Chapter 2

Background

To find a more effective way to represent images, plenty of methods have been proposed in the literature [1]-[38]. We roughly classify some commonly used image representation schemes into boundary-based methods, region-based methods, transform-based methods, multiresolution-based methods, and fractal-based methods. These different approaches are to be briefly introduced in Section 2.1-2.5, respectively. In Section 2.6, some recent works concerning image representation are further mentioned.

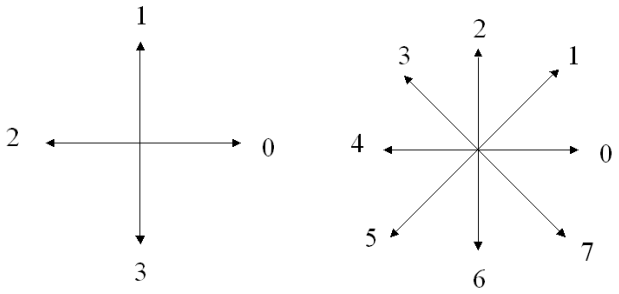
2.1 Boundary-Based Methods

Boundary-based methods, like chain code and signature [3]-[8], have been proposed to describe an image in terms of object boundaries. The destinations of boundary-based representation schemes are to describe shapes of objects or to distinguish objects in different shapes.

2.1.1 Chain Code

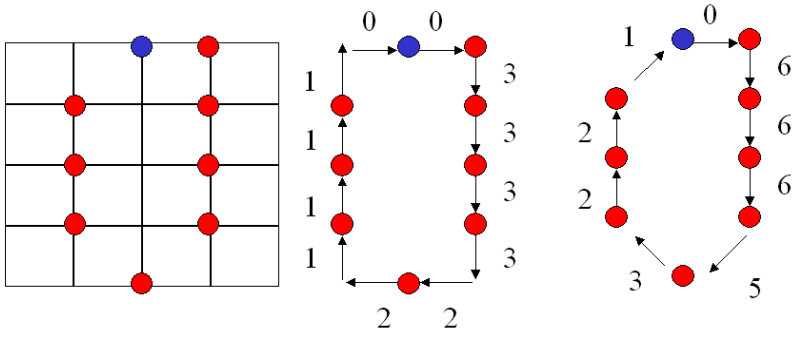
Chain code is designed for permitting the encoding of arbitrary geometric configurations [1], [3]-[5]. An arbitrary curve is represented by a sequence of small vectors of unit length and a limited set of possible directions. On the digital grid, encoding is based on the fact that successive contour points are adjacent to each other.

Depending on whether the 4-connected (or the 8-connected) grid is employed, the chain code is defined as the digits 0~3 (or 0~7) being assigned to the 4 (or 8) neighboring grid points in a clockwise or counter-clockwise counter. An illustration of 4-connected grid and 8-connected grid is shown in Fig. 2.1. A chain is usually coded by the absolute image coordinates of the starting point followed by the relative coordinates of the remaining points with respect to their predecessors. An example of chain code representation is shown in Fig. 2.2, where the starting point is indicated as a blue circle.



(a) (b)

Fig. 2.1. Illustration of 4-connected grid and 8-connected grid



(a) (b) (c)

Fig. 2.2. Examples of chain code. (a) Digital boundary. (b) Chain code using4-connected grid. (c) Chain code using 8-connected grid.

2.1.2 Signature

In general, a signature is a 1-D function representing the 2-D boundary [1], [6]-[8]. Chain codes can also be considered as one form of signatures. A different

form of signature is to encode the radial distance from the centroid of the object as a function of angle. Objects with different shapes in a two-dimensional image, such as triangles and rectangles, can be characterized merely through their signatures. Fig. 2.3 shows an example of signature representation for objects with different shapes. Even though the obtained signatures are rotationally variant and sensitive to scaling, it has the nice property that a scaled version of the same object produces a scaled version of the signature. In general, a signature can be normalized with respect to some other feature of the object, such as maximum radial distance, length of the longest axis, etc.

