COMPUTER VISION AND IMAGE UNDERSTANDING
Vol. 64, No. 3, November, pp. 399–419, 1996
ARTICLE NO. 0068

A Probabilistic Approach to Perceptual Grouping

REBECCA L. CASTAÑO* AND SETH HUTCHINSON†

The Beckman Institute and Department of Electrical and Computer Engineering, University of Illinois, Urbana, Illinois 61801

Received December 5, 1994; accepted September 5, 1995

We present a general framework for determining probability distributions over the space of possible image feature groupings. The framework can be used to find several of the most probable partitions of image features into groupings, rather than just returning a single partition of the features as do most feature grouping techniques. In addition to the groupings themselves, the probability of each partition is computed, providing information on the relative probability of multiple partitions that few grouping techniques offer. In determining the probability distribution of groupings, no parameters are estimated, thus eliminating problems that occur with small data sets and outliers such as the compounding of errors that can occur when parameters are estimated and the estimated parameters are used in the next grouping step. We have instantiated our framework for the two special cases of grouping line segments into straight lines and for grouping bilateral symmetries with parallel axes, where bilateral symmetries are formed by pairs of edges. Results are presented for these cases on several real images.

1. INTRODUCTION

Researchers in the fields of biological and computer vision often conceptualize the overall vision process as comprising a hierarchy of processing levels. At the lowest level, features are extracted from images. At higher levels, these features are grouped together into increasingly more abstract entities, until finally, at the highest levels, scene interpretation occurs [1, 3, 5, 17, 18, 25, 39, 41, 43].

The detection of image features has been studied extensively by both psychologists and computer vision researchers. The study of feature detection in biological vision systems began with the seminal work of Hubel and Wiesel [18]. Since then, psychologists have gathered strong empirical evidence for a variety of feature detection mechanisms in biological visual systems, including neurons for detecting color, motion, depth, and orientation spatial frequency [41, 43]. Computer vision researchers have studied feature detection to determine which features can be reliably extracted from an image as well as which features are useful for higher level processing [17, 34].

Less progress has been made in determining how features should be organized into more abstract structures to be used by higher level visual processes. Much of the work in this area has been driven by the early work of the Gestalt psychologists, who claimed that humans group features based on several principles, including proximity, symmetry, continuation, closure, and familiarity [21, 44]. The work of the Gestalt psychologists has inspired a number of computer vision approaches to perceptual grouping [2, 6, 19, 26, 23, 34, 38]. Typically, these computational approaches rely on thresholds (e.g., a threshold on the difference between orientation of line segments [26], or on the linking radius described in [6]), or on certainty measures that are derived in an ad hoc fashion (e.g., basing the certainty of grouping two line segments on the proximity of their endpoints [23], or using a decaying exponential to define a probability of termination for line segments [10]).

In this paper we present a new, probabilistic approach to the perceptual grouping problem. Our approach allows the representation of every possible grouping of image features, either explicitly or implicitly, along with an associated probability for each grouping. With this approach it is possible, for example, to enumerate several of the most probable groupings of features. This provides more information than the typical approach, in which only a single grouping is determined. Another advantage to representing all possible feature groupings is that no thresholds or arbitrary stopping criteria are needed. Further, no parameter estimation is performed during the process of determining the highly probable feature groups. Thus, our method avoids the problems encountered by parameter estimation schemes when there are outliers in the data or when data sets are small, including the accumulation of errors that can occur when estimated parameters are used in subsequent grouping steps.

The criteria by which our system assesses the probability associated with a feature grouping is based on a simple grouping principle—*features in an image should be grouped together when they participate in a common under-*

* E-mail: becky@ai.uiuc.edu.
† E-mail: seth@uiuc.edu.

1077-3142/96 $18.00

lying geometric structure. Several of the more commonly investigated Gestalt laws for grouping are instances of this principle, for example proximity, symmetry, and continuation. The principle is also consistent with the idea that preference should be given to a perceived structure when that structure has a low probability of occurring only by random chance, thus implying a single cause for the structure [23, 30, 45]. Here, we assume that regular geometric structure in an image does not often occur by chance. Therefore, the existence of such geometric structure provides strong evidence in support of a corresponding feature grouping.

A realization of our criteria within a probabilistic framework requires a parameterized model of geometric structures (which defines a *parameter space*), a characterization of how well a set of observed image features fits to a particular geometric structure (which defines an *observation space*), and a probabilistic model that describes the image formation process and its effects on the distribution of features in the image (which we refer to as the *degradation model*). In Section 3, we derive the parameter space, observation space, and degradation model for the two special cases of straight lines and bilateral symmetries. Although we have chosen these two specific geometric structures for the present work, in principle, our approach generalizes to arbitrary parameterized geometric structures.

Equipped with the models presented in Section 3, in Section 4 we derive the probability that a set of features should be grouped together, conditioned on the observed image data associated with those features. Rather than rely on parameter estimation, we compute marginalizing integrals over the entire parameter space, which has the effect of computing the sum of the probabilities of every possible geometric structure, conditioned on the observed data. Computation of these marginalizing integrals is a difficult numerical problem, which is addressed in the Appendix and in [33].

The formalism that we use for representing all possible feature groupings is a modification of the method originally presented in [22]. The number of possible feature groupings for a typical image will grow to be intractably large; however, most of the groupings will have very low probability. Our formalism, which we briefly describe in Section 5, exploits this property by representing explicitly only those groupings with high probability, while implicitly representing large sets of grouping hypotheses that have small probability. Using this formalism, the probabilities of only the most probable grouping hypotheses are computed. The resulting representation can be used to focus the attention of higher level processes on more likely feature groupings and, in turn, can facilitate a top–down influence on the grouping process by higher level processes.

The ability to handle a top–down influence on the grouping process has recently been advocated as a significant capability in perceptual grouping. While perceptual grouping is often treated as a bottom–up process in which features extracted by the low-level vision system are organized into progressively more abstract structures [35, 47], there is strong evidence from studies of biological vision systems, particularly in the context of visual attention, that the grouping process also involves a top–down component [1, 39, 43]. The necessity of a top–down component in the early- to mid-level visual processing has also been advocated within the computer vision community. For example, Tsotsos has argued that without a top–down component, visual search is NP-complete [42]. Our approach provides a mechanism to fuse bottom–up information with top–down control within a consistent probabilistic framework.

The results of using our new methods and probabilistic models on actual intensity image data are given in Section 6. In the final section, Section 7, conclusions are drawn and directions for future work are suggested. We now turn to a brief review of related research.

2. RELATED WORK

The only work that we are aware of that has similar goals to ours is that of Sarkar and Boyer [36]. Sarkar and Boyer developed Bayesian perceptual inference networks (PINs) as a hierarchical model of the spatial organization of image features. Features represented by nodes higher in the hierarchy are said to be caused by features represented by nodes lower in the hierarchy of the network. The probability of each feature is computed using Bayesian methods in which evidence comes both from nodes that are lower in the hierarchy (bottom–up information) and from nodes that are higher in the hierarchy (top–down information). Lowest level features are determined using a voting method. The problem that we deal with in this paper corresponds to that treated by the lowest levels the hierarchy in [36]; however, our probabilities are derived using mathematical models of image features and feature noise, rather than voting methods.

Below, we briefly review other related research. Although much work has been done in the area of perceptual grouping, both in the biological and computer vision communities, here we restrict our discussion to research that is concerned with grouping straight line segments and finding symmetries in images. For a review of computational approaches to perceptual grouping we refer the reader to [35].

2.1. *Finding Straight Lines in Images*

The process of detecting edges formed by intensity discontinuities in images is a frequent early level image processing task. The extraction of edge contours from an image reduces the amount of data to be processed, while retaining important information about the structure of a

scene or object. A further justification for this simplification is that humans are able to make many interpretations using the edge contours of scenes and objects [4, 23, 38]. A great deal of work in computer image understanding uses image contours. For this reason the extraction of image intensity edges continues to be a topic of active research.

Detection of edges involves identifying edge points and possibly thinning the edges to a width of a single edge point. Edge elements are then grouped and possibly fitted with a desired curve model, such as linear, polynomial, spline, or circular (see, for example, [29]). Edge detection often results in incomplete and broken edges that are difficult for higher-level processing systems to use. To improve the output, edges may be grouped together before a higher process receives the data. This is the first task to which we apply our framework: grouping straight line segments. There have been various approaches to determining straight edges in an image, and we now discuss a few of these.

Burns *et al.* [7] emphasized similarity of gradient orientation in grouping edge pixels into regions called line support regions. These regions were then fitted with straight lines using other image factors.

One of the most well-known procedures for determining straight lines in an image is the Hough transform [11]. The Hough transform involves discretizing the feature (line) parameter space. Each edge element ''votes'' for all lines of which it might be a part. Peaks in the accumulator grid represent lines in the image. A drawback of this method is that there is no facility for representation of proximity of elements or location of the line segment in the image.

Huddleston and Ben-Arie [19] modified the Hough transform to eliminate these shortcomings. Their transform for circular and linear edge segments combines a measure of the Gestalt principle of proximity with a measure of nonaccidentalness to assign weights that are accumulated for each edge element.

Foresti *et al.* [13] presented the *labeled Hough transform*, another modification of the basic Hough transform. Edge points are partially grouped using local information. The groups of points are labeled before points vote in the Hough transform accumulation array. By separating the voting into an array for each grouping, local information is used in the Hough transform. The basic Hough transform with its global information can be obtained by summing the accumulation bins that correspond across the grouping levels.

The Gestalt principles of perceptual grouping are explicitly used by Saund [37] to group edge tokens into coarser level tokens and for grouping edge tokens into structures such as corners and curves. Boldt *et al.* [6] also use perceptual organizational principles to group segments into edges. Scher *et al.* [38] exclusively use the grouping principle of proximity in several splitting and merging algorithms for clustering collinear straight line segments.

A different approach to grouping line segments is taken by Nacken [27]. Nacken constructs a metric for quantifying how well two line segments can be replaced by a single longer segment. Segments are parameterized by center point, length, and angle. Noise is assumed to be reflected as noise in these parameters. In contrast, our method assumes that the noise is in the location of the edge elements and does not estimate parameters in determining whether two segments are part of one longer segment.

Another work that uses probabilistic methods for grouping edge elements is that of Cox *et al.* [10]. They present a probabilistic method of recovering smooth contours from the edge elements of an image, refraining from making a grouping decision until sufficient information has accumulated.

None of the approaches described above allow the representation of multiple feature grouping hypotheses (although [10] incrementally constructs a tree representation of possible groupings, until the moment at which the best grouping is selected). Furthermore, most of these methods do not permit a probabilistic assessment of the quality of a candidate grouping; instead, ad hoc confidence assignments are made based on subjective criteria.

Methods based on the Hough transform suffer when peaks in the accumulator array are blurred or do not exist, sometimes to the point where no feature groupings can be made. With our method, as noise in the input image increases (effectively blurring peaks in the Hough accumulator array), our method continues to enumerate the correct feature groupings, although the resulting probability distributions show increasing entropy and the most probable feature grouping may no longer be the ''correct'' feature grouping. In Section 6, we present an example using synthetic data to illustrate this effect.

Finally, the traditional Hough transform is global in nature and therefore does not readily allow consideration of local effects on the feature grouping process. In Section 6, we present examples of cases for which proximity of line segment endpoints can be used effectively to constrain the grouping process.

2.2. *Detection of Symmetries in Images*

Most of the work on finding symmetries in images assumes as a starting point a figure on a noiseless background. Friedberg [14] finds the axis of symmetry and axis of the skewed plane of a skewed symmetric figure. Marola [24] presents a region-based method for determining the axes of symmetry of a possibly noisy figure. He defines symmetry as invariance with respect to reflection about one or more straight lines. Glachet *et al.* [16] find the axis and plane of bilaterally symmetric planar objects viewed under

perspective projection. Rom and Medioni [31, 32] compute a hierarchical axial description of a planar shape in which a figure is represented by its constituent parts. A work in which the background is cluttered is that of Zielke *et al.* [46] who detect bilaterally symmetric figures with vertical axes in real time using a combination of a region and edge information. We work with bilateral symmetries formed by image edges and do not require elimination of background distractions. Mohan and Nevatia [26] also work with cluttered scenes. They use curvature extrema to identify symmetric curves with axes that are invariant to viewpoint direction. These pairs of curves are used along with additional information obtained using other principles of perceptual organization to segment a scene.

As in the case of grouping straight lines, to date there have been no reported attempts to compute probability distributions over the space of candidate symmetries in an image.

3. PROBABILISTIC MODELS

In this section, the models used for computing grouping probabilities will be formally introduced. The equations for computing the probabilities using these models will be developed in Section 4, and in Section 5 the framework that enables us to selectively compute the probabilities of only the most probable groupings is presented.

We begin by defining parameterizations of geometric structures. We then use the concept of an observation space to formally characterize how well an extracted feature fits to a specific geometric structure. Next we characterize the effect of the imaging process on features by defining a degradation model to be the conditional probability density function of observations, given a particular geometric structure. Finally, we define the prior model as the expected distribution of structures. These types of models have been used by other researchers in the context of region segmentation (e.g., [15, 22]).

