Sampling of Images for Efficient Model-Based Vision

Mohamad Akra, Louay Bazzi, and Sanjoy Mitter, *Fellow*, *IEEE*

Abstract—The problem of matching two planar sets of points in the presence of geometric uncertainty has important applications in pattern recognition, image understanding, and robotics. The first set of points corresponds to the "template." The other set corresponds to the "image" that—possibly—contains one or more deformed versions of the "template" embedded in a cluttered image. Significant progress has been made on this problem and various polynomial-time algorithms have been proposed. In this article, we show how to sample the "image" in linear time, reducing the number of foreground points n by a factor of two-six (for commonly occurring images) without degrading the quality of the matching results. The direct consequence is a time-saving by a factor of 2^p–6^p for an $O(n^p)$ matching algorithm. Our result applies to a fairly large class of available matching algorithms.

Index Terms—Sampling, model-based vision, matching under uncertainty, approximate matching, image understanding.

1 Introduction

The problem of approximate matching of two sets of points T and S has been addressed by researchers from various fields including Pattern recognition, Image Understanding, Computational Geometry, and Robotics. T may represent an ideal 2D or 3D model (template) of an object which one is trying to detect in a 2D image S. Due to sensory errors, T is never exactly replicated in S. Instead, points of T are disturbed by some local perturbation. Furthermore, T may be subject to some kind of deformation such as scaling, rotation, translation, or affine transformation. In general, we denote by $\mathcal{A}^*$ the space of such allowable transformations. Two related approaches of the approximate matching problem have been reported in the literature. The first is to characterize the mappings from T to (subsets of) S that are close to elements in $\mathcal{A}^*$, also called the correspondence space approach. The other approach is to characterize the transformations in $\mathcal{A}^*$ that are close to mappings from T to (subsets of) S, also called the transformation space approach.

Several polynomial-time algorithms have been proposed, handling various cases of $\mathcal{A}^*$. In what follows, we briefly survey the prior work denoting the size of the template $T \times m$ and the size of the image $S \times n$. We present *some* of the research efforts that we feel are related to our work. In particular, we will be concerned with articles that do not pose the restriction that $m = n$ (e.g., Baird [1]), and that use the locations of the foreground points as features, rather than using line segments lengths or orientations (e.g., Yi [10]). For every research work we cite, we specify whether the models used are 2D or 3D, whether the correspondence space version or the transformation space version is considered, the nature of the bounded error model, the space of allowable transformations, and the reported running time complexity.

Cass [2] considered the correspondence space version in the case of 2D models, where $\mathcal{A}^*$ includes translation and rotation. He showed that approximate matching can be done with a crude upper-bound complexity of $O(m^6n^6)$, thereby refuting a long pending claim that this problem is exponential. Basing his analysis on a circular error model, Cass enumerated all the maximal geometrically consistent match sets.

Huttenlocher [7] (see also [9]) addressed the transformation space version for 2D models when $\mathcal{A}^*$ includes only translation and when the error model is square. He was able to do matching in $O(m^2n^2\alpha(mn))$ time.

Breuel [6] enlarged the space of allowable transformations considered in [2] to include scaling. Looking at a convex polygonal error model (of which the square error model is a special case), Breuel provided an $O(m^4n^4)$ average case algorithm for approximate matching. Breuel's approach belongs to the transformation-space class of algorithms.

Later, Cass [4] considered a transformation-version approach to match 3D templates in the case of a square error model and $\mathcal{A}^*$ consisting of rigid motion and scaled orthographic projection. He managed to do matching in $O(m^4n^4\log n)$ time. Notably, his algorithm generates some nonfeasible solutions in addition to all the feasible ones.

But matching is really recognition in the geometric sense, where the features are the locations of the foreground points. In recognition problems, it is customary to reduce the complexity by reducing the number of features. Consequently, it is natural to ask whether or not it is possible to disregard some foreground points (thereby reducing n) from an image S without affecting the matching results.

This paper answers this question in the affirmative. It is possible to replace an image by a subset of it, whose union of error regions is the same as that of the original image, without disqualifying the matching results. This replace-

• *The authors are with the Electrical Engineering and Computer Science Department, Massachusetts Institute of Technology, 77 Massachusetts Ave., Rm. 35-312, Cambridge, MA 02139.*
E-mail: {makra; louay}@mit.edu; mitter@lids.mit.edu.

Manuscript received 4 Sept. 1997; revised 4 Nov. 1998. Recommended for acceptance by P. Flynn.
For information on obtaining reprints of this article, please send e-mail to: tpami@computer.org, and reference IEEECS Log Number 108075.

0162-8828/99/$10.00 © 1999 IEEE

ment can be done in time linear in the number of pixels in the image, thereby throwing a considerable fraction of the foreground points. Experimentally, a reduction in n between two and six was recorded, with an implied time savings between 2^p and 6^p for an $O(n^p)$ matching algorithm. We provide in the Appendix a list of figures showing edge-detected images that typically arise in model-based vision. We illustrate how sampling those images results in a noticeable reduction of foreground points.

In the rest of the article, we formulate the recognition problem in a framework generic enough to include all of the above cited work. We then identify the conditions that a sampling process has to satisfy in order to be of use in the above framework. Based on these conditions, we design a sampling algorithm. We finally conclude by reporting a set of illustrative experiments.

2 Recognition Preliminaries

Let $\mathcal{X}$ be the space whose subsets are the templates, typically $\mathbb{R}^2$ for 2D models or $\mathbb{R}^3$ for 3D models. Let $\mathcal{Y}$ be the space whose subsets are the images, usually $\mathbb{Z}^2$ when the image is a raster image, or—less commonly—$\mathbb{R}^2$, and let d be a metric over $\mathcal{Y}$. In the context of the cited research, d is the L_2 metric for the case of a circular error model, while it is the L_∞ metric for the case of a square error model.

Let $\mathcal{A}^*$ be the space of allowable transformations, a subset of the set of all the mappings from $\mathcal{X}$ to $\mathcal{Y}$. A mapping f from a template T to $\mathcal{Y}$ is called a *transformation* of T. We denote by $\mathcal{D}(T)$ the space of all the transformations of T and by ρ_T a mapping from $\mathcal{D}(T) \times \mathcal{D}(T)$ to $\mathbb{R}$ satisfying

$$\rho_T(f_1, f_2) = \max_{\alpha \in T} d(f_1(\alpha), f_2(\alpha)).$$

It is straightforward to show that ρ_T is a metric norm over $\mathcal{D}(T)$.