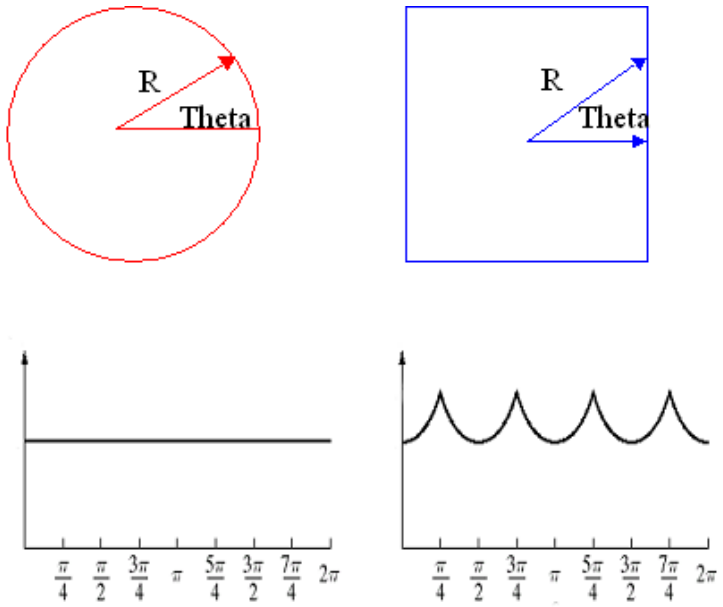


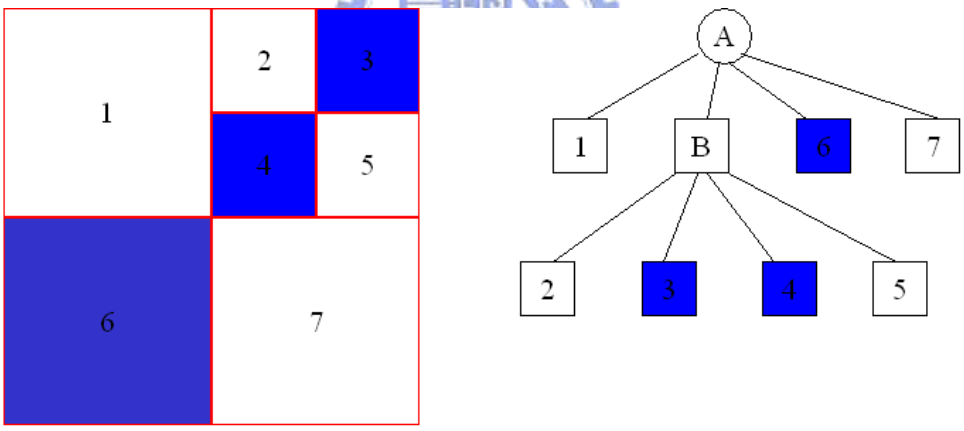
Fig. 2.3. An example of signature representation for objects with different shapes.

2.2 Region-Based Methods

Region-based methods, like quadtree decomposition and polygonal approximation and [2], [9]-[16], have been proposed to describe an image in terms of smooth regions. In region-based methods, images are decomposed into smooth regions in predefined shapes. Then, data structures are designed to record the decomposed smooth regions.

2.2.1 Quadtree decomposition

Quadtree is a hierarchical data structure used for image representation [9]-[11]. Based on successive subdivision of image array into four equal-size quadrants, quadtree structure can be used to represent an image. Each quadrant is then represented by a node in the tree. The root of the quadtree represents the entire stored image while each of its four children represents a one-fourth of the image. If a quadrant is not a homogeneous region, that quadrant is recursively partitioned into four sub-quadrants and that node is split into four grandchildren nodes, and so on. An example for quadtree decomposition is shown in Fig. 2.4, where an object is indicated in blue in Fig. 2.4(a). By quadtree decomposition, the object can be represented by node (3,4,6).



(a) (b)

Fig. 2.4 Concept of quadtree. (a) Partitioned image. (b) Quadtree representation.

2.2.2 Polygon approximation

For polygon approximation, smooth regions in an image are represented by a series of polygons or triangles [12][13]. Triangulation is a widely used form since it is more flexible to represent a given object by arbitrary triangular elements than by

quadrilateral elements. Different types of triangulation algorithms have been proposed. In [14], the authors adjust the position of mesh nodes to describe curved regions with more nodes while to describe smooth regions with fewer nodes. In [15], a mesh construction approach which gradually deforms a primitive model to fit the observed range data is presented. In [16], an image is decomposed using Delaunay triangulation algorithms. This algorithm may insert mesh points and complete a mesh structure based on a set of points formed by the boundary vertices of the input data.