In this paper, we consider two geometric structures representing properties of edges suggested by Biederman [5]: straight lines and bilateral symmetries. Thus, the probability that some set of features should be grouped together is a function of how well those features fit either a straight line or a bilateral symmetry.

3.1. *Parameter Space*

Both straight lines and bilateral symmetries admit a finite unique parameterization (i.e., the parameterization is bijective). For simplicity, we represent a parameter space by a finite vector of random variables, $\mathbf{U_k} = [U_k^1\, U_k^2 \cdots U_k^r]$. A vector value that $\mathbf{U_k}$ can take on is denoted by $\mathbf{u}_k$.

The parameter space (sample space of parameter vectors) we use for the line segment application is $[0, 2\pi) \times \mathbb{R}^+$, representing all straight lines. This space is parameterized by a pair (θ, d), where θ represents the normal of the line and d represents the normal distance from the origin of the reference frame to the line, as shown in Fig. 1.

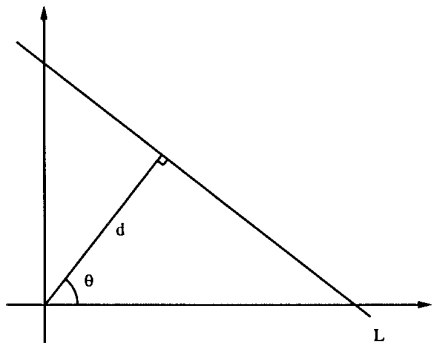

FIG. 1. Normal parameters for line L.

For the application of grouping features into bilateral symmetries we restrict the symmetry contours to straight lines. There are various ways to parameterize the space of all symmetries of this form. We have chosen to use a parameterization in which the axis of symmetry and sweeping rule are explicit and the lines of symmetry are implicit. This parameterization facilitates comparison of axes independent of the sweeping rules. The four parameters are θ, d, α, Δ. The two parameters θ and d define the axis of symmetry while α and Δ define the sweeping rule, as shown in Fig. 2. Thus the parameter space for the symmetries is $\{(\theta, d, \alpha, \Delta) : \theta \in [0, 2\pi),\ d > 0,\ \alpha \in [0, \pi),\ \Delta > 0\}$.

3.2. *Observation Space*

The set of all features in an image is denoted by $\mathcal{S}$. Each feature, S_k, consists of a set of image data points, $S_k = \{\mathbf{x}_1, x_2, \ldots, \mathbf{x}_{|S_k|}\}$, each of which can be represented by $\mathbf{x}_i = \langle x_i, y_i \rangle$. Since noise is introduced into the system during

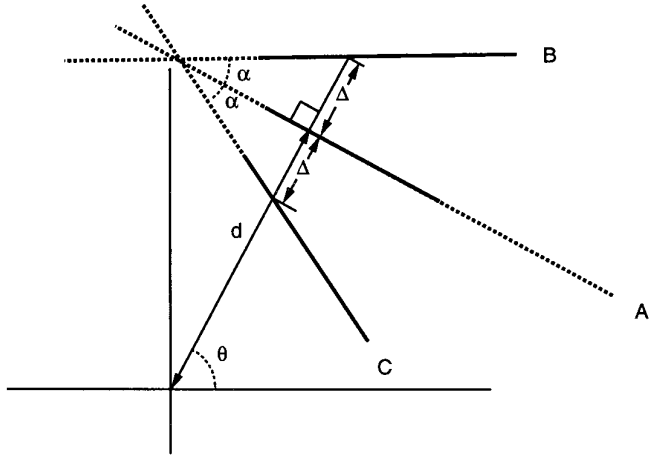

FIG. 2. Symmetry parameterization.

the imaging, digitization, and edge extraction processes, each component of $\mathbf{x}_i$ is modeled as a random variable. Hence, we have $\mathbf{X}_i = \langle X_i = x, Y_i = y \rangle$, where x and y are values that the random variables X_i and Y_i can take on. Appropriate measurability conditions on the random variables as specified in [40] are assumed.

In order to evaluate the probability that a set of features should be grouped together, it is necessary to evaluate how well the features fit to a single geometric structure (in our case, either a straight line or bilateral symmetry). For this purpose we define an observation space as follows.

Let $\mathbf{U}_k$ denote the parameter space for a particular grouping of features that contains S_k. Let $\mathbf{S}_k$ denote the set of all random variables of all data points in feature S_k. Let $\psi_k(\mathbf{S}_k, \mathbf{U}_k)$ be a vector-valued function of the random variables in $\mathbf{S}_k$ and the parameter space, $\mathbf{U}_k$. Each element $\psi_k^i(\mathbf{S}_k, \mathbf{U}_k)$ is a random variable. Define $Y_k^i = \psi_k^i(\mathbf{S}_k, U_k)$. The observation space of the set S_k is defined as

$$\begin{aligned}\mathbf{Y}_k &= [Y_k^1, Y_k^2, \ldots, Y_k^q] \\ &= [\psi_k^1(\mathbf{S}_k, \mathbf{U}_k), \psi_k^2(\mathbf{S}_k, \mathbf{U}_k), \ldots, \psi_k^q(\mathbf{S}_k, \mathbf{U}_k)].\end{aligned}$$

Here, $\mathbf{Y}_k$ is a vector of random variables. Let $\mathbf{y}_k$ denote a vector of values that $\mathbf{Y}_k$ may take on. The number of functions, q, will depend on the class of structures and features that are modeled. We now derive two observation spaces, corresponding to straight lines and bilateral symmetries.

3.2.1. *Straight Lines.* A point, $\mathbf{u}_k = (\theta_k, d_k)$, in the parameter space defines a line in the image plane. Let $\phi(\mathbf{u}_k)$ be the implicit mapping of points in the parameter space to a line in the image plane

$$\phi(\mathbf{u}_k) = \{(x, y): x \cos\theta_k + y \sin\theta_k - d_k = 0\}. \tag{1}$$

Let $\delta(\mathbf{x}_i, \phi(\mathbf{u}_k))$ be the signed distance of the data point $\mathbf{x}_i$ to the line represented by the parameter vector $\mathbf{u}_k$, i.e., the displacement of the point $\mathbf{x}_i$ to the line described by the zero set $\phi(\mathbf{u}_k)$. We can write $\delta(\mathbf{x}_i, \phi(\mathbf{u}_k))$ as

$$\delta(\mathbf{x}_i, \phi(\mathbf{u}_k)) = x_i \cos\theta_k + y_i \sin\theta_k - d_k. \tag{2}$$

Let S_k be an input line segment, represented by a set of data points $\{\mathbf{x}_1 \ldots, \mathbf{x}_{|S_k|}\}$. We define the observation $\mathbf{y}_k(S_k, \mathbf{u}_k)$, to be a vector in which each component is the distance of a data point to the line $\phi(\mathbf{u}_k)$

$$\begin{aligned}&\mathbf{y}_k(S_k, \mathbf{u}_k) \\ &\quad = [\delta(\mathbf{x}_1, \phi(\mathbf{u}_k)), \delta(\mathbf{x}_2, \phi(\mathbf{u}_k)), \ldots, \delta(\mathbf{x}_{|S_k|}, \phi(\mathbf{u}_k))].\end{aligned} \tag{3}$$

3.2.2. *Bilateral Symmetries.* To establish an observation model corresponding to bilateral symmetries we first note that a point $(\theta, d, \alpha, \Delta)$ in the parameter space defines two lines in the image plane,

$$x \cos(\theta + \alpha) + y \sin(\theta + \alpha) - d_B = 0, \tag{4}$$

$$x \cos(\theta - \alpha) + y \sin(\theta - \alpha) - d_C = 0, \tag{5}$$

labeled as lines B and C in Fig. 2. From the figure we see that the point $(x, y) = ((d + \Delta) \cos\theta, (d + \Delta) \sin\theta)$ satisfies (4). Substitute the known point into (4) to obtain

$$\begin{aligned}&(d + \Delta) \cos\theta \cos(\theta + \alpha) \\ &\quad + (d + \Delta) \sin(\theta) \sin(\theta + \alpha) - d_B = 0.\end{aligned} \tag{6}$$

Now solve for d_B to obtain

$$d_B = (d + \Delta) \cos(\alpha). \tag{7}$$

Using an analogous procedure

$$d_C = (d - \Delta) \cos(\alpha). \tag{8}$$

Expressing the two lines implicitly as $\phi_1(\mathbf{u}_k)$ and $\phi_2(\mathbf{u}_k)$ we have

$$\begin{aligned}\phi_1(\mathbf{u}_k) = \{(x, y): &x \cos(\theta_k + \alpha_k) + y \sin(\theta_k + \alpha_k) \\ &- (d_k + \Delta_k) \cos(\alpha_k) = 0\},\end{aligned} \tag{9}$$

$$\begin{aligned}\phi_2(\mathbf{u}_k) = \{(x, y): &x \cos(\theta_k - \alpha_k) + y \sin(\theta_k - \alpha_k) \\ &- (d_k - \Delta_k) \cos(\alpha_k) = 0\}.\end{aligned} \tag{10}$$

Let $\delta(\mathbf{x}_i, \phi_j(\mathbf{u}_k))$ be defined, as before, to be the displacement of the point $\mathbf{x}_i$ from the line $\phi_j(\mathbf{u}_k)$, where $j \in \{1, 2\}$.

Two line segments S_{k_1} and S_{k_2} may be grouped together in a symmetric relationship. An observation is then defined as

$$\begin{aligned}&\mathbf{y}_k(S_{k_1}, S_{k_2}, \mathbf{u}_k) \\ &\quad = [\delta(\mathbf{x}_{1_1}, \phi_1(\mathbf{u}_k)), \delta(\mathbf{x}_{1_2}, \phi_1(\mathbf{u}_k)), \ldots, \delta(\mathbf{x}_{1_{|S_{k_1}|}}, \phi_1(\mathbf{u}_k)), \\ &\qquad \delta(\mathbf{x}_{2_1}, \phi_2(\mathbf{u}_k)), \delta(\mathbf{x}_{2_2}, \phi_2(\mathbf{u}_k)), \ldots, \delta(\mathbf{x}_{2_{|S_{k_2}|}}, \phi_2(\mathbf{u}_k))].\end{aligned} \tag{11}$$

3.3. *Degradation Model*

In the ideal case, a set of features would be grouped together if and only if they exactly fit a single geometric structure. In this case, all observation values $\mathbf{y}$ for either (3) or (11) would be equal to zero. In reality, the imaging process is not perfect. Therefore, we must model how the imaging process affects the observation values. For this purpose a conditional probability density function, $f(\mathbf{y}_k \mid \mathbf{u}_k)$, called the degradation model, is used. The degradation model represents the conditional density of an ob-

servation vector, $\mathbf{y}_k$, given a parameter value, $\mathbf{u}_k$, of the underlying geometric structure

$$f(\mathbf{y}_k \mid \mathbf{u}_k) = f([\delta(\mathbf{x}_1, \phi(\mathbf{u}_k)), \delta(\mathbf{x}_2, \phi(\mathbf{u}_k)), \ldots, \delta(\mathbf{x}_{|S_k|}, \phi(\mathbf{u}_k))] \mid \mathbf{u}_k) = \prod_{\mathbf{x}_i \in S_k} f(\delta(\mathbf{x}_i, \phi(\mathbf{u}_k)) \mid \mathbf{u}_k). \tag{12}$$

The second line follows from the first, since each data point $\mathbf{x}_i$ is assumed conditionally independent of every other data point given the parameter vector $\mathbf{u}_k$. That is, given $\mathbf{u}_k$, every $\mathbf{x}_i \in S_k$ is independent of every $\mathbf{x}_j \in S_k \ \forall i \neq j$.

The degradation model represents the uncertainty in the measured data due to noise and other factors. It is a characterization of the deviation of observations and, thus, features from a specific geometric structure. Various factors could contribute to the deviations of the observation values. Noise in the data may be introduced at any stage of the process: imaging, digitization, or edge detection. Occlusion, shadows, and lack of contrast are all problems that could cause errors when detecting edge contours. The only factor that we model is noise. In this case, the degradation model is a function of the noise in the system.

