

Al-Mustaqbal University

College of Science

Fourth stage

Medical Physical Department



جامعة المستنقب
AL MUSTAQBAL UNIVERSITY

Medical Image Analysis

**SIFT Feature and SURF, Binary Key Point
Descriptor and Detectors, Histogram of
Oriented Gradients- Part II**

By

Asst. Prof. Dr. Mehdi Ebady Manaa

1. Feature Detection

Data-based features in images such as **key point locations or potential parts of object boundaries can be extracted from local image characteristics**. Boundary parts are generated from the results of an edge enhancement **step while key point locations are local extrema of some local object property**. Features may also be computed from samples of an object's boundary or interior. Potential object boundary parts are used for detecting or delineating objects in images. Key points may, in some simple cases, also be used to detect objects. In most cases, however, object characteristics are too complex to be captured by the attributes of a key point. They can be important attributes nonetheless. Key points define an object-dependent reference system in which they may be used to map objects of the same class onto each other.

Concepts, notions and definitions introduced in this chapter

- SIFT and SURF features,
- MSER features,
- local shape context,
- HOG features,
- gist and saliency

1. SIFT and SURF Features

The Scale-Invariant Feature Transform (SIFT) was 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 (testing phase). 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,
- **key point matching**. In key point generation, rotation- and scale-invariant features are generated **by searching for the local extrema of a multiscale blob detector based on a multi-scale Difference of Gaussian** (DoG, see previous section and Fig. 8.1). The detector is insensitive with respect to noise by determining an optimal smoothing for each scale. This is done by computing a sequence of Gaussian smoothings $S(\sigma_i)$ with standard deviations $\sigma_i = \sigma_1, \sigma_2, \dots, \sigma_n$ and computing a sequence of **DoG** $D(\sigma_i) = S(\sigma_i) - S(\sigma_{i-1})$, $i = 2, n$ (see Fig. 8.2). A local extremum is maximal or minimal not only in scale, but also

along the range of different smoothing scales σ_i . Key point generation will create numerous responses from noise and artefacts. In key point reduction, **the contrast at local extrema is used to remove low contrast blob locations** (Lowe 1999).

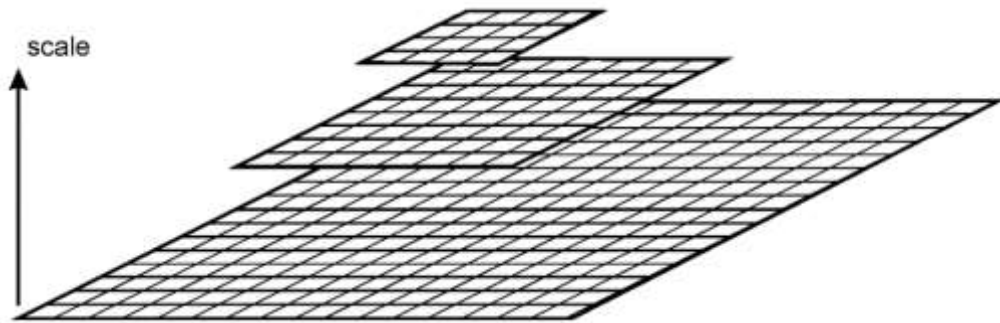


Fig. 8.1 Key point responses are computed at different scales of an image pyramid. Potential key points are local extrema in scale space

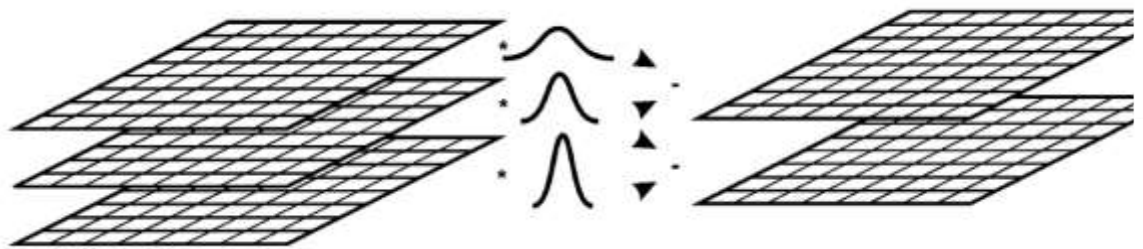


Fig. 8.10 Additionally to being an extremum in scale space it also has to be an extremum for differently smoothing DoG filters

