

# A Survey on Feature Extraction Techniques for Shape based Object Recognition

Mitisha Narottambhai Patel  
Department of Computer Engineering,  
Uka Tarsadia University,  
Gujarat, India

Purvi Tandel  
Department of Computer Engineering,  
Uka Tarsadia University,  
Gujarat, India

## ABSTRACT

Robotics is one of the research area in computer age. So, to make the robots as capable as humans, to allow them to interact with real environment so many algorithms are developed and will be developed. Some of those algorithms are developed in the area of computer vision to allow the robots for accurate recognition. In all those algorithms feature extraction technique is most important part of the algorithm. As the features are robust to different affine transformations like translation, scale, rotation, flipped, etc. the algorithm will be more robust to those transformations. So, feature extraction techniques are one of the important part of the image retrieval systems. Some of those feature extraction techniques, with their invariance properties are discussed here for the image retrieval system.

## General Terms

Pattern Recognition, Digital Image Processing.

## Keywords

feature extraction, object recognition, shape, shape descriptor, survey

## 1. INTRODUCTION

Object recognition in computer vision is the task of determining similar objects in an image or in video sequence. The shape based object recognition is made difficult by many adversaries like variations in pose, occlusion, scaling, poor illumination, rotations, etc<sup>[1]</sup>. A variety of techniques developed for shape based object recognition to make the recognition easier and accurate. Matching is one of the central issues of object recognition and important part of the object recognition system. A common goal is to measure the image data and compare it and verify it for accurate recognition process. Matching process in object recognition involves some type of search and when search takes place the set of extracted features will be compared with the stored features for recognition process. To accomplish this task relevant and comparable information should be extracted from the input data.

There are so many approaches for object recognition process. Some of those approaches are discussed here. In any approach for object recognition rotation and translation invariant properties are crucial in most recognition tasks and should be considered in the features chosen for image retrieval<sup>[2]</sup>.

There are three main stages for invariant shape based object recognition:

- (1) Data Pre-processing
- (2) Feature extraction
- (3) Classification (Match and Search)

In data pre-processing stage, image data are pre-processed to make it noise free or more clearer for feature extraction process. Image processing is the technique that would enhance image quality for preparing images for measurement of the present features<sup>[19]</sup>. In feature extraction stage from pre-processed image features are extracted to make the recognition task easier as well as accurate. Extracted features are then stored in the database for the database images, which will be used to match with the features of the input image to search the similar image. Among all these stages feature extraction is one of the more useful stage for making the object recognition task easier as well as accurate.

## 2. FEATURE EXTRACTION TECHNIQUES

The most of the image retrieval systems are based on the shape, color, texture and object layout<sup>[2]</sup>. The shape of an object refers to its physical structure. Shape can be represented by the boundary, region, moment, etc. These representations can be used for matching shapes, recognizing objects, or for making measurements of shapes. Texture is observed in the structural patterns of surfaces of objects such as wood, grain, sand, grass, and cloth. The term texture generally refers to repetition of basic texture elements called texels. A texel contains several pixels, whose placement could be periodic or random. Natural textures are generally random, whereas artificial textures are often deterministic or periodic. Texture may be coarse, fine, smooth, regular, irregular, or linear. In image analysis, texture is broadly classified into two main categories, statistical and structural. One is Statistical approach in which the textures are random in nature. Second approach is Structural approach in which purely structural textures are deterministic texels, which repeat according to some placement rules, deterministic or random. Third approach: A method that combines the statistical and the structural approaches is based on what have been called mosaic models. These models represent random geometrical processes. VisualSeek, also uses the object layout as an image feature<sup>[2]</sup>. Although shape, colour and texture are important features for image retrieval process they lose their ability when the input image or database image does not have such attributes, for example, whenever the query image is a rough sketch with only black and white lines<sup>[2]</sup>.

In the most object recognition task rotation and translation invariant properties should be considered in the features chosen for the image retrieval. The two main approaches used for the categorization of the invariant methods are as follows:

1. Image alignment
2. Invariant features

**1. Image alignment:** In the object recognition process transformation is applied to the image so that the object in the image is placed in a predefined standard position. This approach is mainly based on the extraction of the geometric properties like boundary curvature. If there are more than one objects exists in the object image then segmentation is required<sup>[2]</sup>.