The model we have selected is i.i.d. Gaussian with zero mean and variance σ^2. Each observed point is assumed to be displaced along a line perpendicular to the ideal line by an amount characterized by the Gaussian distribution. Cox *et al*. [10] also used a Gaussian distribution to characterize the noise of edge points. Using the Gaussian distribution in (12), we can write the degradation model for lines as

$$f(\mathbf{y}_k \mid \mathbf{u}_k) = \prod_{\mathbf{x}_i \in S_k} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(1/2)(\delta(\mathbf{x}_i, \phi(\mathbf{u}_k))/\sigma)^2} \tag{13}$$

$$= \frac{1}{(2\pi\sigma^2)^{|S_k|/2}} e^{-(1/2)\Sigma_{\mathbf{x}_i \in S_k}(\delta(\mathbf{x}_i, \phi(\mathbf{u}_k))/\sigma)^2}. \tag{14}$$

The degradation model for symmetries is

$$f(\mathbf{y}_k \mid \mathbf{u}_k) = f([\delta(\mathbf{x}_{1_1}, \phi_1(\mathbf{u}_k)), \delta(\mathbf{x}_{1_2}, \phi_1(\mathbf{u}_k)), \ldots, \delta(\mathbf{x}_{|1_{S_{k_1}}|}, \phi_1(\mathbf{u}_k)), \delta(\mathbf{x}_{2_1}, \phi_2(\mathbf{u}_k)), \delta(\mathbf{x}_{2_2}, \phi_2(\mathbf{u}_k)), \ldots, \delta(\mathbf{x}_{2_{|S_{k_2}|}}, \phi_2(\mathbf{u}_k))] \mid \mathbf{u}_k)$$

$$= \prod_{\mathbf{x}_i \in S_{k_1}} f(\delta(\mathbf{x}_i, \phi_1(\mathbf{u}_k)) \mid \mathbf{u}_k) \prod_{\mathbf{x}_j \in S_{k_2}} f(\delta(\mathbf{x}_j, \phi_2(\mathbf{u}_k)) \mid \mathbf{u}_k).$$

Applying the Gaussian noise model, we have

$$f(\mathbf{y}_j \mid \mathbf{u}_k) = \frac{1}{(2\pi\sigma^2)^{(|S_{k_1}|+|S_{k_2}|)/2}} e^{g(S_{k_1}, S_{k_2}, \mathbf{u}_k)}, \tag{15}$$

where

$$g(S_{k_1}, S_{k_2}, \mathbf{u}_k) = -\frac{1}{2\sigma^2}\left[\sum_{\mathbf{x}_i \in S_{k_1}} (\delta(\mathbf{x}_i, \phi_1(\mathbf{u}_k)))^2 + \sum_{\mathbf{x}_j \in S_{k_2}} (\delta(\mathbf{x}_j, \phi_2(\mathbf{u}_k)))^2\right]$$

$$= -\frac{1}{2\sigma^2}\left[\sum_{\mathbf{x}_i \in S_{k_1}} (x_i \cos(\theta + \alpha) + y_i \sin(\theta + \alpha) - (d + \Delta)\cos(\alpha))^2 + \sum_{\mathbf{x}_j \in S_{k_2}} (x_j \cos(\theta - \alpha) + y_j \sin(\theta - \alpha) - (d - \Delta)\cos(\alpha))^2\right].$$

Equations (14) and (15) can be used to quantify the deviation of a set of image data points from a line bilateral symmetry, respectively, with given parameter $\mathbf{u}_k$.

3.4. *Prior Model*

The joint pdf of $\mathbf{u}_k$, $f(\mathbf{u}_k)$, is called the prior model, which is a density over the parameter space that represents the expected distribution of features over the space. In this paper, we will assume that all feature parameter values are equally likely. Thus, $f(\mathbf{u}_k)$ will be a uniform distribution over the parameter space.

4. ASSESSING THE PROBABILITY OF FEATURE GROUPINGS

Using the models defined in Section 3, we will now develop an expression for the probability that a set of features should be grouped together. The hypothesis that features $\{S_{\alpha_i}\}$ should be grouped together is represented by $H_{\alpha_1,\ldots,\alpha_n}$, where $S_{\alpha_i} \in \mathcal{S}$. For example, H_{12} represents the hypothesis that S_1 and S_2 should be grouped. We note that the hypothesis H_{12} is equivalent to asserting that S_1 and S_2 fit the same geometric structure and therefore that $\mathbf{u}_1 = \mathbf{u}_2$. We begin by considering the case of grouping two features, and then extend this to the general case of grouping n features.

4.1. *Probability Assignments for Grouping Feature Pairs*

We denote by $P(H_{ij} \mid \mathbf{y}_i, \mathbf{y}_j)$ the probability that the features S_i and S_j associated with a pair of observations $\mathbf{y}_j$ and $\mathbf{y}_j$ should be grouped together. Note that the observation space $\mathbf{Y}_i$ and the parameter space $\mathbf{U}_i$ are continuous, while the hypotheses $H_{\alpha_1,\ldots,\alpha_n}$ are discrete, since they are either *true* or *false*. Our notations are to denote the probability density function (pdf) of continuous random variables by $f(\,)$ and to denote the probability mass function of discrete random variables by $P(\,)$.

We will assume that $\mathbf{Y}_1$ is marginally independent of $\mathbf{Y}_2$ given $\neg H_{12}$, i.e.,

$$f(\mathbf{y}_1, \mathbf{y}_2 \mid \neg H_{12}) = f(\mathbf{y}_1) f(\mathbf{y}_2). \tag{17}$$

By making this assumption, we have chosen to let the observations be independent if they do not correspond to the same geometric structure. This assumption is reasonable under conditions in which features occur independently. For instance, if lines in an image are independent, then segments that are not on the same line are unrelated. There may be conditions under which the independence assumption is not reasonable, such as if it is known that all segments are connected or that intersecting segments are orthogonal. Something about the image would have to be known a priori in order to make the decision that the independence assumption is invalid and to know what model of the relationships would better represent the situation.

If the parameter value associated with a grouping is given, the degradation model completely describes the density of the corresponding observations, i.e.,

$$f(\mathbf{y}_1 \mid \mathbf{u}_1, \mathbf{y}_2, H_{12}) = f(\mathbf{y}_1 \mid \mathbf{u}_1) \tag{18}$$

which implies that

$$f(\mathbf{y}_1, \mathbf{y}_2 \mid \mathbf{u}_1, \mathbf{u}_2) = f(\mathbf{y}_1 \mid \mathbf{u}_1) f(\mathbf{y}_2 \mid \mathbf{u}_2). \tag{19}$$

That is, if the parameter value of the grouping to which a feature belongs is known, no other observations will affect the density of the observation of that feature.

When Eqs. (17), (18), and (19) are satisfied, it can be shown (by derivations analogous to those in [22]) that

$$P(H_{12} \mid \mathbf{y}_1, \mathbf{y}_2) = \frac{1}{1 + \lambda_0 \lambda_1(\mathbf{y}_1, \mathbf{y}_2)}, \tag{20}$$

in which

$$\lambda_0 = \frac{P(\neg H_{12})}{P(H_{12})} \tag{21}$$

represents the prior probability that two features should be grouped together and

$$\lambda_1(\mathbf{y}_1, \mathbf{y}_2) = \frac{\int f(\mathbf{y}_1 \mid \mathbf{u}_1) f(\mathbf{u}_1)\, d\mathbf{u}_1 \int f(\mathbf{y}_2 \mid \mathbf{u}_2) f(\mathbf{u}_2)\, d\mathbf{u}_2}{\int f(\mathbf{y}_1 \mid \mathbf{u}_{12}) f(\mathbf{y}_2 \mid \mathbf{u}_{12}) f(\mathbf{u}_{12})\, d\mathbf{u}_{12}} \tag{22}$$

represents the effect that the observed data have on the probability of a grouping hypothesis. The integrals in the numerator of (22) are used to compute marginal pdfs $f(\mathbf{y}_1)$ and $f(\mathbf{y}_2)$. Any of these integrals can be replaced by a summation for the case of discrete variables. Finally, note that (22) is expressed only in terms of the degradation and prior models previously defined.

4.2. *Straight Lines*

For the line grouping application, we use the degradation model of (14) to obtain

$$\int f(\mathbf{y}_k \mid \mathbf{u}_k) f(\mathbf{u}_k)\, d\mathbf{u}_k$$

$$= \int \frac{1}{(2\pi\sigma^2)^{|S_k|/2}} e^{-(1/2)\Sigma_{\mathbf{x}_i \in S_k} (\delta(\mathbf{x}_i, \phi(\mathbf{u}_k))/\sigma)^2} f(\mathbf{u}_k)\, d\mathbf{u}_k \tag{23}$$

$$= \int\int \frac{1}{(2\pi\sigma^2)^{|S_k|/2}} e^{-(1/2)\Sigma_{\mathbf{x}_i \in S_k} (\delta(x_i \cos\theta + y_i \sin\theta - d/\sigma)^2} f(\theta, d)\, dd\, d\theta. \tag{24}$$

This double integral can be reduced to a single integral (see the Appendix). Using (24) in (20) we can determine the probability that any two sets of points are samples of the same line.

4.3. *Bilateral Symmetries*

To obtain a similar expression for symmetries, (15) can be used to obtain

$$\int f(\mathbf{y}_k \mid \mathbf{u}_k) f(\mathbf{u}_k)\, d\mathbf{u}_k$$

$$= \int_0^{2\pi} \int_{-\infty}^{\infty} \int_{-\pi}^{\pi} \int_{-\infty}^{\infty} f(\mathbf{y}_k \mid \mathbf{u}_k) f(\theta, d, \alpha, \Delta)\, d\Delta\, d\alpha\, dd\, d\theta. \tag{25}$$

The sets of segments with the highest probability of forming a bilateral symmetry can be computed. Each symmetry consists of two groups of segments, each of which forms a straight line. Thus, it may be observed that this is a more complicated (computationally expensive) method for finding collinear line segments. A more interesting problem is that of determining pairs of lines that have the same axis, but not necessarily the same sweeping rule. To compare symmetries with the same axis, but different sweeping rules, it is necessary to generalize the notion of the hypothesis. Consider two observations $\mathbf{y}_1$ and $\mathbf{y}_2$ and their respective parameter values $\mathbf{u}_1$ and $\mathbf{u}_2$, where

$$\mathbf{u}_1 = (\theta_1, \alpha_1, d_1, \Delta_1)$$
$$\mathbf{u}_2 = (\theta_2, \alpha_2, d_2, \Delta_2).$$

Recall that H_{12} represented the hypothesis that $\mathbf{u}_1 = \mathbf{u}_2$, i.e., $\theta_1 = \theta_2$, $\alpha_1 = \alpha_2$, $d_1 = d_2$, and $\Delta_1 = \Delta_2$. We now define H'_{12} to be the hypothesis that $\theta_1 = \theta_2$ and $d_1 = d_2$. If we are interested only in collinear axes, no restrictions will be placed on the sweeping rules, that is, the relationship of α_1 to α_2 or the relationship of Δ_1 to Δ_2:

$$\begin{aligned}
&\int f(\mathbf{y}_1 \mid \mathbf{u}_{12}) f(\mathbf{y}_2 \mid \mathbf{u}_{12}) f(\mathbf{u}_{12})\, d\mathbf{u}_{12} \\
&\quad = \int f(\mathbf{y}_1 \mid (\theta_{12}, d_{12}, \alpha_1, \Delta_1)) f(\mathbf{y}_2 \mid (\theta_{12}, d_{12}, \alpha_2, \Delta_2)) \\
&\qquad f(\theta_{12}, d_{12}, \alpha_1, \Delta_1, \alpha_2, \Delta_2)\, d\mathbf{u}_1\, d\mathbf{u}_2 \\
&\quad = \int_0^{2\pi} \int_{-\infty}^{\infty} \left\{ \int_{-\pi}^{\pi} \int_{-\infty}^{\infty} f(\mathbf{y}_1 \mid (\theta_{12}, d_{12}, \alpha_1, \Delta_1)) \right. \\
&\qquad f(\theta_{12}, d_{12}, \alpha_1, \Delta_1)\, d\alpha_1\, d\Delta_1 \\
&\qquad \int_{-\pi}^{\pi} \int_{-\infty}^{\infty} f(\mathbf{y}_2 \mid (\theta_{12}, d_{12}, \alpha_2, \Delta_2)) \\
&\qquad \left. f(\theta_{12}, d_{12}, \alpha_2, \Delta_2)\, d\alpha_2\, d\Delta_2 \right\} dd_{12}\, \theta_{12}.
\end{aligned} \tag{26}$$

Equation (26) follows, since the sweeping rules are independent, and therefore we can separate the integrals over α and Δ. Equation (26) can be used in (20) to determine $P(H'_{12} \mid \mathbf{y}_1, \mathbf{y}_2)$.

4.4. *Probability Assignments for Grouping n Features*

The expression for the probability of the hypothesis that n features are consistent given the n associated observations is more complicated than the expression for two features. These issues are addressed in [9]. We use the approximation derived in [9],

$$\begin{aligned}
&P(H_{1..n-1} \mid \mathbf{y}_1, \ldots, \mathbf{y}_n) \\
&\quad = 1 \Bigg/ \left[1 + \frac{P(\neg H_{1..n-1} \mid \mathbf{y}_1, \ldots, \mathbf{y}_{n-1})}{P(H_{1..n-1} \mid \mathbf{y}_1, \ldots, \mathbf{y}_{n-1})} \right. \\
&\qquad \left. \times \frac{1 - P(H_{1..n} \mid \mathbf{y}_1, \ldots, \mathbf{y}_n, H_{1..n-1})}{P(H_{1..n} \mid \mathbf{y}_1, \ldots, \mathbf{y}_n, H_{1..n-1})} \right],
\end{aligned} \tag{27}$$

where

$$\begin{aligned}
&P(H_{1..n} \mid \mathbf{y}_1, \ldots, \mathbf{y}_n, H_{1..n-1}) \\
&= \left(1 + \frac{P(\neg H_{1..n} \mid H_{1..n-1})}{P(H_{1..n} \mid H_{1..n-1})} \frac{f(\mathbf{y}_n) f(\mathbf{y}_1, , \ldots, \mathbf{y}_{n-1} \mid H_{1..n-1})}{f(\mathbf{y}_{1,\ldots,} \mathbf{y}_{n-1}, \mathbf{y}_n \mid H_{1..n})} \right)^{-1}.
\end{aligned} \tag{28}$$

5. COMPUTING DISTRIBUTIONS

We could, in principle, compute the probabilities for all possible combinations of features using the methods from Section 4; however, such an endeavor is computationally infeasible and unnecessary in practice. Most feature groupings are so unlikely that effort need not be spent computing their probabilities. In this section, we briefly review methods that were originally presented in [22] for efficiently representing sets of feature groupings (Section 5.1) and partitions of the set of all feature groupings (Section 5.2).