2.3 Transform-Based Methods

Transform-based methods, like DCT (Discrete Cosine Transform) and DFT (Discrete Fourier Transform) [17]-[19], have been used to describe an image in terms of its transform coefficients. DFT uses Fourier orthogonal basis to transform an image from the spatial domain f(x,y) to the frequency domain F_f(u,v), based on the equation:

$$F_f(u,v) = \frac{1}{LW}\sum_{x=0}^{L-1}\sum_{y=0}^{W-1} f(x,y)\exp^{-j2\pi(\frac{ux}{L}+\frac{vy}{W})}, \quad (2.1)$$

where L and W indicate the length and width of an image, respectively. On the other hand, DCT uses cosine orthogonal basis to transform an image from the spatial domain f(x,y) to the frequency domain F_f(s,t), based on the equation:

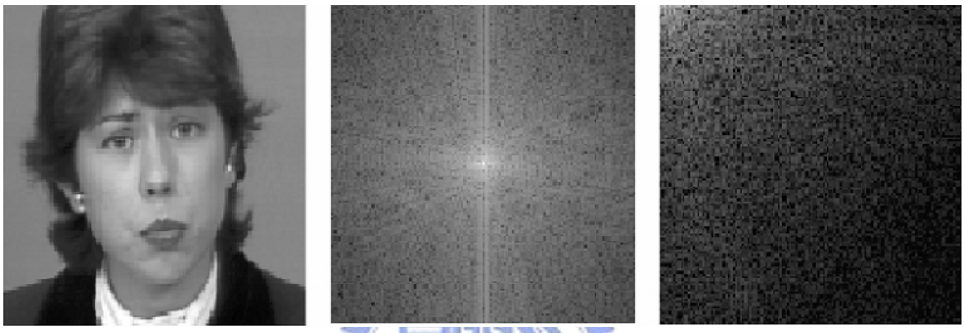
$$F_c(s,t) = \frac{2}{\sqrt{LW}}C(s)C(t)\sum_{x=0}^{L-1}\sum_{y=0}^{W-1} f(x,y)\cos\left[\frac{(2x+1)s\pi}{2L}\right]\cos\left[\frac{(2y+1)t\pi}{2W}\right], \quad (2.2)$$

where C(ξ) is defined as

$$C(\xi) = \frac{1}{\sqrt{2}} \quad \text{if } \xi=0 \quad (2.3)$$

$$= 0 \quad \text{otherwise.}$$

In general, DCT has a better energy compaction capability than DFT. Examples of DFT and DCT transform are shown in Fig. 2.5(b) and (c). For natural images, the high-frequency components are often small enough to be neglected with little visible distortion. Moreover, the DCT transform has been widely adopted for international image/video compression standards, such as JPEG, MPEG-1, MPEG-2, and MPEG-4.



(a) (b) (c)

Fig. 2.5. Examples of transform-based methods. (a) Original image. (b) DFT transform. (c) DCT transform.

2.4 Multi-resolution-Based Methods

Multi-resolution methods, like Gaussian and Laplacian pyramid [20]-[22] and wavelet decomposition [23]-[26], have been proposed to describe an image in hierarchical forms. In multi-resolution methods, an original image is decomposed into multiple images in different resolution or scale. These decompositions are useful in plentiful applications, such as feature detection and image compression.

2.4.1 Gaussian pyramid and Laplacian pyramid decompositions

For Gaussian pyramid construction, the original image I_0 is repeatedly convolved with a Gaussian filter g and subsampled to create the reduced images I_1, I_2, etc. On the other hand, to obtain a Laplacian pyramid H, the r-th level H_r, which is a bandpassed image, is obtained by computing the difference of I_r and I_r*g. The image I_r is then downsampled to form the next image hierarchy. An example of Gaussian pyramid and Laplacian pyramid is shown in Fig. 2.6. In [21], a Laplacian image pyramid is form to represent image. The decomposition (analysis) and reconstruction (synthesis) architecture is shown in Fig. 2.7. The decomposition process is formed in a reduce-and-subtract manner, while the reconstruction process is formed in an expand-and-sum manner.