**2. Invariant features:** An invariant feature means invariant image characteristics which remains unchanged even if the object is translated or rotated. Although this approach is more used it is also based on the geometric properties<sup>[2]</sup>.

The meaning of the word “feature” is in general highly application dependent. A feature is result of some calculations performed on the input data stream. Extracted feature is then matched with the stored feature data to complete the object recognition task accurately. There are so many techniques developed for feature extraction, to make the shape based object recognition easier as well as accurate. Some of those feature extraction techniques are as follows:

### 2.1 Color Histograms

Swain and Ballard introduces the histogram based recognition methods<sup>[3][4]</sup>. For the given color image the color histogram is obtained by discretizing the image colors and counting the number of times each discrete color occurs in the image array. The histograms are invariant to the rotation and translation and change only slowly under change of angle of view, change in scale, and occlusion. Because histograms change slowly with view, a three dimensional object can be represented by a small number of histograms<sup>[3][5]</sup>. In the color histogram method color histogram of the sample object is used for the retrieval of the similar object image<sup>[3]</sup>. The drawback of this method is it is more sensitive to the color and intensity of the light source and color of the input object image.

### 2.2 Edge Pixel Neighbourhood Information (EPNI):

In EPNI method the neighborhood edge pixels are found out that structure of those pixels will be used to make an extended feature vector<sup>[2]</sup>. This feature vector is used for matching process for the image retrieval. This method is scale and translation invariant, but not rotation invariant.

### 2.3 Histograms of Edge Directions (HED)

In computer vision and image retrieval process edge image matching is widely used for the comparison process. In images with the similar color information and in the absence of the color information this histogram of the edge directions is the significant tool for image retrieval. The edge is extracted using the Canny edge operator for this feature extraction and corresponding edge directions are, subsequently, quantized into 72 bins of  $5^0$  each<sup>[2]</sup>. HED is also useful for shape representation.

### 2.4 Edge Histogram Descriptor (EHD)

Edge Histogram Descriptor is the histogram generated using the edge pixels. The edge distribution is a good texture signature and also useful for image to image matching. This approach is not rotation invariant. The MPEG-7 standard defines the edge histogram descriptor (EHD) in its texture part<sup>[2]</sup>. The distribution of edges is useful for image to image matching. But this descriptor is not effective for the rotation invariance.

### 2.5 Angular radial partitioning (ARP)

In ARP method, the images in the stored database are converted to grayscale and edge detection is performed<sup>[2]</sup>. To achieve the scale invariance property, the edge image will be partitioned by the surrounding circles and the intersection points of the edge and surrounding circle are found and angles will be measured for feature extraction process which will be used for the comparison process in image retrieval process. The algorithm uses the surrounding circle of edge of an object and also make the n number of radial partition for that edge of the object image i.e. after making the surrounding circle equidistant circles will be made to extract the features for achieving scale invariance<sup>[2]</sup>.

### 2.6 Eigenvector Approaches

In eigen vector approach each image is represented by the small number of coefficients, which will be stored in the database and searched efficiently for the image retrieval and it is very successful for image retrieval. Though it is successful approach there are some drawbacks of this approach. One drawback is any change in individual pixel value, caused by the transformations like scaling, rotation, translation will change the eigen vector representation of an image. To deal with this difficulty, the eigenspace is calculated under consideration of all possible changes<sup>[6]</sup>.

### 2.7 SIFT-Scale Invariant Feature Transform

Lowe combined the ideas of feature based and histogram based image descriptors, and defined a scale invariant feature transform, SIFT<sup>[4][7]</sup>. Scale Invariant Feature Transform (SIFT), as it transforms image data into scale-invariant coordinates relative to local features. An important aspect of this approach is that it generates large numbers of features that cover the image over the full range of scales and locations. A typical image of size 500×500 pixels will give rise to about 2000 stable features (although this number depends on image content). The quantity of features is particularly important for object recognition, where the ability to detect small objects requires that at least 3 features be correctly matched from each object for reliable identification. For image matching and recognition, SIFT features are first extracted from a set of reference images and stored in a database. A new image is matched by individually comparing each feature from the new image to this previous database and finding matching features based on Euclidean distance of their feature vectors<sup>[7]</sup>.