Responses are also removed if contrast is high, but the localization is unstable. This is the case along edges where the average contrast may be high because the feature is prominent across the edge but the localization accuracy is low because the feature strength varies little along the edge. In Lowe (2004), an improved version of this step was presented where the feature response function was interpolated and the attributes of the interpolated function were used to remove unstable feature 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 (see Fig. 8.3). **Gradient directions** are computed at the scale that was determined for the key point in the first step. **The maximum of this histogram indicates the local key point orientation.** A threshold of 80% of this maximum is set and it is tested whether any other local maximum in the histogram exceeds this threshold. If it is the case, multiple key points with the same location but different orientations are created. **The key point orientation is then used to compute rotation-invariant features.** A local window at the feature scale is defined around the point location. Gradients within this window are weighted with a Gaussian. Then, local gradient histograms in subwindows around the key point location are used as key point features (see Fig. 8.3). The features are organized in a feature vector that is then normalized to make the feature independent from intensity variation. Hence, the feature vector describes the relative gradient length distribution for different gradient directions in the vicinity of the key point.

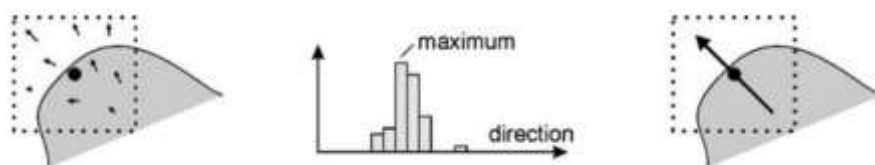


Fig. 8.3 Key point orientation is computed from a histogram of binned gradient directions in the vicinity of the key point

Key point features can then be used for matching model key points with the key points extracted from the image. 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 algorithms. 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 corresponds in the two images. Registration is particularly easy if this correspondence is given by sets of pairs of corresponding feature locations. Since most medical image registration problems are 3D, the SIFT features have been extended to 3D as well. 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. The methodology is patented in the United States and is claimed by the author of not only being faster but also more robust than the SIFT features.

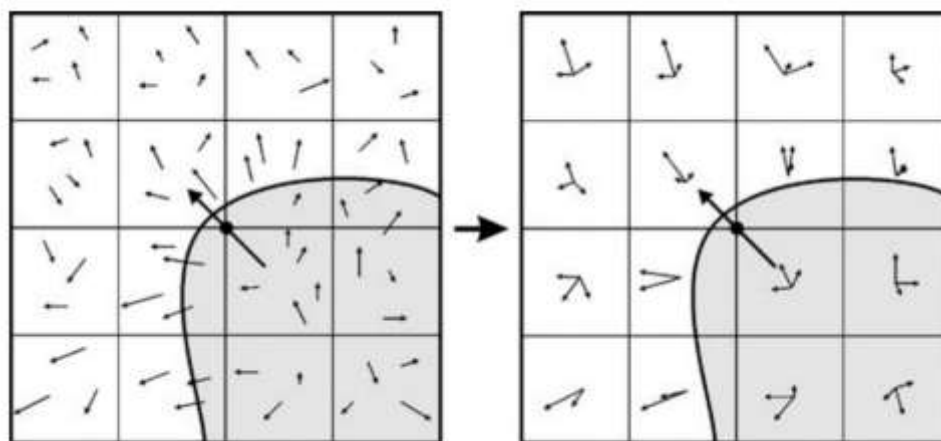


Fig. 8.4 The feature vector for a key point consists of binned histograms of normalized, relative gradient directions with respect to the key point orientation. Hence, key points are represented by a kind of rotation-invariant texture features

2. MSER Features Locations,

such as the center of gravity, generated from Maximally Stable Extremal Regions (MSER) were presented by Matas et al. (2002). The key concept of this approach is to separate an image into local homogeneous regions with maximum Local shape context (Belongie et al. 2002) describes an object by boundary features that need not be—and in most case will not be—key points. The purpose of using *local shape context* is to be able to match two structures based on the shape context information. Local shape context is defined on boundary points that stem from sampling boundaries of an object-of-interest that are generated by an edge detection procedure such as the Canny edge detector. The boundary needs not to be closed. However, edge detection of two similar objects should result in similar sets of boundary parts. Sampling may be arbitrary, but in the absence of further knowledge it should generate point locations distributed evenly over the boundary parts. generate point locations distributed evenly over the boundary parts.