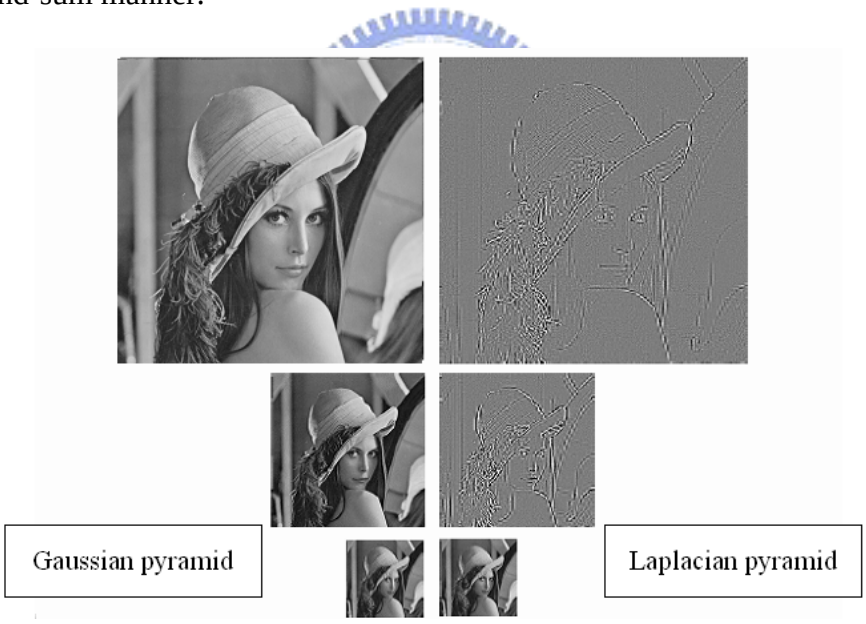


Fig. 2.6. Example of Gaussian pyramid and Laplacian pyramid.

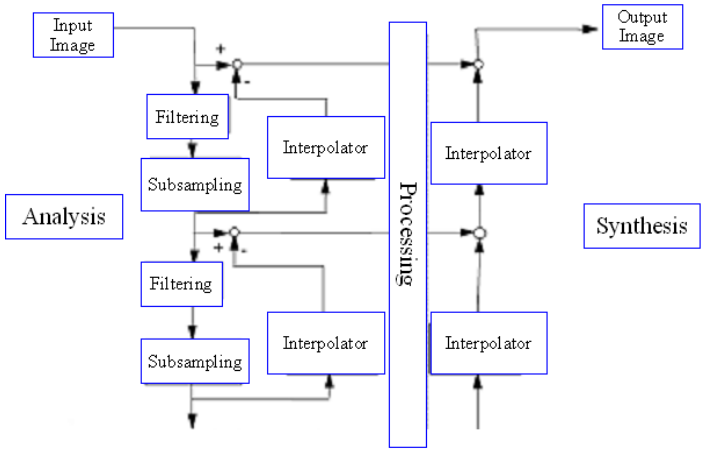
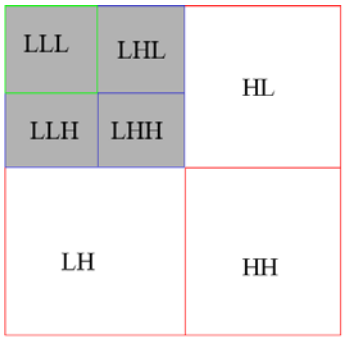


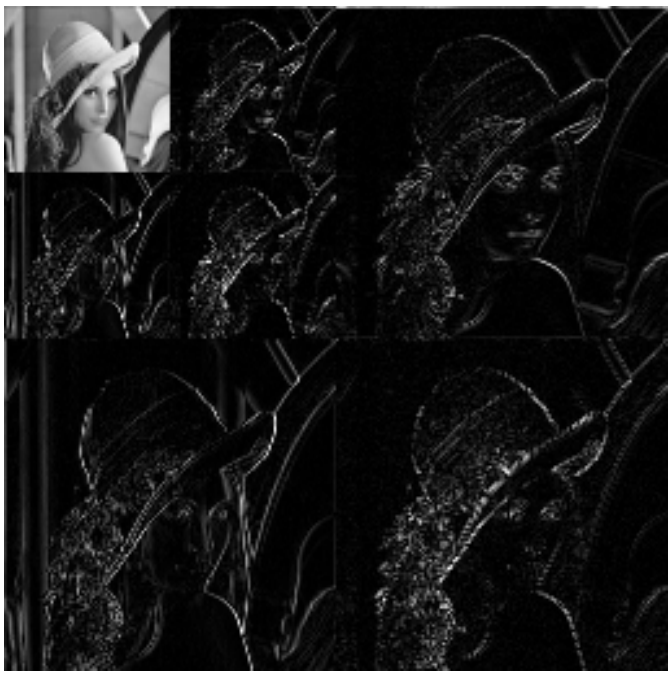
Fig. 2.7. Architecture of image decomposition and reconstruction using Laplacian pyramid.

