gray-Level Images

While these techniques have, so far, been described in terms of binary templates, they can be extended to gray-scale templates by considering the image to be a surface in three dimensions ( $x$ ,  $y$ , and intensity). We will, thus, describe the techniques in terms of the occupied pixels, which are the edges in an edge map or the intensity surface in the three-dimensional representation of a gray-scale image. For grayscale images, the model templates and images

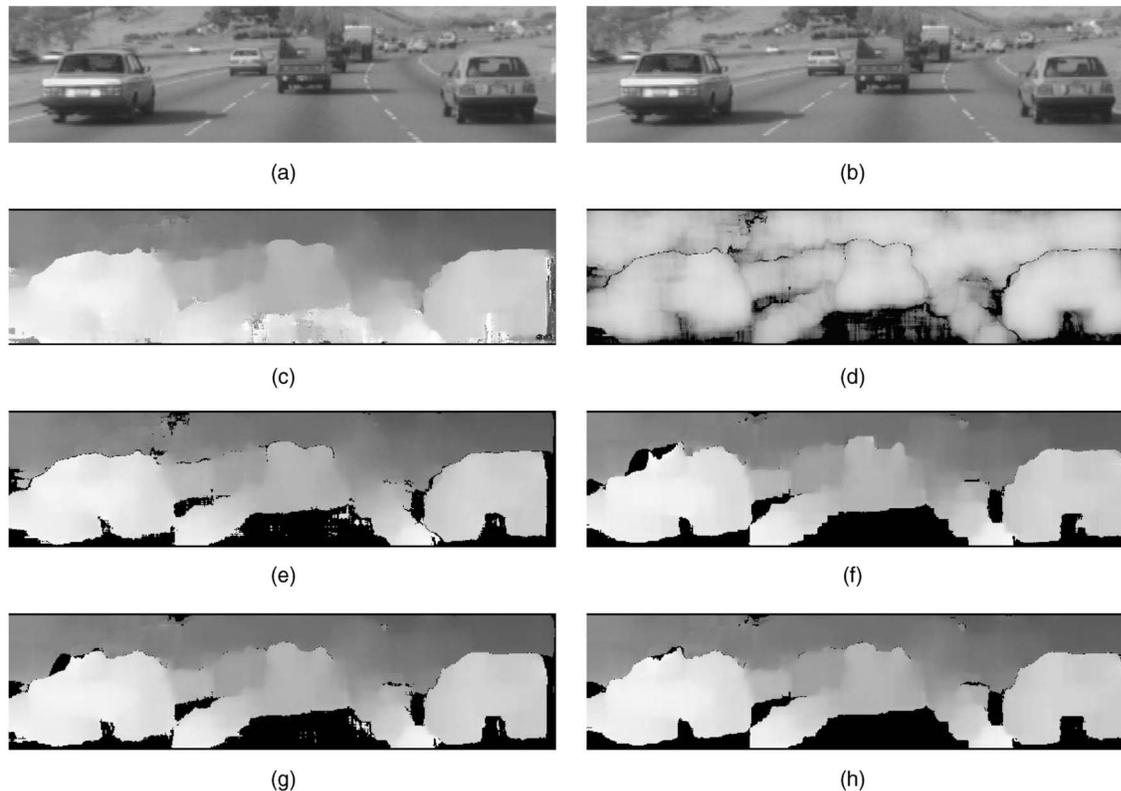


Fig. 1. Stereo matching example. (a) Left image of stereo pair. (b) Right image of stereo pair. (c) Complete disparity map. (d) Uncertainty map (dark values are more uncertain). (e) Disparity map after pruning. (f) Disparity map computed using SSD. (g) Disparity map computed using SAD. (h) Disparity map computed using Lorentzian.

are sets of 3-vectors (corresponding to the occupied pixels in the images). We must now define a distance function over the three dimensions, but the remainder of the methodology is unchanged.

For two pixels  $\mu_i = (x_i, y_i, z_i)$  and  $\nu_j = (x_j, y_j, z_j)$ , where  $z$  is the intensity, we have used a variation of the  $L_1$  distance metric since this makes the distance computations simple:

$$D(\mu_i, \nu_j) = |x_i - x_j| + |y_i - y_j| + \gamma|z_i - z_j|.$$

The value of  $\gamma$  should be chosen such that the expected error in each dimension has the same standard deviation.