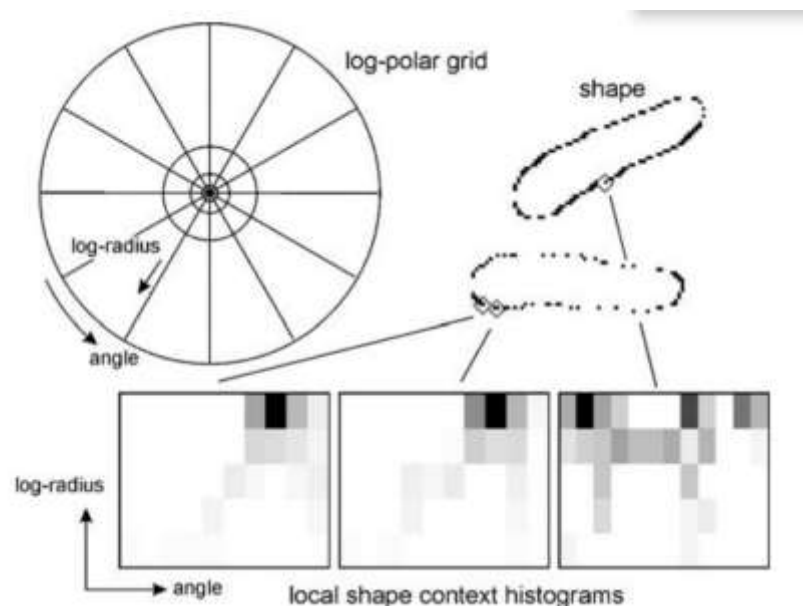


Fig. 8.5 Local shape context is represented by 2D histograms for points that contain frequency of occurrences of other points. The histogram is binned in a log-polar grid. It can be seen that shape contexts

of similar point locations result in similar histograms

3. Saliency and Gist البروز والجوهر: Saliency and gist are image features, which are different to the features discussed above that pertain to structures in the image. **Image features are not immediately useful for the analysis of objects.** However, as they characterize the image, they may be used to **guide image analysis in application such as a feature-based search in an image data base.** *Saliency is an attribute that guides the 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 (مجاورة).* 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.* A biologically plausible method for computing saliency has been presented by Itti et al. (1998). *Several other methods exist as well. Computer vision techniques use saliency in the same manner for quickly directing attention to subparts of an image which may contain objects that are searched.* Such rapid scene analysis is seldom the goal when working with medical images نادرا ما يكون هذا عند تحليل الصور الطبية والذي هو التحليل السريع للمشاهد, but the distribution, extent, and properties of salient locations characterize images in a way that can be useful for

search and comparison tasks between images. This is even more so considering the fact that occlusion problems—which change such configuration—do not occur if the images are 3D.

4. Bag of Features

When features such as the ones discussed above will be used to detect objects they need to be combined to some kind of a superstructure since the semantic of a single feature is usually insufficient for representing the object characteristics. A simple Feature Detection way to arrive at a higher level semantic is to just collect features (points, edges, textures) from a region that shall be recognized. **Feature collection** is done by picking regularly or randomly distributed patches from the region and computing feature attributes in these region. Each patch is assumed to be a “visual word” making up the visual sentence describing the image. It is a bag of features that represents the (main) meaning of the region from which the features are taken. Since the content of the bag is not spatially ordered, the representation is invariant with respect to translation and, if the feature values are invariant to rotation or scale, invariant with respect to these two transformations as well. *The feature values of the different elements in the bag will all be different and their frequency of occurrence will characterize the region recognition is possible,* if the expected feature occurrences for different objects has been trained from the examples.