Note that ρ_T is defined on $\mathcal{D}(T) \times \mathcal{D}(T)$. Therefore, when we need to measure the distance between two functions f_1 and f_2 that may be defined on sets larger than T, we must restrict them to T. We will denote by $f | T$ the restriction of f to T.

We next define the set of all allowable transformations that are *close* to matchings. In the literature on model-based vision, those transformations are termed *feasible*.

Definition 1. *Let T be a template, S be an image, $\epsilon > 0$ be a bound on the noise, and $k < |T|$ be a bound on the number of unmatched points in T. We define $A_\epsilon^k(T, S)$, the set of feasible transformations, to be*

$$A_\epsilon^k(T, S) = \bigcup_{U \subset T / |U| \geq |T| - k} A_\epsilon(U, S),$$

where

$$A_\epsilon(U, S) = \{h \in \mathcal{A}^* / \exists f : U \rightarrow S \text{ where } \rho_U(f, h | U) < \epsilon\}.$$

Observe that the definition allows for noise, occlusion, and spurious features.

Now, we define the set of feasible matchings, in the literature on model-based vision, those matchings are termed *feasible* or *geometrically consistent*.

Definition 2. *Let T be a template, S be an image, and $\epsilon > 0$. We define $M_\epsilon(T, S)$, the set of feasible matchings, to be*

$$M_\epsilon(T, S) = f : T \rightarrow S / \exists h \in \mathcal{A}^* \text{ where } \rho_T(f, h | T) < \epsilon\}.$$

Note that this definition allows for noise and spurious features only. Extending the definition to allow for occlusion is simple but analytically messy, due to the fact that the resulting set will contain functions not sharing the same domain.

The set of feasible matchings may be exponential in size in terms of m, the number of points in the model. On the other hand, the set of feasible transformations is infinite. Rather than directly handling the complexity of these two sets, we turn our attention to representative subsets of them. Checking if a subset is representative may be performed by resorting to the concept of *covering*, which we define next.

Definition 3. *Let T be a template, A and B be subsets of $\mathcal{D}(T)$, and $\epsilon > 0$. A is said to be an ϵ-covering of B if $A \subset B$ and $\forall g \in B$, $\exists f \in A$ such that $\rho_T(f, g) < \epsilon$.*

In other words, the ϵ-neighborhoods of the elements in A cover B. Now we are in a position to formalize the objective of the approximate matching problem in a rather general framework.

Let T be template, S be a image, $k < |T|$ be the maximum number of points in T that are allowed to be unmatched to points in S, and ϵ, $\delta > 0$, where ϵ is a bound on the noise. The objective of the transformation-space version may be stated as: Compute C, a δ-covering of $A_\epsilon^k(T, S)$. In contrast, the objective of the correspondence-space version may be stated as: Compute C, a δ-covering of $M_\epsilon(T, S)$.

Note that we are not allowing for occlusion in the correspondence-space version. The motivation is notational and technical simplification.

Depending on the recognition algorithm, δ may be prespecified or imposed by the algorithm. But in all cases, δ should be relatively small. Note also that the circular error model corresponds to d being L_2, the square error model corresponds to d being L_∞, and the no-occlusion case corresponds to $k = 0$.

Under the assumption that the features of interest are points, the above settings are general enough to be tailored to various matching approaches. Yet, they are necessary and sufficient for algorithmic correctness as we argue next.

Specifically, we require C—the computed set of any matching algorithm—to form a δ-covering of the feasible transformation (matching) space, for some relatively small δ. Not fulfilling this requirement, the algorithm is necessarily leaving correct feasible transformations (matchings) that are not δ-close to any of the computed transformations (matchings), and hence not accounted for in the output.

On the other hand, it is sufficient for an algorithm to enumerate only a δ-covering of the feasible transformation (matching) space for some relatively small δ due to the following reasons. If one were using the correspondence space approach, it would be redundant to enumerate all of the potentially exponential feasible matchings as some of them may just correspond to the same transformation. Such a case occurs, for instance, when trying to match a template to an image containing a higher-density version of this template. Therefore, one is satisfied with a representative part; i.e., a subset that does not contain lots of matchings that are close to each other, yet guarantees that any matching in the original set is represented by (i.e., close to) a matching in the computed set. Formally, one would be satisfied by computing a *covering* of the set of feasible matchings. A similar "sufficiency" argument applies to the feasible transformation space, with the added justification that the space is infinite in size.

In the next section, we define the sampling criteria in a way that preserves the value of the matching algorithm output as long as the objective of the algorithm is consistent with the above formalization of the problem.

3 Sampling Criteria

In the following, we denote by $B_\epsilon(p)$ the ϵ-neighborhood of a point p in X (with respect to the metric d). For a set of points $U \subset \mathcal{Y}$, we extend the definition to $B_\epsilon(U)$, where

$$B_\epsilon(U) = \bigcup_{\alpha \in U} B_\epsilon(\alpha).$$

Definition 4. *Let S be an image and $\epsilon > 0$. An image P is said to be an ϵ-sampling of S if $P \subset S$ and $B_\epsilon(P) = B_\epsilon(S)$.*

Note that sampling differs from covering in two aspects. First, sampling uses = rather than $\subset$. Second, sampling is performed on sets of points, while covering is performed on sets of functions.

Recall that we defined the transformation version of the model-based vision problem as the task of covering the set of feasible transformations. We show below that if we appropriately sample the image, then a solution of the sampled version is also a solution of the original problem.

Theorem 1. *Let T be a template, S be an image, $k \le |T|$, and $\epsilon > 0$. If P is an ϵ-sampling of S, then $A_\epsilon^k(T,P) = A_\epsilon^k(T,S)$.*

By the sampling definition:

1) $P \subset S$.
2) $B_\epsilon(P) = B_\epsilon(S)$.