5.1. *Representing Feature Groupings*

We denote by π_i the set of feature groupings that contain some specific feature, S_i. For example, Fig. 3a shows π_1 (labeled as B_1) for an image in which four features have been detected. In the figure, the set $\{S_1, S_2\}$ represents the hypothesis that S_1 and S_2 should be grouped together, but they should not be grouped with either S_3 or S_4. Note that this does not imply anything about the relationship between S_3 and S_4.

We define a *feature grouping sample space* by the triple $(\pi_i, \mathscr{B}_i, P_i)$, where $\mathscr{B}_i$ is the set of all subsets of π_i and P_i is a probability mapping on $\mathscr{B}_i$. Elements of $\mathscr{B}_i$ are referred to as events. The probability associated with an event is the sum of the probabilities of each of the individual groupings in the event, since the groupings are mutually exclusive and union of all groupings is the space π_i.

We define an implicit representation of events in terms of *include sets* and *exclude sets* by

$$\tau(I, E) = \{T \in \pi_i : I \subseteq T, E \cap T = \varnothing\}. \tag{29}$$

Thus, the event specified by $\tau(I, E)$ corresponds to the set of all feature groupings that contain every feature in the include set I, but none of the features in the exclude set E. Figure 3 illustrates several examples of the mapping from I, E to events in π_1.

A cover, C, of the space π_i is defined to be a set of pairwise disjoint events in $\mathscr{B}_i$ that form a partition of π_i. Figures 3a–e each illustrate one possible cover for π_1. The probability mapping for the events in a cover is an approximation of the probability map on π_i (it is only an approximation, because not all events are represented).

A cover, C_i, is a better approximation to the probability mapping than another cover, C_j, if C_i can be obtained by partitioning an element of C_j. This is accomplished by removing one event from the current cover, partitioning this event into two events and by forming a new cover by including the two new events. We refer to this process as *refinement*. The procedure for obtaining probabilities of feature groupings consists of refining covers until a sufficient approximation to the probability mapping has been obtained. For example, the cover in Fig. 3c is obtained from the cover in Fig. 3b by partitioning the event B_3 into the new events B_4 and B_5.

It is possible, in principle, to determine the exact probability mapping for π_i by refining covers until each singleton event is explicitly represented, as shown in Fig. 3e; however, in practice this is neither desirable or necessary. For example, consider Fig. 3d, in which the event B_6, an individual feature grouping, has a probability of 0.35. No other

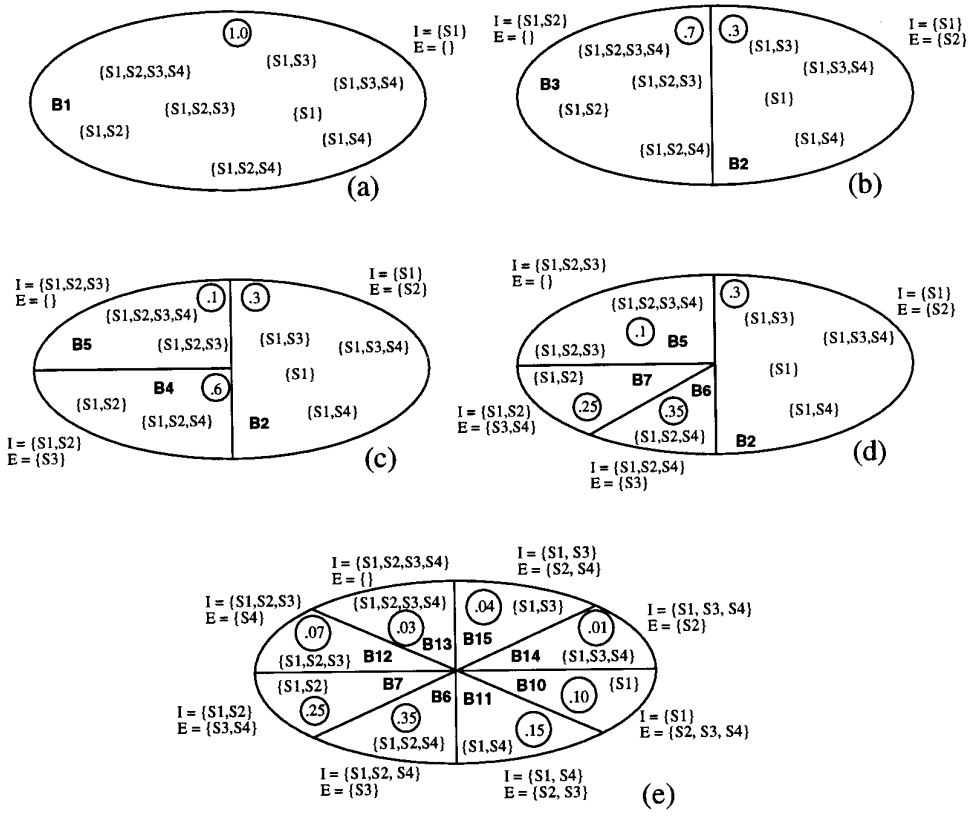

FIG. 3. Covers of π_1.

event in the cover has a probability of 0.35 or greater. Therefore, B_6 must be the most likely feature grouping since the probability of each event in a cover is the sum of the probabilities of each feature grouping in the event. Thus, it is not necessary to refine events B_2 and B_5 further in order to know that no feature grouping in either of these events has a greater probability than that of B_6. We have developed algorithms, reported in [9, 22], that efficiently explore the space of covers and are guaranteed to find the n most probable groupings, where n may be specified according to user preference.

5.2. *Representing Partitions of Feature Groupings*

In many applications, we wish to consider sets of feature groupings rather than individual feature groupings. In this section, we briefly describe how the representation scheme described above can be extended to represent the space of all partitions of image features into groupings.

Let Π be the set of all possible partitions of $\mathcal{S}$. The *feature partition sample space* is defined by the triple (Π, $\mathcal{A}$, P), where $\mathcal{A}$ is the set of subsets of Π, and P is a probability mapping on $\mathcal{A}$. As described in [22], there is an implicit mapping of events on π_i to events on Π. This can be seen by noting that for a particular feature, S_i, every partition in Π contains exactly one feature grouping that includes S_i. This feature grouping is, by definition, an element of π_i. Thus, every partition contains exactly one element of π_i. By extension, each cover on π_i induces a cover on Π. For example, the cover $\{B_2, B_3\}$ on π_1, shown in Fig. 3b, partitions Π into two sets, as shown in Fig. 4a.

We can refine Π indirectly by refining π_i up to the point where a singleton event in $\mathcal{B}_i$ is obtained. For the example shown in Fig. 3, the cover on Π when π_1 has been completely refined to individual groupings is shown in Fig. 4b. As can be seen from the figure, often an individual grouping in π_i will not map to a single partition in Π. In this case, we may further refine the cover of Π by focusing on feature grouping sample spaces for other features. For example, in Fig. 4b, all partitions in the event A_7 contain the feature grouping $\{S_1\}$. We can further refine A_7 by refining covers for π_2 relative to the feature set $\mathcal{S} - \{S_1\}$. (Note that since S_1 has already been assigned to a grouping, we may not now consider it when constructing π_2.) In Fig. 5a π_2 is shown along with the corresponding elements of Π. Successive Π refinements are illustrated in Fig. 5a and b.

5.3. *Selection of Initial and Refinement Features*

The efficiency of the algorithms depends on effective selection of both initial and refinement features. The se-

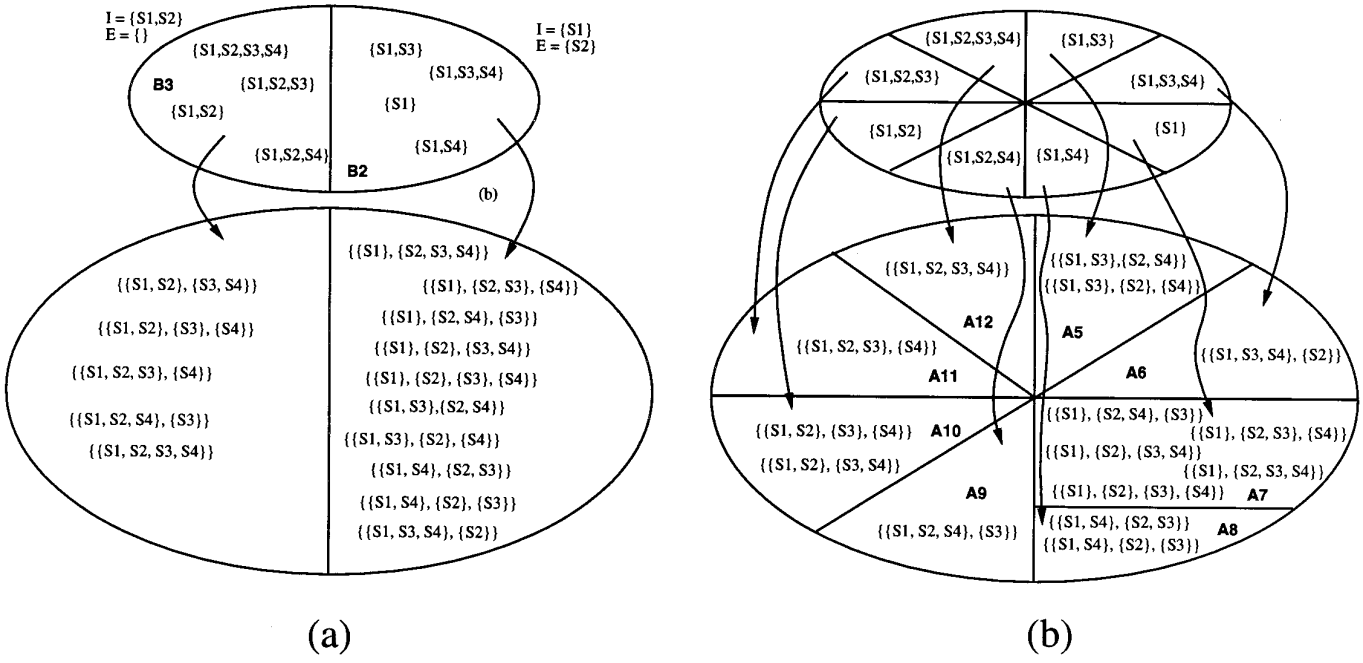

FIG. 4. II cover induced by π_1 cover.

lected feature should have an associated data set that contains as much information as possible. Those sets with little information will not affect probabilities very much. If a set with very little information content is used, the event that includes the set and the event that excludes the set will nearly evenly split the probability of the refinement event.

For example, in the case of line segments, the angle of long segments can vary much less than the angle of short segments. Longer segments contain more information about the segment angle than short segments. When the probability is evenly split across the include event and exclude event, both halves of the space will have to be refined further. However, if one event, either the include event or the exclude event, is very unlikely, the entire half space it represents will never have to be refined. Thus, prudent selection of feature data sets can greatly reduce the space of results that are refined, yielding a reduction in running time and storage space. The effectiveness of the information depends on features, parameterizations, and data samples. In practice, we select features with the most feature data points.

6. EXPERIMENTAL RESULTS

In this section, we show a number of experimental results using the framework. In Section 6.1 the results of our framework on several images with synthetically generated noise are compared with the results of a Hough transform

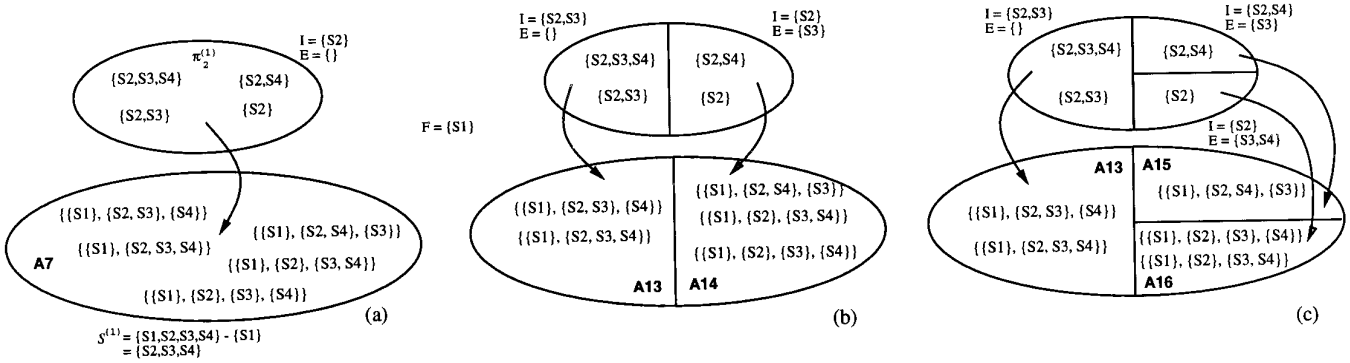

FIG. 5. Refinement of $\pi_2^{(1)}$.

