A New Bayesian Framework for Object Recognition

Yuri Boykov Daniel Huttenlocher

Computer Science Department
Cornell University
Ithaca, NY 14853
{yura,dph} @ cs.cornell.edu

Abstract

We describe a new approach to feature-based object recognition, using maximum a posteriori (MAP) estimation under a Markov random field (MRF) model. The main advantage of this approach is that it allows explicit modeling of dependencies between individual features of an object model. For instance, it can capture the fact that unmatched features due to partial occlusion are generally spatially coherent rather than independent. Efficient computation of the MAP estimate in our framework can be accomplished by finding a minimum cut on an appropriately defined graph. A special case of our framework yields even more efficient method, that does not use graph cuts. We call this technique spatially coherent matching. Our framework can also be seen as providing a probabilistic understanding of Hausdorff matching. We present ROC curves from Monte Carlo experiments that illustrate the improvement of the new spatially coherent matching technique over Hausdorff matching.

1 Introduction

In this paper we present a new Bayesian approach to object recognition using Markov random fields (MRF's). As with many approaches to recognition we assume that an object is modeled as a set of features. The recognition task is then to determine whether there is a match between some subset of these object features and features extracted from an observed image. The central idea underlying our approach is to explicitly capture dependencies between individual features of the object model. Markov random fields provide a good theoretical framework for representing dependencies between features. Moreover, recent algorithmic developments make it quite practical to compute the maximum a posteriori (MAP) estimate for the MRF model that we employ (e.g., [1], [3]).

Our approach contrasts with most feature-based object recognition techniques, as they do not explicitly account for dependencies between features of the object. It is desirable to be able to account for such dependencies, because they occur in real imaging situations. For example, a common case occurs with partial occlusion of objects, where features that are near one another in the image are likely to be occluded together. In our model, we assume that the process of matching individual object features is described *a priori* by a Gibbs distribution associated with a certain Markov random field. This model captures pairwise dependencies between features of the object. We then use *maximum a posteriori* (MAP) estimation to find the match between the object and the scene or to show that there is no such match. While a number of probabilistic approaches to recognition have been reported in the literature (e.g., [8], [7],[10]) these methods do not provide an explicit model of dependencies between features.

We show that finding the best match using the Hausdorff fraction [4], [9] is a special case of our technique, where features in the object model are independent. Therefore, our Bayesian framework can be seen as providing a probabilistic understanding of Hausdorff matching. With this view of Hausdorff matching, it becomes apparent that one of the main limitations of the Hausdorff approach is its failure to take into account the continuity of matches between neighboring features. That is, the Hausdorff approach does not account for the fact that features in a local neighborhood tend to be correlated. From our framework we derive a modification to Hausdorff approach which we call *spatially coherent matching* (SCM). This method requires matching features to be coherent in a given neighborhood system of the model. We present some Monte Carlo experiments demonstrating that this spatially coherent matching measure is a substantial improvement over Hausdorff matching in the case that images are cluttered with many irrelevant features and have substantial occlusion of the object to be recognized.

2 The General MAP-MRF Recognition Framework

In this section we describe our object matching framework in more detail. We represent an object by a set of features, indexed by integers in the set $M = \{1, 2, \ldots, m\}$. Each feature corresponds to some vector M_i in a feature space of the model. Commonly the vectors M_i will simply specify a feature location (x, y) in a fixed coordinate system of the model, although more complex feature spaces fit within the framework.

A given image I is a set of observed features from some underlying true scene. Each feature $i \in I$ corresponds to a vector I_i in a feature space of the image. The true scene can be thought of as some unknown set of features I^T in the same feature space. Similarly, I_i^T is a vector describing the feature $i \in I^T$ in the feature space of the image. We are interested in finding a match between the model M and the true scene I^T, using the observed features I.

A match of the model M to the true scene I^T is described by a pair $\{S, L\}$ where $S = \{S_1, S_2, \ldots, S_m\}$ is a collection of boolean variables and L is a location parameter. If $S_i = 1$ then the ith feature of the model has a matching feature in I^T and if $S_i = 0$ then it does not. In this case we say it is mismatched. For example, the event $\{S_1 = \ldots = S_k = 1,\ S_{k+1} = \ldots = S_m = 0,\ L = l\}$ implies that for $1 \leq i \leq k$, feature i of M has a matching feature $j \in I^T$, such that $I_j^T = M_i \oplus L$. Moreover, the last $(m-k)$ features are mismatched, meaning they have no such matching features. The operation $\oplus$ depends on the type of mapping from the model to the image feature space, which varies for the particular recognition task. In this paper we will use translation (vector summation), but other transformations are possible.

