January 1995

Segmentation Modeling

Anthony J. Hoogs
University of Pennsylvania

Ruzena Bajcsy
University of Pennsylvania

Pamela Lechleider
University of Minnesota

University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-95-16.

Segmentation Modeling

Abstract

A description of a model for accumulating segmentation information over multiple images is presented. The model combines geometric information with segmentation information to improve predicted appearance of objects. An extensive example using polyhedral object representations is described. Statistical results are shown over a set of 40 real images, verifying the underlying assumptions of the segmentation model concerning spatially distributed segmentation information over object edges. Evidence is accumulated in support of the hypothesis that segmentation varies significantly because of the interference between scene objects as well as viewpoint and illumination.

Comments

University of Pennsylvania Department of Computer and Information Science Technical Report No. MS-CIS-95-16.

Segmentation Modeling

MS-CIS-95-16
GRASP 391

Anthony Hoogs
Ruzena Bajcsy
Pamela Lechleider

University of Pennsylvania
School of Engineering and Applied Science
Computer and Information Science Department
Philadelphia, PA 19104-6389

March 1995

Segmentation Modeling

Anthony Hoogs
University of Pennsylvania,
Martin Marietta Corporation
P.O. Box 8048
Philadelphia, PA 19101
hoogs@mdso.vf.ge.com

Ruzena Bajcsy
GRASP Lab
University of Pennsylvania
3401 Walnut Street
Philadelphia, PA 19104
bajcsy@cis.upenn.edu

Pamela Lechleider
Computer Science Department
University of Minnesota
Mineapolis, MN 55455
lechleid@grip.cis.upenn.edu

Abstract

A description of a model for accumulating segmentation information over multiple images is presented. The model combines geometric information with segmentation information to improve predicted appearance of objects. An extensive example using polyhedral object representations is described. Statistical results are shown over a set of 40 real images, verifying the underlying assumptions of the segmentation model concerning spatially distributed segmentation information over object edges. Evidence is accumulated in support of the hypothesis that segmentation varies significantly because of interference between scene objects as well as viewpoint and illumination.

Keywords: Shape and object representation, learning in computer vision, segmentation.

Segmentation Modeling

Abstract

A description of a model for accumulating segmentation information over multiple images is presented. The model combines geometric information with segmentation information to improve predicted appearance of objects. An extensive example using polyhedral object representations is described. Statistical results are shown over a set of 40 real images, verifying the underlying assumptions of the segmentation model concerning spatially distributed segmentation information over object edges. Evidence is accumulated in support of the hypothesis that segmentation varies significantly because of interference between scene objects as well as viewpoint and illumination.

Summary Page

1. The original contribution of this work is its merging of segmentation information into geometric descriptions over multiple images. Information about segmentations is accumulated in geometric models, allowing analysis of segmentation behavior and more accurate predictions of object appearance. To our knowledge, there is no similar treatment of segmentation data anywhere in the literature.

2. This research can potentially influence much research in computer vision. One of the primary bottlenecks preventing progress in the field is segmentation, and we believe that we have developed an effective technique for managing inconsistent segmentations under a broad range of circumstances.

3. This work is relatively different, and we do not know of other research that is pursuing the same course. Efforts under the direction of B. Haralick at the University of Washington are similar, in that they are computing statistics about edges on multiple images. Ram Nevatia's group at USC is using projected model edges to verify object presence in a scene, but they do not use information from multiple images. Many other object recognition systems have used some sort of model verification as a hypothesis confirmation step, but again these methods are much simpler than those we employ.

4. Other researchers may utilize our work by incorporating segmentation modeling into the segmentation steps of their algorithms, if possible. We envision applications in areas that have or develop geometric models over a set of imagery, such as change detection, image registration, object recognition and motion. Given a set of training data, our system can predict when a particular edge is expected to appear in an image. This could be a powerful tool for any system that depends on finding edges of known objects or classes of objects.

1 Introduction

This paper describes an approach to modeling the behavior of image segmentations in an attempt to characterize the effect of various imaging and scene conditions on the performance of segmentations of known objects. As described in our previous work [4], the resulting *segmentation models* indicate under what conditions segmentations of specific scene features are expected to be strong or weak. Higher-level systems may incorporate these segmentation models to improve their performance in such areas as change detection, cloud detection, image registration and object recognition.

The basic idea of segmentation modeling is that the variation of segmentation due to non-geometric image features, and the instability of segmentation caused by non-linear thresholding, can be accounted for by dynamically modeling segmentation behavior over a range of imaging conditions on a per-object basis. These segmentation models are incorporated directly into the geometric representations of objects so that the appearance of an object in an image can be predicted based on previous segmentation information.

Object representations used in computer vision fall generally into two categories, model-based and view-based. Segmentation modeling combines the two, offering the advantages of each. Model-based representations such as superquadrics, polyhedra and generalized cylinders consist of geometric definitions of object shape. Prediction of object appearance is often performed by projecting three-dimensional geometric features such as edges into the image plane. Some systems [13] use object geometry to identify features most likely to appear in a segmentation. However, virtually all of these methods rely purely on geometry to match object features to image features. Other nonstandard model-based systems, such as those based on projective invariants, are also restricted to using object geometry, not observed appearance. These methods all have the same weakness – flaws in object segmentation cause significant performance degradation.

View-based representations such as [10] usually do not maintain geometric models of objects. Object models are learned through an exhaustive set of training data showing all relevant views of an object, possibly including all expected illumination conditions. Thus, view-based systems can implicitly incorporate segmentation information into object models, and some exhibit accurate recognition over sizable object libraries. However, they perform poorly at extracting individual objects from scenes containing multiple objects. Also, they require large image training sets for model acquisition.

Segmentation modeling is a hybrid of the two types of object representation, incorporating view-based information into object geometry. It contains advantages of each, while eliminating some of the major disadvantages. Segmentation models are learned over a set of training imagery, so that significant aspects of object appearance are acquired in actual imagery. This accounts for many factors leading to poor segmentation that the traditional model-based paradigm ignores. Unlike view-based representations, though, the geometry of object models is retained, providing the geometric object descriptions necessary for many applications. In addition, the training set does not need to be particularly large, as discussed below.

Although we do not know of anyone who is pursuing analogous research, there are related efforts that are relevant to our work. Thornton et al. have compiled an exhaustive compilation of statistics for various types of edges in RADIUS imagery [12]. These values are similar to our own computations, although they do not accumulate information across images. Bejanin et al. are studying refined model verification [11], where the image is registered to a model and the algorithm detects the presence of modeled buildings. This work has much in common with our first level of processing, but again it operates on segmentations of only one image.

Segmentation modeling is part of a broader effort in which we are attempting to understand the relationship between scene complexity and image interpretation [4]. Given a specified amount of *a priori* information about a scene of a specified complexity, we desire to know how effectively a model of the imaging process can determine the image features of the scene, and which factors have a strong influence on the accuracy of that interpretation. The imaging process model consists of two levels; the first directly applies knowledge of the scene, and the second constructs of segmentation models attached to geometric representations.

Our results indicate that the formulation of segmentation modeling presented here is sufficient to evaluate the relative effects of different imaging conditions both qualitatively and quantitatively. Results are shown on real and synthetic data in Section 4. We have explored various types of segmentation models, and performed statistical analysis on them as a means of quantitative comparison.

From these results, we conclude that knowledge of scene geometry, viewpoint and illumination is not sufficient to explain the behavior of segmentations. Other scene parameters, such as object albedo and object interactions, must also be considered to achieve complete understanding of extracted image features.

