Chapter 3

FEATURE DETECTION

Concepts Introduced in this Chapter

› Edge tracking;

› Canny edge operator;

› Blobs and vesselness, Hough transform for lines;

› SIFT, SURF, BRIEF, and ORB features, superpixels, texture, MSER features, local shape context, texture features, HOG features, gist, and saliency

Feature Detection Process

- There are a number of reasons to apply a feature detection step in medical image analysis:
- In cases where the object of interest is simple, its attributes may be captured directly by the feature detector. This leads to object detection. More often, it is a preprocessing step.
- Feature locations and attributes may serve as an object-specific reference if the same object is captured by different images that shall be compared. While the content of the images may be different (e.g., if one of the images is CT and the other MRI), the object features such as edges or corners should be extractable in both images.
- Features may also help to define a region of interest that shall be further inspected. A blob detector, for instance, could find potential regions of interest that may comprise the blob-like lymph nodes.
- Features may also help to guide a segmentation process in cases where the data is noisy or of low contrast.
- Finally, features may also be used to characterize deviations from an assumed norm of an anatomic structure. As an example, features to enhance tubular structures may highlight potential sites of aneurisms in the vascular system, since they deviate from the normal, tubelike shape of vessels

Edge Tracking

- Structures are meant to be detectable by a change of appearance between them and the background. Furthermore, the shape given by a structure's outline may be an important characteristic to differentiate it from other objects or to specify object-specific locations.
- Hence, edges are relevant features for several types of analysis tasks.
- Edges can be those of intensity but may also separate different textures.
- For edge detection in medical images, various, rather simple assumptions are used to separate edges from noise:
- • The gradient for edges is often stronger than that caused by noise.
- • The edge direction varies slowly along the edge.
- • The edge strength varies slowly along the edge.

Canny edge operator

- The Canny edge operator takes into account all three assumptions.
- It consists of an edge-enhancement step and an edge-tracking step. Edge enhancement is carried out by taking the maximum response from several one-dimensional edge filters, because—under an idealized edge model—the optimal response for an edge is created by a one-dimensional smoothing differential operator orthogonal to the yet unknown edge direction.
- The local maximum of the gradient length then specifies the edge location.
- In most applications, this step is replaced by a two-dimensional gradient operator (e.g., the derivatives of the Gaussian) combined with a non-maximum suppression step that reduces the edge response to a single pixel in gradient direction.
- Non-maximum suppression can be done, e.g., by computing zero-crossings of the second derivative.

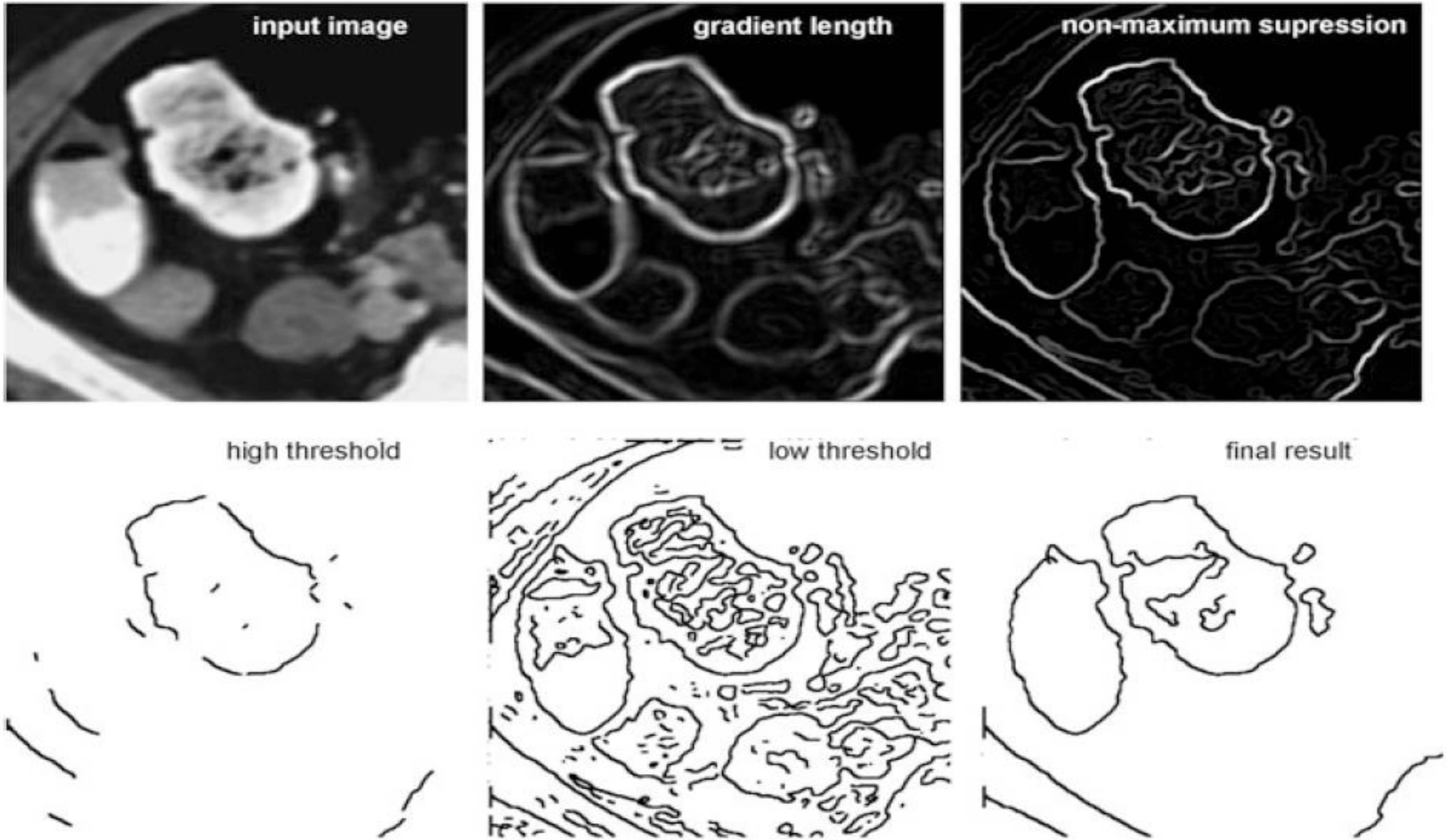

Fig. 5.2 The different steps of the Canny edge operator (the result from non-maximum suppression has been dilated for better visibility)

Edge Model

- An edge model is usually just a local template that is convolved with the image. The template represents the ideal edge.
- The response to the match can be exploited to determine edge locations as well as a measure of confidence for the edge to be present .
- Contour models differ from edge models in that they assume that a set of open or closed contours are searched in the image in a top-down fashion,
- while edge models generate edges in a bottom-up fashion based on a (possibly very elaborate) model of what an edge is in this particular image.

Hough Transform

- The Hough transform computes edge features by comparing image evidence with a very specific edge model.
- Given an image containing edge information of an unknown number of edges of a known kind, the Hough transform finds instances of this kind.
- The Hough transform is a voting scheme that has been first presented to find straight lines in images and then has been extended to find arbitrary kinds of boundaries.
- Each location of a potential boundary—e.g., each location where the gradient length exceeds some threshold—votes for reference points in parameter space that are associated to certain shapes. Parameter combinations that receive the most votes describe likely object instances.
- Being a voting system, the Hough transform keeps its ability to predict structure locations even if some votes are missing because of occlusion or a missing signal.
- The method is robust with respect to noise or artefactual edges that do not follow the edge model.