To determine the values of $\{S, L\}$ we use the maximum a posteriori (MAP) estimate

$$\{S^*, L^*\} \;=\; \arg\max_{S,L} \Pr(S, L|I).$$

Bayes rule then implies

$$\{S^*, L^*\} = \arg\max_{S,L} \Pr(I|S, L)\Pr(S)\Pr(L) \qquad (1)$$

assuming that S and L are a priori independent. The prior distributions $\Pr(S)$ and $\Pr(L)$ are discussed in section 2.1. We assume that the prior distribution of configuration S is described by a certain Markov random field, thus allowing for spatial dependencies among the S_i. The likelihood function $\Pr(I|S, L)$ is discussed in section 2.2.

Let $\mathcal{L}$ denote a set of possible locations of the model in the true scene. Then the range of the location parameter L is $\mathcal{L} \cup \emptyset$ where the extra value $\emptyset$ implies that the model is not in the scene. The basic idea of our recognition framework is to report a match between the model and the observed scene if and only if

$$S^* \neq \bar{0} \quad \text{and} \quad L^* \neq \emptyset. \qquad (2)$$

In section 2.3 we develop the test in (2) for the model specified in 2.1 and 2.2.

2.1 Prior Knowledge

We assume that the prior distribution of the location parameter L can be described as

$$\Pr(L) \;=\; (1-\rho)\cdot f(L) \;+\; \rho \cdot \delta(L = \emptyset) \qquad (3)$$

where $f(L) = \Pr(L|L \in \mathcal{L})$, the parameter ρ is the prior probability that the model is not present in the scene, and $\delta(\cdot)$ equals 1 or 0 depending on whether condition "$\cdot$" is true or false. Generally the distribution function $f(L)$ is uniform over $\mathcal{L}$. However in some applications $f(L)$ can reflect additional information about the model's location. For example, such information might be available in object tracking since the current location of the model can be estimated from previous iterations. The value of the constant ρ may be anywhere in the range $[0, 1)$. In section 2.3 we will see that ρ appears in our recognition technique only as a threshold for deciding whether or not the model is present given the image.

We assume that the collection of boolean variables, S, indicating the presence or absence of each feature, forms a Markov random field independent of L. More specifically, the prior distribution of S is described by the Gibbs[1] distribution

$$\Pr\{S\} \;\propto\; \exp\left\{-\sum_{i\in M} \alpha\cdot(1-S_i) \;-\; \sum_{\{i,j\}} \beta_{\{i,j\}}\cdot\delta(S_i \neq S_j)\right\} \qquad (4)$$

where the second summation is over all distinct unordered pairs of model features.

The motivation for this model is that $\Pr(S)$ captures the probability that features will not be matched even though they are present in the true scene, given some fixed location, L. Such non-matches could be due to occlusion, feature extraction error, or other causes. The parameter $\alpha \geq 0$ is a penalty for such non-matching features. The coefficient $\beta_{\{i,j\}} \geq 0$ specifies a strength of interaction between model features

[1] See [6] for more details on Gibbs distribution.

i and j. For tractability, we consider only pairwise interaction between features. Nevertheless, the pairwise interaction model provided by this form of Gibbs distribution is rich enough to capture one important intuitive property: a priori it is less likely that a feature will be un-matched if other features of the model have a match. Note that if all $\beta_{\{i,j\}} = 0$ then there is no interaction between the features and the S_i's become independent Bernoulli variables with probability of success $\Pr(S_i = 1) = e^{\alpha}/(1 + e^{\alpha}) \geq 0.5$.

2.2 Likelihood Function

The features of the observed image I may appear differently from the features of the unknown true scene I^T due to a number of factors. This includes sensor noise, errors of feature extraction algorithms (e.g. edge detection), and others. It is the purpose of the likelihood function to describe these differences in probabilistic terms.