This paper describes our formulation for segmentation models. The general model described in [4] as applied to polyhedral object representations is reviewed in Section 2. This section is in part a summary of our previous paper, with significant additions to edge segmentation models. Section 4 discusses new experimental results and analysis of constructing segmentation models on real and synthetic data.

2 General Segmentation Models

In our previous work, we describe a general approach to modeling segmentations with geometric representations, as part of a two-level model of the imaging process. The goal of this work is to determine whether it is possible to explain all detected image features in the scene, given the amount of information in the two levels and the allowable error tolerances.

In this paper, we expand upon that work by examining the case of polyhedral object representations in greater detail, concentrating on edge models. Polyhedral representations are a natural choice for segmentation models, since the edges and faces of polyhedra provide a straightforward geometric mapping of segmentation features to object features.

In [4] we define a two-level model of the imaging process. Level I consists of the specified information about the scene and imaging conditions, while Level II consists of segmentation models for objects in the scene developed over a set of images. The amount of image information accounted for by each level of the model depends on the detail of the supplied information used in Level I. If the Level I model is very accurate, containing precise information about complete scene geometry, imaging conditions, surface materials, and so on, then the Level II model may only be useful for estimating segmentation errors and noise. However, having such a model is very unlikely in most applications, including MSE. Usually, the Level II model is necessary for modeling segmentation effects due to unmodeled albedo changes, detailed geometry, and segmentation inconsistencies not accounted for in Level I.

In this paper, we assume that Level I information is approximately that supplied by the Model-Supported Exploitation (MSE) framework [3, 5, 9] developed under the RADIUS project. MSE supplies coarse, CAD-based geometric models of objects in the scene, and images are registered to the scene so that the object models overlay the image data portraying them. Illumination is assumed to originate from the sun, at infinite distance. Occlusions and shadows resulting from modeled objects may be predicted, but this is not implemented in the current system.

Given this Level I information, much feature data extracted from images of the scene can be explained using the geometric model. However, in a typical image there is still a large portion of the extracted image features that are not explained, including those created by segmentation instability. There are also features predicted by the geometric model that do not appear because of poor contrast, occlusion or shadow interference. The detected image features and the missing image features must be accounted for in the Level II model; they include features derived from detailed geometry, albedo changes and various kinds of segmentation errors as well.

Because of their nonlinear nature, segmentations are highly sensitive to minor variations in image intensity. Regardless of the detail of the Level I model, segmentations of the same scene may contain differences that are only explained by modeling segmentation errors in some fashion. Although well-accepted in the community, this belief has been verified on synthetic images of a simple scene (see Section 4).

Subsystems to explain each type of image feature data are defined for each level. Level I Explanation directly uses the geometric information in object models to identify image features corresponding to the model (e.g., geometric edges, shadows) by correlation techniques. Level II Explanation involves the creation and updating of segmentation models to explain the image features unaccounted for in Level I Explanation and to account for segmentation errors.

For the purposes of this paper, we are explicitly modeling the effect on image features of model geometry. We do not predict or identify segmentation elements generated by shadows, shading, albedo changes and noise. Rather, we allow these and any other sources of segmentation features to accumulate in the segmentation model. This situation represents a balance between Levels I and II that is weighted toward Level II, placing more emphasis on the segmentation model for experimentation.

The remainder of this paper focuses on segmentation models as implemented along the edges of polyhedral objects.

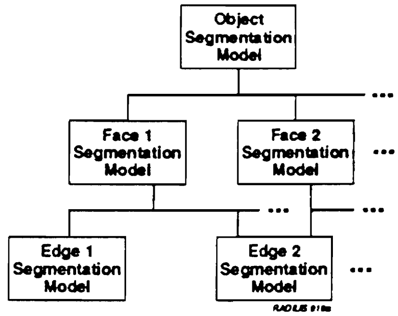

Figure 1: The hierarchical Object Segmentation Model.

2.1 The Polyhedral Object Segmentation Model

Polyhedra are an obvious choice for three-dimensional object representation, since they allow an intuitive approach for incorporating segmentation models. The limitations of polyhedra, however, are well understood by the community, and the segmentation model may be generalized to algebraic or other representations in the future. Polyhedra support reasonable modeling of most buildings and man-made structures.

The segmentation model of a polyhedron, the Object Segmentation Model, is a hierachical organization composed of segmentation models of object features and the relationships between them, as shown in Figure 1. Each object feature, i.e., faces and edges, has its own segmentation model specialized to the feature. Higher level segmentation models combine those models below them, maintaining information about statistical correlations between lower-level models. Depending on the size of the object, there might be one or more high-level models. The current system implements the Object Edge Segmentation Model as described below, and a preliminary version of a higher-order model across multiple edge models.

3 The Object Edge Segmentation Model

The two levels of the imaging process model apply directly to polyhedral representations. When a new image is input to the system, the levels of explanation are applied in order to incorporate or explain the new image data, depending on whether the system is in training or operational mode respectively. The input to Level I is the new image and the object models, and its output is a set of line segments correlated to modeled object edges. Combined with the image, this is input to Level II where it is used to update or validate the segmentation model.

The following subsection treats Level I explanation for the edge-based polyhedral model, and the remaining subsections discuss the object edge segmentation model.

3.1 Level I Explanation

When a registered image is presented to the system, the first stage of processing is to extract salient lines from the image. We use a line finder built upon the Canny edge detector [8, 2] to extract segmentation features.

Level I Explanation for geometric data using polyhedral models is relatively straightforward. To identify extracted image features that correspond to known geometric models, the edges of modeled objects are projected onto the image through the known camera model of the image. A linear correlation technique is applied to associate projected object edges and extracted segmentation edges based on distance and orientation as follows:

1. Given a projected object edge e_o and an observed edge e_s, the minimum distances D_1 and D_2 are computed between the endpoints of e_o and the segment e_s.

2. The sum $D_1 + D_2$ is compared to a threshold; if it is greater than the threshold, then e_s is determined to be not correlated.

3. The angular difference between e_o and e_s is computed. If it is greater than a threshold, then e_s is determined to be not correlated.

4. Otherwise, e_s is determined to be correlated to e_o.

The values used for the two thresholds are 25 pixels and 15 degrees over all imagery. While some benefit might be accrued from manipulating these parameter values, we elected to study the system without doing so to see if it proved effective without manual manipulation. Based on our results, we do not believe that these parameter values need to be adjusted on a per-image basis.

Level I Explanation may cause some errors because of ambiguities in the association of segmentation features to modeled features within the error tolerance of image registration. These ambiguities may be caused by incidental alignments of shadows, object edges, albedo regions or shading effects.

3.2 Level II Explanation

Given the correlated edges output from Level I, Level II uses the edges plus the image to update its segmentation models when in the training phase, or to make some other estimate of the image such as change detection when in the operational phase. Whereas Level I Explanation relies heavily on *a priori* information, Level II Explanation primarily depends on data extracted from images to build statistical models of the expected behavior of segmentations.

The Object Edge Segmentation Model (OESM) is a statistical generalization of observed edge characteristics, or *attributes*, along an object edge. A complete polyhedral segmentation model would include an OESM for every edge of the object. Each object edge is treated as a unit interval. and information computed from image segmentations is mapped into that interval according to the projection of the object edge into the image and the positions of segmentation features.

3.2.1 Edge Attributes

Standard line detection processes are enhanced with a set of edge attributes computed from the image to incorporate more image information into the edge detection framework. The concept of augmenting the dimensionality of edge segmentation is based on edge labeling techniques [6, 1], but unlike some existing systems we do not depend on physics-based models requiring color or specialized imaging. The results presented in Section 4 indicate that certain edge attributes make a strong contribution to the segmentation model.

