

# Image Classification and Object Detection Using Spatial Contextual Constraints\*

Selim Aksoy, R. Gökberk Cinbiş, H. Gökhan Akçay  
Bilkent University  
Department of Computer Engineering  
Bilkent, 06800, Ankara, Turkey  
saksoy@cs.bilkent.edu.tr

## Abstract

Spatial information plays a very important role in high-level image understanding tasks. Contextual models that exploit spatial information through the quantification of region spatial relationships can be used for resolving the uncertainties in low-level features used for image classification and object detection. We describe intuitive, flexible and efficient methods for modeling pairwise directional spatial relationships and the ternary between relationship using fuzzy mathematical morphology. These methods define a fuzzy landscape where each image point is assigned a value that quantifies its relative position with respect to the reference object(s) and the type of the relationship. Directional mathematical dilation with fuzzy structuring elements is used to compute this landscape. We provide flexible definitions of fuzzy structuring elements that are tunable along both radial and angular dimensions. Examples using synthetic images show that our models produce more intuitive results than the competitors. We also illustrate the use of the models described in this chapter as spatial contextual constraints for two image analysis tasks. First, we show how these spatial relationships can be incorporated into a Bayesian classification framework for land cover classification to reduce the amount of commission among spectrally similar classes. Then, we show how the use of spatial constraints derived from shadow regions improves building detection accuracy. The significant improvement in accuracy in these applications confirms the importance of spatial information and the effectiveness of the relationship models described in this chapter in modeling and quantifying this information.

## 1 Introduction

Spatial information plays a fundamental role in the analysis and understanding of remotely sensed data sets. Common ways of incorporating spatial information into classification involve the use of textural, morphological, and object-based features. Features extracted using co-occurrence matrices, Gabor wavelets [1], morphological profiles [2], and Markov random fields [3] have been widely used in the literature to model spatial information in neighborhoods of pixels. However, problems such as scale selection and the detailed content of high-resolution imagery make the applicability of traditional fixed window-based methods difficult for such data sets.

---

\*This work was supported in part by the TUBITAK CAREER Grant 104E074 and European Commission Sixth Framework Programme Marie Curie International Reintegration Grant MIRG-CT-2005-017504.

Another powerful method for exploiting structural information is to perform region-based classification rather than classifying individual pixels. This is also referred to as object-oriented classification in the remote sensing literature. For example, Bruzzone and Carlin [4] performed classification using the spatial context of each pixel according to a hierarchical multi-level representation of the scene. In [5], we proposed an algorithm for selecting meaningful segments that maximize a measure consisting of spectral homogeneity and neighborhood connectivity in a hierarchy of segmentations, and described an algorithm for unsupervised grouping of candidate segments belonging to multiple hierarchical segmentations to find coherent sets of segments that correspond to actual objects. However, image segmentation is still an unsolved problem, and homogeneous regions obtained as a result of segmentation often correspond to very small details in high spatial resolution images obtained from the new generation sensors.

Alternatively, contextual models that exploit spatial information can be used to resolve the ambiguities in the identification of structures having similar low-level spectral and textural properties. Contextual information has long been acknowledged for playing a very important role in both human and computer vision. Consequently, development of context models has become a challenging problem in both statistical and structural pattern recognition. A structural way of modeling context in images is through the quantification of spatial relationships. Typical relationships studied in the literature include geometric (based on size, position, shape, and orientation), topological (based on set relationships and neighborhood structure), semantic (based on similarity and causality), statistical (based on frequency and co-occurrence), and structural (based on spatial configuration and arrangement patterns) relationships [6]. The methods used for computing these relationships depend on the way how objects/regions are modeled. Widely used approaches include grid-based representations, centroids and minimum bounding rectangles. However, even though centroids and minimum bounding rectangles can be useful when regions have circular or rectangular shapes, regions in natural scenes often do not follow these assumptions. Furthermore, fixed sized grids are also not generally applicable as they cannot capture large number of structures with varying sizes and shapes.

When regions are represented as sets of points (pixels), spatial relationships can be modeled in terms of directional and distance information between pixel groups. In particular, adjacency of two regions can be measured as a fuzzy function of the distance between their closest points or using morphological dilations modeling connectivities [7]. Distance-based relationships can also be defined using fuzzy membership functions modeling symbolic classes such as near and far using the distance between boundary pixels. In previous work [8], we developed fuzzy models for pairwise topological spatial relationships such as bordering, invading and surrounding based on overlaps between region boundaries, distance-based relationships such as near and far based on distances between region boundaries, and relative position-based relationships such as right, left, above and below using angles between region centroids. Then, we combined these pairwise relationships into higher order relationship models using fuzzy logic, and illustrated their use in image retrieval [8]. We also developed a Bayesian framework that learns image classes based on automatic selection of distinguishing (e.g., frequently occurring, rarely occurring) relations between regions [9]. Finally, we built attributed relational graph structures to model scenes by representing regions by the graph nodes and their spatial relationships by the edges between such nodes [10], and used relational matching techniques to find similarities between graphs representing different scenes. We demonstrated the effectiveness of these approaches in scenarios that cannot be expressed by traditional approaches but where the proposed models can capture both feature and spatial characteristics of

scenes and model them according to their high-level semantic content. Inglada and Michel also used attributed relational graphs where the relations are computed by using region connection calculus [11], and used these graph for object matching.

This chapter presents an extension of our earlier work on modeling region spatial relationships [8] using relative position-based relationships: binary directional relationships [12] and the ternary between relationship [13]. Most of the existing methods for defining relative spatial positions rely on angle measurements between points of objects of interest where the angle corresponding to a pair of points is computed between the segment joining the points and a reference axis in the coordinate system [14]. For example, Miyajimaya and Ralescu [15] proposed to use a histogram that is constructed using the angles between all pairs of points from both objects where the mean or the maximum angle computed from this histogram can be used to represent the relative position of these objects. Matsakis and Wendling [16] introduced the histogram of forces as an alternative to the histogram of angles. This method computes the degree of satisfaction for a given angle using intersection of longitudinal sections of objects with lines having the desired direction. In [17], Wang et al. proposed the F-templates that incorporate distance information with direction information. Bloch [18] proposed a morphological approach that is based on directional dilations where a fuzzy landscape for a reference object is created at a given angle and other objects are compared to this landscape to evaluate how well they match with the areas having high membership values.

Another relationship that is often used in daily life but has not been studied as thoroughly as the binary relationships is the between relationship ([19] provides an extensive review and a comparative study). Directional dilations are also useful for the between relationship. After obtaining an approximate relative angle between the reference regions, directional dilations are applied to both regions to extend them towards each other to generate the landscape. Angle histogram can be directly used to create the structuring element for dilation.

Intuitively, the influence of the shape of the object (e.g., concavities, extent) and the influence of the distance between objects are important points to be considered in the design of an algorithm for modeling spatial relationships. Mathematical morphology provides a strong basis for such a framework. Furthermore, the ambiguities and subjectiveness inherent in the definitions of the relationships make fuzzy representation a promising approach for modeling the imprecision in both the images and the results. In this chapter, we describe intuitive, flexible and efficient methods for modeling pairwise directional spatial relationships and the ternary between relationship using fuzzy mathematical morphology. These methods define a fuzzy landscape where each point is assigned a value that quantifies its relative position with respect to the reference object(s) and the type of the relationship. Directional mathematical dilation with fuzzy structuring elements is used to compute this landscape. We provide flexible definitions of fuzzy structuring elements that are tunable along both radial and angular dimensions. Furthermore, for the pairwise directional relationships, the definitions for the fuzzy landscape are extended to support sensitivity to visibility to handle image areas that are fully or partially enclosed by a reference object but are not visible from image points along the direction of interest. Given a reference object and a direction of interest that specifies the spatial relationship, the degree of satisfaction of this relation by a target object can be computed by integrating the landscape corresponding to this relation over the support of the target region.

The definitions of the directional spatial relationships are also combined to generate a landscape in which the degree of each image area being located “between” the reference objects is quantified. Our definition also handles the cases where one object is significantly spatially extended relative to the other by taking spatial proximity into consideration. Similarly, the satisfaction of this

ternary relation by a target object relative to two reference objects is computed by integrating the corresponding landscape.

We illustrate the use of the models described in this chapter as spatial contextual constraints for two image analysis tasks. First, we show how these spatial relationships can be incorporated into a Bayesian classification framework for land cover classification. The decision based on the maximum posterior probability rule produces limited accuracy for classes with similar appearance and spectral values when no spatial information is used. However, by constraining the classification of certain classes by using their spatial relationships to other classes as contextual information, we show that the classification accuracy can be significantly improved. Then, we show that detection of buildings with complex shapes and roof structures can be improved by using directional spatial relationships between candidate building regions and shadow regions along the sun azimuth angle.

The rest of the chapter is organized as follows. The fuzzy structuring elements and the morphological approach for quantifying the pairwise directional spatial relationships are described in Section 2. The approach for modeling the ternary between relationship and its computation for spatially extended objects is given in Section 3. Applications of these approaches to land cover classification and building detection are presented in Section 4. Finally, conclusions are given in Section 5.

## 2 Directional spatial relationships

Directional relationships describe the spatial arrangement of two objects relative to each other. Although, it is a common approach to use right (east), left (west), above (north), and below (south) as the directions, for generic modeling purposes it is more convenient and generalizable to use an angle-based definition of these relations where it is possible to calculate the degree of satisfaction of the relation for a given angle.