Hough Transform

- A number of strategies increase the computation speed of the Hough transform and can be applied to most of its variants:
- • The order in which votes are cast does not matter, which makes the method inherently parallelizable.
- • If gradient directions in the edge images are reliable, the number of votes may be reduced by letting every edge point only vote for those solutions for which a is almost perpendicular to the gradient direction.
- • If edge points are selected randomly from the image, intermediate results of the voting process may already be a good estimate for the final outcome.
- There are also some strategies for increasing the robustness of the Hough transform with respect to noise, artifacts, and shape variation:
- • A multi-scale strategy may be applied by computing an initial Hough transform only for large accumulator cells. The result is used as a prediction for ranges of parameters in Hough space that represent potential lines. Accumulation of votes at higher resolution is reduced to these ranges.
- • Vote distribution in parameter space may be smoothed to take variation due to noise and artifacts into account.

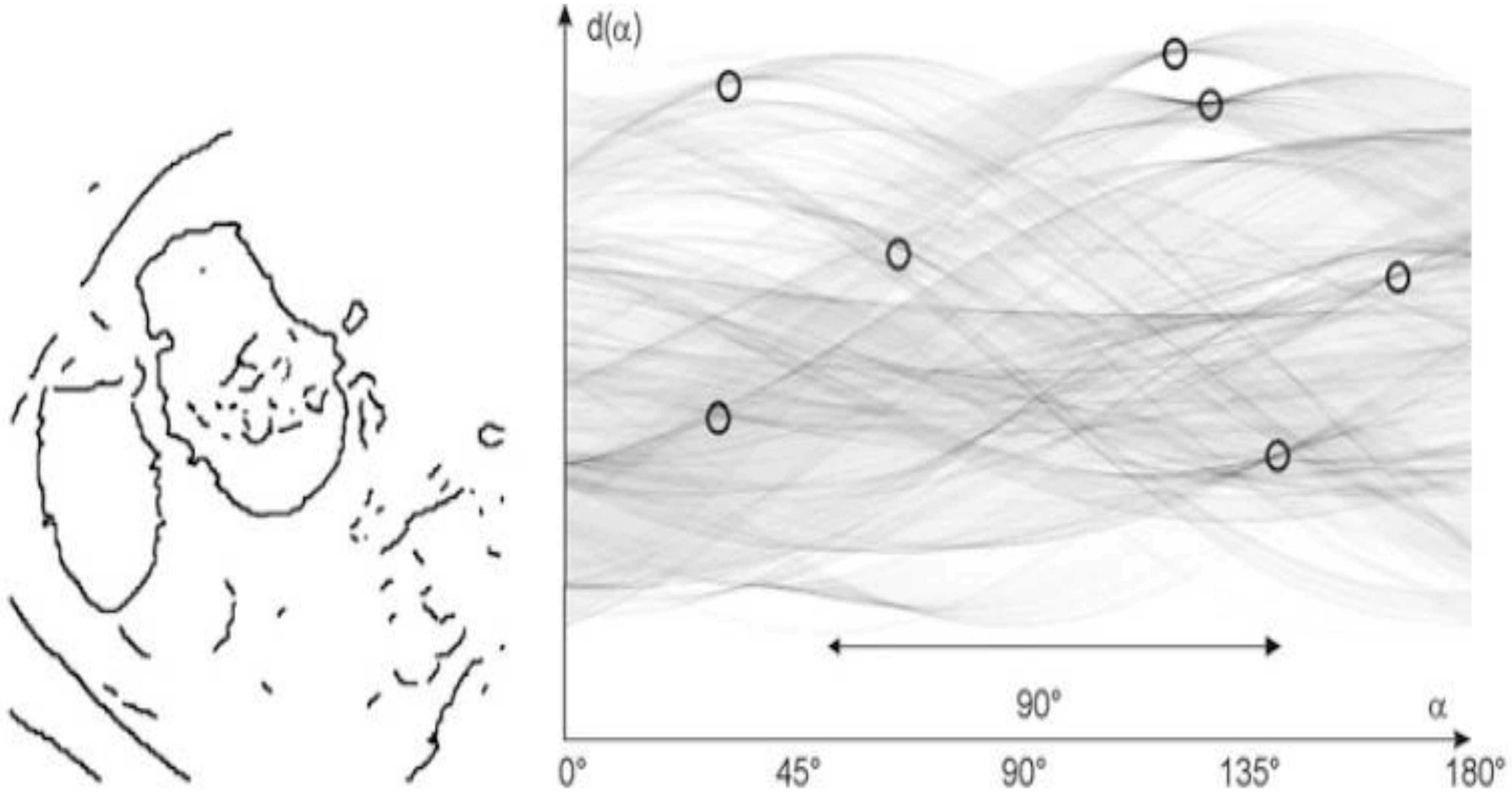

Fig. 5.6 Example of applying the Hough transform on an edge image. The predominance of edges at angles 45° and 135° with respect to the x-axis is visible as local maxima in Hough space

Harris corner detector

- Corners in an image indicate specific locations pertaining to an object. Such locations are important if objects of the same kind are to be compared or if the same object shall be compared in different images (e.g., comparing brain data of a patient from CT and PET).
- Another reason to identify specific locations is to classify an unknown object by its characteristic outline as given by the corner point locations.
- Corners belong to a class of local features that can be computed from images but represent attributes of depicted objects.
- **The Harris corner detector** is based on this assumption and has been shown to have a good performance.
- The detector computes a quantity for a location (x y) that depends on averaged intensity variations in arbitrary directions around (x y). If this variation is high in almost all directions, a point of interest is found .
- The neighborhood across which this variation is averaged constitutes the scale of the corner detector.