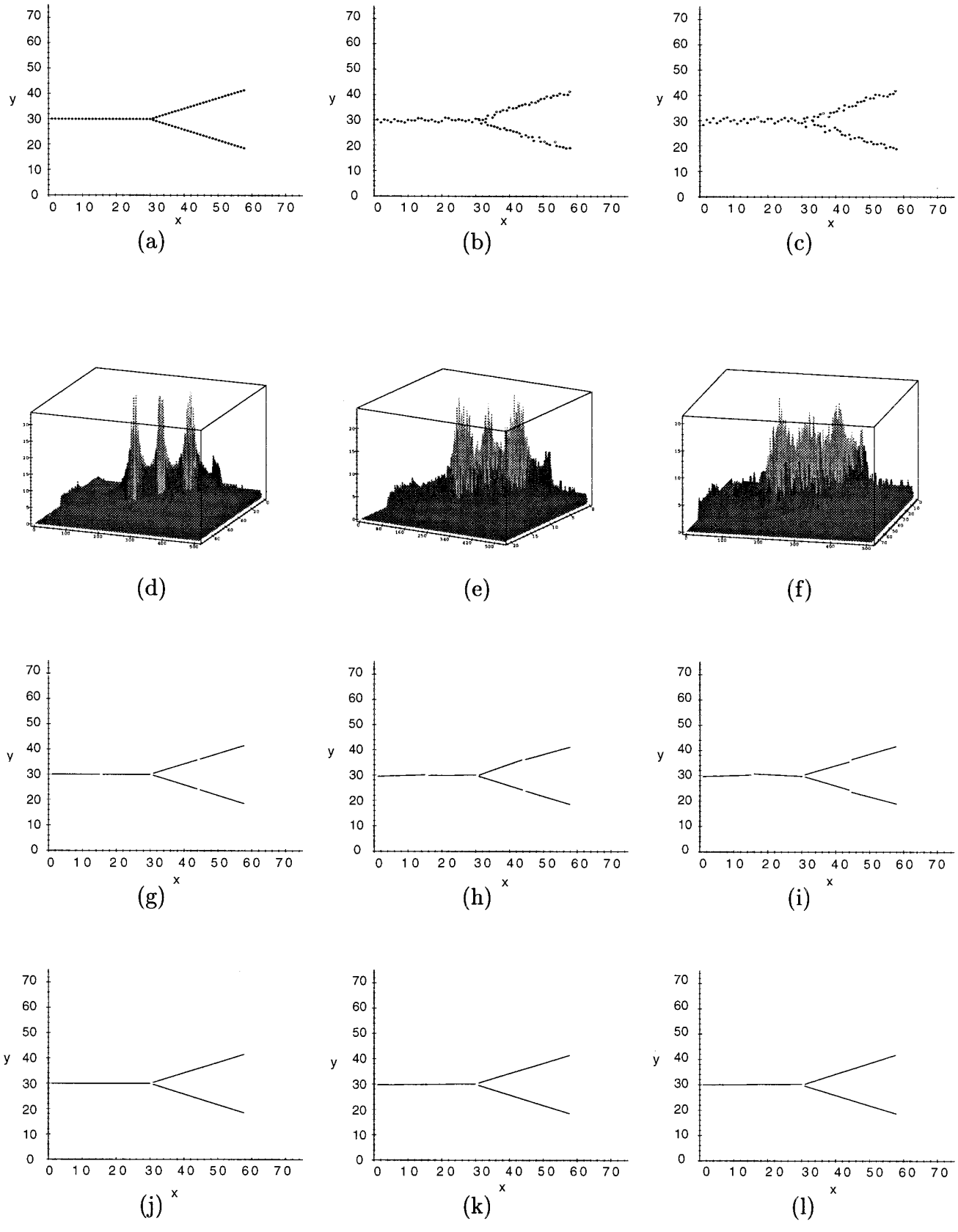

FIG. 6. (a) $\sigma = 0.0$, (b) $\sigma = 0.5$, (c) $\sigma = 0.75$.

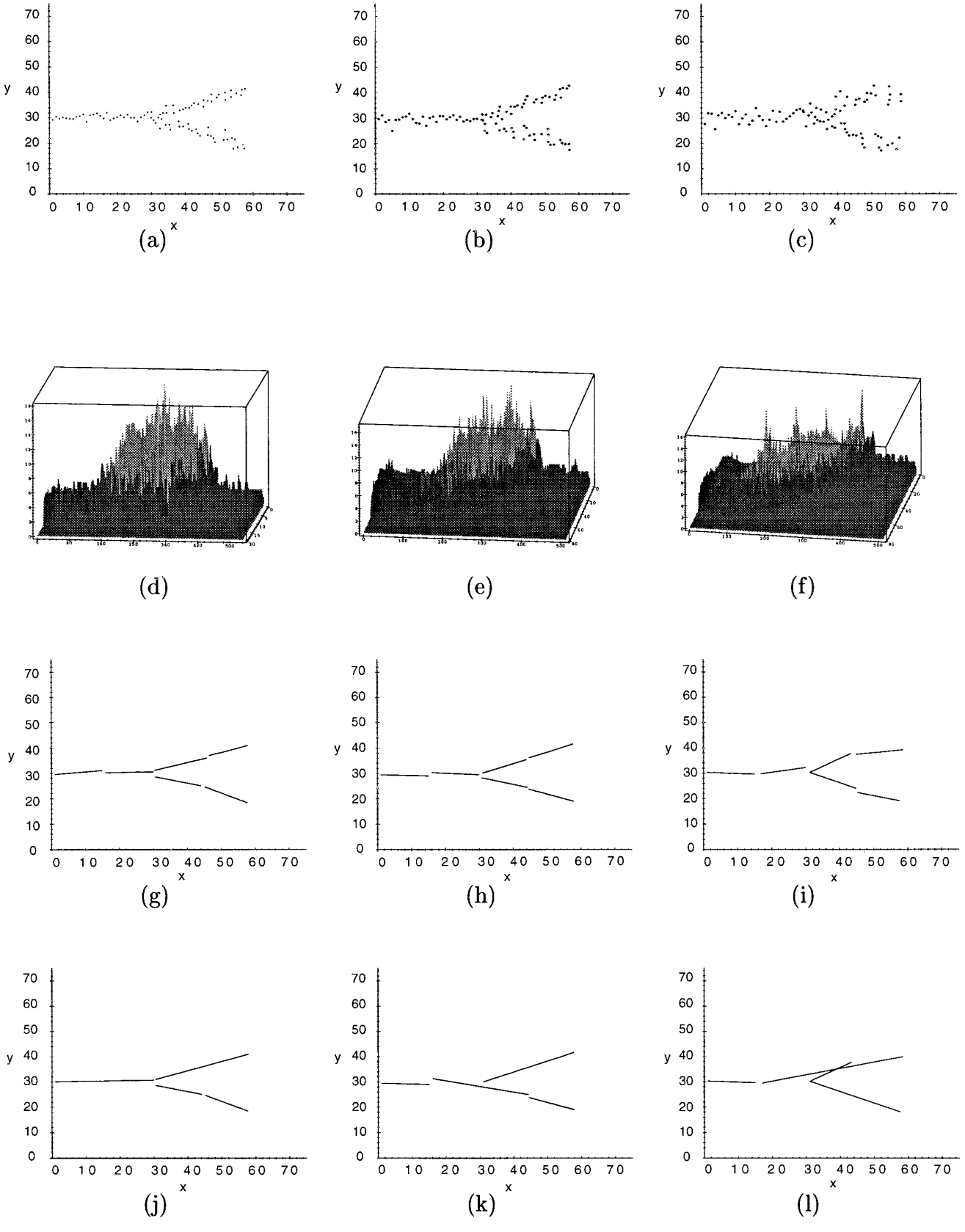

FIG. 7. (a) $\sigma = 1.0$, (b) $\sigma = 1.5$, (c) $\sigma = 2.0$.

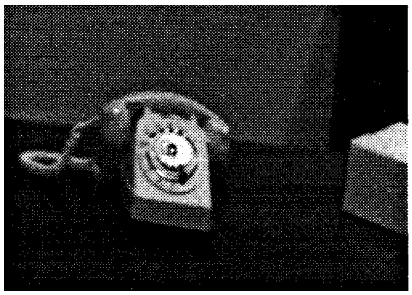

FIG. 8. 512 × 512 image.

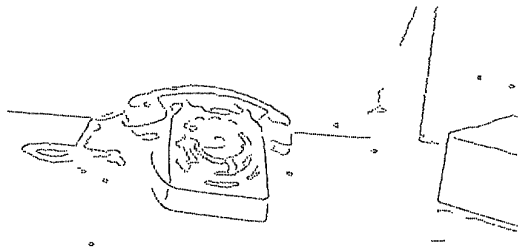

FIG. 10. Edge points found using Canny edge detector.

on the same data. Section 6.2 contains the results of partitioning the edge data of images into straight lines, and Section 6.3 contains the symmetry axis results.

In both cases of image data, the initial input to our algorithms was obtained by first processing the image with a Canny edge detector [8]. The edge detector yields edge contours represented by sets of connected points. These edges were divided into straight segments using an algorithm similar to the iterative endpoint fitting algorithm in [12]. After fitting the edges into straight segments, we have for each segment a set of constituent points. Any deviations of the points in a segment from linearity are treated as noise.

6.1. *Comparison to Hough Transform*

The Hough transform is a very popular method of determining straight lines in an edge image. Here we compare the results of our framework to that of a Hough transform under various noise conditions. Our experimental methodology was as follows. Thirty sample points were generated for each of three lines. Six tests were performed, each with different levels of Gaussian noise added to the data points. These input points are shown in Figs. 6a–c and 7a–c. The surface of the (θ, ρ) accumulation grid of the Hough transform is shown for each test in Figs. 6d–e and 7d–e. For input to our framework, each set of 30 points was divided into two sets of 15 points each. In Figs. 6g–i and 7g–i we show lines fitted to the sets of points that are input to our framework. These are the lines that would be estimated if no grouping were performed.

The most probable grouping of the segments into straight lines, as determined by our algorithms, is shown in Figs. 6j–l and 7j–l. As the noise increases the most probable grouping is no longer what might be considered as the correct grouping. The correct grouping, however, is always among the several most probable groupings. It is not surprising that as the amount of information in the data points is reduced, due to increased noise, the most probable grouping is no longer the ''correct'' grouping.

The Hough transform does not yield meaningful results as the variance of the noise increases past 1.0. In particular, the peaks of the Hough transform become indistinguishable as the noise increases.

In the case of reduced information (high noise), high level information must be used to make a decision among the most probable partitions; however, it should be noted that the partitions computed by our methods will still contain useful information, unlike the Hough transform. Finally, we note that as the noise in the input image is increased, the probability of the most probable groupings decreases from 0.995 for the case of no variance to 0.17 for the case with a variance of 2.0. Thus, our methods provide intuitively satis-

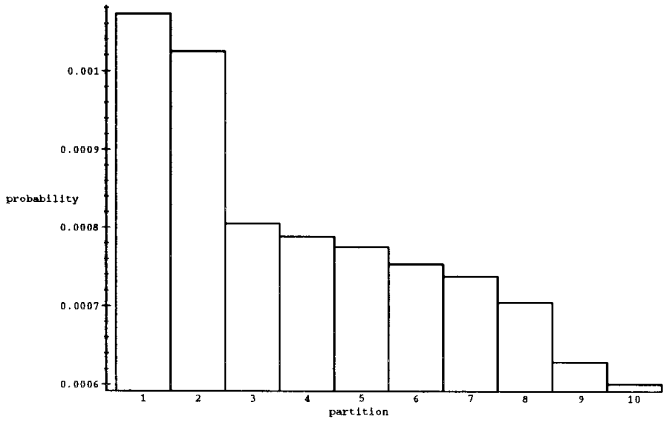

FIG. 9. Probability of 10 most likely partitions.

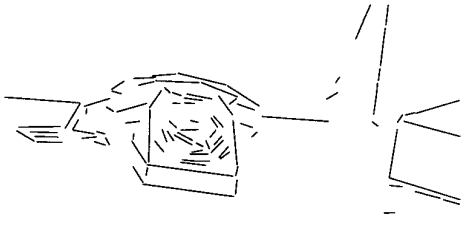

FIG. 11. Edge detector output fitted with straight lines. Lines with five pixels or fewer have been removed.