Given a reference object  $B$  and a direction specified by the angle  $\alpha$ , our goal is to generate a landscape in which the degree of satisfaction of the directional relationship at each image area relative to the reference object is quantified. Then, given a second object, its relation to the reference object can be measured using this landscape. The landscape will be denoted by  $\beta_\alpha(B)$  in the rest of the chapter. Definitions for crisp and fuzzy objects are available in the literature. However, only crisp objects are considered in this chapter.

### 2.1 Morphological approach

The landscape  $\beta_\alpha(B)$  around a reference object  $B$  along the direction specified by the angle  $\alpha$  can be defined as a fuzzy set such that the membership value of an image point corresponds to the degree of satisfaction of the spatial relation under examination where points in areas that satisfy the directional relation with a high degree have high membership values. This relationship can be defined in terms of the angle between the vector from a point in the reference object to a point in the image and the unit vector along the direction  $\alpha$  measured with respect to the horizontal axis. Bloch [18] suggested that the smallest such angle computed for a point in the image considering all points in the reference object corresponds to the visibility of the image point from the reference object in the direction  $\alpha$ . Consequently, the value of the fuzzy landscape at an image point  $x$  can

be computed as a function  $f : [0, \pi] \rightarrow [0, 1]$  of this angle as

$$\beta_\alpha(B)(x) = f \left( \min_{b \in B} \theta_\alpha(x, b) \right) \quad (1)$$

where  $b$  represents a point in  $B$ .  $\theta_\alpha(x, b)$  is the angle between the vector  $\vec{bx}$  and the unit vector  $\vec{u}_\alpha = (\cos \alpha, \sin \alpha)^T$  along  $\alpha$ , and can be computed as

$$\theta_\alpha(x, b) = \begin{cases} \arccos \left( \frac{\vec{bx} \cdot \vec{u}_\alpha}{\|\vec{bx}\|} \right) & \text{if } x \neq b, \\ 0 & \text{if } x = b. \end{cases} \quad (2)$$

Bloch [18] used a function that decreases linearly with  $\theta$  as

$$f(\theta) = \max \left\{ 0, 1 - \frac{2\theta}{\pi} \right\} \quad (3)$$

for (1). It can be shown that this is equivalent to the morphological dilation of  $B$ ,

$$\beta_\alpha(B)(x) = (B \oplus \nu_\alpha)(x) \cap B^c, \quad (4)$$

using the fuzzy structuring element

$$\nu_\alpha(x) = \max \left\{ 0, 1 - \frac{2}{\pi} \theta_\alpha(x, o) \right\} \quad (5)$$

where  $o$  is the origin (center) of the structuring element and  $B$  is removed from the result of dilation in (4) ( $c$  represents complement). The fuzzy morphological dilation of the object  $B$  with the structuring element  $\nu$  is defined as

$$(B \oplus \nu)(x) = \max_y \{ t[\varphi(y), \nu(x - y)] \} \quad (6)$$

where  $\varphi$  is the function representing object  $B$ ,  $t$  is the  $t$ -norm operator for fuzzy intersection, and  $y$  is taken over all points in the image. An example synthetic image and fuzzy landscape examples using morphological dilation are given in Figure 1. In all figures in this chapter, white represents binary 1, black represents binary 0, and gray values represent the fuzziness in the range  $[0, 1]$ .

However, the linear function in (3) and the corresponding structuring element in (5) often lead to a landscape with a large spread and unintuitive transitions when the angle departs from  $\alpha$  particularly at points that are farther away from the reference object. Thus, they may not give realistic results for many cases (see the examples in this section). Instead of using linearly decreasing membership values according to the angle, we developed a more intuitive and flexible structuring element using a nonlinear function with the shape of a Bézier curve

$$\nu_{\alpha, \lambda}(x) = g_\lambda \left( \frac{2}{\pi} \theta_\alpha(x, o) \right) \quad (7)$$

where  $\lambda \in (0, 1)$  determines the inflection point of the curve and increasing  $\lambda$  increases the spread around  $\alpha$  (see the Appendix for the derivation). The nonlinear function enables different definitions

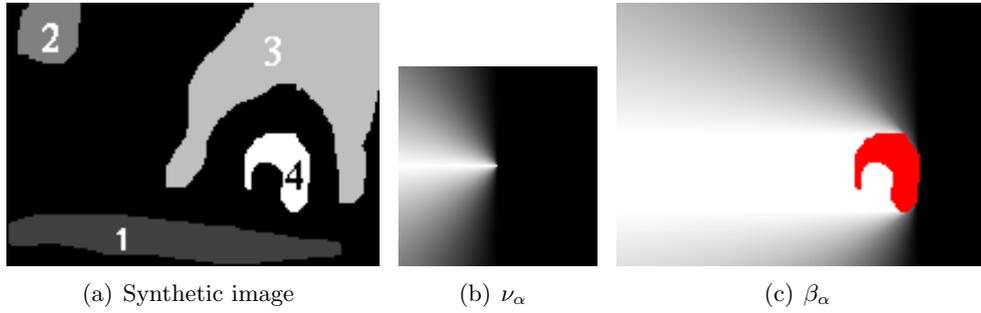


Figure 1: An example synthetic image and the directional landscape  $\beta_\alpha$  for object labeled 4 using the structuring element  $\nu_\alpha$  defined in (5) for  $\alpha = \pi$ .

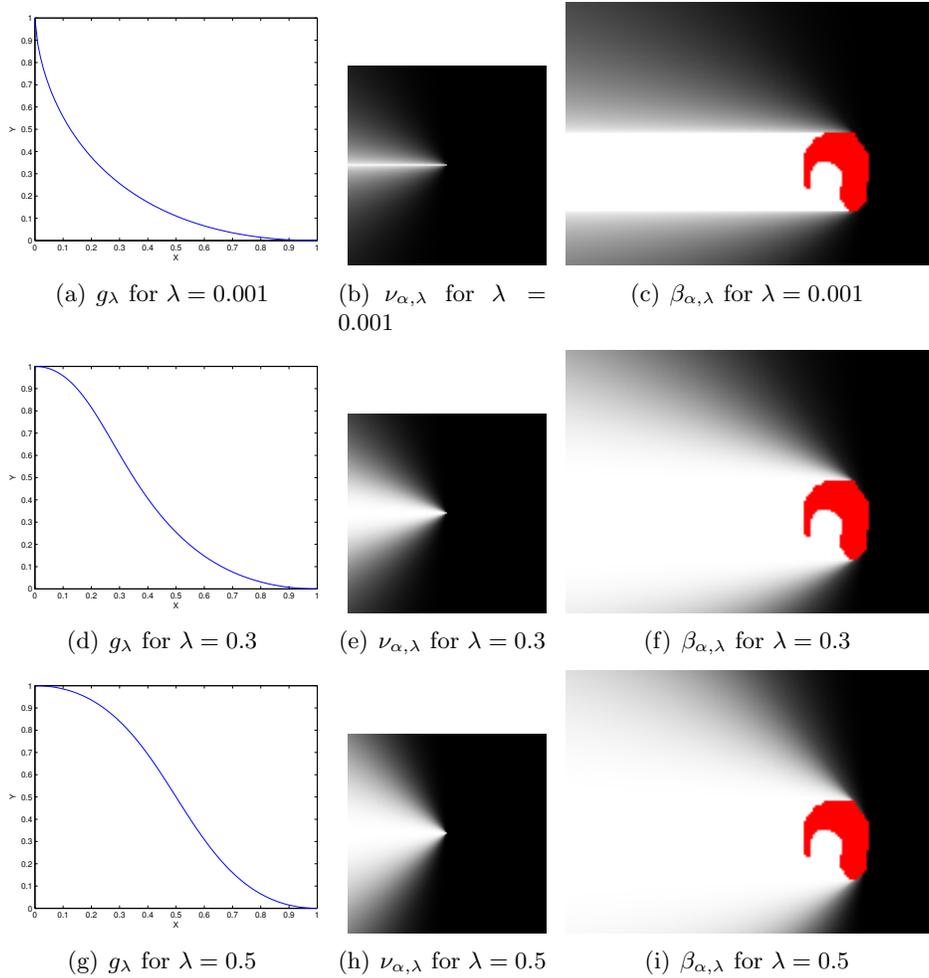


Figure 2: Structuring element  $\nu_{\alpha, \lambda}$  defined in (7) and directional landscape  $\beta_{\alpha, \lambda}$  of object 4 (from Figure 1) for  $\alpha = \pi$  and different values of  $\lambda$ .

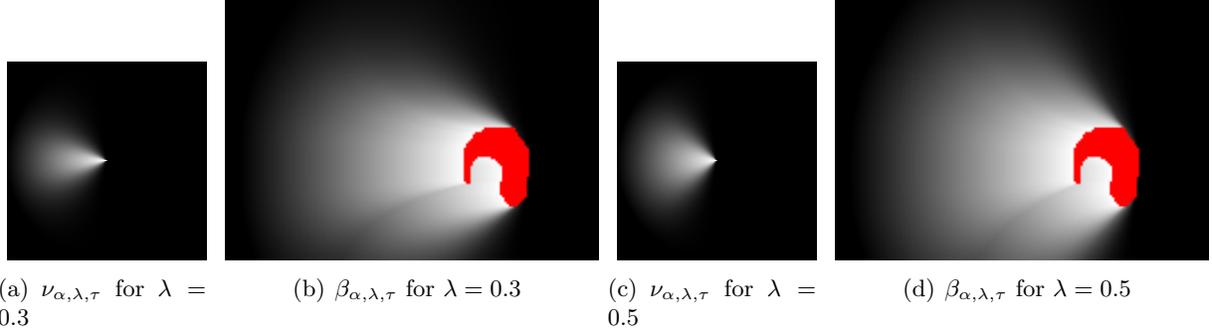


Figure 3: Structuring element  $\nu_{\alpha, \lambda, \tau}$  defined in (8) and directional landscape  $\beta_{\alpha, \lambda, \tau}$  of object 4 (from Figure 1) for  $\alpha = \pi$ ,  $\tau = 100$ , and different values of  $\lambda$ .

of fuzziness for different cases. Fuzzy landscape examples using this structuring element definition are given in Figure 2.