We assume that the likelihood function is given by

$$\Pr(I|S, L) \propto \prod_{i \in M} g_i(I|S_i, L) \tag{5}$$

where $g_i(\cdot)$ is a likelihood function corresponding to the ith feature of the model. If $S_i = 0$ or $L = \emptyset$ then $g_i(I|S_i, L)$ is the likelihood of I given that the true scene does not contain the ith feature of the model. We assume that all cases of mismatching feature have the same likelihood. That is, for any $i \in M$ and $L \in \mathcal{L}$

$$g_i(I|1, \emptyset) = g_i(I|0, \emptyset) = g_i(I|0, L) = C_0 \tag{6}$$

where C_0 is a positive constant.

If $L \in \mathcal{L}$ then $g_i(I|1, L)$ is the likelihood of observing image I given that the i-th feature of the model is at location $(L \oplus M_i)$ in the feature space of the true scene I^T. The choice of $g_i(I|1, L)$ for $L \in \mathcal{L}$ will depend on the particular application.

Example 1. (Recognition based on edges) Consider an edge-based object matching problem, where all features of the model are edge pixels. We observe a set of image features I obtained by an intensity edge detection algorithm. One reasonable choice of $g_i(I|1, L)$ for $L \in \mathcal{L}$ is

$$g_i(I|1, L) = C_1 \cdot g(d_I(L \oplus M_i)) \tag{7}$$

where $d_I(\cdot)$ is a distance transform of the image features I. That is, the value of $d_I(p)$ is the distance from p to the nearest feature in I. The function $g(\cdot)$ is some probability distribution that is a function of the distance to the nearest feature. Normally, g is a distribution concentrated around zero. The underlying intuition is that if the true scene I^T has an edge feature located at $(L \oplus M_i)$ then the observed image I should contain an edge nearby. Thus the distance transform $d_I(L \oplus M_i)$ will be small with large probability. A number of existing feature based recognition schemes use functions of this form, including Hausdorff matching [4].

2.3 MAP Estimation

By substituting (3), (4), (5) into (1) and then taking the negative logarithm of the obtained equation we can show that MAP estimates $\{S^*, L^*\}$ minimize the value of the posterior energy function

$$E(S, L) = \begin{cases} H_L(S) - \ln f(L) - \ln(1 - \rho) & \text{if } L \in \mathcal{L} \\ H_L(S) \qquad\qquad - \ln \rho & \text{if } L = \emptyset \end{cases}$$

where

$$\begin{aligned} H_L(S) &= \sum_{\{i,j\}} \beta_{\{i,j\}} \cdot \delta(S_i \neq S_j) \\ &+ \sum_{i \in M} \left(\alpha \cdot (1 - S_i) - \ln g_i(I|S_i, L)\right). \end{aligned} \tag{8}$$

Our goal is to find $\{S^*, L^*\}$. The main technical difficulty is to determine $\{\hat{S}, \hat{L}\}$ that minimize $H_L(S) - \ln f(L)$ for $L \in \mathcal{L}$. In general this can be done using graph cut techniques[2] developed in [1] and [3]. In section 3 we consider some special cases where no sophisticated algorithmic scheme is needed. For the moment assume that $\{\hat{S}, \hat{L}\}$ are given.

Consider $H_L(S)$ for $L = \emptyset$. Equation (6) implies that $H_\emptyset(S)$ is minimized by the configuration $S = \bar{1}$ where all $S_i = 1$. If $E(\hat{S}, \hat{L}) > E(\bar{1}, \emptyset)$ then $\{S^*, L^*\} = \{\bar{1}, \emptyset\}$. According to (2), in this case we report that the model is not recognized in the scene. If $E(\hat{S}, \hat{L}) \leq E(\bar{1}, \emptyset)$ then $\{S^*, L^*\} = \{\hat{S}, \hat{L}\}$. In this case $L^* \in \mathcal{L}$. Nevertheless, if $\hat{S} = \bar{0}$ we would still report the absence of the model in the scene.

Finally, our recognition framework can be summarized as follows. The match between the model and the observed scene is reported if and only if $\hat{S} \neq \bar{0}$ and

$$H_{\hat{L}}(\hat{S}) - \ln f(\hat{L}) \leq m \cdot \ln \frac{1}{C_0} + \ln \frac{1 - \rho}{\rho} \tag{9}$$

where (9) is derived from the inequality $E(\hat{S}, \hat{L}) \leq E(\bar{1}, \emptyset)$. The right hand side in (9) is a constant that represents a certain decision threshold. Note that this decision threshold depends on two things: first, the prior probability of occlusion, ρ; and second, the product of the number of model features, m, with the log-likelihood of a mismatch, C_0.