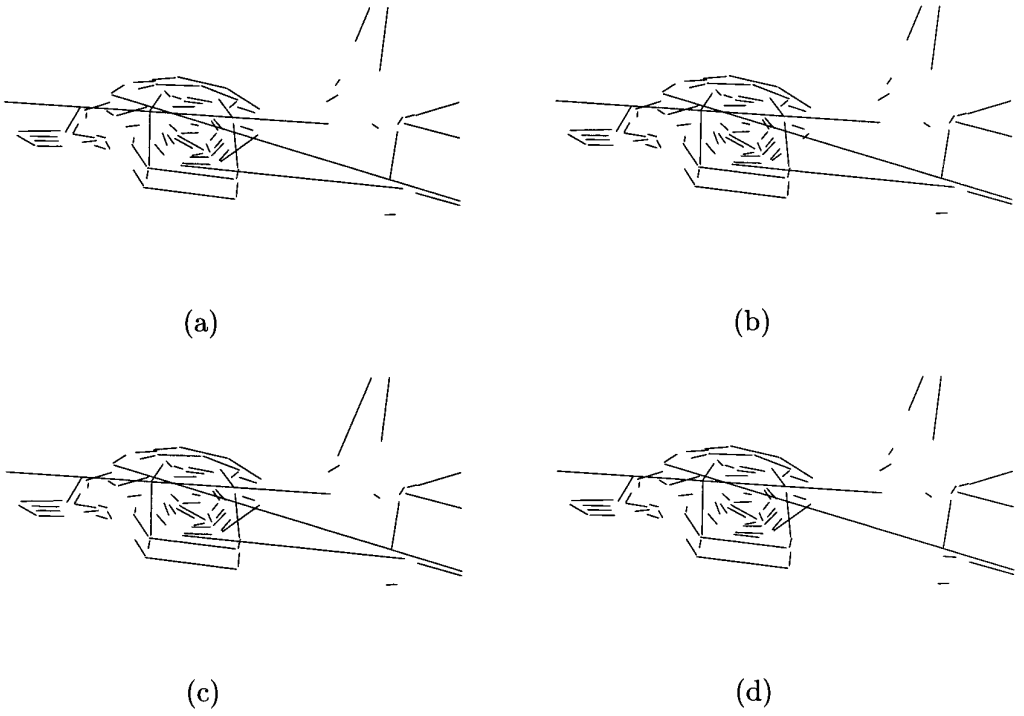

FIG. 12. Four most probable partitions (not using distance factor).

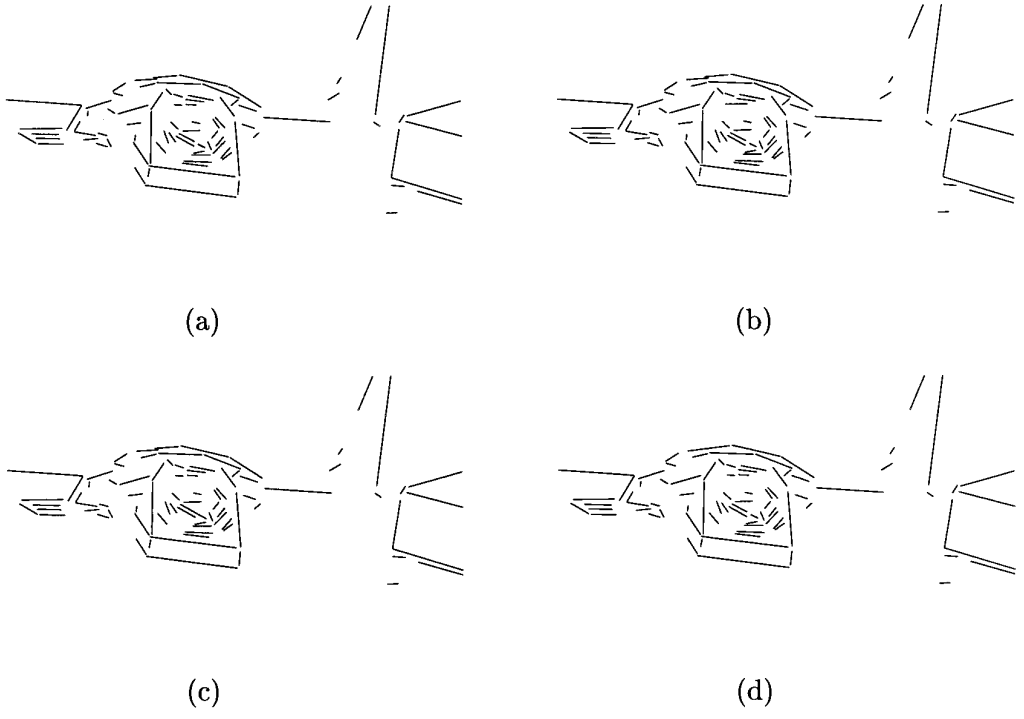

FIG. 13. Four most probable partitions (using distance factor).

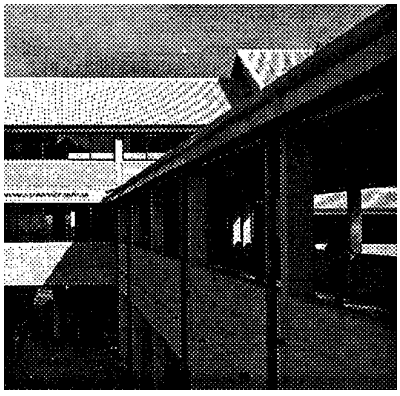

FIG. 14. 440 × 431 image.

fying results: as noise increases, the entropy in the probability distribution increases and the probability assigned to the most probable grouping decreases.

6.2. *Line Segments*

Line segment grouping results are shown for two images (Fig. 8 and Fig. 14). For each image the four most probable partitions are shown (Figs. 12, 13, and 17), and the distribution of the 10 most probable partitions are shown (Figs. 9 and Fig. 15). The distribution of 10 most probable partitions gives an indication of how many partitions have nearly the same probability and how rapidly the probability of alternate partitions decreases. The edges found by the Canny edge detector (Fig. 10 and Fig. 16) and the edge detector edges fitted with straight lines (Fig. 11) that are used as input to the probabilistic framework are shown. For each figure of partitions (i.e., Figs. 12, 13, and 17) the partition with the highest probability is shown in (a) and the partition with the fourth highest probability is shown in (d). A variance of $\sigma = 0.5$ was used for the phone image and a variance of $\sigma = 1.0$ was used for the building image.

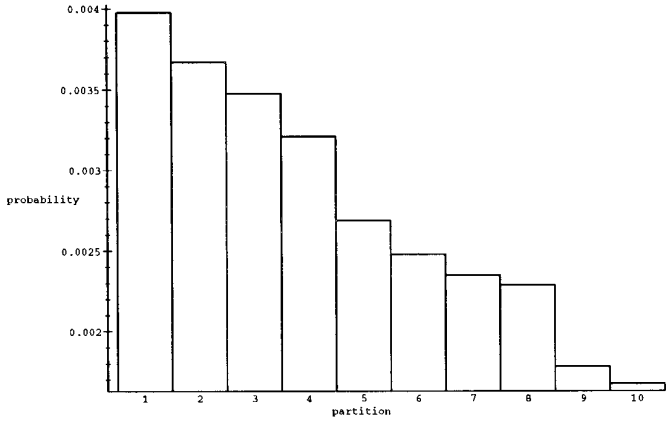

FIG. 15. Probability of 10 most likely partitions.

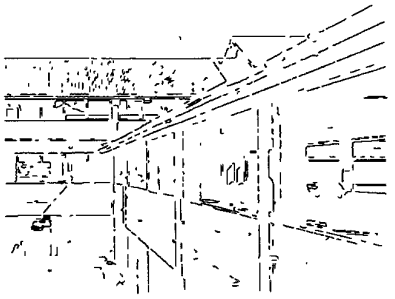

FIG. 16. Edges found by Canny edge detector.

Figures 12 and 13 demonstrate how proximity can be incorporated as a factor in grouping. We used a proximity factor to reduce the grouping of distant lines in the image for the results in Figs. 13 and 17. The factor was ignored if the distance between the closest pair of points was less than 25% of the sum of the number of points in each line. If the distance between segments was greater than this cutoff, the proximity factor was $e^{-(\text{dist}-\text{weight})^2}$, where the weight is the sum of the number of pixels in each segment. This factor was incorporated into the joint density $f(\mathbf{y}_1, \mathbf{y}_2 \mid H_{12})$ of (20). These figures illustrate the value of being able to consider local properties in the grouping process.

6.3. *Symmetry Axes*

The partitioning of the symmetry axes of two images are shown. The images were first processed by grouping line segments. The most probable partition of line segments was used as input for the symmetry axis partitioning. All pairs of segments were initially considered. Those pairs whose projections onto an estimated axis overlapped less than a threshold were thrown away. A variance of $\sigma = 3.0$ was used for all experiments in this section. We show the two most probable partitions of the axes into sets of parallel axes.

Partitions of parallel axes are shown as a series of figures. Groupings are shown as individual figures. For example, Fig. 21 represents one partition with four sets of groups.

All axes in an individual figure belong to the same group. Each figure shows the axes in a group (black) and, for reference, the input lines are also shown (gray). The axes shown are the estimated axes. The algorithm groups sets of data, but does not estimate any symmetry axes. We have estimated the projection of the lines of symmetry onto the axis for visualization purposes.

The first example, shown in Figs. 18–22, while very simple demonstrates the information that can be obtained from the probabilities of the top several partitions. From the distribution of the top five partitions, it can be seen that the most likely partition is significantly more likely

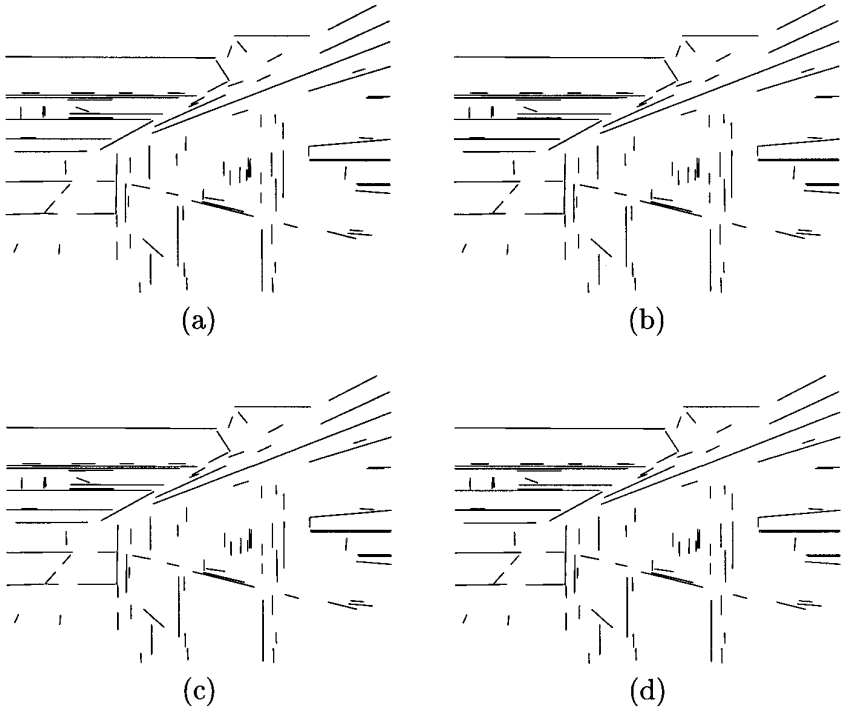

FIG. 17. The four most probable partitions.

than any other partition. This is due to the fact that every input line carries significant information and leads us to conclude that no other partitions need be considered as viable alternative explanations of the data,

The final example is shown in Figs. 23–26. By looking at the groupings of parallel axes, one can infer a great deal about the symmetry in the input image. For example, Fig. 26d contains many parallel axes. From this, a higher level process could infer the existence of specific symmetric structure in the input image; likewise for Figs. 26e, f, g, and k. We do not intend to imply that every grouping of symmetry axes is significant. Indeed, such inferences should be made only by higher level grouping processes, not by low-level grouping algorithms such as those that we propose.

7. CONCLUSIONS AND FUTURE RESEARCH

We have presented a framework for determining the most probable partitions of data sets representing image features into consistent feature groupings. Probabilistic models of noise and image features are used to compute

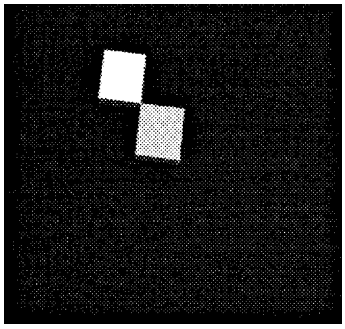

FIG. 18. 512 × 512 image.

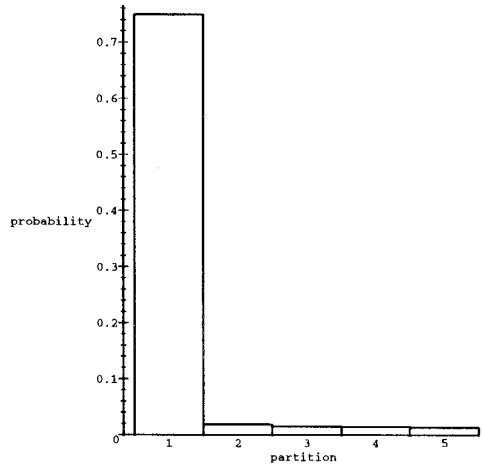

FIG. 19. Probability of five most likely partitions.

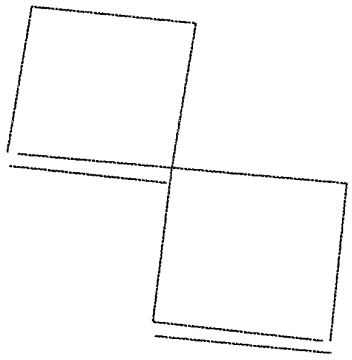

FIG. 20. Straight edge segments. There are nine lines.

the probability that features are consistent, given data sets representing the features. The framework has been demonstrated to perform well for partitioning line segments into collinear sets and bilateral symmetry axes into sets of parallel axes.