The definition of the structuring element can be further extended to decrease the degree of a point's spatial relation to the reference object according to its distance to that object by introducing a new term

$$\nu_{\alpha, \lambda, \tau}(x) = g_{\lambda} \left( \frac{2}{\pi} \theta_{\alpha}(x, o) \right) \max \left\{ 0, 1 - \frac{\|\vec{o}x\|}{\tau} \right\} \quad (8)$$

where  $\|\vec{o}x\|$  is the Euclidean distance of point  $x$  from the structuring element's center. In this definition, a point's spatial relation to the reference object decreases linearly with its distance to the object where  $\tau$  corresponds to the distance where a point is no longer visible from the reference object. This definition also has a computational advantage because in the previous definitions (5) and (7) the structuring element must be at least twice as large as the landscape of interest in the image space whereas in definition (8) a structuring element with size of at most  $2\tau \times 2\tau$  is sufficient. In fact, the resulting operations in (8) have linear time complexity with respect to image size as opposed to the quadratic complexity in (5) and (7), leading to dramatic improvements in the efficiency of the algorithm. Fuzzy landscape examples using this structuring element definition are given in Figure 3.

## 2.2 Visibility

In the directional dilation of (4), the areas that are fully or partially enclosed by the reference object but are not visible from image points along the direction of interest may have high values as shown in Figures 3 and 4. To overcome this problem, we introduced the following definition

$$\beta_{\alpha, \lambda, \lambda', \tau}(B)(x) = (B \oplus \nu_{\alpha, \lambda, \tau})(x) \cap (B \oplus \nu_{\alpha + \pi, \lambda'})(x)^c \quad (9)$$

where the first dilation uses the structuring element defined in (8) and the second dilation uses the structuring element defined in (7). We compute fuzzy intersection using multiplication as the  $t$ -norm operator and compute fuzzy complement by subtracting the original values from 1.  $\lambda'$  can be set to a very small number to consider only the image points along  $\alpha + \pi$ . The proposed definition of visibility is illustrated in Figure 4.

Figure 5 shows additional examples of directional landscapes of objects using the definition in [18] (equation (5) in this chapter) and our definition (9) on a synthetic image that was also used

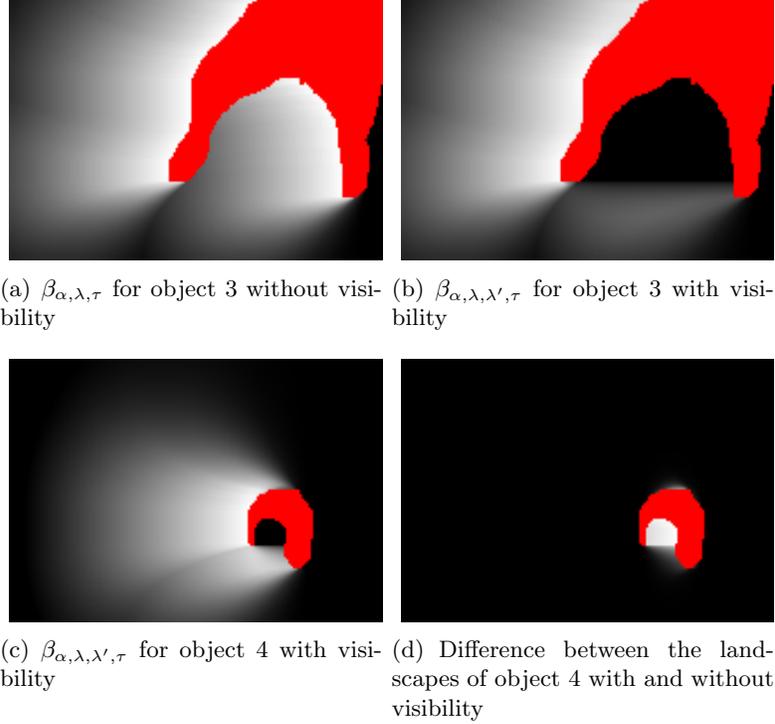


Figure 4: Directional landscapes  $\beta_{\alpha,\lambda,\tau}$  and  $\beta_{\alpha,\lambda,\lambda',\tau}$  for objects 3 and 4 (from Figure 1) without and with the visibility extension, respectively, for  $\alpha = \pi$ ,  $\lambda = 0.3$ ,  $\lambda' = 0.001$  and  $\tau = 100$ . (a) uses the structuring element definition in (8) without visibility, (b) and (c) use the definition in (9) with visibility, (d) illustrates the difference between the landscapes with and without visibility.

in [18, 14, 7, 19]. Figures 5(b) and 5(c) illustrate the differences between the fuzzy directional landscapes obtained using the definition in [18] and our structuring element definitions. The latter is sensitive to the distance to the object according to the constant  $\tau$  and the landscape’s fuzziness is more centralized along the main direction of interest by the help of the constant  $\lambda$ . Figures 5(d) and 5(e) present the importance of the support for visibility in our definition for directional relationships. Although both landscapes for the direction “right” have similar distributions to the right and above of the reference object, the first one also has nonzero values on the left of the object, which contradicts the intuition.

### 3 Between relationship

Between relationship is a ternary relationship defined by two reference objects and a target object. Given two reference objects  $B$  and  $C$ , our goal is to generate a landscape in which the degree of each image area being located between the reference objects is quantified. Then, given a third object, its relation to the reference objects can be determined using this landscape. The landscape will be denoted by  $\beta_{\mathcal{B}}(B, C)$  in the rest of the chapter.

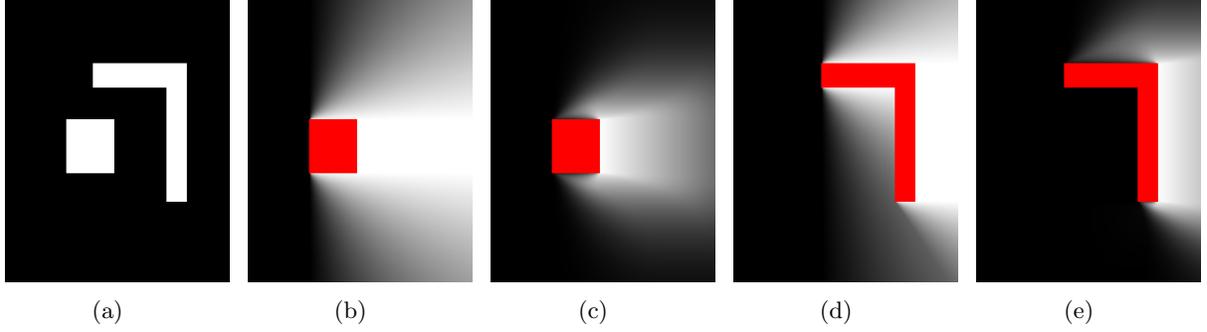


Figure 5: Directional landscape examples. (a) Synthetic image with two objects: square (A) and L-shaped (B). (b)  $\beta_\alpha(A)$  for  $\alpha = 0$  using the definition in [18]. (c)  $\beta_{\alpha,\lambda,\lambda',\tau}(A)$  for  $\alpha = 0$  using our definition. (d)  $\beta_\alpha(B)$  for  $\alpha = 0$  using the definition in [18]. (e)  $\beta_{\alpha,\lambda,\lambda',\tau}(B)$  for  $\alpha = 0$  using our definition. The constants are set as  $\lambda = 0.3$ ,  $\lambda' = 0.001$ , and  $\tau = 200$ .

### 3.1 Morphological approach

Similar to the directional spatial relationships described in Section 2.1, the landscape  $\beta_{\theta_{\tilde{\gamma}}}(B, C)$  between two reference objects  $B$  and  $C$  can be defined as a fuzzy set such that image points with a high degree of the spatial relation have high membership values. This landscape can be computed as the intersection of the directional dilations of the reference objects along the directions  $\alpha = \theta_{\tilde{\gamma}}$  and  $\alpha = \theta_{\tilde{\gamma}} + \pi$  where  $\theta_{\tilde{\gamma}}$  is the relative position of the reference objects. This relative position can be calculated using the maximum or mean value in the histogram of angles between all pairs of points of the reference objects [19]. Using the horizontal axis as the axis of reference, the histogram of angles for the objects  $B$  and  $C$  can be computed as

$$h_{B,C}(\theta) = |\{(b, c) | b \in B, c \in C, \angle(\vec{bc}, \vec{u}_{\alpha=0}) = \theta\}| \quad (10)$$

and normalized as

$$H_{B,C}(\theta) = \frac{h_{B,C}(\theta)}{\max_{\theta'} h_{B,C}(\theta')}. \quad (11)$$

Then, using  $\theta_{\tilde{\gamma}}$  as the relative position obtained from this histogram (as the maximum or mean value), the landscape between the reference objects  $B$  and  $C$  is computed as

$$\beta_{\theta_{\tilde{\gamma}}}(B, C)(x) = \beta_{\alpha=\theta_{\tilde{\gamma}},\lambda,\lambda'}(B)(x) \cap \beta_{\alpha=\theta_{\tilde{\gamma}}+\pi,\lambda,\lambda'}(C)(x) \quad (12)$$

where the directional landscape  $\beta_{\alpha,\lambda,\lambda'}$  is computed as

$$\beta_{\alpha,\lambda,\lambda'}(B)(x) = (B \oplus \nu_{\alpha,\lambda})(x) \cap (B \oplus \nu_{\alpha+\pi,\lambda'})(x)^c \quad (13)$$

using the structuring element definition in (7). Since the landscape should include only the areas that are visible from both reference objects, the notion of visibility defined in Section 2.2 is used in the computation. Fuzzy landscape examples for the between relationship using this definition are given in Figure 6.