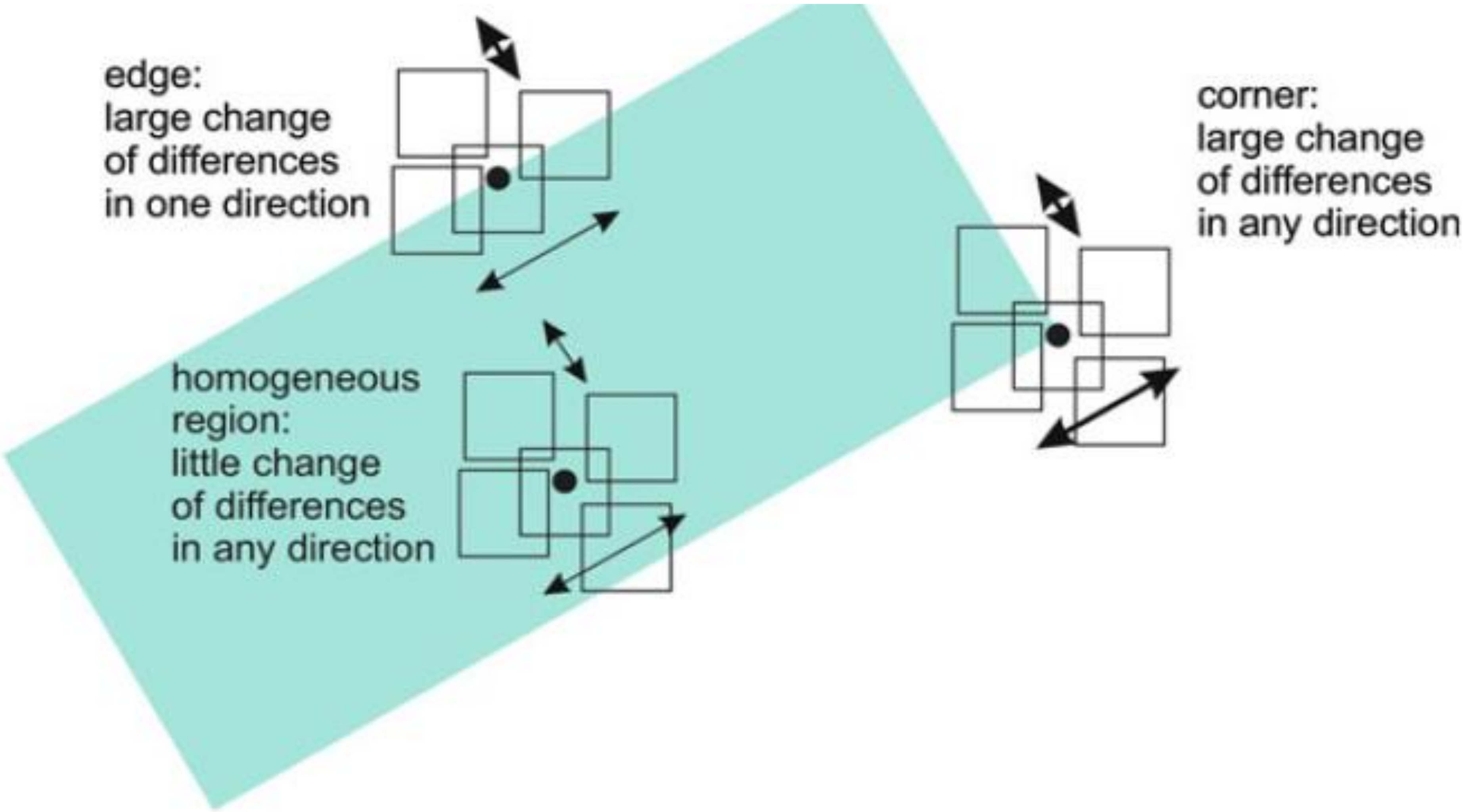

Fig. 5.7 The Harris corner detector measures changes of intensity of a region around a point. A corner causes significant changes for displacement in arbitrary direction, an edge only if displaced orthogonal to it

The corner feature is computed from the eigenvalue of the *Harris matrix*. It consists of the partial derivatives, weighted with some function w of the image function $I(x, y)$ in a neighborhood k around a direction $(x_0\ y_0)$ in which the image is displaced and then subtracted. It is defined as follows

$$H(x_0,\ y_0) = \sum_{x=x_0-k}^{x_0+k} \sum_{y=y_0-k}^{y_0+k} w(x,y) \begin{pmatrix} \left(\frac{\partial I}{\partial x}\right)^2 & \left(\frac{\partial I}{\partial x}\right)\left(\frac{\partial I}{\partial y}\right) \\ \left(\frac{\partial I}{\partial x}\right)\left(\frac{\partial I}{\partial y}\right) & \left(\frac{\partial I}{\partial y}\right)^2 \end{pmatrix}. \qquad (5.2)$$

three different cases can be derived from inspecting the eigenvalues k1 and k2 of H:

- A corner has been found if k1 and k2 are large.
- An edge has been found if either k1 or k2 is large.
- The region is locally homogeneous if k1 and k2 are low.

Texture

Textures are difficult to define, as they are a kind of microstructure in a structured world.

Textures have two properties in common:

1. A texture has a repeating pattern. It may be an exact repetition as in manufactured surfaces or a repetition with a random component as in many textures of natural objects.

2. Computation of texture features requires a texture-specific, minimum size of the window in which all scene elements belong to the same texture.

- Deterministic texture patterns on an object are usually found on manufactured objects. Representation of such textures is by their constituting, deterministic texture elements (texels). In medical imaging applications, textures on objects are caused by the structure of organs imaged. Constituting elements certainly exist, but they are usually much smaller than the imaging resolution.
- **Visible textures** are caused from interaction between the measured signal and these structures. The textures can be described statistically in the spatial or the frequency domain.

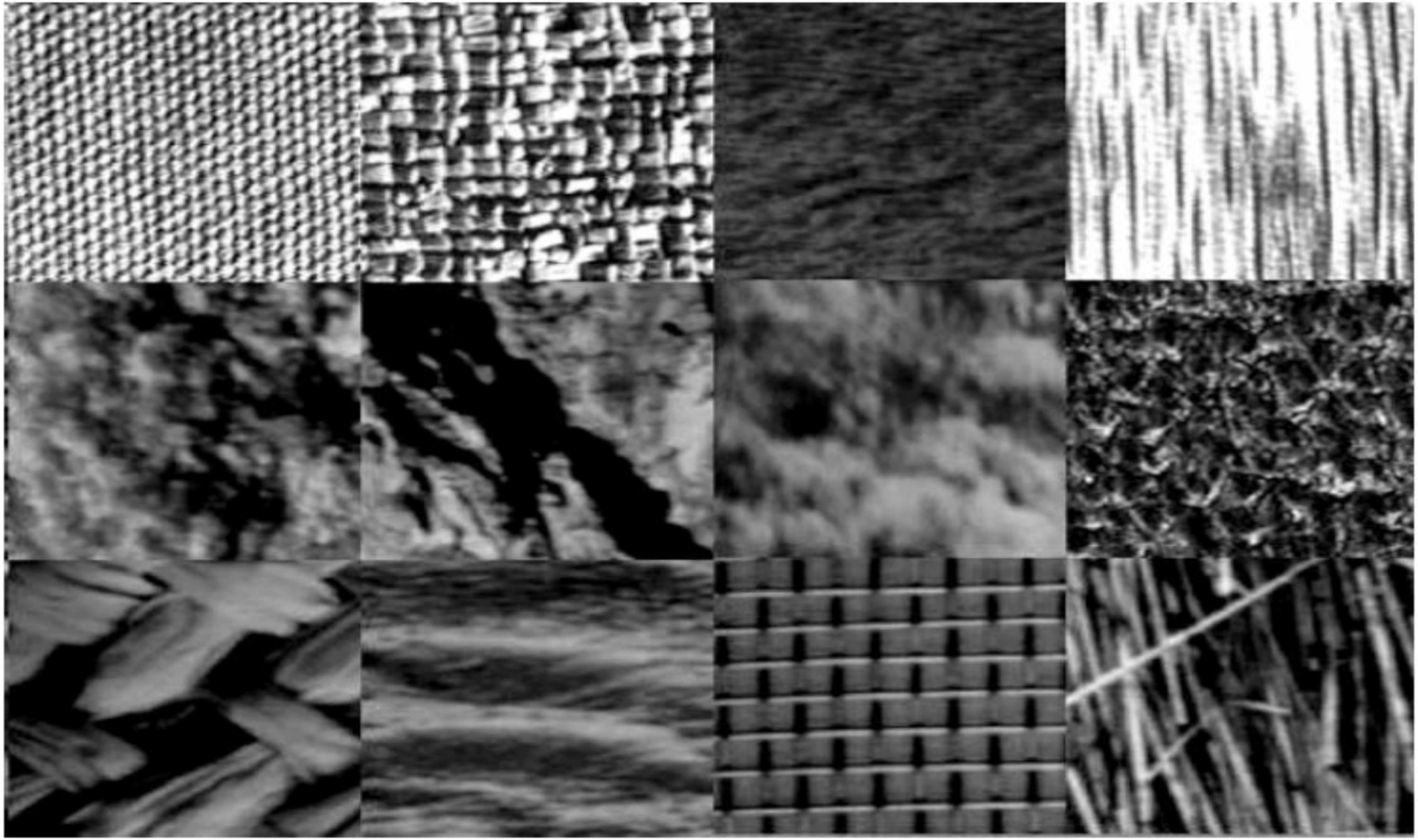

Fig. 5.19 Samples from the Brodatz textures. It can be seen that it will be difficult to define feature sets that capture the characteristics of all these textures