Many of the most probable partitions determined were very similar and nearly equally likely. For the image of Fig. 18, however, one partition of the symmetry axes was significantly more likely than any others. The four next most likely partitions of symmetry axes were almost equally probable. In the other figures there were less marked drop-offs in the probabilities of the most likely partitions. This information could be useful for a high level process. For the case of the image in Fig. 18, it may be that additional processing to distinguish between the most probable partitions is unnecessary, while in other cases, the similarities of the partitions may be noticed and used

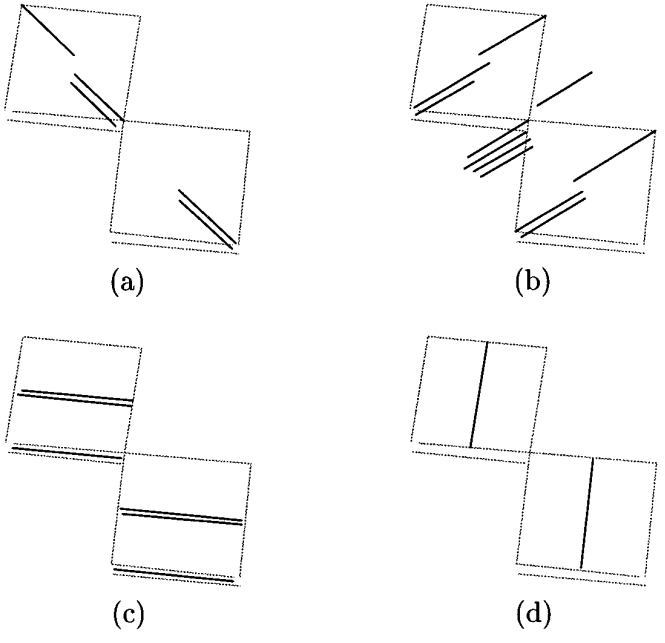

FIG. 21. Parallel symmetry axes. This is the most probable partition. This partition divides the set of axes into four groupings as shown in (a)–(d).

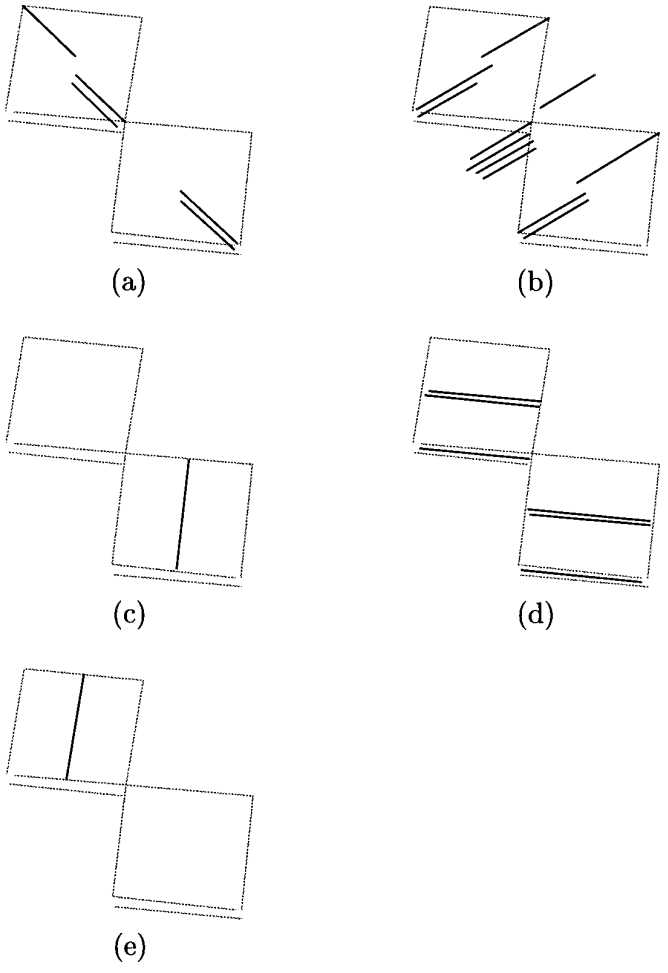

FIG. 22. Parallel symmetry axes. This is the second most probable partition. This partition divides the set of axes into four groupings as shown in (a)–(e).

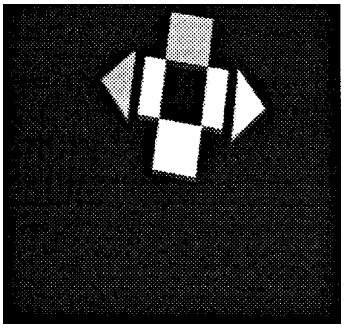

FIG. 23. 512 × 512 image.

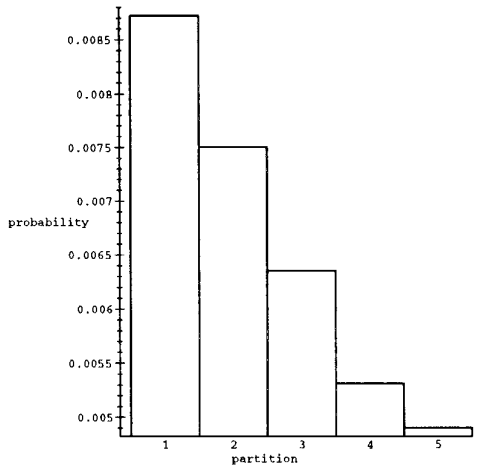

FIG. 24. Probability of five most likely partitions.

in conjunction with additional information to make distinctions between the most probable partitions.

It may be desirable to determine sets of collinear segments that are connected. Features would be segments with endpoints, rather than the infinite line model we used. Factors that could be used to influence the probability that two segments are collinear and connected are the gradient direction and magnitude, distance between the segments, and length of the segments. Modeling features as finite segments rather than infinite lines would eliminate the need for the heuristic distance factor that was used.

The symmetry axis application can be easily extended to skewed symmetries by adding a skew parameter. Other contour models such as B-splines could be used.

Other models of features and noise can be used in the framework. Features such as circles and corners could be used if appropriate parameter spaces and models can be determined. For more complex models and high dimension parameter spaces, the required integrals become difficult to calculate, so simplifications and approximations may become necessary.

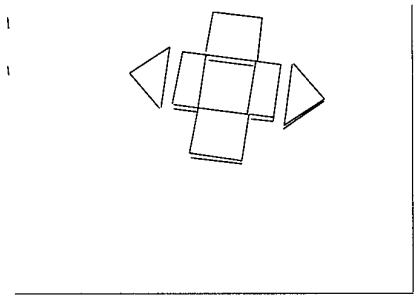

FIG. 25. Straight edge segments.

A. APPENDIX

A.1. *Line Segments*

We now make an observation to reduce the double integral in (24) to a single integral. The best fitting line (minimizing the sum of squares of perpendicular distance to the line) passes through the center of gravity $(\bar{x}, \bar{y}) = (1/n)\sum x_i, (1/n)\sum y_i)$ of the data points [12]. Noise introduces uncertainty in the angle of the line about the center of gravity $(\bar{x}, \bar{y})$. There is virtually no noise in the direction perpendicular to the best fitting line at the center of gravity [20]. Therefore, we can assume $\delta((\bar{x}, \bar{y}), \phi(\mathbf{u}_k)) = 0$. We can use this assumption to eliminate d from (24), thus simplifying it to a single integral rather than a double integral. We are *not* eliminating d by estimating its value. We are assuming that the conditional density of the degradation model over the portion of the parameter space representing lines that do not pass through the center of mass of the data set is negligible. We use this assumption to formulate one parameter, d, in terms of the second line parameter, θ. We can then substitute for d and obtain an expression dependent only on θ and the data.

We obtain an expression for d of the best fitting line by minimizing the sum of squares distance to the line. This is accomplished by setting the derivative of the expression for the sum of squares distance to the line with respect to d to zero. We have

$$\frac{\partial \sum (x_i \cos\theta + y_i \sin\theta - d)^2}{\partial d} = -2\sum (x_i \cos\theta + y_i \sin\theta - d) = 0.$$

We multiply by the constant $-2/n$ (where n is the number of data points) and solve for d to obtain

$$d = \bar{x}\cos\theta + \bar{y}\sin\theta. \tag{30}$$

We now have an expression for the line parameter d in terms of the center of gravity, $(\bar{x}, \bar{y})$, of the data points and the line parameter θ. Substituting (30) into the exponent of (24), we have

$$\begin{aligned} -\frac{1}{2\sigma^2}\sum [\delta(\mathbf{x}_i, \phi(\mathbf{u}_k))]^2 &= -\frac{1}{2\sigma^2}\sum (x_i \cos\theta + y_i \sin\theta - d)^2 \\ &= -\frac{1}{2\sigma^2}\sum ((x_i - \bar{x})\cos\theta_k + (y_i - \bar{y})\sin\theta_k)^2 \qquad (31) \\ &= -\frac{1}{2\sigma^2}\sum ((x_i - \bar{x})^2 \cos^2\theta_k \\ &\quad + (x_i - \bar{x})(y_i - \bar{y})\cos\theta_k \sin\theta_k + (y_i - \bar{y})^2 \sin^2\theta_k). \end{aligned}$$

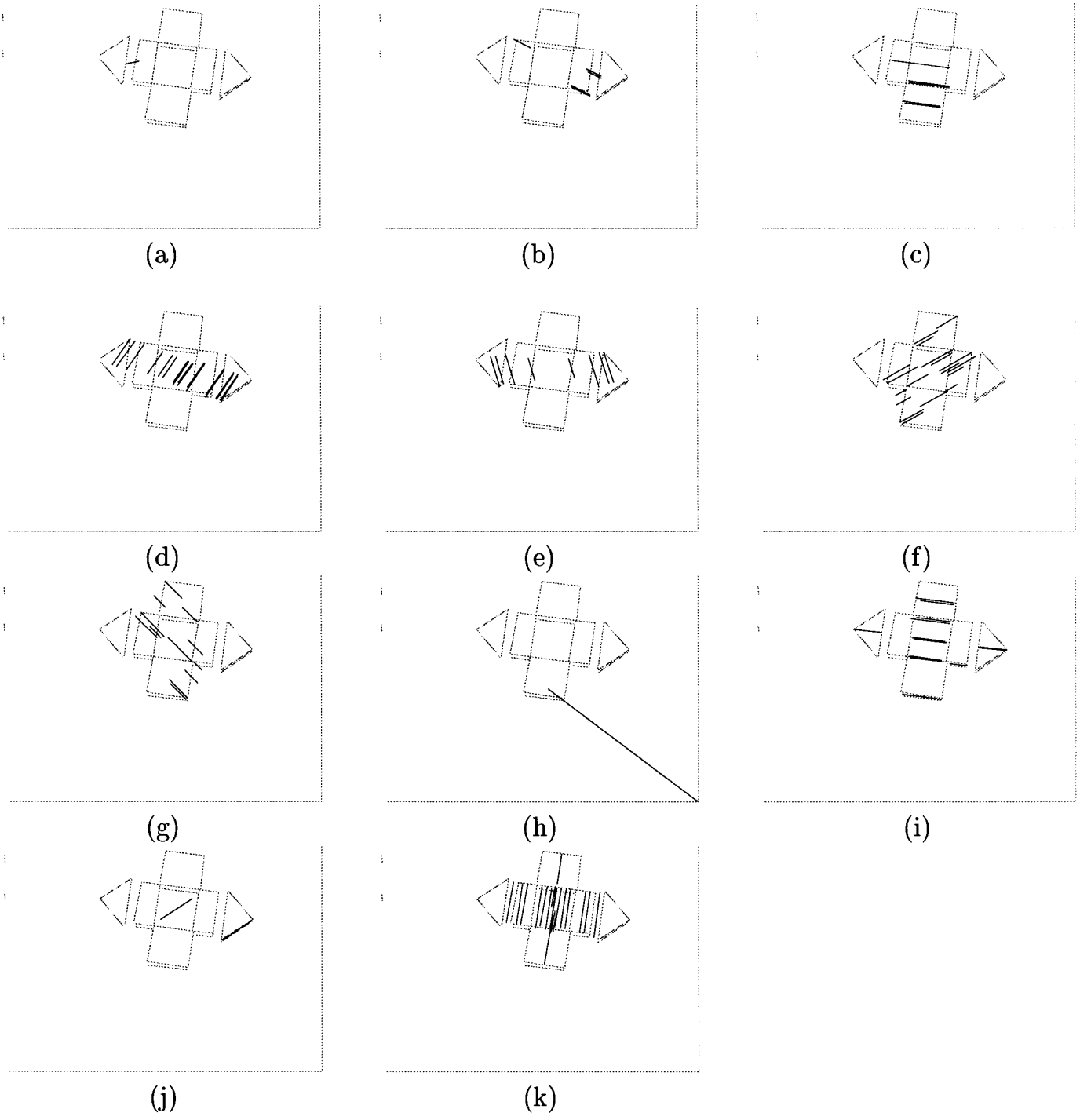

FIG. 26. Parallel symmetry axes. Most probable partition. This partition divides the set of axes into 11 groups shown in (a)–(k).