It is interesting to note that we could achieve the same results as the SSD measure by assuming that there are no outliers ( $\alpha = 1$ ) and using:

$$D(\mu_i, \nu_j) = \begin{cases} z_i - z_j, & \text{if } x_i = x_j \text{ and } y_i = y_j, \\ \infty, & \text{otherwise.} \end{cases}$$

The maximum-likelihood measure gains robustness by explicitly modeling the possibility of outliers and allowing matches against pixels that do not precisely overlap the template pixel. The SAD and robust measures, such as the Lorentzian [1], [2], can be similarly placed into this framework.

### 3 RESULTS

This section applies the maximum-likelihood matching techniques to stereo matching and feature tracking, including the use of uncertainty estimation techniques to select the best features for tracking. For these experiments, we used  $\alpha = 0.75$ , which is lower than we expect the actual value to be. We have found the method is insensitive to this value, as long as it is not close to one or zero. We estimate  $p_{exp}$  by sampling random locations from the image.

#### 3.1 Stereo Matching

Our method can be applied to tracking and stereo matching applications in a manner similar to correlation-based methods, where small windows from one image are used as the templates, and they are matched against windows in another image from a slightly different viewpoint (for stereo) or after motion has taken place.

Fig. 1 shows an example where stereo matching was performed using these techniques. The disparity, uncertainty, and probability of a false match were computed for each template in the left image of the stereo pair by matching against the right image. The complete disparity map, which is shown in Fig. 1c, contains outliers and inaccuracies due to occlusions and image regions with low texture. Darker values correspond to smaller disparities. Fig. 1d displays an uncertainty image, where the darker pixels show the image locations that are more likely to produce large errors in the disparity computation. The disparity map, after pruning the locations for which the uncertainty estimate or probability of failure is large (Fig. 1e), contains no remaining outliers. Note that the disparity varies smoothly within objects, but changes abruptly along object boundaries. This can be attributed to the robustness of the matching measure with respect to data arising from multiple sources (e.g., a car and the background). For the windows containing edges from multiple depths, the window is automatically assigned the disparity corresponding to the majority of the textured data.

For comparison, Figs. 1f, 1g, and 1h show the results of applying these same techniques using the SSD measure, the SAD measure, and the Lorentzian, respectively. For these cases, we can still compute an uncertainty and probability of failure using our techniques. However, the results are less accurate with these methods. A small number of outliers remain in the disparity map for each case. In addition, this measure yields lower quality results

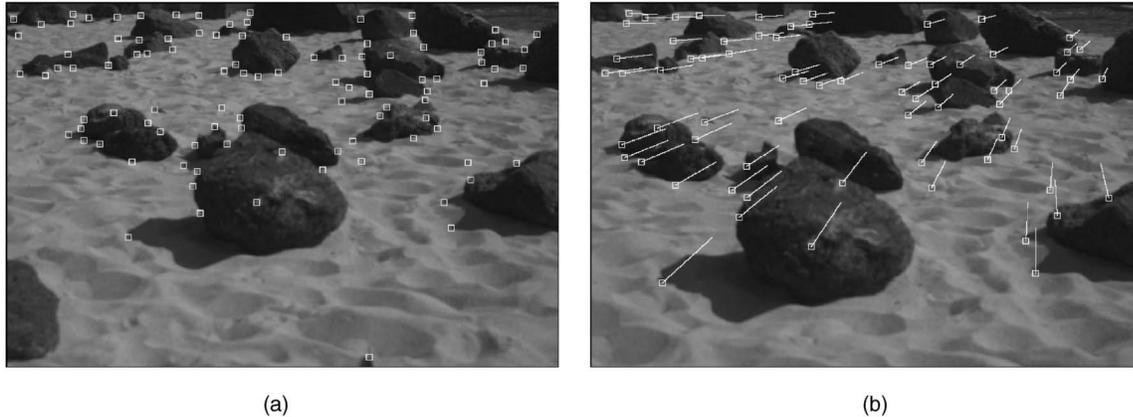


Fig. 2. Optimal feature selection for grayscale matching. (a) Features selected. (b) Tracked features.

in the neighborhoods of occlusion boundaries since they are less robust to changes in the composition of the template.

### 3.2 Feature Selection and Tracking

Since our techniques provide estimates of the uncertainty for matching a template, they can be adapted to perform feature selection for tracking applications. This is performed by estimating the uncertainty of matching each possible feature and selecting those with low uncertainties. We generalize our previous work on landmark selection for robot localization here [13].