Each computed attribute has its own set of models that is maintained separately from the other attributes. In other words, each OESM has a dimension for each attribute. Currently, the attributes included in the segmentation models are *correlation* and *gradient magnitude*. We anticipate adding a larger attribute suite in the future, but these two attributes have proven to be sufficient to show the utility of the segmentation modeling paradigm.

The correlation attribute is the proportion of the object edge covered by observed edges, i.e., how well segmentation features match the object edge geometrically. Correlation is computed using the output from Level I, the correlated extracted line segments, and mapping them onto the object edge.

The gradient magnitude is also computed from the extracted segments. For a segmentation feature f, the gradient is measured by taking the difference of the average intensities of rectangular regions on either side of the edge. Each region is defined to be along the length of edge, with a width of 0.1 times the edge length. This method was chosen to utilize the output of the line extractor, rather than the output of the Canny edge detector which would produce a gradient value for each pixel. The technique used to combine the gradient attribute across multiple extracted segments correlated to a common object edge is the average.

3.2.2 The Segmentation Model

Once the segmentation attributes are chosen, the segmentation model can be defined in terms of those attributes. The segmentation model is formulated as an enhancement of the geometric object model that incorporates statistical segmentation information for each attribute. Each geometric feature of an objeet, such as a face or edge, has an associated segmentation model specialized to that feature.

To accurately predict segmentations, the segmentation model must take into account all factors that strongly influence segmentation behavior. One of the well-known problems with segmentations is that they vary significantly across images of the same scene taken with different viewpoints and/or illumination conditions. Other factors also influence segmentations of images showing natural scenes, such as weather conditions, seasonal variations, etc. For the purposes of this paper, we assume that viewpoint and illumination orientation, termed the *imaging parameters*, are the dominant factors causing differences in segmentations. The generalization of the model to account for additional imaging parameters is straightforward in most cases, but our experiments provide empirical support for the hypothesis that viewpoint and illumination are two of the most important factors.

To account for the effect of viewpoint and illumination on segmentation, the segmentation model is parameterized by the imaging parameters. Each segmentation model is a mapping M that maps imaging parameters into segmentation values for an object feature (edge). If the segmentation values are represented as a vector $\mathbf{A}$ of feature attributes (or distributions over each attribute), then the mapping M is defined as

$$M : \mathbf{S} \times \mathbf{V} \rightarrow \mathbf{A}$$

where $\mathbf{V}$ and $\mathbf{S}$ are imaging parameters (viewpoint and solar orientation) affecting the segmentation.

In the case of the OESM, $\mathbf{A}$ is defined to be $(c, g)^T$ where c is the correlation attribute and g is the gradient attribute. Both c and g may be single values or distributions along the object edge, as discussed below.

The purpose of M is to predict the values of $\mathbf{A}$ given previously unseen values of $\mathbf{V}$ and $\mathbf{S}$. For each object feature, M must be learned dynamically over a set of images. Given a set T of images showing an object feature f from various viewpoints and illumination positions, we desire to recover a segmentation model $M_f(T; \mathbf{V}, \mathbf{S})$ describing the characteristic segmentation of f. The segmentation model of a feature f can be thought of as the generalization of segmentations of f found in multiple images, e.g., a vector of representative edge attribute values.

The number of parameters influencing M_f is considerable. $\mathbf{S}$ is a vector, described by two angles θ and ϕ. $\mathbf{V}$ is the viewpoint position and orientation, described by three angles, α, β, γ, and a position x_v, y_v, z_v. For object edge f over image set T, we then have

$$M_f(T; \mathbf{V}, \mathbf{S}) = M_f(T; \alpha, \beta, \gamma, x_v, y_v, z_v, \theta, \phi)$$

This level of parameterization can be significantly reduced by reformulating the problem slightly. The viewpoint position is not significant relative to a single edge, but the angle between the viewpoint and the edge is important. In a perspective projection model, the viewpoint position will affect this angle. However, if a viewpoint vector is computed for the edge given the projection model, then the type of projection model used is no longer needed. Hence, by defining the segmentation model in terms of a vector (α_v, β_v) indicating the viewpoint angle relative to the object edge the dependence on viewpoint position is removed. This simplifies M_f to four degrees of freedom, $M_f(T; \alpha_v, \beta_v, \theta, \phi)$.

The OESM may be defined in a number of ways, depending on how the OESM is related to the imaging parameters and how observed edge information (segmentation features) is mapped to the object edge. For the dependence of the OESM on imaging parameters, there are two general possibilities:

Independence – The OESM assumes that all edge attributes are independent of imaging parameters. One profile (unified or distributed) is maintained for predicting edge attribute values for all imaging parameters.

Dependence – The OESM assumes that edge attributes are dependent on (parameterized by) imaging parameters.

The utility of the independent model depends on how well the independence assumption holds, or how widely the attribute values vary across the range of imaging parameters to be used. If the values are somewhat uniform, then applying the independent model would be more efficient in terms of storage and execution time than using a more complicated parameterization. For most applications, it is expected that the independent model will not be sufficient to model segmentation behavior.

If the OESM depends on the all of the imaging parameters, then it must be parameterized by at least four variables, $\alpha_v, \beta_v, \theta$ and ϕ. While each of these is known for each image, a straightforward tesselation of the complete parameter space may require a large image training set, depending on the resolution of tesselation. Note, however, that each image is used to update the segmentation model for each scene feature visible in that image.

The OESM may be simplified by reducing the partitioning of the $(\mathbf{S}, \mathbf{V})$ space based on the geometry of polyhedra and knowledge of the imaging process. For an edge of a building in a scene, the values of segmentation attributes are expected to change gradually or remain relatively constant as the viewpoint is moved over most of the viewing hemisphere. At certain points, though, as the edge transitions from being an occluding edge to an interior edge or vice versa, the edge attribute values should change rapidly. These transition regions correspond to borders on the aspect graph

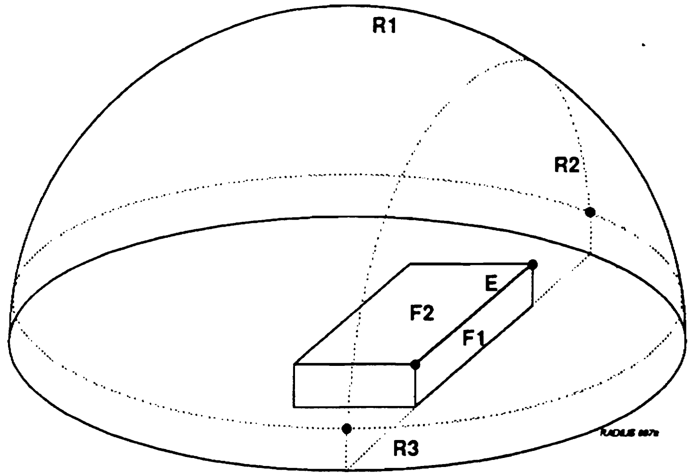

Figure 2: The partitioning of $(\mathbf{S}, \mathbf{V})$ space using the aspect graph. The edge **E** is used to define regions **R1**, **R2** and **R3**.

of the object. The illumination orientation should produce a similar effect on the attribute values.

The $(\mathbf{S}, \mathbf{V})$ space is then partitioned into nine regions, three regions each for **S** and **V**, as shown in Figure 2. Furthermore, in the MSE framework the number of valid partitions can be reduced to four for horizontal edges such as the edges of rooftops. The situations where only the vertical face bordering the horizontal edge is visible or illuminated are very unusual, and can be ignored. This simplifies the space to four partitions, or *imaging modes* for these types of edges, which are often the most salient features in a scene.