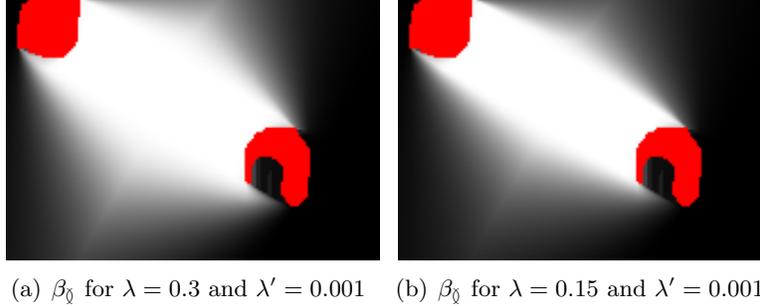


Figure 6: Between landscape  $\beta_{\theta}$  of objects 2 and 4 (from Figure 1) using the definition in (12) with different values of  $\lambda$  and  $\lambda'$ . The relative angle for these objects is found as  $\theta_{\theta} = -30.04^{\circ}$ .

### 3.2 Myopic vision

Although the histogram of angles generally provides a good approximation to the relative position of two objects, it fails in the cases where one object is significantly spatially extended relative to the other [19] (see Figure 7 for examples). We can solve this problem by taking into account only the part of the spatially extended object close to the other object. (Bloch et al. [19] called this the “myopic vision” and suggested to use the distance map to find close parts of objects.)

Spatial proximity for handling spatially extended objects is incorporated into our morphological approach using a weighted histogram of angles where the contribution of the angle between each point pair in the histogram is weighted by the term  $\max\{0, 1 - \|\vec{bc}\|/\tau_{\text{myopic}}\}$  (instead of a constant weight of 1 in (10) as in [19]) where  $\|\vec{bc}\|$  is the Euclidean distance between the points  $b$  and  $c$ , and  $\tau_{\text{myopic}}$  is the threshold for the maximum distance between two points for allowing them to contribute to the histogram. The proposed definition of myopic vision is illustrated in Figure 7 using objects 1 and 4 where object 1 is spatially extended relative to object 4.

Figures 8 and 9 show additional examples of between landscapes of objects using the definition in [19] and our definition (12). The former calculates the landscape using dilation by a structuring element derived from the histogram of angles as defined in (17) in [19]. These examples illustrate the differences in the two definitions when the objects have different spatial extents. For example, the landscape in Figure 8(b), which is generated according to the definition in [19], is spatially too extended in the upper and lower parts of the image. It also includes non-smooth transitions that are unintuitive. On the other hand, the landscape in Figure 8(c), which is generated using (12), is more compact and is fully covering the expected between area.

## 4 Applications

We illustrate the use of the models described in this chapter as spatial contextual constraints for two image analysis tasks. The image classification task, described in Section 4.1, uses a Bayesian framework to incorporate contextual information in land cover classification to reduce the amount of commission among spectrally similar classes and improve the classification accuracy [12]. The object detection task, described in Section 4.2, illustrates the use of spatial constraints derived from shadow regions in improving the building detection accuracy [20]. As the  $t$ -norm operator, minimum is used in all definitions, except for visibility in directional relationships where multiplication is used

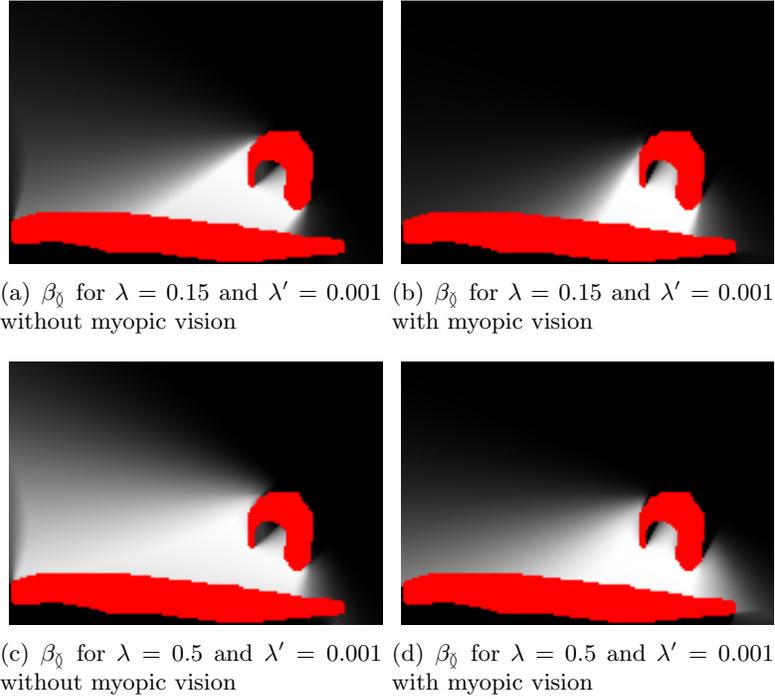


Figure 7: Between landscape  $\beta_{\emptyset}$  of objects 1 and 4 (from Figure 1) without and with myopic vision for different values of  $\lambda$  and  $\lambda'$  where object 1 is spatially extended relative to object 4.  $\tau_{\text{myopic}}$  is taken as the half of the width of the image. The relative angles are  $42.28^\circ$  and  $63.40^\circ$  for the figures without and with myopic vision, respectively. For larger values of  $\lambda$ , error in landscape without myopic vision becomes more significant.

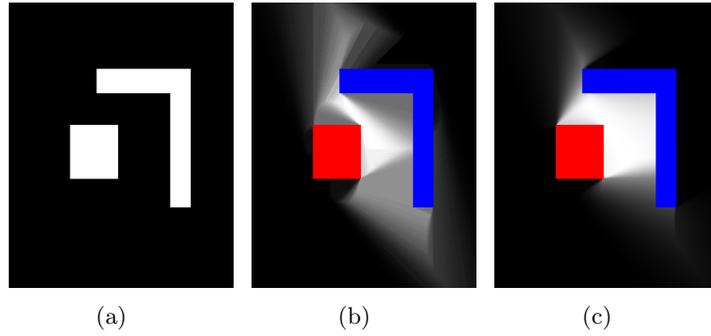


Figure 8: Between landscape examples. (a) Synthetic image with two objects: square (A) and L-shaped (B). (b)  $\beta_{\emptyset}(A, B)$  using the definition in [19]. (c)  $\beta_{\emptyset}(A, B)$  using our definition. The constants are set as  $\lambda = 0.3$  and  $\lambda' = 0.001$ .

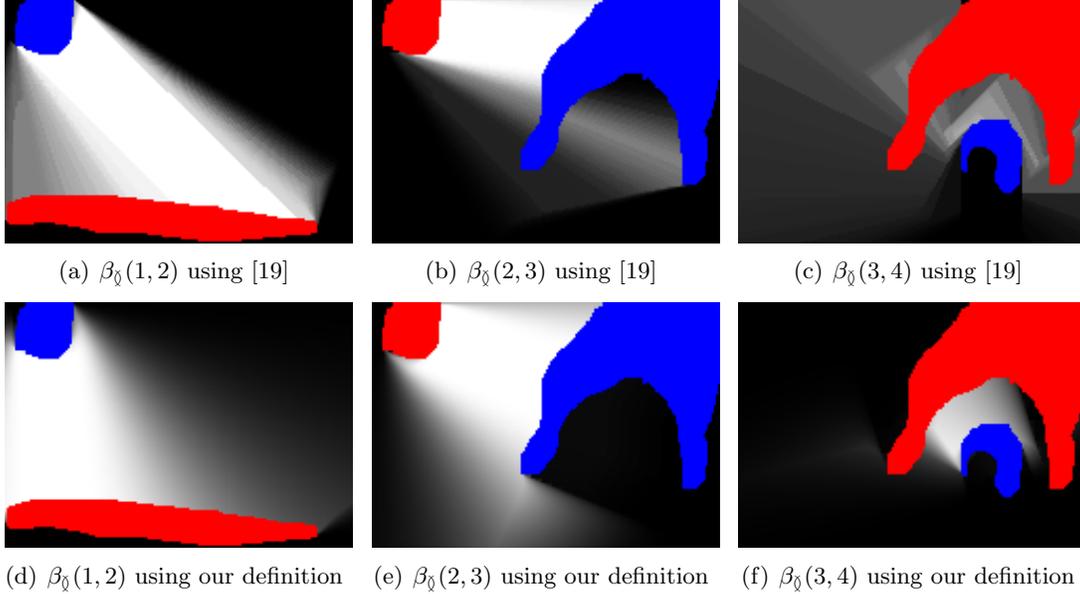


Figure 9: Between landscape examples for the synthetic image in Figure 1.

as suggested in Section 2.2. After calculating the landscape  $\beta$  for a spatial relation as in Sections 2 or 3, the degree of satisfaction of this relation by a target object  $A$  is computed as

$$\mu(A) = \frac{1}{\text{area}(A)} \sum_{a \in A} \beta(a). \quad (14)$$

#### 4.1 Image classification

The conventional method for automatically producing a land cover map in the remote sensing literature is to use a statistical classifier for supervised classification of pixels based on their spectral values. Even though these classifiers improve the processing time compared to manual digitization, their accuracy is limited with the discrimination ability of the spectral values of individual pixels as pixel-based classification does not take into account any spatial context.