[2] More details about computing $\{\hat{S}, \hat{L}\}$ in the general case can be found in [2].

3 Spatially Coherent Matching

In this section we consider models where certain pairs of features can be viewed as local neighbors. One simple kind of model with a natural local neighborhood system is successive points in an edge chain, as illustrated in Figure 1. In Section 3.1 we introduce a simple matching technique that captures dependencies between features in a local neighborhood. We call this method *spatially coherent matching* (SCM) because it takes into account the fact that feature mismatches generally occur in coherent groups (e.g., due to partial occlusion of an object).

In fact, SCM is a special case of our general result in Section 2. The reduction is shown in Section 3.2. SCM technique identifies some interesting properties of our general recognition framework. SCM technique can also be seen as a natural generalization of the Hausdorff matching. Section 3.3 shows how Hausdorff matching relates both to SCM technique and to our general framework.

3.1 SCM Algorithm

Both Hausdorff matching and SCM consider model features that are within some distance r of the nearest image feature. Let $M_L = \{i \in M : d_I(L \oplus M_i) \leq r\}$ denote the subset of model features lying within distance r of image features, when the model is positioned at L. We think of M_L as a set of *matchable* model features for a given location L. In addition, we define a subset of *unmatchable* model features $U_L = \{i \in M \mid d_I(L \oplus M_i) > r\}$ that also corresponds to a fixed location L. The set U_L consists of model features that are greater than distance r from any image features. Note that $U_L = M - M_L$.

The main idea of the SCM scheme is to require that matching features should form large connected groups. There should be no isolated matches. Let $B_L \subset M_L$ denote the subset of features in M_L that are "near" features of U_L. That is, $B_L = \{i \in M_L \mid u_L(i) \leq R\}$, where R is a fixed integer parameter and $u_L(i)$ is a distance[3] from i to the set U_L. We will refer to B_L as a *boundary* of the set of matchable features M_L. In the example of Figure 1 the boundary features B_L are shown in gray color.

The locally coherent matching technique works as follows. The main task is to find

$$L_{scm} \;=\; \arg\max_{L \in \mathcal{L}} \left(|M_L| - |B_L| + \frac{\ln f(L)}{\lambda} \right)$$

[3] In Section 3.2 we assume that $u_L(i)$ is the number of chains in the shortest sequence $\{i, i_1\}, \{i_1, i_2\}, \ldots, \{i_{k-1}, u\}$ of neighboring features that connect $i \in M_L$ to some unmatchable feature $u \in U_L$. In practice, L_1 or L_2 distances may be used.

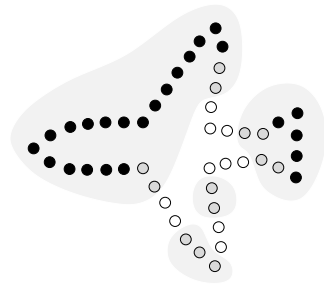


Figure 1: The pairs of neighboring features are connected by edges. The features of M_L (for some fixed L) are highlighted by shading. The unmatchable features U_L are white. The boundary features B_L for $R = 2$ are shown in gray. The non-boundary features, that is the elements of the set $M_L - B_L$, are black.

where $\lambda \geq 0$ is some constant. Note that $|M_L| - |B_L|$ is the number of non-boundary features in M_L. Thus, SCM seeks a location in the image where matchable features form large coherent groups. As illustrated in Figure 1, isolated matches are disregarded since they lie completely inside the boundary. The prior distribution $f(L)$ is also taken into consideration. The SCM technique matches the model to the image at the location L_{scm} if

$$|M_{L_{scm}}| - |B_{L_{scm}}| + \frac{\ln f(L_{scm})}{\lambda} \;>\; K \qquad (10)$$

where K is a decision threshold. Efficient implementation of SCM algorithm is discussed in Section 4.1.

3.2 Derivation of SCM

The SCM technique can be derived analytically from the results of Section 2. In fact, SCM is an optimal solution for a certain class of models where features interact only in a local neighborhood. In this section we discuss the corresponding special case of our general framework. The method of section 2 requires minimization of the function $H_L(S) - \ln f(L)$ where $f(L)$ is a prior distribution of possible locations and $H_L(S)$ is defined in (8). The following assumptions specify our particular choice of $H_L(S)$.