Figure 3 shows the geometry of the four modes. Instead of giving ranges on values of **S** and **V**, the imaging parameters have been translated into geometric scene parameters according to the enhanced edge E.

Irrespective of whether the OESM is dependent or independent, the attributes computed from observed edges may be associated to the corresponding object edge in one of two ways:

Unified Model - The OESM contains a single profile of a typical observed edge segment along the object edge. The observed edge attributes from each correlated observed edge are mapped to

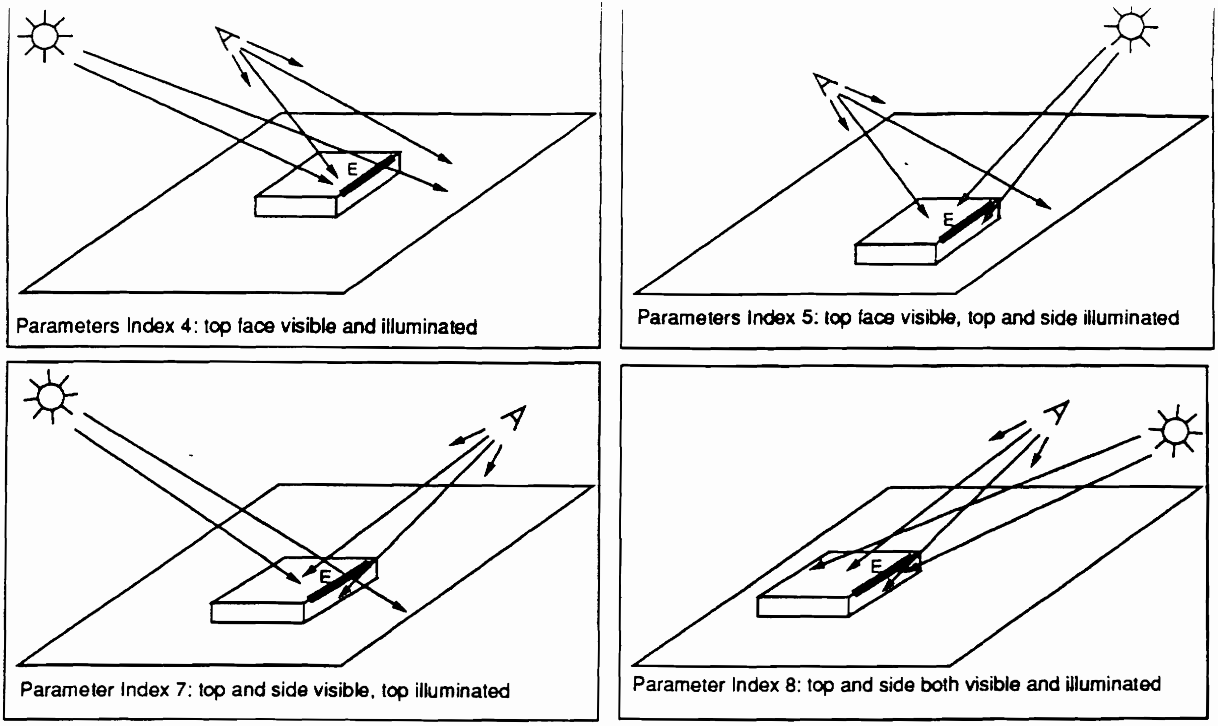

Figure 3: Prototypical imaging conditions for the four common imaging modes. The edge of interest is enhanced and labeled E.

the complete object edge without regard to where the observed edge is relative to the object edge (beyond the fact that it is correlated).

Distributed Model – The OESM contains a partitioning of edge attribute values along the object edge, i.e., the object edge is discretely quantized into equally-sized intervals, each containing its own attribute vector.

The unified model is considerably easier to build and maintain, since only a single value or distribution is stored for each attribute for each mode, but it is most appropriate only if the attribute values for observed edges are relatively uniform along the object edge. The distributed model can be thought of as a set of unified models, one for each interval defined over the object edge.

Any combination of the two OESM factors is possible. An OESM may be unified and dependent, such that each of the modes has only one value stored per attribute. Conversely, the distributed independent model maintains a single profile along the edge for each attribute regardless of $\mathbf{S}$ and $\mathbf{V}$.

The most general case is the distributed dependent model, in which each value of $\mathbf{S}$ and $\mathbf{V}$ maps into a set of attribute values, one per attribute for each interval. This implies that for each object edge f divided into n intervals and having k attributes, the storage required is $n \times k \times N$, where N is the number of partitions of the $(\mathbf{S}, \mathbf{V})$ space and each attribute is represented by a single value. $k \times N$ segmentation models are maintained, each with n values per attribute along the object edge.

3.2.3 Dynamic Construction of the OESM

With the segmentation model definition specified, the process of dynamically determining the relationships between imaging parameters and segmentation attributes can be addressed. As a segmentation model develops over a set of training imagery, information on the segmentation features generated by the corresponding scene feature is accumulated. The segmentation model learns the characteristics of segmentations of the object.

The technique used to build the segmentation model should perform well on small sample sizes and should allow incremental updates as new images are acquired. Using statistical methods for developing segmentation models has the advantage that segmentation errors and spurious, atypical segmentations tend to be filtered out over a set of imagery. Because segmentation information is combined across multiple images, there is very little dependence on specific parameter values used in segmentations – the same parameter values can be used on every image in the training set. Also, dynamic construction allows objects with arbitrarily complex surface features to be modeled, providing robustness on images of real buildings in outdoor settings.

Given an image I taken from viewpoint $\mathbf{V}_I$ with solar orientation $\mathbf{S}_I$, the segmentation model M for feature f is updated as follows. Let A_M be the representative attribute vector corresponding to $\mathbf{V}_I$ and $\mathbf{S}_I$. The attribute vector A_I for image I is computed, and used to update A_M:

$$A_M = \Delta(A_M, A_I)$$

where $\Delta(A_M, A_I)$ is a function defined by the learning paradigm. If I is the first image in the training set, then A_M is initialized to A_I.

The Δ function used to update the OESM is a weighted average,

$$A_M = \frac{i}{i+1} A_M + \frac{1}{i+1} A_I$$

where i is the number of images used to build A_M, not including the new image I. The average was chosen over other Δ functions for a number of reasons. A Kalman filter approach is sensitive to the initial image used to create the segmentation model, giving it an inappropriate importance. Over time this effect disappears, but within the data sets we used (40 images) the influence of a poor initial image was still apparent. The median was also tested, but it proved to be less informative because the correlation attribute tends to be exactly zero or one in most edge intervals derived from I. Thus the correlation profiles contained mostly zeroes and ones, which did not provide accurate estimates of probabilities.

Once the segmentation model is constructed, it may be applied and updated simultaneously as part of a system performing a segmentation-related task. Training and usage may occur on the same image, provided that some feedback is given regarding the true state of any new image introduced to the system.

The complete process for building segmentation models on object edges can now be outlined for the distributed dependent OESM. Given a training set T containing m images showing the same objects without physical change, but from a variety of viewpoints and illumination conditions,

1. For each image I in T, perform line extraction yielding a set of extracted segments S_x.

2. For each object edge E_o, find the set of segments S_o, $S_o \subseteq S_x$, that are correlated with the projection of E_o in the image plane of I.

3. For each segment in S_o, compute the gradient attribute.

4. Determine the imaging mode q based on the viewpoint and illumination of I.

5. For each segment in S_o, compute the gradient attribute.

6. For each interval i in E_o, compute the combined correlation attribute c_i across S_o.

7. Using the combined correlation attribute, for each i compute the combined gradient attribute g_i across S_o.

8. For each interval i in E_o, update each attribute in mode q of E_o using the Δ function:

$$E_{o,q}(i,c) = \frac{N}{N+1} E_{o,q}(i,c) + \frac{1}{N+1} c_i$$

$$E_{o,q}(i,g) = \frac{N}{N+1} E_{o,q}(i,g) + \frac{1}{N+1} g_i$$

where $E_{o,q}(i,g)$ is the value of the i^{th} interval of the gradient attribute of edge E_o in mode q.