The marginal density of an observation is now

$$f(\mathbf{y}_k) = \int \frac{1}{(2\pi\sigma^2)^{|S_k|/2}} e^{g(S_k,\mathbf{u}_k)} f(\mathbf{u}_k)\, d\theta, \qquad (32)$$

where

$$g(S_k, \mathbf{u}_k) = -\frac{1}{2\sigma^2} \sum_{\mathbf{x}_i \in S_k} ((x_i - \bar{x})^2 \cos^2 \theta_k + (x_i - \bar{x})(y_i - \bar{y}) \cos \theta_k \sin \theta_k + (y_i - \bar{y})^2 \sin^2 \theta_k).$$

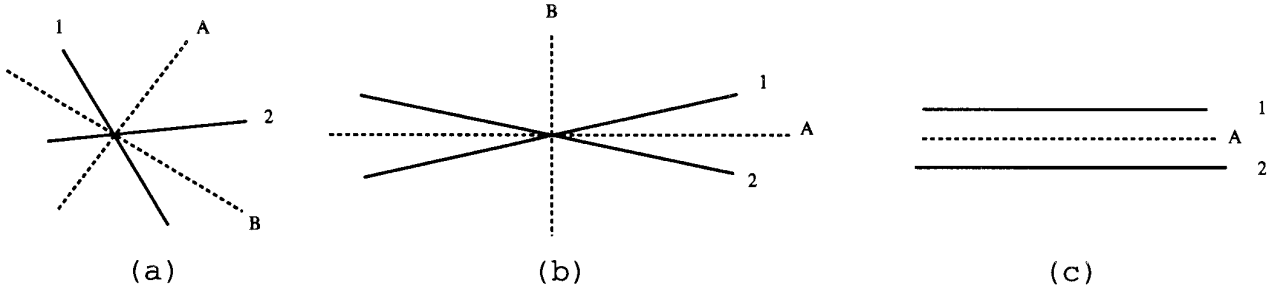

FIG. 27. Lines A and B are the bilateral symmetry axes for lines 1 and 2.

The models described in this section are not affected by the distance between segments. In particular, the model determines only the probability that segments are collinear, not how probable it is that the segments are connected.

A.2. *Approximations for Computing Symmetry Integral*

The quadruple integral in (26) must be calculated thousands of times for a typical image; thus, it is advantageous to find methods of reducing the complexity of computation. We now describe the simplifications that we have implemented.

The first reduction is the same simplification that was made in Section A.1 for lines. The best fitting line will always pass through the center of gravity of the sample points. We can eliminate d and Δ from the integral (26) in the same manner that we eliminated d in Section A.1. The new expression for $f(\mathbf{y}_1, \mathbf{y}_2 \mid H'_{12})$ is

$$f(\mathbf{y}_1, \mathbf{y}_2 \mid H'_{12}) = \int \left[\int f(\mathbf{y}_1 \mid \mathbf{u}_1) f(\mathbf{u}_1)\, d\alpha_2 \int f(\mathbf{y}_2 \mid \mathbf{u}_2) f(\mathbf{u}_2)\, d\alpha_2 \right] d\theta. \tag{33}$$

Our next approximation involves reducing the number of initial symmetry axes considered. Bilateral symmetry axes can be generated from every pair of lines by bisecting the angle between the lines, as shown in Fig. 27. In Figs. 27a and b axes A and B can be generated from lines 1 and 2. Notice that the axes pass through the point of intersection of lines 1 and 2 and that the two axes are perpendicular to each other. In Fig. 27c, we see that the parallel lines 1 and 2 yield only one axis, A.

In an image, most line segments will not intersect as in Fig. 28a. If we project segments 1 and 2 onto each axis, A and B, the projections do not overlap on A. Sometimes the projections do not overlap on either axis, as in Fig. 28b. For n input segments there are $2\binom{n}{2} = n(n+1)$ initial axes of symmetry. The axis onto which the projection of two symmetry lines does not overlap will not be a perceptually significant axis of symmetry. We can eliminate many initial axes by considering only those axes onto which the line segments overlap.

Our final approximation is to concentrate the integration on the range of angles over which there is a significant probability of the axis. For symmetries whose data sets have many points, the range of possible angles with significant probability will be very small. Many integrations can be avoided by observing that the ranges of axis angles do not overlap.

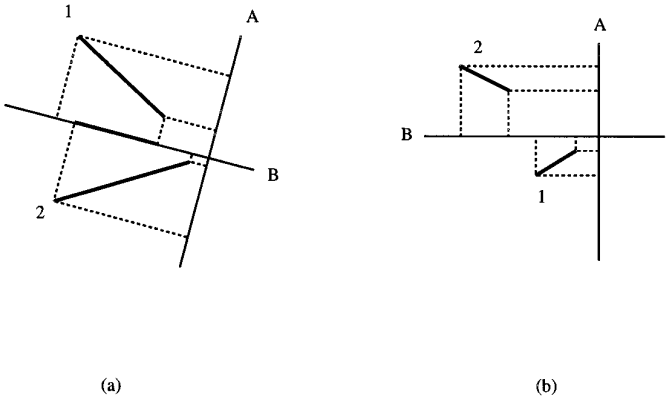

FIG. 28. Projections of segments 1 and 2 onto axes A and B.

REFERENCES

1. S. Ahmad, *VISIT: An Efficient Computational Model of Human Visual Attention*, Technical Report TR-91-049, Int'l Computer Science Institute, Berkeley, CA, 1991.
2. N. Ahuja and M. Tuceryan, Extraction of early perceptual structure in dot patterns: Integrating region, boundary, and component gestalt, *Comput. Vision Graph Image Process.* **48,** 1989, 304–356.
3. D. H. Ballard and C. M. Brown, *Computer Vision*, Prentice-Hall, Englewood Cliffs, NJ, 1982.
4. H. G. Barrow and J. M. Tenebaum, Interpreting line drawings as three-dimensional surfaces, *Artif. Intell.* **17,** 1981, 75–116.
5. I. Biederman, Recognition-by-components: A theory of human image understanding, *Psych. Rev.* **94**(2), 1987, 115–147.
6. M. Boldt, R. Weiss, and E. Riseman, Token-based extraction of straight lines, *IEEE Trans. Systems ManCybern.* **19**(6), 1989, 1581–1594.

7. J. B. Burns, A. R. Hanson, and E. M. Riseman, Extracting straight lines, *IEEE Trans. Pattern Anal. Mach. Intell.* **8**(4), 1986, 425–455.
8. J. Canny, A computational approach to edge detection, *IEEE Trans. Pattern Anal. Mach. Intell*. **8**(6), 1986, 679–698.
9. R. L. Castaño, *A Probabilistic Framework for Grouping Image Features*, Master's thesis, University of Illinois at Urbana-Champaign, 1994.
10. I. J. Cox, J. M. Rehg, and S. Hingorani, A Bayesian multiple hypothesis approach to edge grouping and contour segmentation, technical report, NEC, 1993.
11. R. O. Duda and P. E. Hart, Use of the Hough transform to detect lines and curves in pictures, *Commun. ACM* **15,** 1972, 11–15.
12. R. O. Duda and P. E. Hart, *Pattern Classification and Scene Analysis*, Wiley, New York, 1973.
13. G. Foresti, V. Murino, C. S. Regazzoni, and G. Vernazza, Grouping of rectilinear segments by the labeled Hough transform, *Comput. Vision Graph. Image Process*. **58,** 1994, 22–42.
14. S. A. Friedberg, Finding axes of skewed symmetry, *Comput. Vision Graph Image Process*. **34,** 1986, 138–155.
15. S. Geman, Experiments in Bayesian image analysis, *Bayesian Statist.* **3,** 1988, 159–172.
16. R. Glachet, J. T. Lapreste, and M. Dhome, Locating and modelling a flat symmetric object from a single perspective image, *CVGIP: Image Understanding*, **57**(2), 1993, 219–226.
17. R. M. Haralick and L. G. Shapiro, *Computer and Robot Vision*, Vol. 1, Addison–Wesley, Reading, MA, 1992.
18. D. H. Hubel and T. N. Wiesel, Receptive fields and functional architecture of monkey striate cortex, *J. Physiol.* **195,** 1968, 215–242.
19. J. N. Huddleston and J. Ben-Arie, Grouping edgels into structural entities using circular symmetry, the distributed Hough transform, and probabilistic non-accidentalness, *CVGIP: Image Understanding* **57**(2), 1993, 227–242.
20. K. Kanatani, *Geometric Computation for Machine Vision*, Oxford Univ. Press, Oxford, 1993.
21. G. Kanizsa, *Organization in Vision*, Praeger, New York, 1979.
22. S. M. LaValle and S. Hutchinson, A framework for constructing probability distributions on the space of image segmentations, *Comput. Vision Image Understanding* **61**(2), 1995, 203–230.
23. D. G. Lowe, Three-dimensional object recognition from single two-dimensional images, *Artif. Intell.* **31,** 1987, 355–395.
24. G. Marola, On the detection of the axes of symmetry of symmetric and almost symmetric planar images, *IEEE Trans. Pattern Anal. Mach. Intell.* **11**(1), 1989, 104–108.
25. D. Marr, *Vision*, Freeman, New York, 1982.
26. R. Mohan and R. Nevatia, Perceptual organization for scene segmentation and description, *IEEE Trans. Pattern Anal. Mach. Intell.* **14**(6), 1992, 616–635.
27. P. F. M. Nacken, A metric for line segments, *IEEE Trans. Pattern Anal. Mach. Intell.* **15**(12), 1993, 1312–1318.
28. V. S. Nalwa, Line-drawing interpretation: Bilateral symmetry, *IEEE Trans. Pattern Anal. Mach. Intell.* **11**(10), 1989, 1117–1120.
29. R. Nevatia and K. Ramesh Babu, Linear feature extraction and description, *Comput. Vision Graph Image Process.* **13,** 1980, 257–269.
30. I. Rock, *The Logic of Perception*, MIT Press, Cambridge, MA, 1983.
31. H. Rom and G. Medioni, Hierarchical decomposition and axial shape description, in *Proceedings, IEEE Conference on Computer Vision and Pattern Recognition, 1992*, pp. 49–55.
32. H. Rom and G. Medioni, Hierarchical decomposition and axial shape description, technical report, University of Southern California, Los Angeles, CA, 1992.
33. K. J. Moroney, S. M. LaValle, and S. A. Hutchinson, Methods for numerical integration of high-dimensional posterior densities with application to statistical image models, in *Proceedings, SPIE Conf. on Neural and Stochastic Methods in Image and Signal Processing, 1993*, pp. 292–303.
34. S. Sarkar and K. Boyer, Integration, inference, and management of spatial information using bayesian networks: Perceptual organization, *IEEE Trans. Pattern Anal. Mach. Intell.* **15**(3), 1993, 256–274.
35. S. Sarkar and K. Boyer, Perceptual organization in computer vision: A review and a proposal for a classificatory structure, *IEEE Trans. Systems Man Cybern.* **23**(2), 1993, 382–399.
36. S. Sarkar and K. Boyer, *Computing Perceptual Organization in Computer Vision*, World Scientific, Teaneck, NJ, 1994.
37. E. Saund, Symbolic construction of a 2-d scale-space image, *IEEE Trans. Pattern Anal. Mach. Intell.* **12**(8), 1986, 156–173.
38. A. Scher, M. Shneier, and A. Rosenfeld, Clustering of collinear line segments, *Pattern Recognit*. **15**(2), 1982, 85–91.
39. L. Spillman and J. S. Werner, *Visual Perception: The Neurophysiological Foundations. Academic Press, San Diego, 1990.*
40. H. Stark and J. W. Woods, *Probability, Random Processes, and Estimation Theory for Engineers*, Prentice–Hall, Englewood Cliffs, NJ, 1994.
41. A. Treisman and S. Gormican, Feature analysis in early vision: Evidence from search asymmetries, *Psychol. Rev.* **95**(1), 1988, 15–48.
42. J. Tsotsos, Analyzing vision at the complexity level, *Behav. Brain Sci.* **13,** 1990, 423–469.
43. D. Van Essen and C. H. Anderson, Information processing strategies and pathways in the primate retina and visual cortex, in *An Introduction to Neural and Electronic Networks* (S. F. Zornetzer, J. L. Davis, and C. Lau, Eds.), Academic Press, San Diego, 1990.
44. M. Wertheimer, Principles of perceptual organization, in D. Beardslee and M. Wertheimer, editors, *Readings in Perception*, pp. 115–135, VanNostrand, Princeton, NJ, 1958.
45. A. Witkin and J. Tenebaum, On the role of structure in vision, in *Human and Machine Vision* (J. Beck, B. Hope, and A. Rosenfeld, Eds.), pp. 481–543, Academic Press, San Diego, 1983.
46. T. Zielke, M. Brauckmann, and W. von Seelen, Intensity and edge-based symmetry detection with an application to car-following, *Comput. Vision Graph Image Process.* **58**(2), 1993, 177–190.
47. S. W. Zucker, The diversity of perceptual grouping, in *Vision, Brain, and Cooperative Computation*, pp. 231–261, MIT Press, Cambridge, MA, 1990.