Texture

- It has been shown that three properties are necessary for differentiating adjacent regions by texture:
- 1. **Texture orientation** can best be explained in frequency space. Textures possessing one or more pronounced orientations will have most of the energy in frequency space in sectors along those orientations.
- 2. **Texture periodicity** relates to the smallest region necessary to represent all properties of a texture.
- 3. **Texture complexity** relates to the composition of it. A texture which is composed of basic elements of the same size is less complex than one that consists of basic elements of different sizes. A texture which consists of white noise is less complex than one consisting of colored noise.
- Segmentation by texture plays a minor role in medical image analysis, as most properties from differently textured organs are too subtle to be exploited. Examples for exemptions are as follows:
- • MRI, where artifacts from tissue-specific magnetic field inhomogeneities contribute to noise.
- • Ultrasound, where tissue-specific refraction artifacts influence the signal.
- • Photographic images of the skin, where, e.g., scars exhibit macroscopic textures. • Microscopic images, where cell textures can be spatially resolved.

texture measures

- A multitude of different texture measures exists, but in view of the rather limited use of texture as a feature we will only list some that are exemplary for different ways of measuring statistical properties of texture:
- • Haralick's features of the co-occurrence matrix measure second-order statistical attributes of the gray level co-occurrence matrix (GLCM).
- **Spectral features** are created from integrating amplitudes over a partitioning of the texture representation in the frequency domain. They are a direct measure of orientation and periodicity of a texture in the frequency or spatial domain Gabór filters are a specific variant of this as they combine spatial and frequency characteristics by a bank of differently oriented, windowed frequency transforms.
- **Law's filter** masks are a set of orthogonal convolution kernels to measure periodicity and orientation of textures in the spatial domain.

Local binary patterns (LBPs)

- Local binary patterns (LBPs) compute for each pixel p in a given window the relative intensity w.r.t. neighboring pixels in an 8-neighborhood.
- pixels in the neighborhood are ordered along a circle for the pixel in question (see Fig. 5.20). The local feature is a binary vector with eight entries (stored in a byte). Each entry is set to 1 if the respective pixel is brighter than p and it is set to 0 otherwise. The feature vector consists of all features of all pixel p in the texture window. This is the base version of LBP.
- Textures should not be computed across object boundaries if the goal is the determination of object characteristics.
- If the object is not segmented, it may be useful to split the image into superpixels and compute the texture in the superpixels.

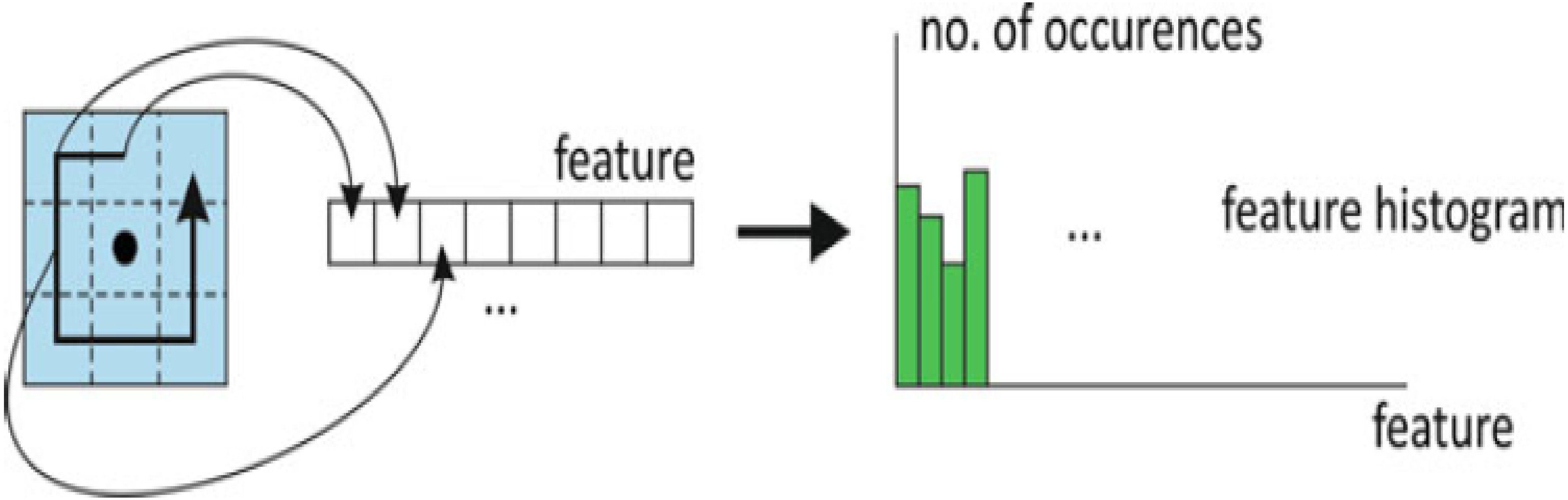

Fig. 5.20 LBPs are generated by comparing the intensity at some location with intensities of its neighbors. If it is greater, a "1" is entered into the bit string; otherwise, a "0" is entered. With 8-neighbors, 256 different feature values are possible. The histogram of features of a texture patch constitutes the feature vector of the texture. Variants of this method use different neighborhoods, remove irrelevant features, or use differential features

Texture descriptors based on histograms

- Measures of texture computed using only histograms carry no information regarding spatial relationships between pixels, which is important when describing texture.
- One way to incorporate this type of information into the texture-analysis process is to consider not only the distribution of intensities, but also the *relative positions* of pixels in an image.
- Let Q be an operator that defines the position of two pixels relative to each other, and consider an image, f , with L possible intensity levels. Let **G** be a matrix whose element g_{ij} is the number of times that pixel pairs with intensities z_i and z_j occur in image f in the position specified by Q, where $1 \le i, j \le L$. A matrix formed in this manner is referred to as a ***graylevel* (or *intensity*) *co-occurrence matrix***. When the meaning is clear, **G** is referred to simply as a *co-occurrence matrix*.