The fact that $P \subset S$ implies that $A_\epsilon^k(T,P) \subset A_\epsilon^k(T,S)$. So, we only have to show that $A_\epsilon^k(T,S) \subset A_\epsilon^k(T,P)$. Consider any $h \in A_\epsilon^k(T,S)$. By definition, there exists $U \subset T$ and $f : U \rightarrow S$, such that $|U| \ge |T| - k$ and $\rho_U(f, h \mid U) < \epsilon$. $\rho_U(f, h \mid U) < \epsilon$ means that $h(\beta) \in B_\epsilon(f(\beta))$ for each $\beta \in U$. Consider any $\beta \in U$, we have $h(\beta) \in B_\epsilon(f(\beta)) \subset B_\epsilon(S) = B_\epsilon(P)$. In other words, there exists a $\gamma \in P$, such that $d(h(\beta), \gamma) < \epsilon$. This is true for any $\beta \in U$, so there exists $g : U \rightarrow P$ such that $\rho_U(g, h \mid U) < \epsilon$. It follows that $h \in A_\epsilon^k(T,P)$.

Let us move to the correspondence version of the problem. Recall that we defined the correspondence version of the model-based vision problem as the task of covering the set of feasible matchings. In the following theorem, we show that if we appropriately sample the image, then an acceptable solution of the sampled version is also an acceptable solution of the original problem, albeit with a slightly different covering radius. The new covering radius is greater by 2ϵ, which is considered a slight difference since ϵ—the recognition error—is set to a small value. More formally, see Theorem 2.

Theorem 2. *Let T be a template, S be an image, $k \le |T|$, and $\epsilon > 0$. If P is an ϵ-sampling of S, then a δ-covering of $M_\epsilon(T, P)$ is a $(2\epsilon + \delta)$-covering of $M_\epsilon(T, S)$.*

Let A be a δ-covering of $M_\epsilon(T, P)$. We show that A is a $(2\epsilon + \delta)$-covering of $M_\epsilon(T, S)$. First, we have

$$A \subset M_\epsilon(T, P) \subset M_\epsilon(T, S),$$

where the first inclusion follows from the definition of A and the second from the fact that $P \subset S$. Hence, we only need to show that for each $g \in M_\epsilon(T, S)$, there exists $f \in A$, such that $\rho_T(f, g) < 2\epsilon_1 + \delta$.

Consider any element g of $M_\epsilon(T, S)$. By definition, there exists $t \in \mathcal{A}^*$, such that $\rho_T(g, t \mid T) < \epsilon$. This means that $t \in A_\epsilon^0(T,S)$. But, according to the previous theorem, $A_\epsilon^0(T,S) = A_\epsilon^0(T,P)$ because P is an ϵ-sampling of S. So, there exists $h : T \rightarrow S$ such that $\rho_T(h, t \mid T) < \epsilon$. This means that $h \in M_\epsilon(T, S)$. But, P is a δ-covering of $M_\epsilon(T, S)$. So, P must contain an element f with $\rho_T(f, h) < \delta$. Summing it up, we conclude that P contains an element f satisfying

$$\rho_T(f, g) \le \rho_T(f, h) + \rho_T(h, t \mid T) + \rho_T(g, t \mid T) < \delta + e + \epsilon,$$

which completes the proof.

To put things in perspective, we assumed that the features of interest are point sets. We considered two versions of the model-based vision problem, the transformation space version and the correspondence space version. We indicated that previously reported research work falls in either one of those two versions. We then formalized, in a rather general setting, the criteria of an acceptable solution. We argued that an acceptable solution should consist of a covering (whose radius is relatively small) of the space of feasible transformations (or matchings). Then, we defined sampling and showed that if it is used to reduce the size of the image the quality of the recognition output will not be affected. The solution that results from running the matching algorithm on the sampled image will either satisfy the same acceptance criteria as the results from running it on the original image (in the transformation space version) or a practically equivalent criteria (in the correspondence space version).

In the next section, we present a sampling algorithm. The ideal objective is to compute a sampling that is optimal in terms of the resulting number of points. To reduce the computational complexity of the problem, however, we will be satisfied with a locally optimal solution that we define next.

Definition 5. *Let S be an image and $\epsilon > 0$. P is said to be a locally optimal ϵ-sampling of S if P is an ϵ-sampling of S, and P is the only ϵ-sampling of P.*

4 Sampling Algorithm

In this section, we present a general version of the sampling algorithm that computes a locally optimal sampling of an image S and that applies to a large class of error models. Next, we turn our attention to the case of a square error model and conclude by adapting the sampling algorithm for the case of L_∞ metric and images that are subsets of $\mathbb{Z}^2$. The algorithm turns out to run in time linear in the number of pixels in the image.

THEOREM 3. *Let S be an image and $\epsilon > 0$, and consider the following algorithm.*

```
SAMPLE(S, ε)
1. P ← S
2. for all α ∈ S
3.     if  B_ε(α) ⊂ B_ε(P − {α})
4.         P ← P − {α}
5. return P
```

Then SAMPLE *computes P, a locally optimal ϵ-sampling of S.*

Let $\alpha_1, \alpha_2, \ldots, \alpha_n$ be the elements of S in the order considered by the loop in line 2. Let P_i be the value of P in line 3 for $\alpha = \alpha_i$, and let $P_{n+1} = P$. Clearly,

$$P = P_{n+1} \subset P_n \subset \ldots \subset P_i \subset \ldots \subset P_1 = S.$$

We use induction to prove that $B_\epsilon(P_i) = B_\epsilon(S)\ \forall i = 1 \ldots n + 1$.

Obviously, $B_\epsilon(P_1) = B_\epsilon(S)$. Now assume that $B_\epsilon(P_i) = B_\epsilon(S)$ for $i = k$. If $B_\epsilon(\alpha_i) \not\subset B_\epsilon(P_i - \{\alpha_i\})$, then $P_{i+1} = P_i$. Else if $B_\epsilon(\alpha_i) \subset B_\epsilon(P_i - \{\alpha_i\})$, then

$$\begin{aligned} B_\epsilon(P_{i+1}) &= B_\epsilon(P_i - \{\alpha_i\}) \\ &= B_\epsilon(P_i - \{\alpha_i\}) \cup B_\epsilon(\alpha_i) \\ &\quad (\text{because } B_\epsilon(\alpha_i) \subset B_\epsilon(P_i - \{\alpha_i\})) \\ &= B_\epsilon(P_i) \end{aligned}$$

It follows that $B_\epsilon(P_i) = B_\epsilon(S)$ for $i = k + 1$. So for $i = n + 1$, we obtain $B_\epsilon(P) = B_\epsilon(S)$. Since $P \subset S$, we conclude that P is an ϵ-sampling of S.