2.4.2 Wavelet-based methods

In [23], a mathematical model for wavelet multi-resolution signal decomposition is presented. A signal is decomposed into a coarser-scale component using a scaling function and a detail component using a wavelet function. The coarser scale component is further decomposed to form a wavelet representation. As compared with the Fourier basis functions used in Fourier transform, wavelets are localized in both frequency/scale domain and spatial domain, while Fourier basis functions are only localized in frequency domain but not in spatial domain [25]. Hence, small frequency changes in the Fourier transform will produce changes everywhere in the spatial domain. In wavelet transform, this localization may be useful in several applications. An example of wavelet decomposition with a 3-layer hierarchy is shown in Fig. 2.8(a) and (b).



(a)



(b)

Fig. 2.8. An example of wavelet decomposition with a 3-layer hierarchy (a) Coefficient notation. (b) Wavelet decomposition of a real image.

2.5 Fractal-Based Methods

A fractal is self-similar across different scales [27]-[30]. Fractal image representation is a technique to represent an image by a contractive transform on the vector space of images [27]. For this contractive transform, the fixed point is the original image. If a concise configuration that will reproduce the image is found, then we can store the configuration and produce a reconstructed output with a small

amount of information. Image decomposition and reconstruction can be done through iterative function system (IFS), which is a collection of transformations (including geometric and contrast/brightness manipulations) that yield the fixed point after an arbitrarily large number of iterations. For an IFS to converge to a fixed point, a contractivity condition must be satisfied. More specifically, each successive iteration of the transformations in an IFS must approach closer to the fixed point, where "closer" is defined by a metric on the vector space of images under consideration. In [29], Jacquin proposed partitioned IFS to reduce the calculation time. In [30], Barthel et al proposed a method applying the discrete cosine transform (DCT) to each range and domain block prior to the finding of the contraction mapping.

2.6 Recent Works

In recent years, new methods are still emerging, trying to offer new ways to effectively represent images. In [31], an image is represented in an edge-based approach by parametrically modeling relevant image surface variations. A Gaussian smoothed intensity edge model is used to fit several edge parameters, like contrast, width, and edge center. An approximation of the original image can be reconstructed under the framework of regularization theory. Furthermore, a multi-scale smooth model in the regularization is used to speed up image reconstruction. For storing the edge data, chain code technique modeled by a second order Markov chain is applied, and Huffman table is used to perform entropy coding.

In [32], image is represented using irregular distributed samples in space. In this work, image surface is decomposed into N transversal layers along intensity axis, where N is specified by users to determine the required representation accuracy. Then, the irregular samples are chosen along contours with the spacing determined by the

contour density in the neighborhood. To reconstruct an original image surface, triangular mesh is used based on these irregular samples.

In [33], an image is represented by multi-resolution decomposition and Gaussian Markov random field simultaneously. More specifically, an image is decomposed using multi-resolution decomposition to form a Gaussian pyramid or Laplacian pyramid, together with an estimated set of Gaussian Markov random field parameters. These parameters increase representation overhead slightly, but benefit the optimal estimation process for reconstruction under Bayesian statistical inference.

In [34], the components obtained by multi-scale differential operators are used to represent images. The authors use B-spline functions to approximate the Gaussian function to provide efficient algorithms for multiscale kernel computation. Images are represented by these differential operators such as Canny operator-like wavelet and LoG-like wavelets. Moreover, multi-scale and multidirectional operators are proposed to represent texture image. The proposed representations provide geometric information for further applications like edge detection and image enhancement.

In [35], multi-scale edges, accompanied with a suitable wavelet model, are used to decompose an image. The authors presented that mutliscale edges can be detected by local maxima using quadratic spline wavelet. The edges can be characterized by the evolution of wavelet maxima across scales. Image reconstruction is achieved via an iterative projection algorithm. A pipeline hardware architecture is also built to perform real time image reconstruction.

In [36], a set of block pattern models that satisfy certain image variation constraints are used to represent images. In the proposed method, three block types, constant, oriented, and irregular, map to different features, like shades, edges, and textures, in an image. Constant blocks are identified based on just noticeable

difference (JND). Oriented blocks are identified using principal orientation basis. Irregular models are classified into mixed model, such as corners or curved lines, and texture model (without line-like features). Based on these blocks with parameters, image coding, image smoothing, and image zooming can be achieved.