a b c

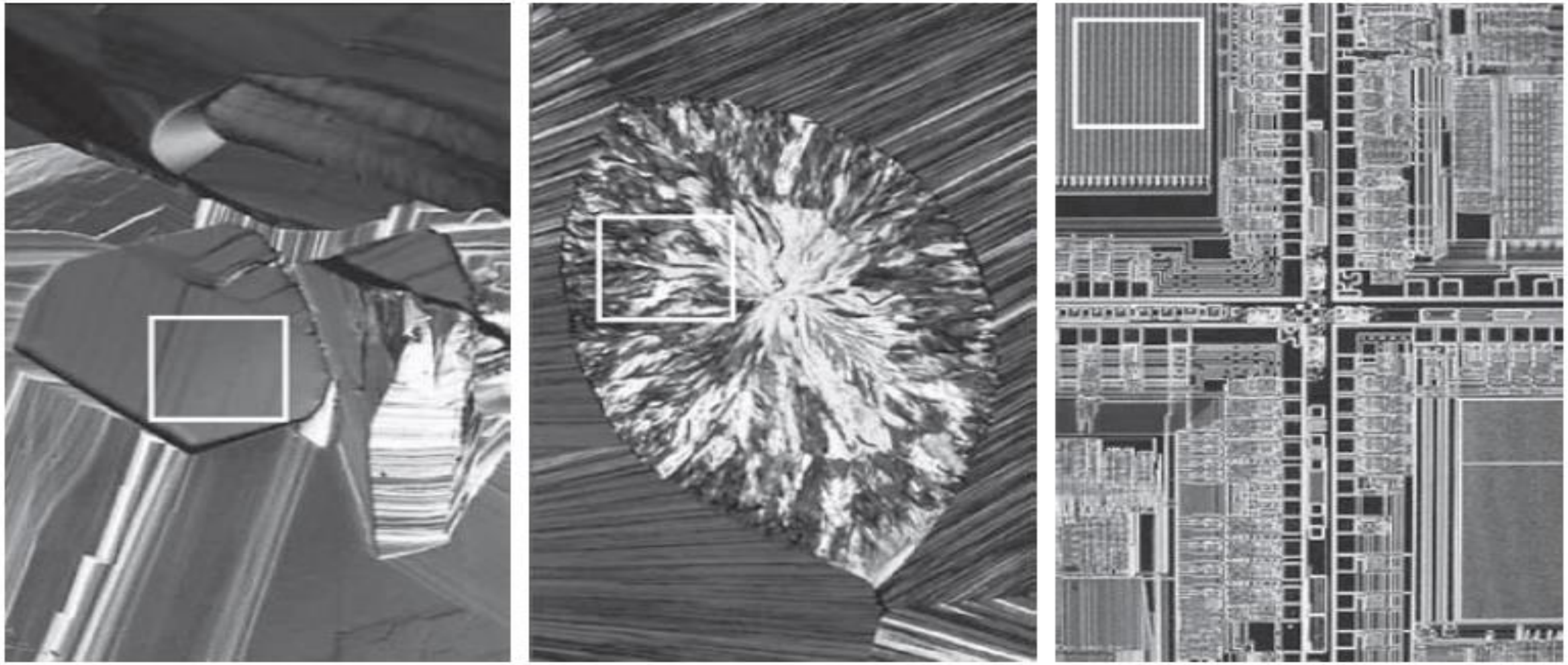

FIGURE 11.29 The white squares mark, from left to right, smooth, coarse, and regular textures. These are optical microscope images of a superconductor, human cholesterol, and a microprocessor. (Courtesy of Dr. Michael W. Davidson, Florida State University.)

TABLE 11.2
Statistical texture measures for the subimages in Fig. 11.29.

Texture	Mean	Standard deviation	R (normalized)	3rd moment	Uniformity	Entropy
Smooth	82.64	11.79	0.002	−0.105	0.026	5.434
Coarse	143.56	74.63	0.079	−0.151	0.005	7.783
Regular	99.72	33.73	0.017	0.750	0.013	6.674

FIGURE 11.30
How to construct a co-occurrence matrix.

1	1	7	5	3	2
5	1	6	1	2	5
8	8	6	8	1	2
4	3	4	5	5	1
8	7	8	7	6	2
7	8	6	2	6	2

Image f

	1	2	3	4	5	6	7	8
1	1	2	0	0	0	1	1	0
2	0	0	0	0	1	1	0	0
3	0	1	0	1	0	0	0	0
4	0	0	1	0	1	0	0	0
5	2	0	1	0	1	0	0	0
6	1	3	0	0	0	0	0	1
7	0	0	0	0	1	1	0	2
8	1	0	0	0	0	2	2	1

Co-occurrence matrix **G**

- how to construct a co-occurrence matrix using $L = 8$ and a position operator Q defined as "one pixel immediately to the right.
- The array on the left is a small image and the array on the right is matrix **G**. We see that element (1,1) of **G** is 1, because there is only one occurrence in *f* of a pixel valued 1 having a pixel valued 1 immediately to its right
- Similarly, element (6,2) of **G** is 3, because there are three occurrences in *f* of a pixel with a value of 6 having a pixel valued 2 immediately to its right.
- If we had defined Q as, say, "one pixel to the right and one pixel above," then position (1,1) in **G** would have been 0 because there are no instances in *f* of a 1 with another 1 in the position specified by Q.
- On the other hand, positions (1,3), (1,5), and (1, 7) in **G** would all be 1's,

TABLE 11.3 Descriptors used for characterizing co-occurrence matrices of size $K \times K$. The term p_{ij} is the ij-th term of $\mathbf{G}$ divided by the sum of the elements of $\mathbf{G}$.

Descriptor	Explanation	Formula
Maximum probability	Measures the strongest response of $\mathbf{G}$. The range of values is $[0, 1]$.	$\max_{i,j}(p_{ij})$
Correlation	A measure of how correlated a pixel is to its neighbor over the entire image. The range of values is 1 to -1 corresponding to perfect positive and perfect negative correlations. This measure is not defined if either standard deviation is zero.	$\sum_{i=1}^{K}\sum_{j=1}^{K}\frac{(i-m_r)(j-m_c)p_{ij}}{\sigma_r \sigma_c}$ $\sigma_r \neq 0;\ \sigma_c \neq 0$
Contrast	A measure of intensity contrast between a pixel and its neighbor over the entire image. The range of values is 0 (when $\mathbf{G}$ is constant) to $(K-1)^2$.	$\sum_{i=1}^{K}\sum_{j=1}^{K}(i-j)^2 p_{ij}$
Uniformity (also called Energy)	A measure of uniformity in the range $[0, 1]$. Uniformity is 1 for a constant image.	$\sum_{i=1}^{K}\sum_{j=1}^{K}p_{ij}^2$
Homogeneity	Measures the spatial closeness to the diagonal of the distribution of elements in $\mathbf{G}$. The range of values is $[0, 1]$, with the maximum being achieved when $\mathbf{G}$ is a diagonal matrix.	$\sum_{i=1}^{K}\sum_{j=1}^{K}\frac{p_{ij}}{1+\lvert i-j\rvert}$
Entropy	Measures the randomness of the elements of $\mathbf{G}$. The entropy is 0 when all p_{ij}'s are 0, and is maximum when the p_{ij}'s are uniformly distributed. The maximum value is thus $2\log_2 K$.	$-\sum_{i=1}^{K}\sum_{j=1}^{K}p_{ij}\log_2 p_{ij}$

Template Matching

- Template matching may be used to detect or highlight structures if their shape is known and simple. Two examples are blobness and vesselness filters.
- Blobs are circular structures in the images. Blob detection either detects such structures (e.g., in counting microorganisms in cell microscopy) or highlights blobs as important features of an object to be detected, e.g., in the detection of lung nodules.
- Formally speaking, a blob is a local maximum or minimum of a radially symmetric intensity distribution.
- Blobs are characterized by their size and location. If size is known, blobs can be enhanced by a matched filter of the same size and shape.
- The Laplacian of Gaussian with standard deviation r according to the blob size is often used.
- Alternatively, the Difference of Gaussian (DoG) filter may be used. The response of a DoG filter is computed by subtracting the filter response of the image with two Gaussians with different standard deviation from each other.