A popular approach for incorporating spatial information into the classification process is to base the decisions on image regions using image segmentation techniques that automatically group neighboring pixels into contiguous regions based on similarity criteria on pixels' properties [21]. Even though image segmentation has been heavily studied, it is still an unsolved problem, especially for images with a very complex content with thousands of objects as in the examples that we study in this chapter.

Alternatively, spatial relationships can be used to model and quantify context in images. A commonly observed problem in pixel-level classification using spectral information is the confusion between the land covers with similar spectral values. In particular, a significant amount of confusion may occur between water pixels and shadow pixels and between asphalt pixels and shadow pixels that all appear dark in the image, as well as between snow and cloud classes that both have bright color values close to white. In this section, we illustrate the use of spatial relationships for improving the land cover classification accuracy.

Let  $\mathbf{x}$  denote the feature vector of a pixel or an object at location  $x$  in a binary classification problem with two classes  $w_1$  and  $w_2$ . As a widely used solution, the Bayesian classifier makes a decision using the posterior probabilities as

$$\text{Decide } \begin{cases} w_1 & \text{if } \frac{P(w_1|\mathbf{x})}{P(w_2|\mathbf{x})} > 1 \\ w_2 & \text{otherwise} \end{cases} \quad (15)$$

which is equivalent to

$$\text{Decide } \begin{cases} w_1 & \text{if } \frac{P(\mathbf{x}|w_1)}{P(\mathbf{x}|w_2)} > \frac{P(w_2)}{P(w_1)} \\ w_2 & \text{otherwise} \end{cases} \quad (16)$$

using the Bayes rule with the class-conditional and prior probabilities. The equal priors assumption ( $P(w_1) = P(w_2)$ ) is often used when no additional information is available.

Assume that there is a third class  $w_3$  that is related to  $w_2$ . The pixels/objects that are assigned to  $w_3$  can be used as spatial constraints for improving the discrimination between  $w_1$  and  $w_2$ . First, the directional landscape  $\beta_\alpha(w_3)$  is computed for the whole scene by using  $w_3$  as the reference. Then, the fuzzy landscape value in the range  $[0, 1]$  at each image location is used as the spatial prior for  $w_2$  at that location, i.e.,  $P(w_2) = \beta_\alpha(w_3)(x)$  and  $P(w_1) = 1 - P(w_2)$ . The resulting contextual decision rule becomes

$$\text{Decide } \begin{cases} w_1 & \text{if } \frac{P(\mathbf{x}|w_1)}{P(\mathbf{x}|w_2)} > \frac{\beta_\alpha(w_3)(x)}{1 - \beta_\alpha(w_3)(x)} \\ w_2 & \text{otherwise} \end{cases} \quad (17)$$

using these spatial priors. The extension of (17) for multi-class classification with multiple reference classes and multiple priors is straightforward.

This classification scheme is applied to two data sets. The first one consists of a LANDSAT scene covering the Washington State in the U.S.A. and the British Columbia in Canada. The multispectral image has 6 bands with 30 m spatial resolution and  $7,680 \times 10,240$  pixels. We trained Bayesian classifiers for the water, shadow, and cloud classes as described in [21]. These classifiers produce the posterior probabilities  $P(w_i|\mathbf{x}), i = 1, 2, 3$  for the water ( $w_1$ ), shadow ( $w_2$ ) and cloud ( $w_3$ ) classes, respectively. A threshold is applied to the posterior probabilities to allow some of the pixels not being assigned to any class if none of the corresponding probabilities is high enough. Then, the maximum posterior probability rule is used for the final classification. Figure 10(a) shows a  $2,115 \times 1,070$  pixel section of the scene. The resulting classification map is shown in Figure 10(c). Binary classification between the water and shadow classes is also performed as in (15). Table 1(a) shows the corresponding confusion matrix where Figure 10(b) is used as the ground truth (independent from the training set). The resulting 79.88% accuracy shows a significant amount of confusion between the water and shadow pixels when only pixel-based information is used.

To incorporate the spatial context into the decision process, we use the pixels classified as clouds ( $w_3$ ) as reference objects, and compute the directional landscape  $\beta_{\alpha,\lambda,\tau}(w_3)$  using the parameters  $\alpha = 135^\circ$ ,  $\lambda = 0.3$ , and  $\tau = 1.6$  km. The  $\alpha$  value of  $135^\circ$  measured from the horizontal axis in counter-clockwise direction approximates the sun angle that can be obtained from the metadata of the image. The  $\lambda$  and  $\tau$  values are determined empirically. The resulting landscape is shown in Figure 10(d). The classification map and the confusion matrix resulting from the use of clouds as reference objects in the decision rule in (17) are shown in Figure 10(e) and Table 1(b), respectively. Constraining the decision for classifying a pixel as shadow by requiring a high degree of directional

Table 1: Confusion matrix for water versus shadow classification.

(a) Using the decision rule (15) without spatial information. Overall accuracy is 79.88%. (b) Using the decision rule (17) with spatial information. Overall accuracy is 96.48%.

|      |        | Assigned |        |      |        | Assigned |        |
|------|--------|----------|--------|------|--------|----------|--------|
|      |        | water    | shadow |      |        | water    | shadow |
| True | water  | 51,502   | 20,533 | True | water  | 68,736   | 3,299  |
|      | shadow | 244      | 31,008 |      | shadow | 337      | 30,915 |

spatial relationship with respect to clouds at a particular angle results in 96.48% accuracy, which corresponds to a net 16.60% improvement over the case where no spatial information is used. Figure 11 illustrates the results for the upper right portion of the scene in more detail.

The second data set consists of a  $490 \times 199$  pixel section of a well-known hyperspectral image of Pavia, Italy obtained by the ROSIS sensor and has 102 spectral bands and 2.6 m spatial resolution. A pixel-based Bayesian classification was performed using spectral and textural features as described in [21]. The output of the Bayesian classifier is a probability value for each class at each pixel as in the LANDSAT case. Figures 12(a) and 12(b) show the true-color image and the corresponding classification map, respectively. Table 2(a) shows the confusion matrix for the pixels where either the asphalt class ( $w_1$ ) or the shadow class ( $w_2$ ) has the highest probability. Figure 12(e) is used as the ground truth (independent from the training set). The 63.66% accuracy shows a significant amount of commission between these two classes when only pixel-based information is used.

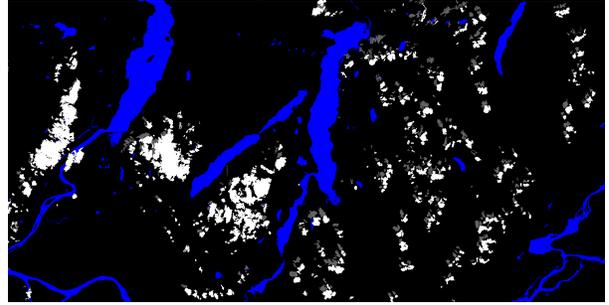
The spatial context can be incorporated into the decision process by using the tiles (building roofs) and trees as additional information. The directional landscape is computed for the pixels classified as tiles using the parameters  $\alpha = -50^\circ$ ,  $\lambda = 0.3$ , and  $\tau = 25$ . The  $\alpha$  value measured from the horizontal axis in counter-clockwise direction is visually determined from the image to approximate the sun angle. The  $\lambda$  and  $\tau$  values are determined empirically. Similarly, the directional landscape for the trees class is computed using the parameters  $\alpha = -50^\circ$ ,  $\lambda = 0.3$ , and  $\tau = 10$ . The two landscapes, shown in Figures 12(c) and 12(d), are combined using the “max” operator (which is the equivalent of the Boolean “or” operator). Then, the contextual decision rule in (17) is used to update the classification at each pixel by using tiles and trees as reference ( $w_3$ ). Figure 12(f) and Table 2(b) show the classification results when spatial information is used. The updated contextual decision gives an 86.16% accuracy that corresponds to a net 22.50% improvement by classifying a pixel with shadow-like feature values as shadow only when it also has a high degree of directional spatial relationship with respect to buildings or trees at a particular angle. The significant improvement in accuracy for both the LANDSAT data set and the ROSIS data set confirms the importance of spatial information in classification and the effectiveness of the relationship models described in this chapter in modeling and quantifying this information.