To show that P is locally optimal, assume that there exists P', a proper subset of P, that is also an ϵ-sampling of S. Let i be the index of an element α in $P - P'$. Then, according to Line 3, $B_\epsilon(\alpha) \not\subset B_\epsilon(P_i - \{\alpha\})$. But $P' \subset P_i - \{\alpha\}$, so $B_\epsilon(\alpha) \not\subset B_\epsilon(P')$. Hence, $B_\epsilon(P') \neq B_\epsilon(S)$. As a result, P is the only ϵ-sampling of P. In other words, P is a locally optimal ϵ-sampling of S.

From here onwards, we turn our attention to the case where $\mathcal{Y} = \mathbb{R}^2$ and $d = L_\infty$. Namely, for p and q in $\mathcal{Y}$

$$d(p, q) = max\{\,|p_x - q_x|,\ |p_y - q_y|\,\}.$$

This is the case corresponding to research efforts that use the square error model as in [4]. The following lemma is used in Theorem 5, which is our roadmap to move from the general version of the sampling algorithm to the more specific version handling the case of the L_∞ metric.

LEMMA 4. *Let S be a finite subset of $\mathbb{Z}^2$, $t \in \mathbb{Z}_+/2$, and $\eta \in (0, 0.5)$. Let $q \in \mathbb{R}^2$, such that $B_\eta(q) \subset B_t(S)$. Then there exists $\beta \in S$ such that $B_\eta(q) \subset B_t(\beta)$.*

See the Appendix for the proof.

THEOREM 5. *Let S be a finite subset of Z^2, $\alpha \in Z^2$, and $\epsilon > 0$. Then the following statements are equivalent,*

1) $B_{\lceil 2\epsilon \rceil/2}(\alpha) \subset B_{\lceil 2\epsilon \rceil/2}(S)$.

2) $B_\epsilon(\alpha) \subset B_\epsilon(S)$.

3) $B_\epsilon(\alpha) \cap (\mathbb{Z}/2)^2 \subset B_\epsilon(S) \cap (\mathbb{Z}/2)^2$.

See the Appendix for the proof. The previous theorem is used next to customize the sampling algorithm—presented in Theorem 3—for the L_∞ metric and images that are subsets of $\mathbb{Z}^2$. The customized version computes a locally optimal sampling in time linear in the number of pixels in the image.

THEOREM 6. *Let S be a finite subset of $\mathbb{Z}^2$. Let N and M be the respective height and width of the smallest enclosing rectangle of S. Let $\epsilon > 0$. Then, a locally optimal ϵ-sampling of S can be computed in $O(MN\lceil 2\epsilon \rceil^2)$.*

Without loss of generality, we may assume that the smallest enclosing rectangle of S is $[1, N + 1] \times [1, M + 1]$.

Let $\mathcal{S}[1 \ldots N + 1]\,[1 \ldots M + 1]$ be the binary matrix representing S, where $\mathcal{S}[m]\,[n] = 1$ if the point $m \times n \in S$ and $\mathcal{S}[m]\,[n] = 0$ elsewhere.

Consider the condition in Line 3 of SAMPLE in Theorem 3:

$$B_\epsilon(\alpha) \subset B_\epsilon(S - \{\alpha\}). \tag{1}$$

According to Theorem 5, we know that for $\beta \in \mathbb{Z}^2$, $B_\epsilon(\beta) \subset B_\epsilon(S)$ if and only if $B_\epsilon(\beta) \cap (\mathbb{Z}/2)^2 \subset B_\epsilon(S) \cap (\mathbb{Z}/2)^2$. So, (1) is equivalent to

$$B_\epsilon(\alpha) \cap (\mathbb{Z}/2)^2 \subset B_\epsilon(S - \{\alpha\}) \cap (\mathbb{Z}/2)^2. \tag{2}$$

Let $i = \alpha_x$, $j = \alpha_y$, and $e = (\lceil 2\epsilon \rceil - 1)/2$. Let $A[1 \ldots 2(M + 2e)]\,[1 \ldots 2(N + 2e)]$ be a matrix defined by

$$A[k][l] = \text{card}\{\alpha \in S \,/\, (k/2 - e) \times (l/2 - e) \in B_\epsilon(\alpha)\}.$$

Accordingly, (2) is equivalent to $A[k][l] > 1$, for $k = 2i - 1, \ldots, 2(i + e) - 1$, $l = 2j - 1$ to $2(j + e) - 1$. Therefore, the following realization of SAMPLE computes a locally optimal ϵ-sampling of S in $O(MN\lceil 2\epsilon \rceil^2)$.

```
SAMPLE(S, ε)
e ← (⌈2ε⌉ − 1)/2
P ← S
for i = 1 to M + 1, j = 1 to N + 1
    if 𝒮[i][j] = 1
        for k = 2i − 1 to 2(i + e) − 1,
            l = 2j − 1 to 2(j + e) − 1
                A[k][l] ← A[k][l] + 1
for i = 1 to M + 1, j = 1 to N + 1
    if 𝒮[i][j] = 1
        if A[k][l] > 1 for k = 2i − 1 to 2(i + e) − 1,
                          l = 2j − 1 to 2(j + e) − 1
            P[i][j] ← 0
            for k = 2i − 1 to 2(i + e) − 1,
                l = 2j − 1 to 2(j + e) − 1
                    A[k][l] ← A[k][l] − 1
return P
```

TABLE 1
SAMPLING OF THREE IMAGES: SAVINGS FOR DIFFERENT ϵ

ϵ	Kitchen	Chair	Telephone
1.5	2.46	2.23	2.07
2.0	3.47	3.13	2.69
3.0		5.78	

Note that all the entries of A are assumed to have been initially set to zero. When ϵ is given or assumed to be bounded (as is the case in model-based vision), the above running time becomes $O(MN)$. Consequently, the algorithm becomes linear in the terms of the number of pixels in the image.