5. HOG Feature Descriptor

HOG, or Histogram of Oriented Gradients, is a feature descriptor that is often used to extract features from image data. It is widely used in computer vision tasks for object detection. An Important aspects of HOG that makes it different from other feature descriptors:

- The HOG descriptor focuses on the structure or the shape of an object. This is done by extracting **the gradient and orientation** (or you can say magnitude and direction) of the edges
- Additionally, these orientations are calculated in **'localized' portions**. This means that the complete image is broken down into smaller regions and for each region, the gradients and orientation are calculated.
- Finally, the HOG would generate a Histogram for each of these regions separately. The histograms are created using the gradients and orientations of the pixel values, hence the name 'Histogram of Oriented Gradients'

To put a formal definition to this:

The HOG feature descriptor counts the occurrences of gradient orientation in localized portions of an image.

Multiple-Choice Questions (MCQs):

1. What is the primary objective of the SIFT (Scale-Invariant Feature Transform) method in medical image analysis?

- A) To enhance image intensity.
- B) To identify objects using key features.
- C) To increase image blurring.
- D) To eliminate local extrema.
- E) To compute gradient directions.

Correct Answer: B) To identify objects using key features.

2. Which step in the SIFT method involves determining and storing orientation attributes for each key point?

- A) Key point generation.
- B) Key point reduction.
- C) Feature computation.
- D) Key point matching.
- E) Local shape context.

Correct Answer: C) Feature computation.

3. What is a key characteristic of SURF (Speeded-up Robust Features) compared to SIFT?

- A) SURF uses a slower convolution process.
- B) SURF is less robust.
- C) SURF replaces slow convolutions with faster approximations.
- D) SURF does not use key points.
- E) SURF does not involve key point matching.

Correct Answer: C) SURF replaces slow convolutions with faster approximations.

4. What is the main application of SIFT features in medical image analysis?

- A) To eliminate features for efficient processing.
- B) To enhance image contrast.
- C) To guide image analysis in a feature-based search.
- D) To compute saliency.
- E) To support feature-based registration algorithms.

Correct Answer: E) To support feature-based registration algorithms.

5. What does the MSER (Maximally Stable Extremal Regions) feature focus on in terms of image analysis?

- A) Identifying key points.
- B) Extracting gradient directions.
- C) Separating an image into local homogeneous regions.
- D) Determining object orientation.
- E) Matching two structures based on shape context.

Correct Answer: C) Separating an image into local homogeneous regions.

6. How is local shape context defined in medical image analysis?

- A) It focuses on image intensity.
- B) It uses key points for matching.
- C) It samples boundaries generated by a blob detector.
- D) It describes an object using boundary features.

- E) It generates histograms for gradient directions.

Correct Answer: D) It describes an object using boundary features.

7. What is the purpose of saliency in image analysis, particularly in biologically inspired computer vision?

- A) To compute key points efficiently.
- B) To guide the attention of human vision to certain locations.
- C) To separate an image into local regions.
- D) To match structures based on orientation.
- E) To eliminate image features for efficient processing.

Correct Answer: B) To guide the attention of human vision to certain locations.

8. What is the "Bag of Features" concept used for in medical image analysis?

- A) To eliminate features for efficient processing.
- B) To describe an object using boundary features.
- C) To focus on image intensity.
- D) To separate an image into local regions.
- E) To collect and combine features for region recognition.

Correct Answer: E) To collect and combine features for region recognition.

9. What is the focus of the HOG (Histogram of Oriented Gradients) feature descriptor?

- A) Image intensity.
- B) Object rotation.
- C) Structure or shape of an object.
- D) Key point reduction.
- E) Feature matching.

Correct Answer: C) Structure or shape of an object.

10. How does the HOG descriptor achieve localization in image analysis?

- A) By computing saliency.

- B) By extracting gradient and orientation in localized portions.
- C) By generating histograms for gradient directions.
- D) By replacing slow convolutions with faster approximations.
- E) By computing orientation attributes for each key point.

Correct Answer: B) By extracting gradient and orientation in localized portions.