Template Matching

- The local minima or maxima in this space represent the blob location and the size of the blob.
- A different approach is to use the determinant of the Hessian matrix. It evaluates the differential properties of the point-like shape of blobs.
- The vesselness filter finds tubular structures. The basic idea is similar to the definition of the blobness filter.
- If an elongated structure can be discriminated in the image (brighter or darker than the background), then the preferred direction can be determined from an eigen decomposition of the Hessian matrix.
- **The vesselness filter was presented to find vessels in MRA images but may also be used to find other tubular structures in the human body.**

SIFT Feature and SURF

- The scale-invariant feature transform (SIFT) has been developed by Lowe (1999, 2004) and patented by the University of British Columbia.
- SIFT generates and uses features to detect and identify objects in images. Local features are identified and represented in a descriptor. Objects are identified by comparing expected feature configurations with all possible subsets of configurations from features detected in an image. The object is detected if a sufficiently large correspondence has been found. The method proceeds in several steps:
- • key point generation,
- • key point reduction,
- • feature computation, and
- • key point matching.

SIFT Feature

- In key point generation, rotation- and scale-invariant features are generated by searching for local extrema of a multi-scale blob detector based on a multi-scale Difference of Gaussian .The detector is insensitive with respect to noise by determining an optimal smoothing for each scale.
- Key point generation will create numerous responses from noise and artifacts.
- In key point reduction, the contrast at local extrema is used to remove low-contrast blob locations In feature computation, orientation attributes are determined and stored for each remaining key point. A histogram of gradient directions is computed for locations in the neighborhood of a key point.
- Key point features can then be used for matching model key points with key points extracted from the image.

SURF Features

- Although the objective for developing the SIFT procedure was to identify objects by key features, its main application in medical image analysis is to support feature-based registration.
- Registration finds a transformation to map two images of the same object onto each other.
- The reason for using SIFT features is that it can be assumed that this mapping will only be successful if a sufficiently large number of features correspond in the two images. A registration is particularly easy if this correspondence is given by sets of pairs of corresponding feature locations.
- A faster variant to SIFT is SURF [speeded-up robust features (Bay et al. 2006)]. It uses essentially the same mechanism, except for the fact that the slow convolutions in SIFT are replaced by faster approximations.

Binary Key Point Descriptor and Detectors

- Key point matching plays an important role in tracking, registration, and matching tasks as their invariance w.r.t. several image acquisition-related artifacts makes them good candidates for object-specific, local features.
- BRIEF (binary robust independent elementary features) is a binary descriptor that can be computed given a key point.
- The key point could be generated, e.g., from a corner detector such as SUSAN or FAST or it could be the key points from a SIFT computation where the descriptor computation is omitted.
- Similar to local binary patterns, BRIEF produces binary features from comparing intensities at pixel locations.

BRIEF

- For BRIEF, these are locations pi and qi in an S S neighborhood around key point k (see Fig. 5.13).
- For a given number N of pairs (pi, qi) of sample locations in this neighborhood, the binary feature bi for each pair (pi, qi) is computed from smoothed intensities Ir (Gaussian smoothing with standard deviation r) as follows
- The features are then concatenated in a bit string of length N/8 bytes. The Hamming distance is used for matching key point locations. It measures the difference between two strings of equal size by counting the number of different characters at the same position in the string.

$$b_i = \begin{cases} 1, & \text{if } I_\sigma(p_i) < I_\sigma(q_i) \\ 0, & \text{otherwise} \end{cases}$$

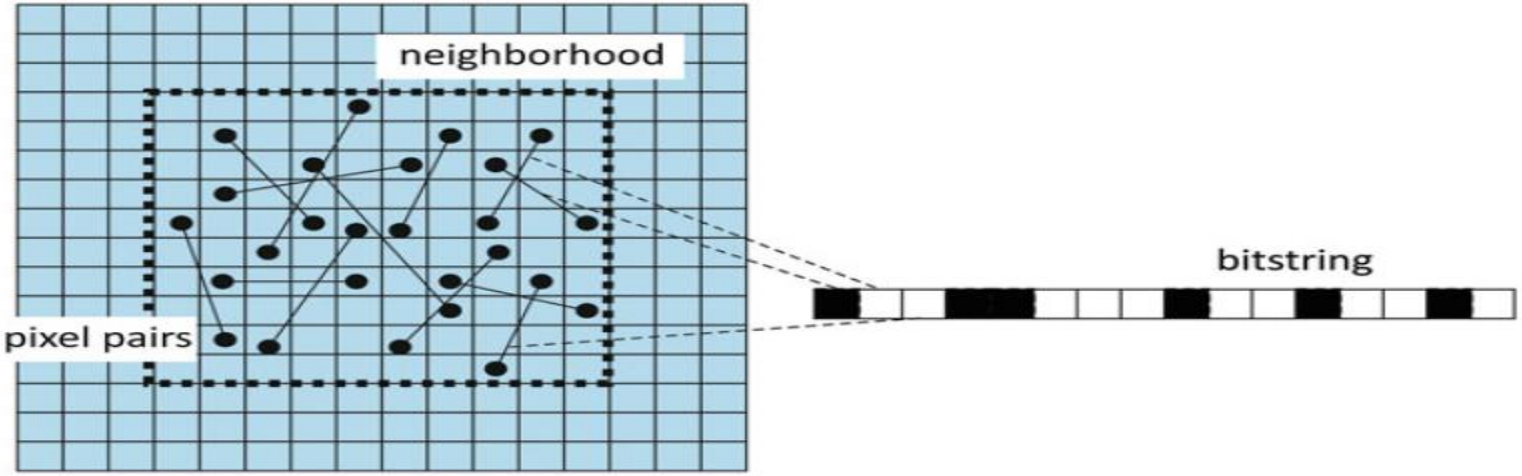

Fig. 5.13 The BRIEF descriptor is constructed from a set of pixel pairs in a predefined neighborhood around each key point. Each pair is represented by a bit in a bit string that is then interpreted as integer

MSER Features

- Locations, such as the center of gravity, generated from maximally stable extremal regions (MSER) .
- The key concept of this approach is to separate an image into local homogeneous regions with maximum contrast. Opposed to SIFT feature and SURF, this detector defines key points based on intensity and not on gradients.
- The definition for an MSER can be described informally as follows:
  - A region r_t is any connected set of scene elements **s** with intensity $f(\mathbf{s}) > t$, where t is an arbitrary threshold.
  - An instability measure $s_\Delta(r_t, t)$ (see Fig. 5.17) rates the size change of r_t when the threshold is varied by some value Δ.

    where $|\cdot|$ indicates computing the volume of the respective region.
  - A region r_t is maximally stable if $s_\Delta(r_t, t)$ is minimum for t.