Let $\mathcal{N}_M$ denote a set of all pairs of neighboring features for a given object M. We assume that $\beta_{\{i,j\}} = \beta$ if the features $\{i, j\} \in \mathcal{N}_M$ are neighbors and $\beta_{\{i,j\}} = 0$ if the features $\{i, j\} \notin \mathcal{N}_M$ are not neighbors. The nonnegative constant β describes dependency between the neighboring features. Intuitively, it is reasonable to expect that neighboring features of the model are more likely to interact than a pair of features isolated from each other.

As in Example 1 we assume that $g_i(I|1, L) = C_1 \cdot g(d_I(L \oplus M_i))$, and moreover we use the particular function,

$$g(d) \;=\; \begin{cases} \frac{1}{r} & \text{if} \quad d \le r \\ 0 & \text{if} \quad d > r \end{cases}$$

where r is the distance to the nearest model feature used in the definition of matchable features M_L. In fact, this likelihood function prohibits assigning matches to features not in M_L.

Now all terms in (8) are specified. The next step is to minimize $H_L(S)$ for a fixed location L. Theorem 1 provides the necessary technical result. It works under the assumptions stated above. In addition, we consider $\lambda = \alpha + \ln \frac{C_1}{rC_0}$.

Theorem 1 *If the neighborhood system $\mathcal{N}_M$ forms a chain and the level of interaction between the neighboring features is $\beta = R \cdot \lambda$ then*

$$\min_S \; H_L(S) \;=\; m \cdot (\alpha - \ln C_0) - \lambda \cdot (|M_L| - |B_L|)$$

and the optimal $S \neq \bar{0}$ iff $|M_L| > |B_L|$. ■

Due to space limitations we do not give the proof of this theorem here. Recall that our final goal is to minimize $H_L(S) - \ln f(L)$ for $L \in \mathcal{L}$. As follows from Theorem 1, the optimum is achieved at the location

$$\hat{L} \;=\; \arg\min_{L \in \mathcal{L}} \; \Big(-\lambda \cdot (|M_L| - |B_L|) - \ln f(L)\Big).$$

Obviously, $\hat{L} = L_{scm}$. The corresponding optimal value $H_{\hat{L}}(\hat{S}) - \ln f(\hat{L})$ equals

$$m \cdot (\alpha - \ln C_0) - \lambda \cdot (|M_{L_{scm}}| - |B_{L_{scm}}|) - \ln f(L_{scm}).$$

Substituting this into (9) gives (10) with

$$K \;=\; \frac{1}{\lambda} \cdot \left(m\alpha - \ln \frac{1-\rho}{\rho}\right).$$

3.3 Relation to Hausdorff Matching

The classical Hausdorff distance is a max-min measure for comparing two sets for which there is some underlying distance function on pairs of elements, one from each set. The application of Hausdorff matching in computer vision has used a generalization of this classical measure [4], based on computing a quantile rather than maximum of distances.

One form of the generalized Hausdorff measure counts the number of matchable features, $|M_L|$, when the model is positioned at L. The model is matched at the location $L_h = \arg\max_{L \in \mathcal{L}} |M_L|$ if and only if the number of matched features, $|M_{L_h}|$, is larger than some critical fraction of the total number of model features, m.

SCM reduces to Hausdorff matching if $R = 0$ and $f(L) = const$. In fact, $R = 0$ implies that the boundary B_L of the set of matchable features is always empty. Then

$$L_{scm} \;=\; \arg\max_{L \in \mathcal{L}} \left(|M_L| - 0 + \frac{const}{\lambda}\right) \;=\; L_h$$

and the test in (10) reduces to $|M_{L_h}| \ge K'$ which is exactly the Hausdorff test described above. As follows from Theorem 1, $R = 0$ corresponds to $\beta = 0$. Therefore, Hausdorff matching is a special case of our general framework when the features are independent.

SCM technique generalizes Hausdorff matching in an interesting way. Note that the size of the boundary $|B_L|$ is small if the features in M_L are grouped in large connected blobs and $|B_L|$ is large if the matchable features are isolated from each other. Therefore, SCM technique is reluctant to match if the features in M_L are scattered in small groups even if the size of M_L is large. In contrast, the Hausdorff matching cares only about the size of M_L and ignores connectedness. Besides, SCM technique naturally incorporates prior knowledge represented by the distribution $f(L)$.