Finally, images that are subsets of $\mathbb{Z}^2$ exhibit a nice property under sampling. This property is clarified in the following corollary (of Theorem 5). In words, it implies that the output of the sampling algorithm changes only at values of ϵ that lie on the $\mathbb{Z}/2$ grid.

COROLLARY 7. *Let S be a finite subset of $\mathbb{Z}^2$ and $\epsilon > 0$. Then P is a $(\lceil 2\epsilon \rceil / 2)$-sampling of S if and only if P is an ϵ-sampling of S.*

Because $\epsilon \leq \lceil 2\epsilon \rceil / 2$, it is clear that if P is an ϵ-sampling of S, then P is a $\lceil 2\epsilon \rceil / 2$-sampling of S. We still have to prove the converse.

Let P be a $\lceil 2\epsilon \rceil / 2$-sampling of S, then $P \subset S$ and $B_{\lceil 2\epsilon \rceil / 2}(S) = B_{\lceil 2\epsilon \rceil / 2}(P)$. Let α be an element in S, then we have $B_{\lceil 2\epsilon \rceil / 2}(\alpha) = B_{\lceil 2\epsilon \rceil / 2}(P)$. According to the previous theorem, we obtain $B_\epsilon(\alpha) = B_\epsilon(P)$. As a result, $B_\epsilon(S) \subset B_\epsilon(P)$. On the other hand, $B_\epsilon(P) \subset B_\epsilon(S)$ because $P \subset S$. Consequently, $B_\epsilon(S) = B_\epsilon(P)$ and, therefore, P is an ϵ-sampling of S.

5 EXPERIMENTAL VERIFICATION

In Section 2, we argued that we should sample the image at a value ϵ equal to the error bound, in order to guarantee preservance of recognition information.

In this section, we use the previous algorithm to sample several images at different values of ϵ (under the L_∞ metric). We observe the effect of ϵ on the saving factor, which we define to be the ratio of the number of points in the original image to that of the sampled one.

We consider four images that were subject to edge detection. Those images are typical of the ones that arise in model based vision and are representative of the kind of savings that one may achieve by sampling. The results of the sampling algorithm are reported in Table 1.

While the saving factor can be as low as one (when the image is already sampled), and as high as $(\lceil 2\epsilon \rceil - 1)^2$ (when the image is a large black rectangle), these cases are not common in practice. Commonly occurring cases are more like the ones displayed in Fig. 1, Fig. 4, and Fig. 7. Their sample versions are shown in Fig. 2, Fig. 3, Fig. 5, Fig. 6, Fig. 8, Fig. 9, and Fig. 10. For such cases, it is justifiable to state that the saving factor falls—roughly speaking—in the range of two to six. Finally, we emphasize that a saving factor of k results in a time gain of k^p for an $O(n^p)$ matching algorithm, where n is the number of points in the image.

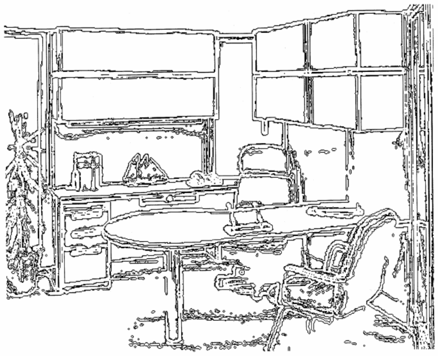


Fig. 1. Original image.

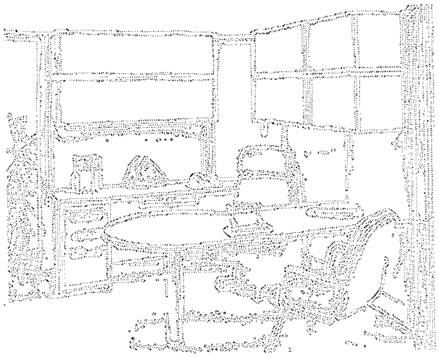


Fig. 2. Sampling for ϵ equals 1.5.

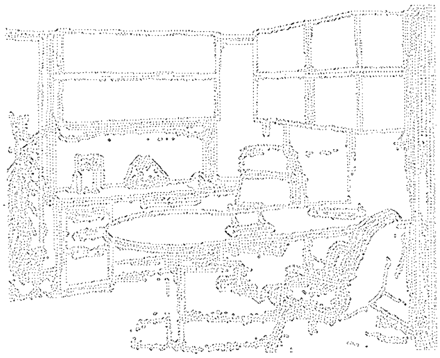


Fig. 3. Sampling for ϵ equals 2.0.

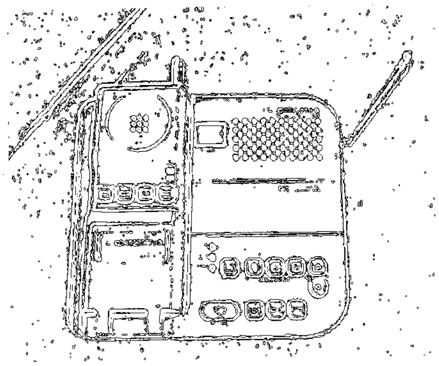


Fig. 4. Original image.

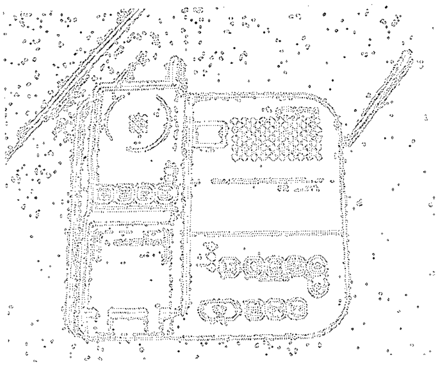


Fig. 5. Sampling for ϵ equals 1.5.