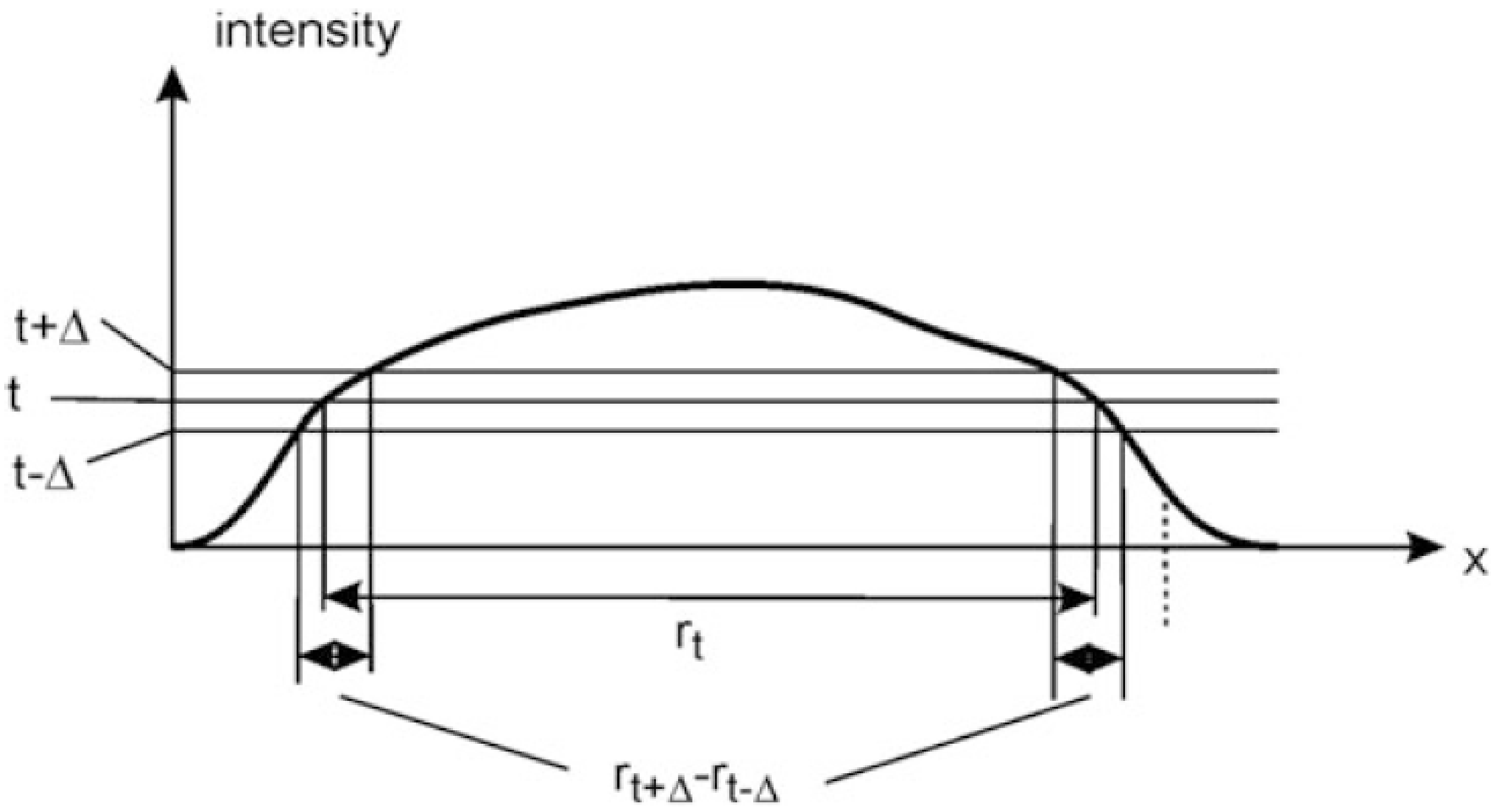

Fig. 5.17 The instability measure rates the size change of a region for some threshold variation against the region size. It is minimal for a maximally stable extremal region

$$s_\Delta(r_t, t) = |r_{t+\Delta} - r_{t-\Delta}|/|r_t|, \qquad (5.20)$$

Superpixel

- If features relate to the interior of regions rather than to edges and corners, these regions have to be computed prior to feature computation.
- Pixels belong to the same region if some homogeneity criterion applies to them. Hence, a necessary step for computing region-based features is a segmentation of the image to find the regions.
- To deal with the separation of medically relevant structures (organs, pathological processes, etc.) from the images, the segmentation only needs to extract large enough segments that enable computation of region features.
- Computation of superpixels is a means to arrive at such regions.
- Each superpixel consists of a set of connected pixels from the original image. Superpixels are non-overlapping and space-filling.

Superpixel

- The motivation for the computation of superpixels is twofold:
- • Stable separation of information from noise is often impossible at the pixel level. Hence, the true resolution of the image, i.e., the level of detail of the information that is represented by the image, is lower than the actual spatial resolution.
- • The searched level of detail, such as segment membership, is lower than the image resolution. Regional features may vary across a segment, but this variation may be captured much better by few superpixels than by the pixels themselves.
- superpixels can be generated by a data-driven segmentation technique such as the watershed transform or normalized graph cuts.
- A good overview on methods to generate superpixels is **SLIC (simple linear iterative clustering) method** as a fast and simple alternative.

- SLIC carries out a local clustering in a feature space of pixels where each pixel is attributed to its location and its intensity (or color, if color images are to be processed).
- Superpixel centers are initially seeded evenly over the image.
- The spacing between seeds depends on the desired number of superpixels.
- Then, pixels are assigned to the most similar superpixel center.