The combined correlation attribute is computed using the geometry of the extracted edges. When multiple segmentation edges are correlated to a common object edge, their correlation intervals may overlap, or there may be gaps in coverage of the object edge. The combined correlation attribute is computed by merging the correlation intervals of all of the extracted edges correlated to one object edge.

Similarly, the combined gradient attribute is computed by taking the correlation-weighted average of the gradient values of any overlapping extracted edges. For example, if an interval i is correlated to two extracted edges which have correlation values of 1.0 and 0.72, with gradient values of 23.2 and 36.1, then the combined gradient value of i is $((1.0)(23.2) + (0.72)(36.1))/(1.0 + 0.72) = 28.6$.

4 Experimental Results

To validate the concept of segmentation modeling, we performed experiments on a set of real images and a smaller set of synthetic images. For each data set, distributed, dependent segmentation models were computed for a number of object edges using two edge attributes:

1. correlation between the observed edges and projected object edges;
2. gradient magnitude of observed, correlated edges.

Note that the gradient attribute is dependent on the correlation attribute.

The real images are the 40 images of the Model Board 2 (MB2) site from the RADIUS dataset. A representative image giving an overview of the site is shown in Figure 4. This site contains a number of buildings of various sizes and shapes, some close together and others isolated. We selected five buildings such that three are close enough to interfere with one another, and computed the segmentation models of their rooftop edges, yielding 28 edge segmentation models each using the 40 images. The chosen buildings and edges are shown in Figure 5; each building and edge is labeled for cross-referencing. The direction of each edge is specified to indicate the ordering of edge elements in the segmentation models.

The imaging parameters of the MB2 images vary widely over the viewing hemisphere. Illumination orientations are restricted to a broad arc because of solar geometry, but the viewpoints cover a broad range. All images maintain the aerial imagery paradigm, in that there are no ground-level

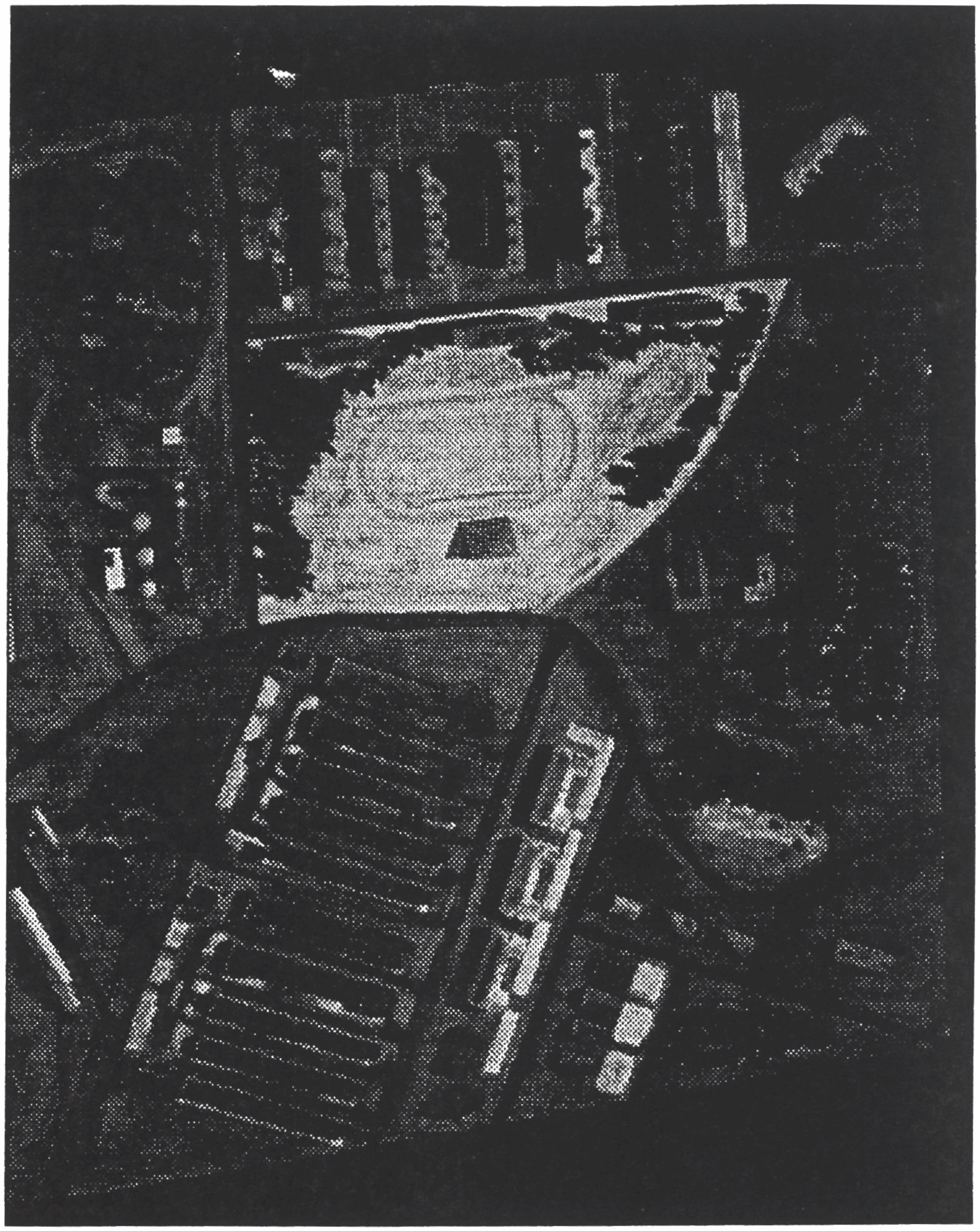

Figure 4: The Model Board 2 site.