In [37], in which the concept of using high-curvature points to represent images is first proposed, Liu used two one-dimensional morphological operators, dilation gradient and erosion gradient, to detect high-curvature points on image surfaces. Given an intensity profile f(x) and a structure element b(x), dilation track d(s) and erosion track e(s) are defined as

$$d(s) \equiv (f \oplus b)(s) = \max(\{f(s-x) + b(x) \mid (s-x) \in D_f ; x \in D_b\}, \quad (2.4)$$

and

$$e(s) \equiv (f - b)(s) = \min(\{f(s+x) - b(x) \mid (s+x) \in D_f ; x \in D_b\}. \quad (2.5)$$

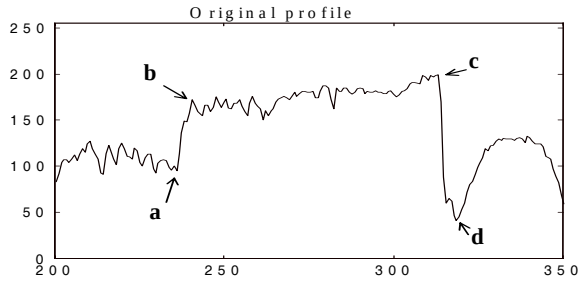
where D_f , D_b are the domains of f(x) and b(x). The dilation gradient and erosion gradient are expressed as

Dilation gradient = d(s)-f(s)-r, (2.6)

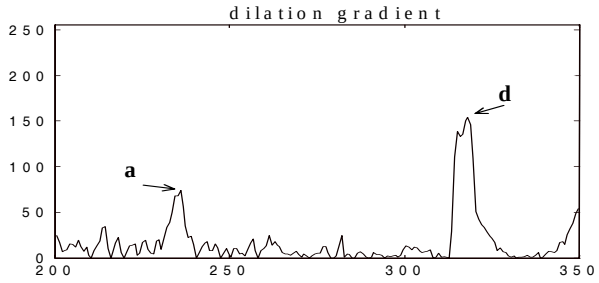
and

Erosion gradient = f(s)-e(s)-r, (2.7)

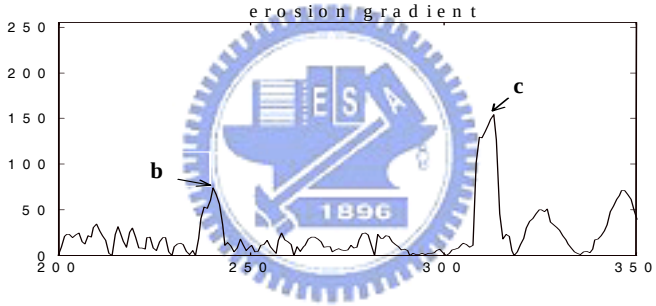
where the radius r is subtracted from the gradient values for the purpose of normalization. Fig. 2.9 shows the dilation and erosion gradients of an image profile. The local maxima in dilation and erosion gradients indicate the places of high-curvature points.



(a)



(b)



(c)

Fig. 2.9 Feature extraction by dilation gradient and erosion gradient. (a) Original profile. (b) Dilation gradient. (c) Erosion gradient.

In [38], based on Liu's work, Jong used the 2-dimentional operator $\begin{bmatrix} \frac{\partial^2 f}{\partial x^2} & \frac{\partial^2 f}{\partial x \partial y} \\ \frac{\partial^2 f}{\partial y \partial x} & \frac{\partial^2 f}{\partial y^2} \end{bmatrix}$ to detect high-curvature points on image surfaces. The eigenvectors of this matrix indicate the direction the principal curvatures and the eigenvalues are proportional to the value of the principal curvatures. Surfaces with large eigenvalues represent more curved surfaces. The positions of these high-curvature points can be extracted by detecting the places with large eigenvalues. In Jong's work, the behavior

of the selected high curvature operator is not extensively discussed, and the curvature thresholds are determined manually.

In this thesis, based on the prototypes developed in [37] and [38], we further derive a comprehensive and versatile representation framework. We have revisited the one-dimensional case using one-dimensional differential estimators, comprehensively reinvestigated the two-dimensional verge point detectors, developed the theoretic parts of these operators, and explored the selection of thresholds. In addition, we have developed frameworks for image compression, image feature detection, and image editing. These explorations will be presented in Chapter 3, Chapter 4, and Chapter 5.