4 Experimental Results

In order to evaluate the recognition measures developed in this paper, we have run a series of experiments using Monte Carlo techniques to estimate Receiver Operating Characteristic (ROC) curves for each measure. A ROC curve plots the probability of detection along the y-axis and the probability of false alarm along the x-axis. Thus, the ideal recognition algorithms would produce results near the top left of the graph (low false alarm and high detection probabilities).

We use the experimental procedure reported in [5], where it was shown that Hausdorff matching works better than a number of previous binary image matching methods including correlation and Chamfer matching. For that reason we are mainly interested in comparing the algorithms developed here with Hausdorff matching, because it has already been shown to have better performance than these other techniques. Thus we contrast Hausdorff matching with the SCM technique. In Section 4.1 we explain some extra details about implementing SCM technique. In 4.2 we discuss the Monte Carlo technique used to estimate the ROC curves and present the results.

a) An object b) A simulated image

Figure 2: The simulated image above contains 4% of clutter. The perturbed and partly occluded (30% occlusion) instance of the object is located in the center.

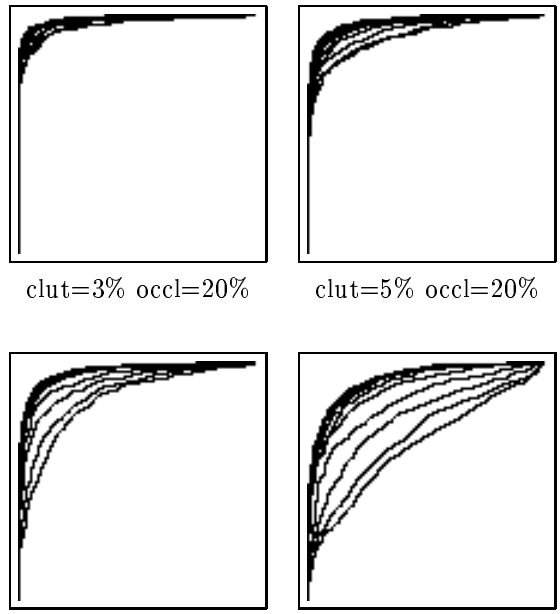
4.1 Implementation of SCM

In this section we provide some details of our implementation of the SCM technique from Section 3.1. The SCM technique is simple to implement using image morphology. Given the set of model features, M, and location, L, the set of matchable features M_L are those within distance r of image features. This can be computed by dilating the set of image features I by radius r (replacing each feature point with a disc of radius r). Now the set M_L is simply the intersection of M with this dilated image. The next step is to compute the boundary B_L which is the subset of features in M_L that are within distance R of some feature in U_L, the set of unmatchable features. Recall that $U_L = M - M_L$. Again, we can find features in one set near the features in some other set using dilation. Dilating the set U_L by R, and taking the intersection with M_L yields B_L, the points of M_L within distance R of points in U_L.

The quality of the match produced by the SCM technique at each location L is determined by the number of non-boundary matchable features, that is, by $|M_L| - |B_L|$. Note that the search for the best match over all values of $L \in \mathcal{L}$ can be accelerated using the same pruning techniques that were developed for the Hausdorff measure [9]. This follows from a simple fact that if the Hausdorff measure gives no match at L then the spatially coherent matching technique can not match at L either. It is easy to see that $|M_L| < K$ implies that the test in (10) is necessarily false.

4.2 ROC Curves

We have estimated ROC curves by performing matching in synthetic images and using the matches found in these images to estimate the curve over a range of possible parameter settings. 1000 test images were used in the experiments, and were generated according to the following procedure. Random chains of edge pixels with a uniform distribution of lengths between 20 and 60 pixels were generated in a 150×150 image until a predetermined fraction of the image was covered with such chains. Curved chains were generated by changing the orientation of the chain at each pixel by a value selected from a uniform distribution between $-\frac{\pi}{8}$ and $+\frac{\pi}{8}$. An instance of the object was then placed in the image, after rotating, scaling, and translating the object by random values. The scale change was limited to $\pm 10\%$ and the rotation change was limited to $\pm\frac{\pi}{18}$. Occlusion was simulated by erasing the pixels corresponding to a connected chain of the model image pixels. Gaussian noise was added to the locations of the model image pixels ($\sigma = 0.25$). The pixel coordinates were finally rounded to the closest integer. This procedure was also used in [5].



clut=3% occl=20% clut=5% occl=20%

clut=3% occl=40% clut=5% occl=40%

Figure 3: ROC curves.