## 4.2 Building detection

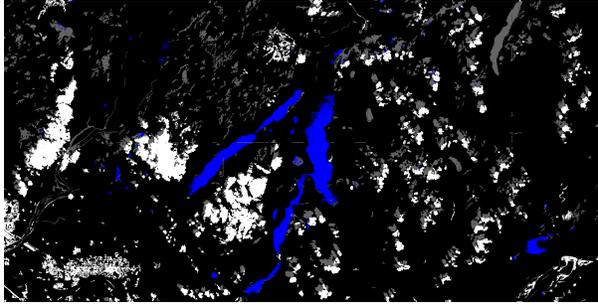
Automatic detection of buildings in very high spatial resolution remotely sensed imagery has been an important problem because the detection results can be used in many applications such as change detection, urbanization monitoring, and digital map production. There is an extensive literature on building detection where both pixel level and object/region level processing have been used. However, most of the previous methods try to solve the problem for specific settings such as images



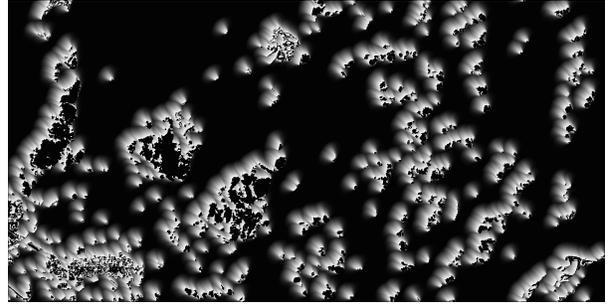
(a) True-color image



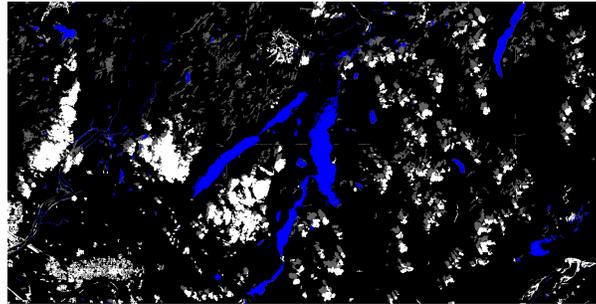
(b) Ground truth map



(c) Classification map using decision rule (15) without spatial information



(d) Directional landscape with respect to clouds



(e) Classification map using decision rule (17) with spatial information

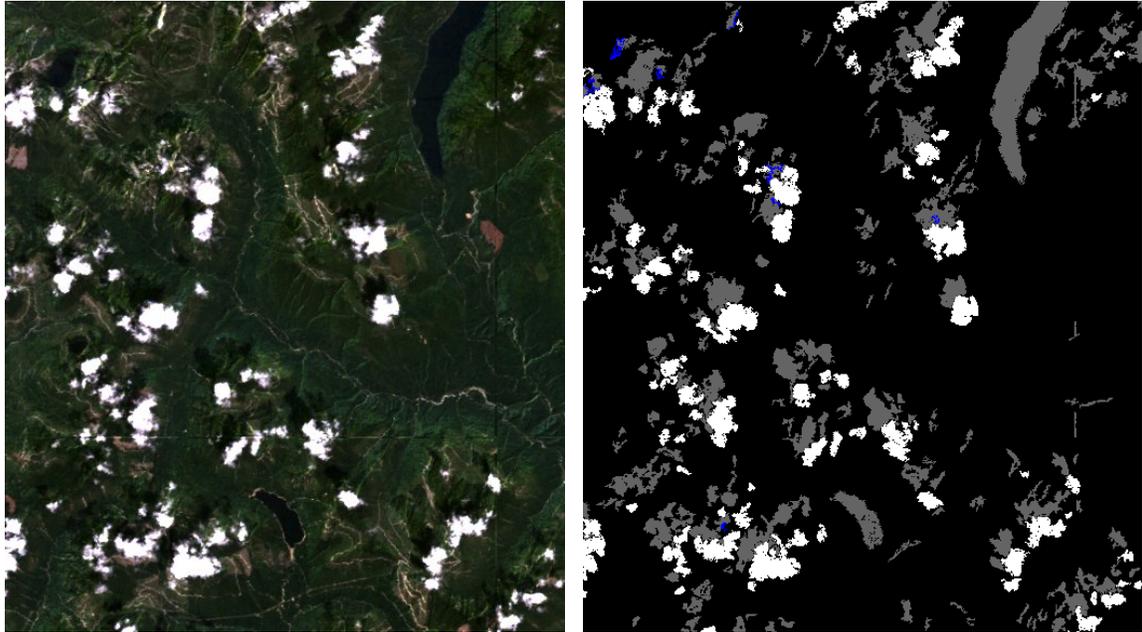
Figure 10: A  $2,115 \times 1,070$  section of the LANDSAT scene and its classification without and with using spatial contextual information. The classes in the classification and ground truth maps are water (blue), shadow (gray), and cloud (white). The ground truth is produced by careful visual inspection.

Table 2: Confusion matrices for asphalt versus shadow classification.

(a) Using the decision rule (15) without spatial information. Overall accuracy is %63.66. (b) Using the decision rule (17) with spatial information. Overall accuracy is %86.16.

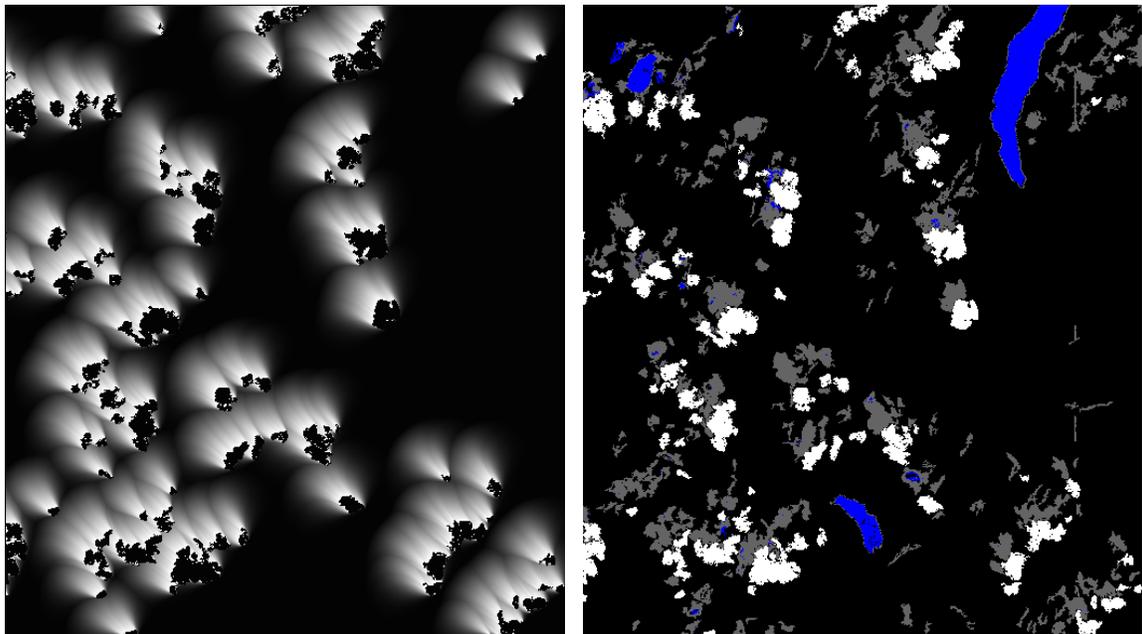
|      |         | Assigned |        |
|------|---------|----------|--------|
|      |         | asphalt  | shadow |
| True | asphalt | 3,302    | 2,000  |
|      | shadow  | 1,028    | 2,003  |

|      |         | Assigned |        |
|------|---------|----------|--------|
|      |         | asphalt  | shadow |
| True | asphalt | 5,054    | 248    |
|      | shadow  | 905      | 2,126  |



(a) True-color image

(b) Classification map using decision rule (15) without spatial information



(c) Directional landscape with respect to clouds

(d) Classification map using decision rule (17) with spatial information

Figure 11: A  $520 \times 587$  section in the upper right part of Figure 10(a) and its classification without and with using spatial contextual information. The classes in the classification map are water (blue), shadow (gray), and cloud (white). The classification in (b) has 61.32% accuracy where all water pixels are misclassified as shadow. The classification in (d) achieves perfect detection (100%).