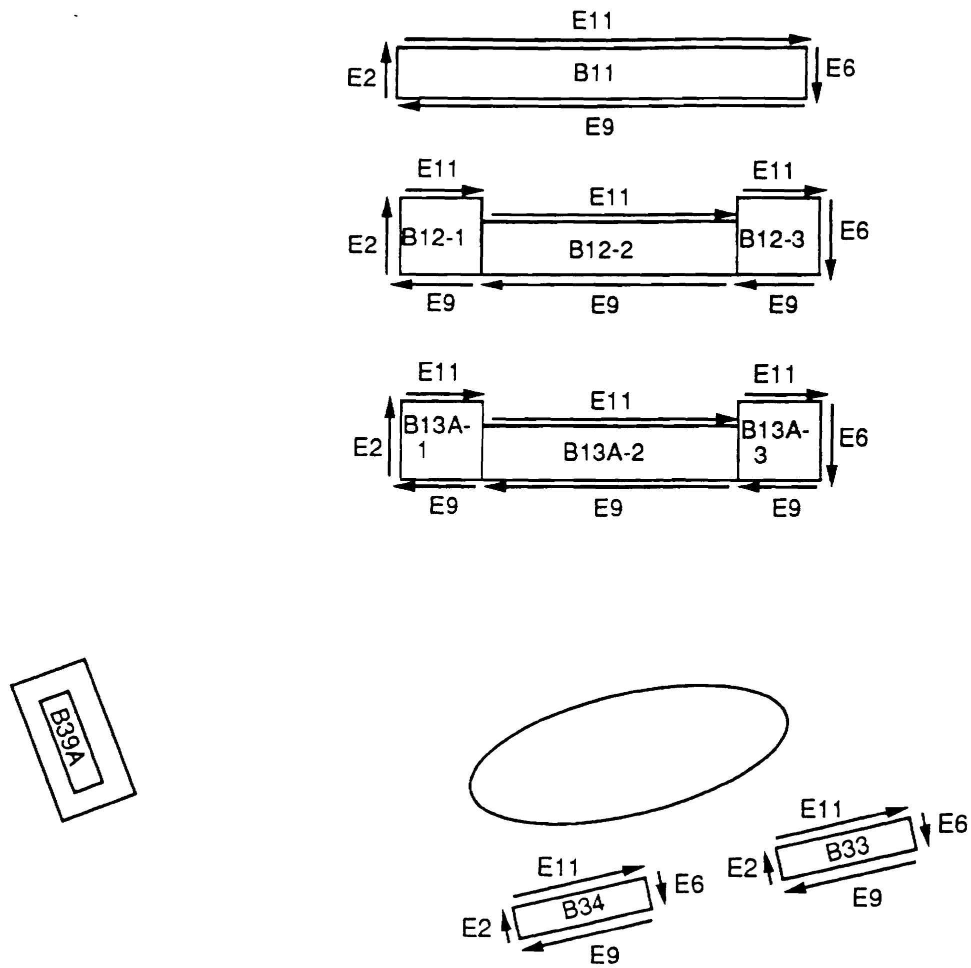

Figure 5: The Model Board 2 buildings and edges used in the experiments.

viewpoints and the maximum angle from the vertical of any view vector is approximately 60 degrees.

As expected, the 40 images resulted in the construction of distributed segmentation models in at most four of the possible nine modes for each edge. Most of the 28 edge models had four modes with at least four images in the "smallest" mode; the "largest" mode typically contained 16 images.

For both of the observables, edge correlation and gradient magnitude, we have computed statistics characterizing the different types of edge segmentation models – uniform vs. distributed and independent vs. dependent. These results on the MB2 data are presented in the next two sections. The combined results on the synthetic data are discussed in the third section.

4.1 The Edge Correlation Attribute

The value of the correlation attribute may be interpreted as the probability that an interval of an object edge will match an extracted segmentation edge. In other words, it is an estimate of the probability that the object edge will give rise to a segmentation edge in an image.

The correlation data gathered from the 40 MB2 images is presented graphically along modeled object edges of Buildings 11, 12 and 13 in Figure 6. Each edge segmentation model contains up to four "occupied" modes; these are shown in the figure in the same order that they appear in Figure 3. Note that the ranges of the axes of the plots is non-uniform (the correlation attribute ranges between 0 and 1).

The distributions along the edge within each mode make a strong argument for maintaining a distributed model over a uniform model. The uniform model is only appropriate if the attribute values are constant along the profile, and it is obvious that in most cases they are not. Combining the non-uniform values into a single number would lose important information by making an unsupported assumption of uniformity, and the resulting number may not be meaningful.

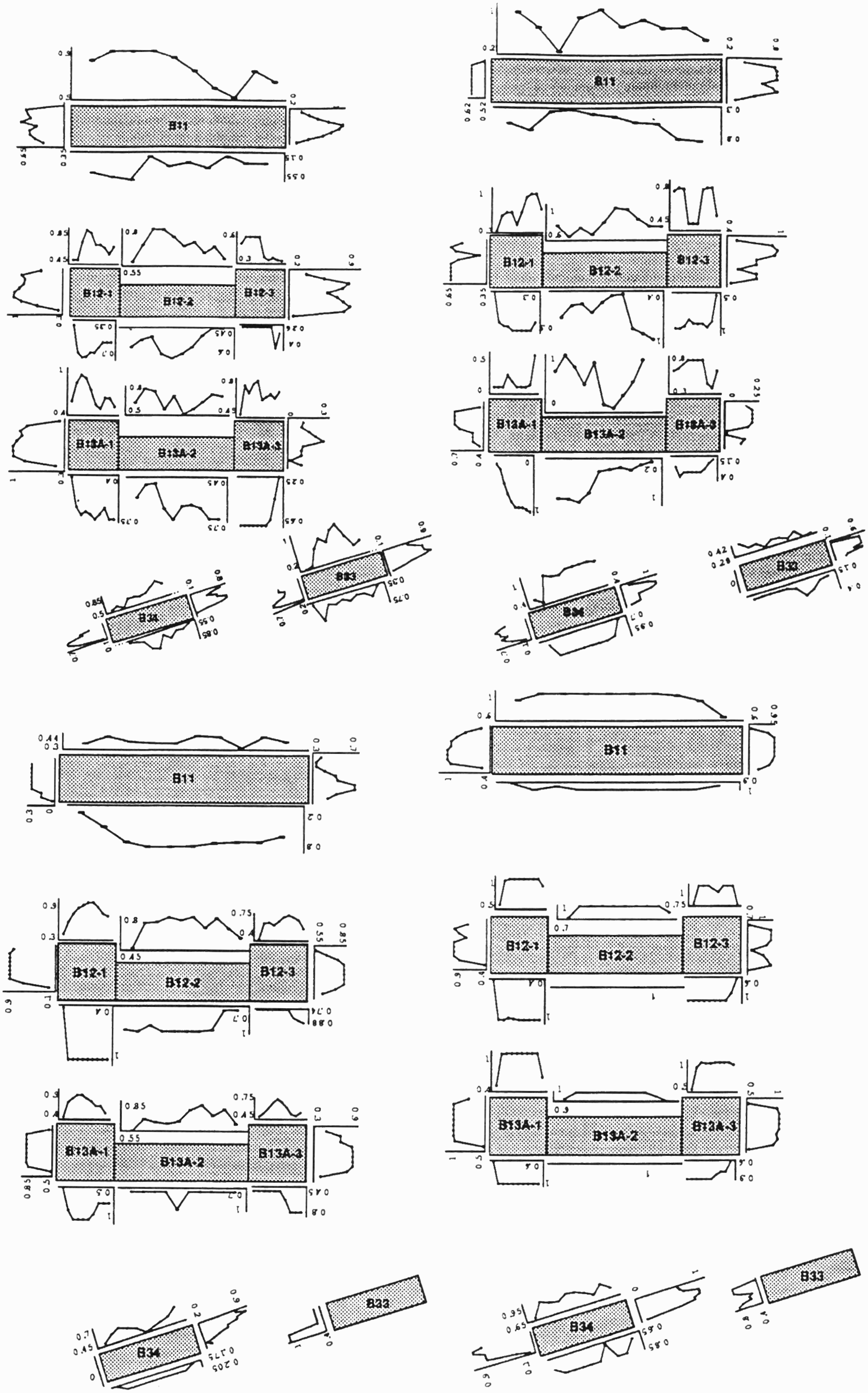

Figure 6: The correlation attribute for four modes of selected edge segmentation models.