We first compute a distribution over the image intensities that captures the template variations as the camera moves. This distribution models only the changes in pixel intensity, ignoring the translation of the pixels in the image, since the template matching procedure searches over the translations. To estimate this distribution, we initially take the intensity surface of the template to have probability 1. (This initial distribution implies that we would observe the same intensity value at each pixel as in the template, with no possibility of a different value.) This distribution is next smoothed in both position and intensity to model noise and warping as the camera moves. We then perform uncertainty estimation for each possible template by matching against the computed distribution (which is treated as a three-dimensional image). The locations with low uncertainties are selected as the optimal features to track [13].

Fig. 2 shows an example of the feature selection techniques applied to an image of rocky terrain. In this case,  $1007 \times 7$  feature templates were selected for tracking. We then performed tracking in a subsequent image, after the camera had undergone motion. For each selected feature, we searched over the entire postmove image for a match.<sup>1</sup> Fig. 2b shows the 72 features that survived pruning using the uncertainty and probability of failure measures. No false positives remain in the tracked features for this case.

To compare against other techniques, the same procedure was applied to a sequence of images similar to Fig. 2. The results for our method, as well as the SSD, SAD, and Lorentzian, can be found in Table 1. Our maximum-likelihood method achieves both a higher tracking rate and lower fraction of outliers than the other methods, owing the higher robustness to template variations at occluding boundaries, where many of the features occur.

1. In practice, the search space is usually reduced using prior knowledge about the camera motion. We performed the search over the entire image in this experiment to demonstrate the robustness of the matching techniques.

## 4 RELATED RESEARCH

There has been considerable research on topics related to this work, including probabilistic recognition, robust image matching, and distance transform-based edge matching. In this section, we describe the most closely related work and its relationship to our work.

Robust measures for motion estimation have been developed by Black and Anandan [1], Black and Rangarajan [2], and Black et al. [14]. Rather than the squared difference between pixel intensities, they have used robust measures (usually the Lorentzian) that have an influence function that tends to zero for large distances. Unlike our work, they have also examined methods for improving models for appearance change, rather than treating such changes as outliers.

Improvements to previous edge-based matching methods using the Hausdorff distance (or a variation) were given by Huttenlocher et al. [4]. The classical Hausdorff distance uses the distance of the worst matching point in calculating the distance between two sets of points. However, this does not improve the robustness of the measure since a single outlier in either set may result in an arbitrarily large error in the measure. In order to allow outliers, Huttenlocher et al. define a partial distance between point sets, which uses a predetermined quantile to eliminate some fraction of the worst matching points and, thus, allows outliers. Another variation is to determine the maximum fraction of the points that can be used in one or both sets such that the distance is below some threshold [5], [10]. Our method can be considered a generalization of these techniques. Instead of simply counting the number of points that lie within some distance, all of the points contribute a probabilistic score to the overall likelihood. Some methods based on the Hausdorff distance yield a reverse distance that measures how well the image matches the template (where the forward distance measures how well the template matches the image). Our approach does not yield such a reverse distance.

Viola and Wells [15] describe an information theoretic approach based on maximizing the mutual information between the

TABLE 1  
Comparison of the Tracking Rate and Outlier Rate of Four Methods

Function	Fraction tracked	Fraction outliers
SSD	0.380	0.023
SAD	0.630	0.021
Lorentzian	0.600	0.019
MLE	0.706	0.016

template and the image. Rather than working with features extracted from an image, this method works directly on the image intensities and is able to gain greater robustness to lighting variation than correlation-based methods. Their method uses an iterative optimization method to detect the optimal template positions, so an initial estimate of the position is necessary. While this method is able to achieve robust recognition, applications of the method usually examine only a fraction of the image data (using sampling), since the computation is time consuming.

Rucklidge [16] has given an alternative strategy for robust matching of gray-level images. He first generalizes previously developed search strategies to apply to gray-level image matching. This strategy is applicable to the matching measure described in this paper. In addition, he describes a new matching measure based on neighborhood difference functions. In this method, each pixel is allowed to match any gray level spanned by a neighborhood of pixels, with a nonzero distance only if the value is outside this span. While this improves upon the brittleness of the SSD measure, it still imposes a sharp boundary in the image space between allowable matches and nonallowable matches, rather than using a continuous likelihood function.