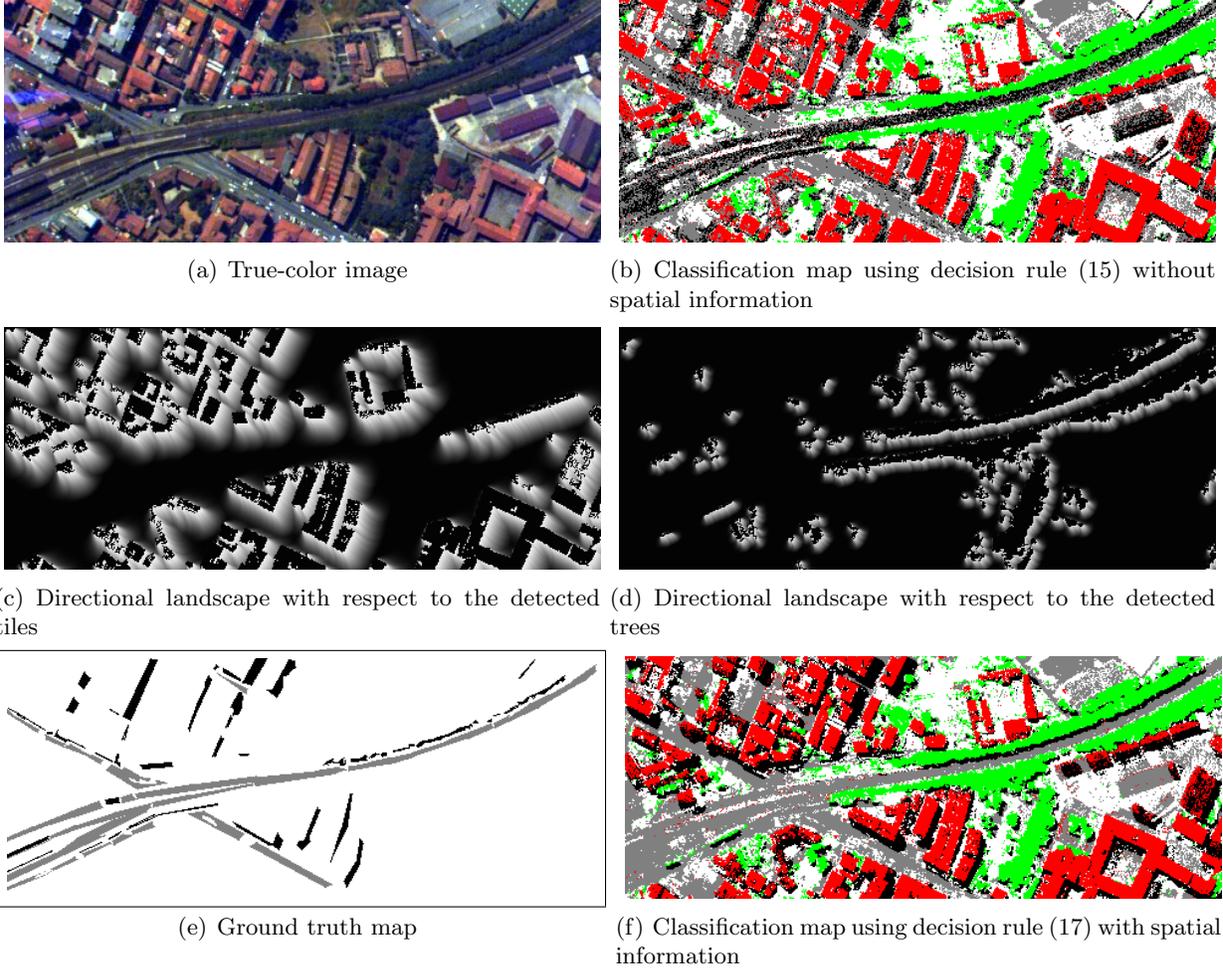


Figure 12: Classification of the Pavia image without and with using spatial contextual information. The classes in the classification and ground truth maps are asphalt (gray), shadow (black), tiles (red) and trees (green). The ground truth is produced by visual inspection.

having buildings with the same type of appearance and images where the buildings are isolated and have simple roof structures. With the increase in the spatial details in the images obtained from new generation sensors with meter and sub-meter spatial resolution, the buildings may have very complicated appearances and may have complex structures with very different spectral signatures.

Even though different buildings may appear in significantly different colors and shapes, a common property of such buildings can be the existence of shadows. The relationship between buildings and shadows has actually been exploited in earlier works [22, 23]. More recently, Sirmacek and Unsalan [24] detected buildings with red roofs using color information and verified their existence with the occurrences of shadow-like nearby regions. However, the assumption of red roofs is limiting and there may be other sources of shadows in the image.

In this section, we describe a method for detection of buildings with complex shapes and roof structures in very high spatial resolution images by exploiting spectral, structural, and contextual information. The input to the method is a satellite image consisting of a panchromatic and



Figure 13: Examples from an Ikonos image of Antalya, Turkey and the corresponding watershed segmentation results. The segmentation boundaries are overlaid as white.

multispectral data pair. First, the watershed segmentation algorithm is used to partition the panchromatic band into spectrally homogeneous regions. The results contain oversegmented regions because the test areas in this study include buildings with complex roof structures as shown in Figure 13. Then, among all regions, the ones that are likely to belong to shadows are selected using their spectral properties. This selection uses the normalized difference vegetation index (NDVI) that is computed using the pan-sharpened image where the regions whose average brightness values are lower than a brightness threshold and average NDVI values are lower than an NDVI threshold are denoted as shadow regions.

Next, candidate building regions are identified using the directional spatial relationships of all regions with respect to the detected shadow regions along the sun azimuth angle. Given the sun azimuth angle, we can find the directional landscapes of the shadow regions along this direction by using (4). The resulting directional landscapes give high responses in areas close to the shadow regions along the sun azimuth angle. These areas correspond to the locations where the probability of the presence of buildings is high. Figures 14(a) and 14(c) show the shadow regions and the

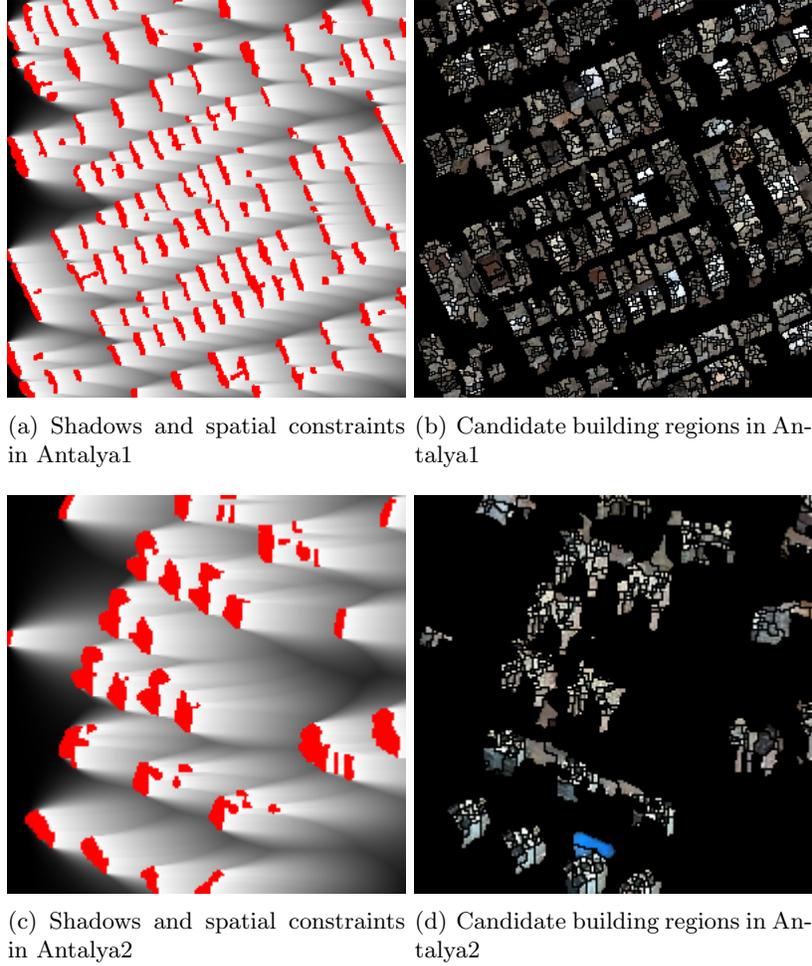


Figure 14: Examples of shadow regions, directional landscapes, and candidate building regions.

corresponding landscapes. Consequently, the regions whose average satisfaction degrees are higher than a satisfaction threshold, average NDVI values are lower than the NDVI threshold, and sizes are lower than a size threshold are identified as candidate building regions. Figures 14(b) and 14(d) show examples for candidate regions. As can be seen from the figures, most of the regions are correctly identified with a small number of misdetections and several false alarms.

Finally, the building regions are selected by clustering the oversegmented regions that satisfy the spatial constraints using minimum spanning trees. An important observation is that regions forming a building are densely located whereas regions separating different buildings are found far from their neighbors. The distance between two regions is measured as the distance between their centroids. This seems to be a valid assumption because the regions are obtained from oversegmentation and mostly have compact shapes. Hence, we construct a graph where the graph nodes correspond to the candidate regions' centroids and the edges are created between two neighboring nodes with a weight corresponding to their spatial distance. What we expect is that the nodes representing parts of building regions will form dense subgraph components. These dense components are found by constructing the minimum spanning tree of the graph, and by eliminating some of the remaining