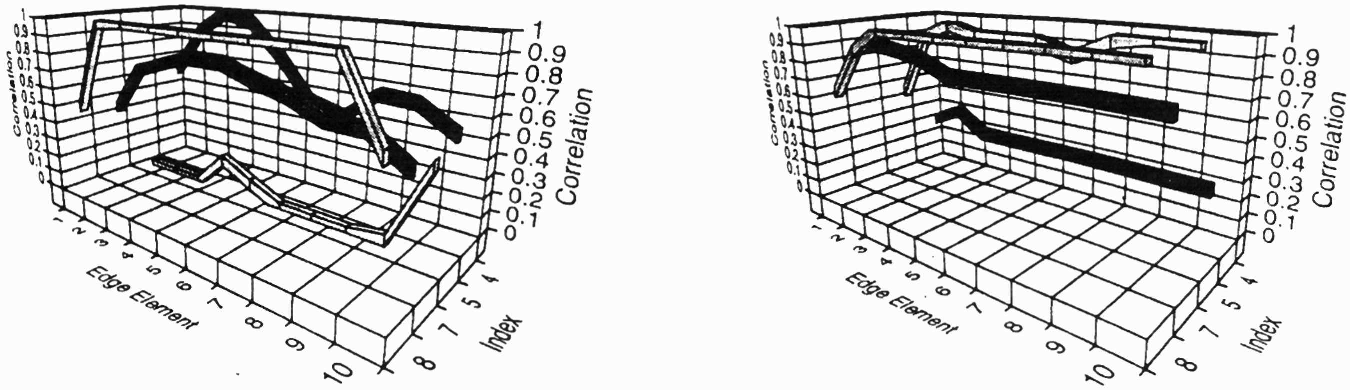

Figure 7: The four modes of Edge 11, Building 13A-1 and Edge 9, Building 12-3.

Statistically, the level of non-uniformity present in the correlation attribute of the 28 MB2 edges can be measured by analyzing the range along the edge profiles. In the distributed, independent case, where only one profile is maintained for each edge attribute, the range of correlation values across each edge is computed. Over the 28 edges, the minimum range in correlation along one edge is 0.12, the maximum is 0.71 and the average is 0.33. Thus, the data indicates that on average, the correlation attribute varies by 33% of the allowable correlation values. Clearly, this attribute should not be treated as a single uniform value.

The efficacy of using a dependent model over an independent model is also an important issue, since the validity of the dependent model is a direct indication of the dependence of segmentation on imaging parameters. Figure 7 shows plots of the four "occupied" modes for Edge 11 of Building B13A-1 and Edge 9 of Building 12-3. It is obvious from the graph of Edge 11 that the modes are significantly different, in form as well as magnitude. Geometrically, this means that the segmentation of this edge will behave very differently from different viewpoints and illumination orientations. Note, however, that within the modes the difference in attribute values is not nearly as great as it is between modes. For Edge 9, the modes are similar in form, but distinctly different in magnitude.

While these are only two examples of 28 computed edge models, they are representative of the types of profiles we have observed. In both of these cases, an independent edge model that averages all correlation attributes into a single mode would not be an accurate model of the data. It is possible that the data across modes may be behaving according to some distribution, but given the large variety in the complete data set we believe this to be unlikely.

Statistical data supporting dependent models across the complete set of segmentation models is discussed in the next section.

4.2 The Edge Gradient Attribute

The gradient attribute depends on the correlation attribute, because it is computed from extracted edge segments that have already been correlated with an object edge. If an interval on an object edge does not have a correlation value in a particular image, i.e., no segmented edge matched that piece of the object edge, then there will be no gradient value for that interval on that image.

In theory, gradient values can range from 0 to 255, but in our case the maximum gradient was measured to be 90. This value may be used as a normalizing factor if values ranging between 0 and 1 are desired.

Using the 40 MB2 images on the 28 object edges, we computed the dependent, distributed segmentation models for the gradient attribute. Figure 8 shows three pieces of data attached to each edge. The first is the average gradient for each edge, as if the segmentation model were independent and uniform. The second number is a measure of the variance *between* modes, computed as follows: for each interval i of Edge E, the standard deviation s_i of the values of the gradient across the modes is computed. Then average of the s_i values is taken. In mathematical terms, the computed value is

$$\frac{1}{n}\sum_{i=1}^{n}\sqrt{\frac{1}{k^2}\sum_{j=1}^{k}(E_j(i)-\overline{E_j(i)})^2}$$

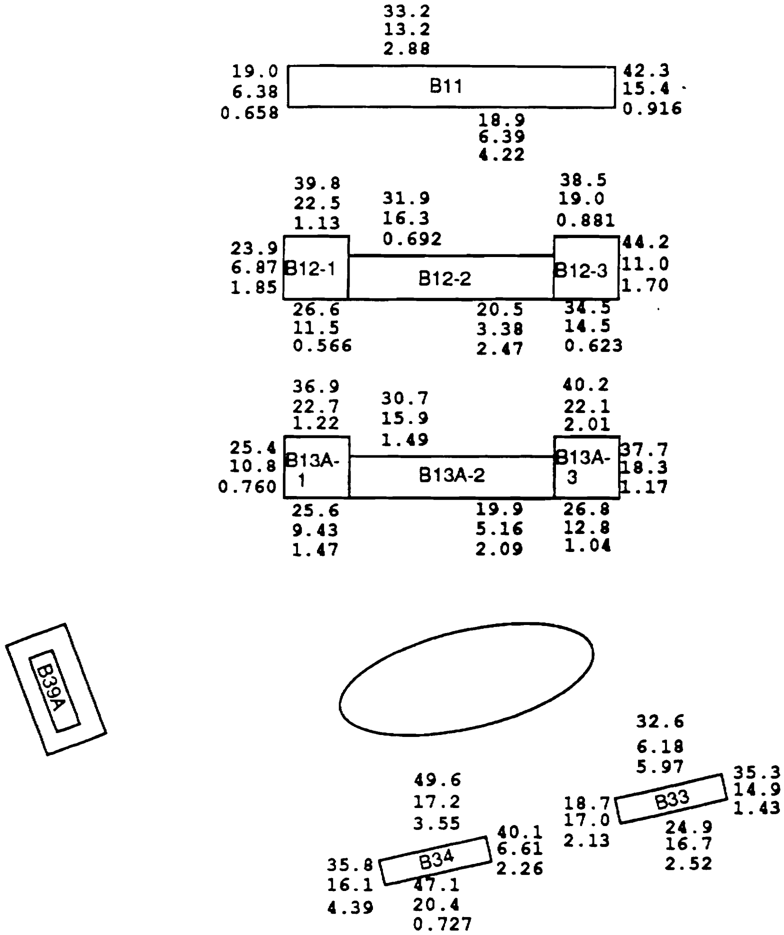

Figure 8: Statistics on the strength attribute for selected edge segmentation models.

where n is the number of intervals, k is the number of modes, $E_j(i)$ is the value of the i^{th} interval in mode j of Edge E, and $\overline{E_j(i)}$ is the average over k of $E_j(i)$, or $\frac{1}{k}\sum_{j=1}^{k} E_j(i)$.

The third number on each edge in Figure 8 is the weighted average of the standard deviations of the gradient values *within* modes, or

$$\frac{1}{\sum_{j=1}^{k} c_j} \sum_{j=1}^{k} c_j \sqrt{\frac{1}{n^2}\sum_{i=1}^{n}(E_j(i) - \overline{\overline{E_j(i)}})^2}$$

where $\overline{\overline{E_j(i)}}$ is the average over n of $E_j(i)$, or $\frac{1}{n}\sum_{i=1}^{n} E_j(i)$; and c_j is the number of images used to computed E_j.