## 5 SUMMARY

In this paper, we have described a new formulation for image matching in terms of maximum-likelihood estimation. This formulation seeks local maxima in the likelihood function of the position of the template with respect to the image. While this implicitly assumes that the template appears once in the image, this formulation can be applied when the template does not appear in the image (or appears multiple times), if an appropriate threshold is used to determine which locations are output as likely template positions.

This formulation yields advantages over previous methods. The measure is more robust to template variations than the conventional SSD measure. It is also superior to improved measures, such as SAD and the Lorentzian function. In contrast to measures based on the Hausdorff distance, we use a smoothly varying probability density function that eliminates the sharp boundary inherent in the conventional two-valued support function.

We have applied this technique to stereo matching and feature tracking with optimal feature selection. Improved results were obtained in both domains.

## ACKNOWLEDGMENTS

The research described in this paper was carried out in part at the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California, under a contract with the US National Aeronautics and Space Administration. An earlier version of this work appears in the 2000 IEEE Conferences on Computer Vision and Pattern Recognition [3].

## REFERENCES

- [1] M.J. Black and P. Anandan, "The Robust Estimation of Multiple Motions: Parametric and Piecewise-Smooth Flow Fields," *Computer Vision and Image Understanding*, vol. 63, no. 1, pp. 75-104, Jan. 1996.
- [2] M.J. Black and A. Rangarajan, "On the Unification of Line Processes, Outlier Rejection, and Robust Statistics with Applications in Early Vision," *Int'l J. Computer Vision*, vol. 19, no. 1, pp. 57-91, July 1996.
- [3] C.F. Olson, "Maximum-Likelihood Template Matching," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, vol. 2, pp. 52-57, 2000.
- [4] D.P. Huttenlocher, G.A. Klanderman, and W.J. Rucklidge, "Comparing Images Using the Hausdorff Distance," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 15, no. 9, pp. 850-863, Sept. 1993.
- [5] D.P. Huttenlocher, R.H. Lilien, and C.F. Olson, "View-Based Recognition Using an Eigenspace Approximation to the Hausdorff Measure," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 21, no. 9, pp. 951-955, Sept. 1999.
- [6] C.F. Olson, "Probabilistic Self-Localization for Mobile Robots," *IEEE Trans. Robotics and Automation*, vol. 16, no. 1, pp. 55-66, Feb. 2000.
- [7] A. Rosenfeld and J. Pfaltz, "Sequential Operations in Digital Picture Processing," *J. ACM*, vol. 13, pp. 471-494, 1966.
- [8] G. Borgefors, "Distance Transformations in Digital Images," *Computer Vision, Graphics, and Image Processing*, vol. 34, pp. 344-371, 1986.
- [9] D.W. Paglieroni, "Distance Transforms: Properties and Machine Vision Applications," *CVGIP: Graphical Models and Image Processing*, vol. 54, no. 1, pp. 56-74, Jan. 1992.
- [10] D.P. Huttenlocher, "Using Two-Dimensional Models to Interact with the Three-Dimensional World," *Proc. Int'l NSF-ARPA Workshop Object Representation in Computer Vision*, pp. 109-124, 1995.
- [11] D.P. Huttenlocher and W.J. Rucklidge, "A Multi-Resolution Technique for Comparing Images Using the Hausdorff Distance," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 705-706, 1993.
- [12] W.J. Rucklidge, "Efficiently Locating Objects Using the Hausdorff Distance," *Int'l J. Computer Vision*, vol. 24, no. 3, pp. 251-270, Sept. 1997.
- [13] C.F. Olson, "Landmark Selection for Terrain Matching," *Proc. Int'l Conf. Robotics and Automation*, pp. 1447-1452, 2000.
- [14] M.J. Black, D.J. Fleet, and Y. Yacoob, "Robustly Estimating Changes in Image Appearance," *Computer Vision and Image Understanding*, vol. 78, no. 1, pp. 8-31, Apr. 2000.
- [15] P. Viola and W.M. Wells, "Alignment by Maximization of Mutual Information," *Int'l J. Computer Vision*, vol. 24, no. 2, pp. 137-154, 1997.
- [16] W.J. Rucklidge, "Efficient Guaranteed Search for Gray-Level Patterns," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, pp. 717-723, 1997.

► For more information on this or any other computing topic, please visit our Digital Library at <http://computer.org/publications/dlib>.