For the experiments reported here, we performed recognition using the 56×34 object shown in Figure 2(a). This object contains 126 edge features. An example of a synthetic image generated using this object and the procedure described above is shown in Figure 2(b). In each trial, a given matching measure with a given parameter value was used to find all the matches of the object to the image. A trial was said

to find the correct object if the position (considering only translation) of one of the matches was within three pixels of the correct location of the object in the image. A trial was said to find a false positive if any match was found outside of this range (and that match was not contiguous with a correct match position). Thus note that the test images were formed by slight rotation and scaling of the object model, but the searched was only done under translation. Any non-translational change to the object was not modeled by the matching process.

Figure 3 shows the ROC curves corresponding to experiments with different levels of occlusion and image clutter. For these tests we assumed that all locations in the image are a priori equaly likely, that is, $f(L) = const$. The black curve shows the best results we could obtain from the general method of Section 2 where we applied the graph-cut techniques explained in [2]. The gray curves correspond to the SCM technique for various values of $R \in [0, 25]$. As R gets larger, up to 20 or 21, the results improve, so the curves closer to the top left are for larger values of R. For even larger values of R, which we do not show, the ROC curves rapidly deteriorate. It is interesting to note that given this particular object, a distance of $R = 25$ corresponds approximately to the height of the object. Thus the performance does not deteriorate until the coherence region begins connecting together disconnected pieces of the object.

The case of $R = 0$ corresponds to Hausdorff matching. Thus the spatial coherence approach plays a large role in improving the quality of the match, because $R = 0$ has the worst matching performance. Note that in [5], using the same Monte Carlo framework, it was shown that Hausdorff matching works better than a number of other methods including binary correlation and Chamfer matching. Thus these results indicate that SCM is a substantial improvement over several commonly used binary image matching techniques.

It should be noted that the value of R does not make a big difference for lower clutter or occlusion cases (top row of the figure), but makes a very large difference when these are larger (bottom row of the figure). Thus we see that for "easy" recognition problems, the spatial coherence of the matches is less important (though still offers a slight improvement). However as the object becomes more occluded and as there are more distractors, it becomes quite important to consider the spatial coherence of the matches. It should also be noted that in real imaging situations there would likely be small gaps in the instance of an object for which it would be undesirable that the SCM technique penalize such gaps. Recall that the parameter r can be used to cause features of the object model to match across small gaps in the image. Any larger gaps would then be subject to penalty based on the value of R.

References

[1] Y. Boykov, O. Veksler, and R. Zabih. Markov random fields with efficient approximations. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 648–655, 1998.

[2] Yuri Boykov and Daniel P. Huttenlocher. A new bayesian framework for object recognition. *Technical Report*, ncstrl.cornell/TR98-1713, 1998.

[3] D. Greig, B. Porteous, and A. Seheult. Exact maximum a posteriori estimation for binary images. *Journal of the Royal Statistical Society, Series B*, 51(2):271–279, 1989.

[4] D. P. Huttenlocher, G. A. Klanderman, and W. J. Rucklidge. Comparing images using the Hausdorf distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(9):850–863, September 1993.

[5] Daniel P. Huttenlocher. Monte carlo comparison of distance transform based matching measures. In *DARPA Image Understanding Workshop*, 1997.

[6] S. Z. Li. *Markov Random Field Modeling in Computer Vision*. Springer-Verlag, 1995.

[7] Clark F. Olson. A probabilistic formulation for Hausdorff matching. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 150–156, 1998.

[8] Arthur Pope and David G. Lowe. Learning probabilistic appearance models for object recognition. In Shree K. Nayar and Tomaso Poggio, editors, *Early Visual Learning*, pages 67–98. Oxford University Press, 1996.

[9] William Rucklidge. *Efficient Visual Recognition Using the Hausdorff Distance*. Number 1173 in Lecture Notes in Computer Vision. Springer-Verlag, 1996.

[10] Jayashree Subrahmonia, David B. Cooper, and Daniel Keren. Practical reliable bayesian recognition of 2D and 3D objects using implicit polynomials and algebraic invariants. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(5):505–519, May 1996.