Looking first at the last of the three numbers, it is apparent that the gradient attribute has very little variance along the edge within each mode. However, the difference between modes, the

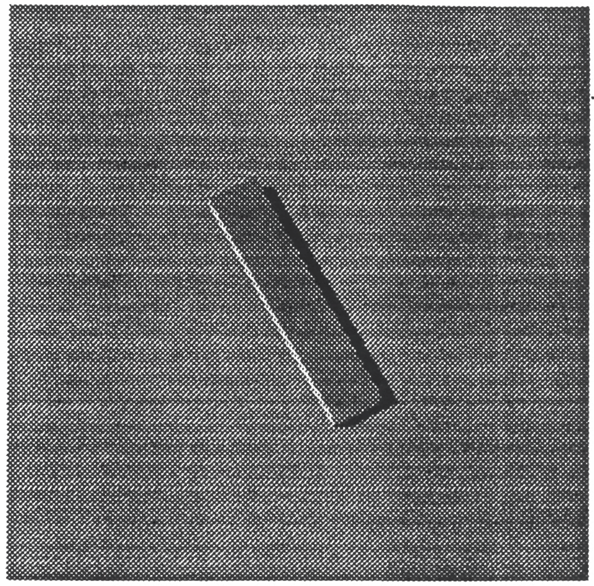

Figure 9: A representative synthetic image.

second number, is typically very large in relative terms (both measures are in the same units, pixel intensities), making a strong case for a dependent model for the gradient attribute. Within each mode, however, the model could reasonably represent the gradient data as a uniform edge model.

Also note that the average gradient (the first number) on corresponding edges of Buildings 12 and 13 are very similar – for example, Edge 2 on both Buildings B13A-1 and B12-1, or similarly Edge 11 on Buildings B12-1, B12-2 and B12-3. This indicates that even an independent segmentation model can capture similarity between two buildings in terms of their segmentations, under certain circumstances.

4.3 Results on Synthetic Data

Finally, we have validated our approach by applying the same segmentation modeling techniques to synthetic data. The synthetic data set consists of ten images of one rectilinear solid with different albedoes for the object roof, sides and the background. Over the ten images, the illumination is held fixed while the viewpoint ranges over a semicircle centered at the center of the object. One of the synthetic images is shown in Figure 9.

The four edges are the top edges of the object, as follows:

- Edge 2 is short edge with its face in shadow.
- Edge 9 is the long edge with its side face in shadow.
- Edge 11 is the long edge with its side face illuminated.
- Edge 6 is the short edge with its side face illuminated.

As observed in Figure 10, the gradient profile shows a much stronger consistency and smaller variance than the real data as expected. The variation at the corners is due to the property of the edge detector which is known to be weak on corners.

The importance of the synthetic data analysis is the validation in the ideal case of the statistical analysis performed on real data. The segmentation models obviously capture the segmentation information on the near-perfect images. However, the profiles reveal that even under ideal cirsumstances the segmentations will not be perfect; as expected, the corners of the object are not recovered well by the Canny-based segmentation system. The synthetic data has proved useful for seperating the image processing component and the enviromental component in the segmentation model.

5 Conclusions

This work presented in this paper has concentrated on developing a meaningful model of segmentations that can be used to analyze the effects of imaging and scene parameters on segmentations. It is hoped that this will be lead to a greater understanding of the interactions between segmentations and higher level vision systems.

We have concentrated on verification of our hypothesis of how such a segmentation model should

edge	mode	# images	correlation profile
2	4	4	0.601 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000 0.819
2	7	5	0.299 0.773 0.973 1.000 1.000 1.000 1.000 1.000 0.766 0.000
9	4	5	0.797 0.964 1.000 1.000 1.000 1.000 1.000 1.000 1.000 0.927
9	7	5	0.880 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000 0.492
11	5	4	0.889 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000 0.879
11	8	5	0.878 1.000 1.000 1.000 1.000 1.000 1.000 1.000 1.000 0.925
6	5	5	0.151 0.804 1.000 1.000 1.000 1.000 1.000 1.000 1.000 0.739
6	8	5	0.779 0.800 0.800 0.800 0.988 1.000 1.000 0.803 0.975 0.563

edge	mode	# images	gradient profile
2	4	4	18.990 18.990 18.990 18.990 18.990 18.990 18.990 18.990 18.990 18.990
2	7	5	13.508 8.259 6.778 6.778 6.778 6.763 6.089 6.089 6.293 0.000
9	4	5	14.434 14.386 14.343 14.343 14.343 14.343 14.343 14.343 14.343 14.386
9	7	5	19.334 19.351 19.351 19.351 19.351 19.351 19.351 19.351 19.351 19.305
11	5	4	20.069 20.069 20.069 20.069 20.069 20.069 20.069 20.069 20.069 20.069
11	8	5	49.209 49.279 49.279 49.279 49.279 49.279 49.279 49.279 49.279 49.539
6	5	5	15.408 15.373 15.373 15.373 15.373 15.373 15.373 15.373 15.373 15.373
6	8	5	28.389 27.769 27.769 27.769 30.858 30.858 30.858 30.858 27.889 32.271

Figure 10: The profiles for the correlation and gradient attributes of the object edges from the synthetic dataset.

be built. Based on our analysis, we argue that distributed model, as opposed to the unified model, is necessary in most cases, and that the dependency on the image acquisition conditions must be encoded explicitly in the segmentation model.

Furthermore, we expect that scene properties, such as the number of objects and the complexity of the background, must be included in the model as well. Also, image processing parameters, such as the particular edge detectors and region growers, probably have a very strong influence on the segmentation model. However, the effect of image processing may be isolated and studied by analysis of synthetic data.

The current formalization of the imaging process model is directly applicable to change detection. In fact, we believe that without segmentation modeling in some form one will never be able to detect change systematically and reliably. In addition, we hope that this analysis and understanding of segmentation process has potential to improve the performance of other types of systems using segmentation. Any system that requires extraction of linear features from a known model or class of models may be able to benefit from segmentation modeling.

References

[1] G.J. Brelstaff and A. Blake. Detecting Specular Reflections Using Lambertian Constraints. *Proceedings of IEEE ICCV*, pp. 297-302, December 1988.

[2] J.F. Canny. A Computational Approach to Edge Detection. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. PAMI-8, No. 6, pp. 679-698, November 1986.

[3] D. Gerson and S. Wood. RADIUS Phase 2: The RADIUS Testbed System. *Proceedings of the ARPA IU Workshop*, Nov. 1994.

[4] A. Hoogs and R. Bajcsy. Segmentation Modeling for Change Detection. *Proceedings of the ARPA IU Workshop*, Nov. 1994.

[5] A. Hoogs and D. Hackett. Model-Supported Exploitation as a Framework for IU. *Proceedings of the ARPA IU Workshop*, Nov. 1994.

[6] S.W. Lee. *Understanding of Surface Reflection in Computer Vision by Color and Multiple Views.* PhD Thesis, University of Pennsylvania, 1991.

[7] J. Mundy and P. Vrobel. The Role of IU Technology in RADIUS Phase 2. *Proceedings of the ARPA IU Workshop*, Nov. 1994.

[8] V. Venkateswar and R. Chellappa. Extraction of Straight Lines in Aerial Images. *IEEE Trans. on Pattern Analysis and Machine Intelligence* bf 16, pp. 1111-1116, Nov. 1992.

[9] J. Mundy and P. Vrobel. The Role of IU Technology in RADIUS Phase 2. *Proceedings of the ARPA IU Workshop*, Nov. 1994.

[10] S.A. Nene, S.K. Nayar and H. Murase. Software Library for Appearance Matching (SLAM). *Proceedings of the ARPA IU Workshop*, Nov. 1994.

[11] M. Bejanin, A. Huertas, G. Medioni and R. Nevatia. Model Validation for Change Detection. *Proceedings of the ARPA IU Workshop*, Nov. 1994.

[12] K. Thornton, D. Nadadur et al. Ground Truthing the RADIUS Model Board. *Proceedings of the ARPA IU Workshop*, Nov. 1994.

[13] M. Wheeler and K. Ikeuchi. Sensor Modeling, Markov Random Fields and Robust Localization for Recognizing Partially Occluded Objects. *Proceedings of the ARPA IU Workshop*, Apr. 1993.