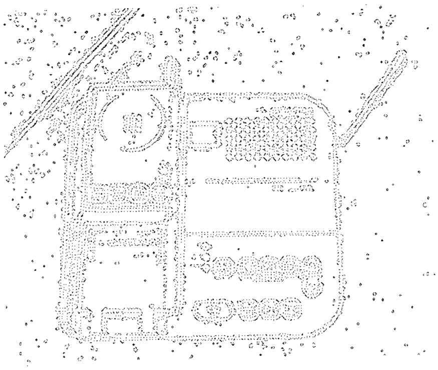


Fig. 6. Sampling for ϵ equals 2.0.

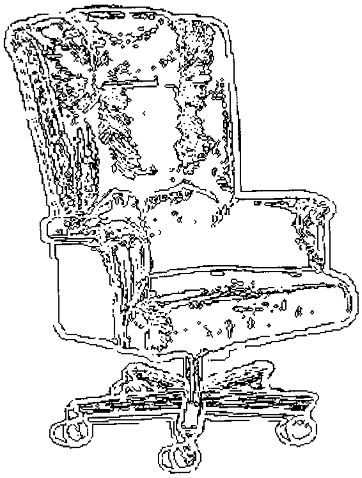


Fig. 7. Original image.

6 CONCLUSION

In this article, we presented a general formalization of the model-based vision problem or the problem of matching with geometric uncertainty. Based on this formalization, we designed a powerful tool to reduce the number of foreground points, thereby speeding up a large class of matching algorithms.

APPENDIX

LEMMA 5. *Let S be a finite subset of* Z^2, $t \in Z_+/2$, *and* $\eta \in (0, 0.5)$.

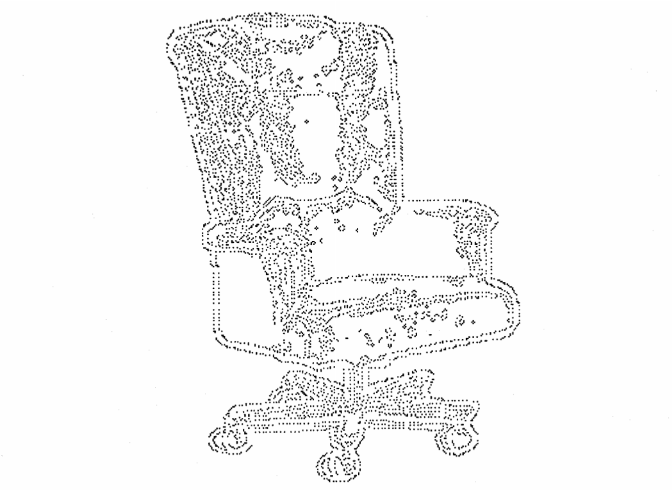


Fig. 8. Sampling for ϵ equals 1.5.

Fig. 9. Sampling for ϵ equals 2.0.

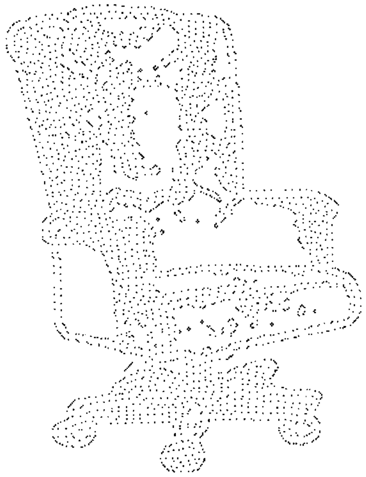


Fig. 10. Sampling for ϵ equals 3.0.

Let $q \in R^2$ *such that* $B_\eta(q) \subset B_t(S)$. *Then there exists* $\beta \in S$, *such that* $B_\eta(q) \subset B_t(\beta)$.

Suppose that

$$\forall \beta \in S, B_\eta(q) \not\subset B_t(\beta). \quad (3)$$

Let H be a minimal subset of S whose t-neighborhood covers $B_\eta(q)$, i.e.,

$$B_\eta(q) \subset B_t(H) \text{ and } \quad (4)$$

$$B_\eta(q) \not\subset B_t(H - \{\alpha\}) \ \forall \alpha \in H. \quad (5)$$

According to (3), H contains more than one element. Let α be an element of H. We know that $H - \{\alpha\} \neq \phi$. Equations (3), (4), and (5) imply that

$$B_\eta(q) \not\subset B_t(\alpha), B_\eta(q) \not\subset B_t(H - \{\alpha\}), \text{ and}$$

$$B_\eta(q) \subset B_t(\alpha) \cup B_t(H - \{\alpha\}).$$

Accordingly, let

$$x_1 \times y_1 \in (B_t(\alpha) - B_t(H - \{\alpha\})) \cap B_\eta(q) \quad \text{and}$$

$$x_2 \times y_2 \in (B_t(H - \{\alpha\}) - B_t(\alpha)) \cap B_\eta(q).$$

We have $x_1 \times y_2, x_2 \times y_1 \in B_\eta(q)$ because $B_\eta(q)$ is an *open rectangular region*. As a result, $x_1 \times y_2, x_2 \times y_1 \in B_t(\alpha) \cup B_t(H - \{\alpha\})$.

Consider $x_1 \times y_2 \in B_t(\alpha) \cup B_t(H - \{\alpha\})$. Then either

$$x_1 \times y_2 \in B_t(\alpha) - B_t(H - \{\alpha\}), \quad (6)$$

$$x_1 \times y_2 \in B_t(H - \{\alpha\}) - B_t(\alpha), \text{ or} \quad (7)$$

$$x_1 \times y_2 \in B_t(\alpha) \cap B_t(H - \{\alpha\}). \quad (8)$$

In the last case, namely, if $x_1 \times y_2 \in B_t(\alpha) \cap B_t(H - \{\alpha\})$, then it is not plausible to have $x_2 \times y_1 \in B_t(\alpha)$. Because if this were true, we would obtain $x_1 \times y_2, x_2 \times y_1 \in B_t(\alpha)$ which, due to the fact that $B_\eta(q)$ is an open rectangular region, implies that $x_2 \times y_2 \in B_t(\alpha)$, which is not true by definition.