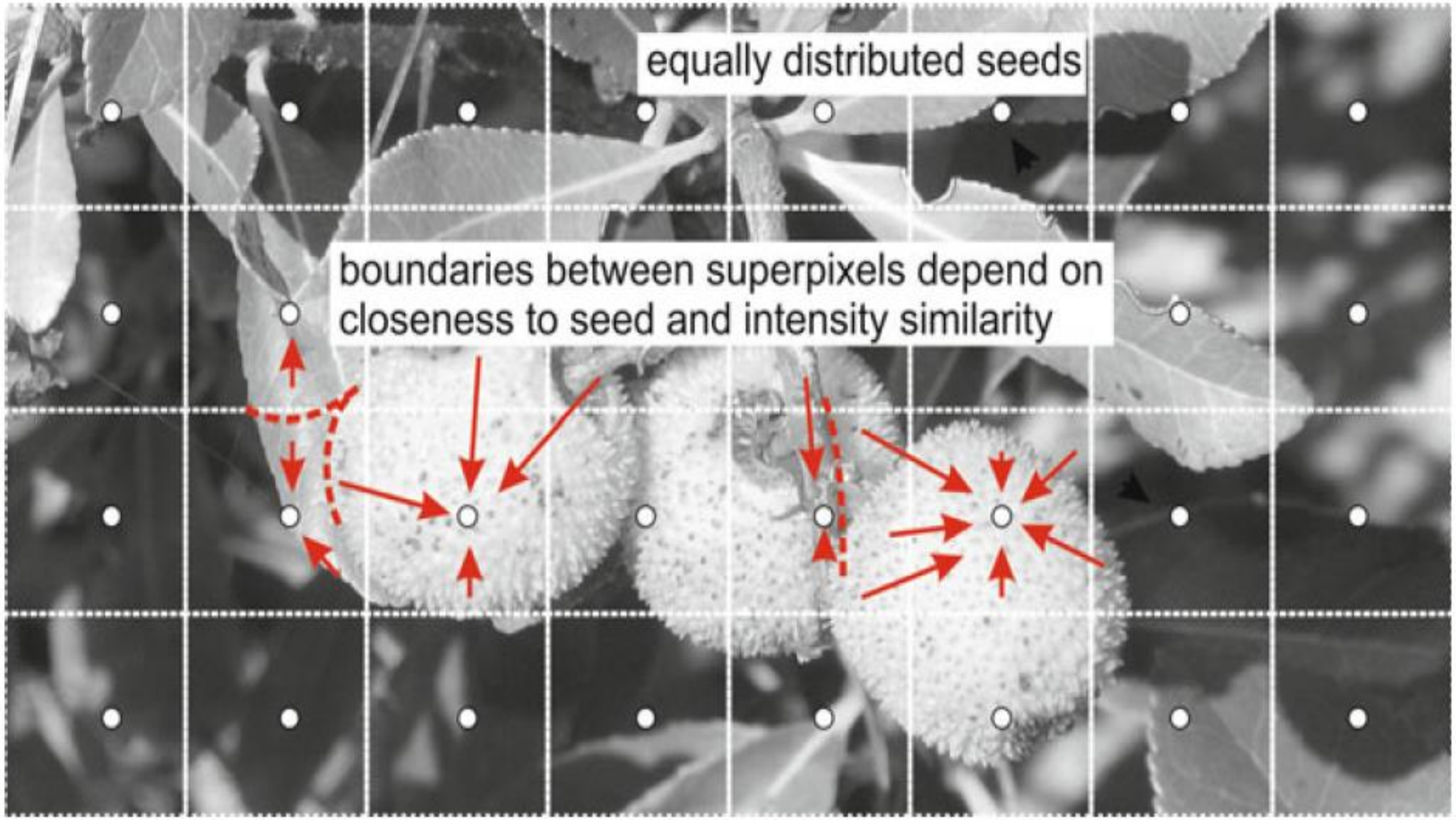

Fig. 5.15 The number of SLIC superpixels depends on the number of seeds that are evenly distributed over the image. All other pixels are assigned to their most similar seed. Similarity depends on intensity and distance between the pixel in question and the seed

Histogram of Oriented Gradients

- The histogram of gradients (HOG) is a feature detector that does not rely on boundaries as it samples a dense gradient texture map and uses this to generate the features.
- Hence, HOG features are always computed from a gridded region of interest that is assumed to be mostly occupied by the structure to be analyzed.
- A HOG cell is defined at each gridline intersection. The authors suggest two types of cells. R-HOG cells have a rectangular shape, while C-HOG cells have a circular shape. Gradients are computed for each pixel. Each pixel within a cell votes for a direction in a binned histogram of gradient directions. The vote is weighted with the strength of the gradient.

HOG

- In order to make the response insensitive to intensity variation, cells are grouped into larger blocks. The gradient strength is the normalized with the average gradient in this block.
- LBP features have been combined with HOG features for pedestrian detection, and its superiority w.r.t. HOG features alone has been shown by Wang et al. (2009).
- HOG features, either the original or variants, have been used for classification and detection of objects in medical and biological images .
- Examples are the lesion detector in lung CT of Song et al. (2012) who also combine LBP features with HOG features for classification or Sanroma et al. (2014) who use HOG features to classify the similarity of atlases to an image to be segmented in a multi-atlas segmentation approach.

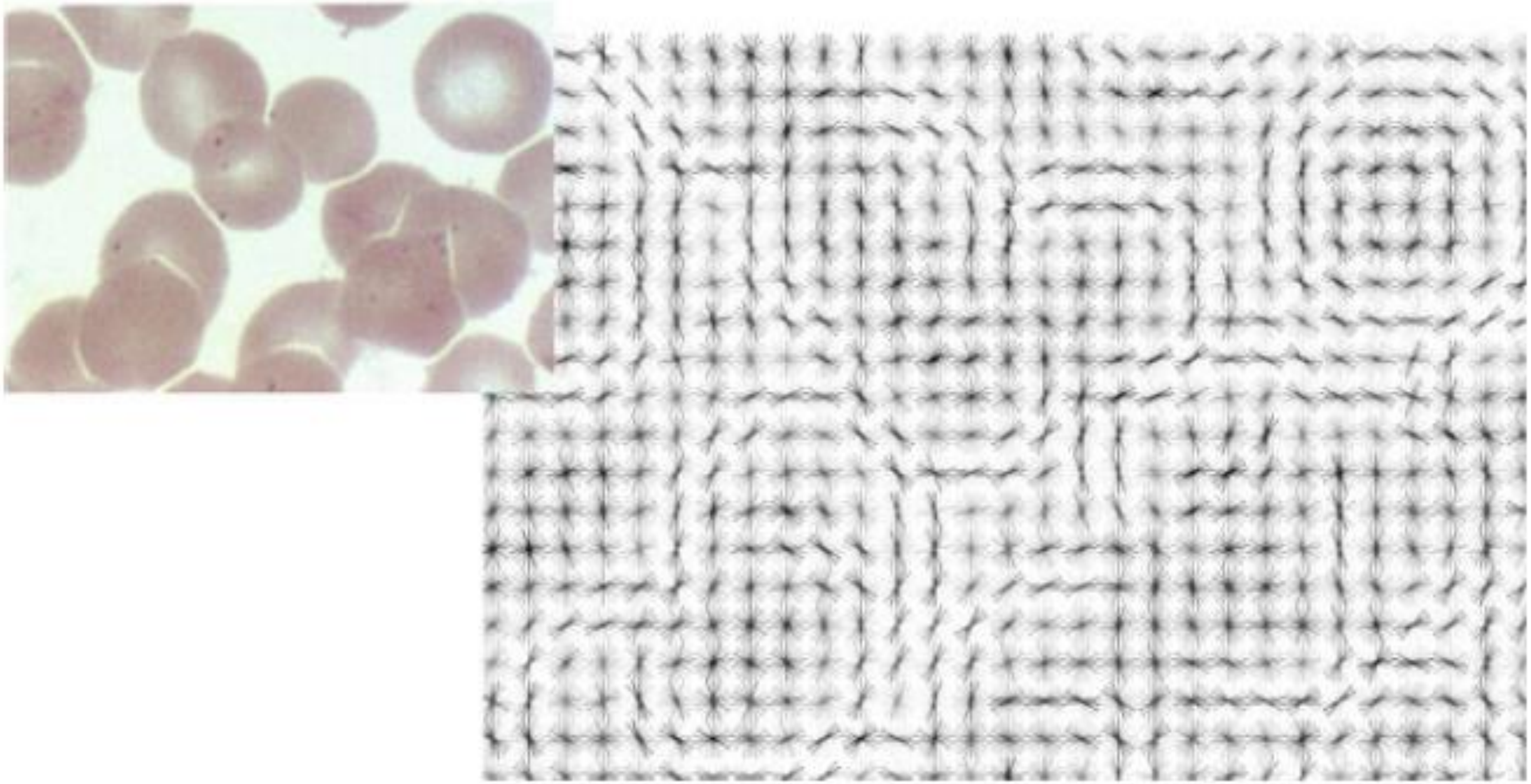

Fig. 5.21 HOG features computed from a microscopic image. The directional histogram in each HOG cell is given by the intensity of a line in the respective direction. Local orientation preferences are clearly recognizable

Saliency and Gist

- Saliency is an attribute that guides attention of human vision to certain locations in the image.
- Focusing on certain parts in an image is an integral part of the perception process of a human operator who searches, recognizes, and categorizes structures in an image as it allows the operator to single out subsets of the image and focus on attributes of this subset.
- A salient location in an image is a region where features differ significantly from features in the vicinity. Since saliency is a concept of biologically inspired computer vision, features are those perceived at early stages of visual perception.
- Examples are image intensity, image color, and local orientation.

GIST

- The gist of an image is an overall categorization of an image (e.g., by characterizing an image as a CT image of a head depicting a subdural hemorrhage).
- In human vision, deriving the gist of an image allows to pick a constraining category for further analysis. Being able to compute the gist of an image can substantially speed up image analysis as it reduces the number of possible explanations for depicted objects.
- Since gist summarizes information from all parts of the image to come up with an overall categorization, methods to compute gist are integrations from feature values everywhere in the image.
- gist features may help in the automatic categorization of images in a data basis of medical images.