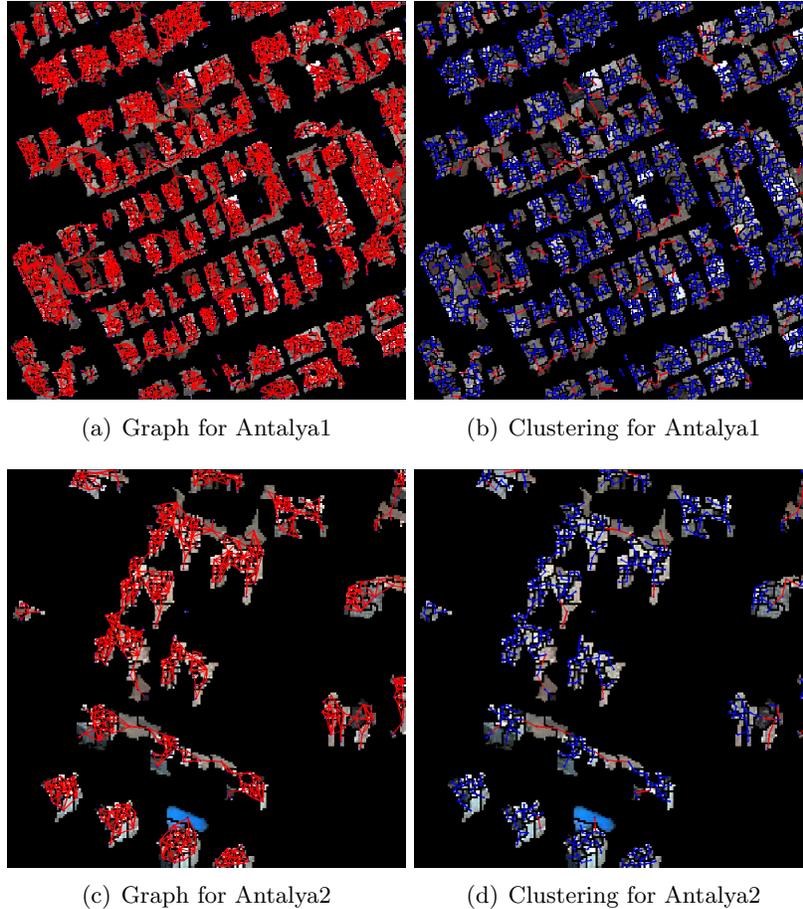


Figure 15: Examples of graph construction and minimum spanning tree-based clustering. The removed edges are colored in red.

edges that are longer than a length threshold. As a result, the nodes that are spatially close enough remain in the same cluster. Figure 15 shows examples for graph construction and clustering.

Six sub-scenes of 1 m spatial resolution Ikonos images of Antalya, Turkey are used to qualitatively evaluate the algorithm. Figure 16 shows example detection results. It can be seen that most of the building regions that cannot be obtained by traditional spectral segmentation methods that cannot incorporate structural and contextual information are correctly extracted.

## 5 Conclusions

We presented new, intuitive, flexible and efficient definitions for modeling pairwise directional spatial relationships and the ternary between relationship using fuzzy mathematical morphology techniques. Our contributions included flexible definitions for the fuzzy directional structuring elements that are tunable along both radial and angular dimensions, support for the notion of visibility for handling image areas that are partially enclosed by objects and are not visible from image points along the direction of interest, and handling of the cases where one object is signifi-



Figure 16: Building detection results. The detected buildings are highlighted in red.

cantly spatially extended relative to the other. Illustrations using synthetic data showed that our models produce more intuitive results than the state-of-the-art techniques. We also presented two applications with real data. First, we showed that incorporating the spatial relationships as contextual information in a Bayesian classification framework results in a significant improvement in land cover classification accuracy by reducing the amount of commission among spectrally similar classes in multispectral and hyperspectral data. Then, we showed that the use of spatial constraints derived from shadow regions improves building detection accuracy in very high spatial resolution imagery. The significant improvement in accuracy in these applications confirms the importance of spatial information in classification and the effectiveness of the relationship models described in this chapter in modeling and quantifying this information. Future work includes investigating ways of automating the selection of the parameters for different applications.

## References

- [1] S. Bhagavathy and B. S. Manjunath. Modeling and detection of geospatial objects using texture motifs. *IEEE Transactions on Geoscience and Remote Sensing*, 44(12):3706–3715, December 2006.

- [2] M. Pesaresi and J. A. Benediktsson. A new approach for the morphological segmentation of high-resolution satellite imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 39(2):309–320, February 2001.
- [3] F. Melgani and S. B. Serpico. A Markov random field approach to spatio-temporal contextual image classification. *IEEE Transactions on Geoscience and Remote Sensing*, 41(11):2478–2487, November 2003.
- [4] L. Bruzzone and L. Carlin. A multilevel context-based system for classification of very high spatial resolution images. *IEEE Transactions on Geoscience and Remote Sensing*, 44(9):2587–2600, September 2006.
- [5] H. G. Akcay and S. Aksoy. Automatic detection of geospatial objects using multiple hierarchical segmentations. *IEEE Transactions on Geoscience and Remote Sensing*, 46(7):2097–2111, July 2008.
- [6] S. Steiniger and R. Weibel. Relations among map objects in cartographic generalization. *Cartography and Geographic Information Science*, 34(3):175–197, 2007.
- [7] I. Bloch. Fuzzy spatial relationships for image processing and interpretation: A review. *Image and Vision Computing*, 23(2):89–110, February 2005.
- [8] S. Aksoy, C. Tusk, K. Koperski, and G. Marchisio. Scene modeling and image mining with a visual grammar. In C. H. Chen, editor, *Frontiers of Remote Sensing Information Processing*, pages 35–62. World Scientific, 2003.
- [9] S. Aksoy, K. Koperski, C. Tusk, G. Marchisio, and J. C. Tilton. Learning Bayesian classifiers for scene classification with a visual grammar. *IEEE Transactions on Geoscience and Remote Sensing*, 43(3):581–589, March 2005.
- [10] S. Aksoy. Modeling of remote sensing image content using attributed relational graphs. In *Proceedings of IAPR International Workshop on Structural and Syntactic Pattern Recognition*, pages 475–483, Hong Kong, August 17–19, 2006. Lecture Notes in Computer Science, vol. 4109.
- [11] J. Inglada and J. Michel. Qualitative spatial reasoning for high-resolution remote sensing image analysis. *IEEE Transactions on Geoscience and Remote Sensing*, 47(2):599–612, February 2009.
- [12] S. Aksoy and R. G. Cinbis. Image mining using directional spatial constraints. *IEEE Geoscience and Remote Sensing Letters*, 7(1):33–37, January 2010.
- [13] R. G. Cinbis and S. Aksoy. Relative position-based spatial relationships using mathematical morphology. In *Proceedings of IEEE International Conference on Image Processing*, volume II, pages 97–100, San Antonio, Texas, September 16–19, 2007.
- [14] I. Bloch and A. Ralescu. Directional relative position between objects in image processing: A comparison between fuzzy approaches. *Pattern Recognition*, 36(7):1563–1582, July 2003.
- [15] K. Miyajima and A. Ralescu. Spatial organization in 2D segmented images: Representation and recognition of primitive spatial relations. *Fuzzy Sets and Systems*, 65(2-3):225–236, August 1994.

- [16] P. Matsakis and L. Wendling. A new way to represent the relative position between areal objects. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(7):634–643, July 1999.
- [17] X. Wang, J. Ni, and P. Matsakis. Fuzzy object localization based on directional (and distance) information. In *Proceedings of IEEE International Conference on Fuzzy Systems*, 2006.
- [18] I. Bloch. Fuzzy relative position between objects in image processing: A morphological approach. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 21(7):657–664, July 1999.
- [19] I. Bloch, O. Colliot, and R. M. Cesar. On the ternary spatial relation “between”. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 36(2):312–327, April 2006.
- [20] H. G. Akcay and S. Aksoy. Building detection using directional spatial constraints. In *Proceedings of IEEE International Geoscience and Remote Sensing Symposium*, Honolulu, Hawaii, July 25–30, 2010.
- [21] S. Aksoy. Spatial techniques for image classification. In C. H. Chen, editor, *Signal and Image Processing for Remote Sensing*, pages 491–513. Taylor & Francis Books, 2006.
- [22] A. Huertas and R. Nevatia. Detecting buildings in aerial images. *Computer Vision, Graphics, and Image Processing*, 41(2):131–152, 1988.
- [23] R. B. Irvin and D. M. McKeown Jr. Methods for exploiting the relationship between buildings and their shadows in aerial imagery. *IEEE Transactions on Systems, Man, and Cybernetics*, 19(6):1564–1575, 1989.
- [24] B. Sirmacek and C. Unsalan. Building detection from aerial images using invariant color features and shadow information. In *Proceedings of International Symposium on Computer and Information Sciences*, 2008.

## A Bézier curves

Bézier curve is a parametric curve defined using a number of reference points. Four points  $a_0, a_1, a_2, a_3$  on a plane define a cubic Bézier curve where the curve starts at  $a_0$  going toward  $a_1$  and arrives at  $a_3$  coming from the direction of  $a_2$ . The parametric form of the curve is

$$b(t) = (1-t)^3 a_0 + 3t(1-t)^2 a_1 + 3t^2(1-t) a_2 + t^3 a_3 \quad (18)$$

where  $t$  is the parameter having values in  $[0, 1]$ .

To construct a one-dimensional function that has the shape of a Bézier curve and maps each  $x \in [0, 1]$  to a  $y \in [0, 1]$ , we set the reference points  $a = (x, y)^T$  as

$$a_0 = (0, 1)^T, \quad a_1 = (\lambda, 1)^T, \quad a_2 = (\lambda, 0)^T, \quad a_3 = (1, 0)^T \quad (19)$$

where  $\lambda \in (0, 1)$  so that the cubic curve has only one parameter. Then, equation (18) reduces to

$$b_x(t) = 3t(1-t)^2 \lambda + 3t^2(1-t) \lambda + t^3 \quad (20)$$

$$b_y(t) = (1-t)^3 + 3t(1-t)^2 + 3t^2(1-t) \quad (21)$$

and for any  $x \in [0, 1]$ ,  $b_x(t)$  can be solved for  $t$ , and the corresponding  $y \in [0, 1]$  can be computed using  $b_y(t)$ .

In this chapter, this function/mapping is denoted as  $g_\lambda(x)$ . The function has an inflection point at  $x = \lambda$ . Examples of  $g_\lambda(x)$  for different  $\lambda$  values are shown in Figure 2.