Now, we add some given facts to (6) and (7) and to a consequence of (8) to obtain

$$\begin{cases} x_1 \times y_2 \in B_t(\alpha) - B_t(H - \{\alpha\}) & \text{and} \\ x_2 \times y_2 \in B_t(H - \{\alpha\}) - B_t(\alpha) \end{cases} \quad (9)$$

$$\text{or} \begin{cases} x_1 \times y_2 \in B_t(H - \{\alpha\}) - B_t(\alpha) & \text{and} \\ x_1 \times y_1 \in B_t(\alpha) - B_t(H - \{\alpha\}) \end{cases} \quad (10)$$

$$\text{or} \begin{cases} x_2 \times y_1 \in B_t(H - \{\alpha\}) - B_t(\alpha) & \text{and} \\ x_1 \times y_1 \in B_t(\alpha) - B_t(H - \{\alpha\}) \end{cases} \quad (11)$$

We prove that each of (9), (10), and (11) is not possible. Hence, (3) would be contradicted and the proof would be complete.

First, consider (9). We define $I_\tau(x)$ and $I_\tau(V)$, for $x \in R$, $V \subset R$, and $\tau > 0$, to be

$$I_\tau(x) = (x - \tau, x + \tau)$$

$$I_\tau(V) = \bigcup_{x \in V} I_\tau(x).$$

Let $H_x = \{x / x \times y \in H \text{ and } R \times y_2 \cap B_t(x \times y) \neq \varnothing\}$. Clearly, $\alpha_x \in H_x$, $I_t(H_x - \alpha_x) \times y_2 = B_t(H - \{\alpha\}) \cap R \times y_2$, and $I_t(\alpha_x) \times y_2 = B_t(\alpha) \cap R \times y_2$. So (9) implies that,

$$x_1 \in I_t(\alpha_x) - I_t(H - \{\alpha_x\}) \text{ and}$$

$$x_2 \in I_t(H_x - \{\alpha_x\}) - I_t(\alpha_x).$$

As a result,

$$I_\eta(q_x) \not\subset I_t(a), \forall a \in H_x \quad (12)$$

and

$$I_\eta(q_x) \subset I_t(H_x). \quad (13)$$

Let $a_1, a_2, \ldots, a_n$ be the elements of H_x ordered such that, for $i, j \in \{1, ..., n\}$, $a_i > a_j$ if $i > j$.

We claim that $a_1 < q_x < a_n$. If this were not true, in other words if $q_x \le a_1$ or $q_x \ge a_n$, then (13) would imply that $I_\eta(q_x) \subset I_t(a_1)$ or $I_t(a_n)$, which contradicts (12). Let

$$s = \max_{a_i \le q_x} i.$$

Because $a_1 < q_x < a_n$, we know that $1 \le s, s + 1 \le n$. We further claim that $I_\eta(q) \subset I_t(a_s) \cup I_t(a_{s+1})$; we have $q_x \in [a_s, a_{s+1}]$, so $I_\eta(q_x) \subset I_\eta([a_s, a_{s+1}])$. In addition, $\eta \in (0, 0.5)$ and $t \in Z_+/2$, so $\eta < t$. Consequently, $I_\eta(q_x) \subset I_t([a_s, a_{s+1}])$. Using (13), we obtain $I_\eta(q_x) \subset I_t([a_s, a_{s+1}]) \cap I_t(H_x)$. But, $I_t(H_x) \subset I_t(R - (a_s, a_{s+1}))$ and $I_t([a_s, a_{s+1}]) \cap I_t(R - (a_s, a_{s+1})) = I_t(a_s) \cup I_t(a_{s+1})$, so $I_\eta(q) \subset I_t(a_s) \cup I_t(a_{s+1})$.

As a result, we have

$$I_\eta(q_x) \not\subset I_t(a_s),$$

$$I_\eta(q_x) \not\subset I_t(a_{s+1}), \text{ and}$$

$$I_\eta(q_x) \subset I_t(a_s) \cup I_t(a_{s+1}).$$

So $I_t(a_s) \cap I_t(a_{s+1}) \neq \phi$ and $I_t(a_s) \cap I_t(a_{s+1}) \subset I_\eta(q_x)$. In other words, $(a_s + t) - (a_{s+1} - t) > 0$ and $(a_s + t) - (a_{s+1} - t) \le 2\eta$. So $0 < 2t - (a_{s+1} - a_s) \le 2\eta$. But $S \subset Z^2$ and $t \in Z_+/2$. So $\{a_s, a_{s+1}, 2t\} \in Z$. Consequently, $2\eta \ge 1$, which contradicts the fact that $\eta \in (0, 0.5)$. As a result, (9) is not possible. Similarly, we can prove that (10) and (11) are not possible. Hence, (3) is not possible and the proof is complete.

THEOREM 6. *Let S be a finite subset of Z^2, $\alpha \in Z^2$, and $\epsilon > 0$. Then the following are equivalent*:

1) $B_{\lceil 2\epsilon \rceil/2}(\alpha) \subset B_{\lceil 2\epsilon \rceil/2}(S)$.

2) $B_\epsilon(\alpha) \subset B_\epsilon(S)$.

3) $B_\epsilon(\alpha) \cap (\mathbb{Z}/2)^2 \subset B_\epsilon(S) \cap (\mathbb{Z}/2)^2$.

Clearly, (2) implies (3). We still have to prove that (1) implies (2) and that (3) implies (1).

If (1) were true, it could be written as

$$B_{(\lceil 2\epsilon \rceil/2) - \epsilon}(B_\epsilon(\alpha)) \subset B_{\lceil 2\epsilon \rceil/2}(S).$$

Let q be any element of $B_\epsilon(\alpha)$, so we have

$$B_{(\lceil 2\epsilon \rceil/2) - \epsilon}(q) \subset B_{\lceil 2\epsilon \rceil/2}(S).$$

Since $\lceil 2\epsilon \rceil / 2 \in \mathbb{Z}_+ / 2$ and $(\lceil 2\epsilon \rceil / 2) - \epsilon \in (0, 0.5)$, then using the previous lemma, we conclude that

$$\exists \beta \in S \ / \ B_{(\lceil 2\epsilon \rceil / 2) - \epsilon}(q) \subset B_{\lceil 2\epsilon \rceil / 2}(\beta).$$

In other words, $\exists \beta \in S / q \in B_\epsilon(\beta)$. Hence, $q \in B_\epsilon(S)$, which implies that $B_\epsilon(\alpha) \subset B_\epsilon(S)$. Consequently, (1) implies (2).

If (3) were true, i.e., if $B_\epsilon(\alpha) \cap (\mathbb{Z}/2)^2 \subset B_\epsilon(S) \cap (\mathbb{Z}/2)^2$, then we would obtain that

$$B_{1/2}(B_\epsilon(\alpha) \cap (Z/2)^2) \subset B_{1/2}(B_\epsilon(S) \cap (\mathbb{Z}/2)^2).$$

In other words,

$$B_{1/2}\left(B_\epsilon(\alpha) \cap (\mathbb{Z} / 2)^2\right) \subset \bigcup_{\beta \in S} B_{1/2}\left(B_\epsilon(\beta) \cap (\mathbb{Z} / 2)^2\right). \qquad (14)$$

But for $\gamma \in \mathbb{Z}^2$, $B_{1/2}\left(B_\epsilon(\gamma) \cap (\mathbb{Z} / 2)^2\right) \subset B_{\lceil 2\epsilon \rceil / 2}(\gamma)$. Replacing in (14), we get

$$B_{\lceil 2\epsilon \rceil / 2}(\alpha) \subset \bigcup_{\beta \in S} B_{\lceil 2\epsilon \rceil / 2}(\beta).$$

In other words, $B_{\lceil 2\epsilon \rceil / 2}(\alpha) \subset B_{\lceil 2\epsilon \rceil / 2}(S)$. Consequently, (3) implies (1).

ACKNOWLEDGMENTS

Research supported by ARO grant DAA L03-92-G-0115 and DAA H04-96-10445.

REFERENCES

[1] H. Baird, *Model-Based Image Matching Using Locations*. Cambridge, Mass.: MIT Press, pp. 12-13, 1984. Also in *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 8, no. 6, pp. 334-343, 1986.

[2] T. Cass, "Feature Matching for Object Localization in the Presence of Uncertainty," *Proc. Third Int'l Conf. on Computer Vision*, pp. 360-364, Dec. 1990.

[3] T. Cass, "Polynomial-Time Geometric Matching for Object Recognition," PhD dissertation, Massachusetts Institute of Technology, 1992.

[4] T. Cass, "Robust Affine Structure Matching for 3D Object Recognition," *Fourth European Conf. on Computer Vision*, vol. 1, pp. 492-503, Apr. 1996.

[5] B. Chazelle, "The Polygon Containment Problem," *Advances in Computing Research*, vol. 1, pp. 1-33, 1983.

[6] T. Breuel, "Geometric Aspects of Visual Object Recognition," PhD dissertation, Massachusetts Institute of Technology, 1992.

[7] D. Huttenlocher and K. Kedem, "Computing the Minimum Hausdorff Distance of Point Sets Under Translation," *Sixth ACM Symp. Computational Geometry*, vol. 1, pp. 340-349, 1990.

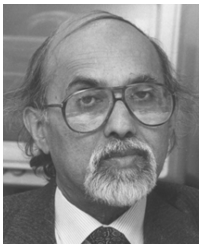
[8] D. Huttenlocher and S. Ullman, "Recognizing Solid Objects by Alignment With an Image," *Int'l J. Computer Vision*, vol. 5, no. 2, pp. 195-212, 1990.

[9] D. Huttenlocher, G. Klanderman, and W. Rucklidge "Comparing Images Using the Hausdorff Distance Under Translation," *Proc. IEEE Computer Vision and Pattern Recognition*, pp. 654-656, 1992.

[10] X. Yi and O. Camps, "Line Feature-Based Recognition Using Hausdorff Distance," *Proc. Int'l Symp. Computer Vision*, pp. 79-84, 1995.

Mohamad Akra received the PhD degree from the Massachusetts Institute of Technology in September 1993. He worked as a postdoctoral fellow at MIT after his graduation. During the summer of 1994, he was also appointed as a visiting scientist in the Center for Intelligent Control Systems at MIT. In October 1993, he joined the American University of Beirut, where he is currently an assistant professor of electrical and computer engineering. He is also the director for the Multimedia Center and the head for the technical advisory committee for the faculty of engineering and architecture. His research interests are pattern recognition, algorithms, and multimedia. Dr. Akra is a visiting scientist at MIT for the years 1996-1998.

Louay Bazzi received the BE degree from the American University of Beirut, Lebanon, in June 1996. He is currently a graduate student at the Electrical Engineering and Computer Science Department at the Massachusetts Institute of Technology. His research interests include computation theory and computer vision.



Sanjoy Mitter received the PhD degree from the Imperial College of Science and Technology, University of London, in 1965. He had previously worked as a research engineer at Brown Boveri & Co. Ltd., Switzerland (now ASEA Brown Boveri) and Battelle Institute in Geneva, Switzerland. He taught at Case Western Reserve University from 1965 to 1969 and joined MIT in 1969 first as a visiting professor and then in 1970 as associate professor in the Department of Electrical Engineering and Computer Science. He is currently a professor of electrical engineering and codirector of the Laboratory for Information and Decision Systems. He is also director of the Center for Intelligent Control Systems, an interuniversity (Brown/Harvard/MIT) center for research on the foundations of intelligent systems. He had held visiting positions at the Tata Institute of Fundamental Research, Bombay, India; Scuola Normale Superiore, Pisa, Italy; Imperial College of Science and Technology; Institut National de Recherche en Informatique et en Automatique, France; University of Gronengen, the Netherlands; and several universities in the United States. His research has spanned the areas of systems, communication, and Control. Although his primary contributions have been on the theoretical foundations of the field, he has also contributed to significant engineering applications, notably in the control of interconnected power systems and automatic recognition and classification of electrocardiograms. His current research interests are theory of stochastic dynamical systems, nonlinear filtering, stochastic and adaptive control; mathematical physics and its relationship to system theory; image analysis and computer vision; and structure, function, and organization of complex systems. Professor Mitter has served on several advisory committees and editorial boards for IEEE, SIAM, AMS, NSF, and ARO. He is currently associate editor of *Acta Applicandae Mathematicae; Circuits, Systems, and Signal Processing; Journal of Applied Mathematics and Optimization; SIAM Review*; and the *Ulam Quarterly*. In 1988, he was elected to the National Academy of Engineering.