### 2.8 Harris Corner

Harris also showed its value for efficient motion tracking, and the Harris corner detector has since been widely used for many other image matching tasks<sup>[7]</sup>. While these feature detectors are usually called corner detectors, they are not selecting just corners, but rather any image location that has large gradients in all directions at a predetermined scale. The Harris corner detector is very sensitive to changes in image scale, so it does not provide a good basis for matching images of different sizes.

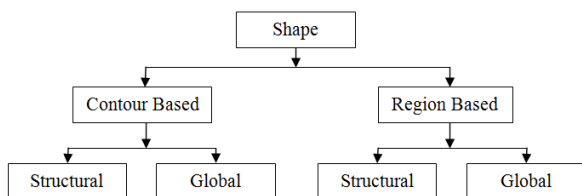
### 2.9 Shape Descriptor

Shape is an important basic feature which is used to describe the image content. But because of the noise, occlusion, arbitrary distortion shape is often corrupted and object recognition problem became more complex. Shape representation is mainly based on the shape features which are either based on the shape boundary information or boundary plus interior content. Various types of shape features are designed for object recognition which are evaluated on the

basis of how accurately those shape features allow one to retrieve the similar shapes from the database<sup>[8]</sup>.

For the good retrieval accuracy shape descriptor should be able to effectively find the similar shapes from the database whether they are affinely transformed shapes like rotated, translated, flipped, scaled, etc. The shape descriptor should also be able to effectively find the defective shapes, noise affected shapes, which are tolerated by human beings for the comparison and retrieval of the shapes. This is known as the robustness requirement. A shape descriptor should be able to perform image retrieval for maximum types of shapes not only for certain types of shapes, so it should be application independent. One of the important characteristic of the shape descriptor is low computation complexity. By involving the fewer properties of image in the computation process, computation can be minimized and lower computation complexity achieved and shape descriptor became robust. Here low computation complexity means clarity and stability. There are so many shape representation and description techniques have been developed for the shape retrieval applications.

Classification of the shape representation and description techniques:



**Figure 1: Classification of shape representation and description techniques<sup>[8]</sup>**

Based on whether the shape features are extracted from the contour only or are extracted from the complete shape region, the shape representation and description techniques are classified into two categories<sup>[8]</sup>.

1. Contour based methods.
2. Region based methods.

Each method is further divided into two approaches, structural approach and global approach. Structural approaches and global approaches are based on whether the shape is represented as a whole or by segments. Some of those techniques are discussed here<sup>[8]</sup>.

### 2.9.1 Contour based shape representation and description techniques:

By using the Contour based techniques boundary information will be extracted. The contour based shape representation technique will be further divided into global approach and structural approach. Global approach will not divide the shape into subparts and the complete boundary information will be used to derive the feature vector and for matching process, so it is also known as the continuous approach. Structural methods breaks the shape boundary information into segments(subparts), called primitives, so this method is also known as the discrete approach. Generally the final representation of the structural method is a string or a graph(or tree), which will be used for matching for image retrieval process<sup>[8]</sup>.

*Global methods:*

From the shape boundary information multi-dimensional numeric feature vector is generated which will be used for the matching process. Matching process is done using calculating the Euclidean distance or by point to point matching<sup>[8]</sup>.

*Simple Shape Descriptor:*

Area, circularity, eccentricity, major axis orientation are the simple shape descriptors<sup>[8]</sup>. From the boundary information area, circularity, maximum length, minimum length, etc. information are calculated for the feature extraction. Simple shape descriptors can only discriminate shapes with large differences, so sometimes they combined with other shape descriptors for accurate retrieval process. So, accuracy point of view they are not so suitable descriptor for image retrieval<sup>[8]</sup>.

*Correspondence based shape matching:*

In correspondence based shape matching every point of the shape boundary information is treated as the feature point, which will be used for point to point matching for image retrieval process. Because it is point to point matching process it is not translation, scaled, rotation invariant<sup>[8]</sup>.

*Shape signature:*

In the Shape signature, shape is represented by the one dimensional function obtained from the extracted boundary information. Centroid distance, area, chord length (maximum length, minimum length) are shape signatures<sup>[8][9]</sup>. Shape signatures are translation and rotation invariant, because though the object of the given image is rotated or translated centroid distance, area, chord length will not be changed for the same object image. Shape signatures are sensitive to slight changes in the boundary, because slight change in the boundary information causes large errors in the matching. Centroid distance is the distance between the edge points and its center is found, which is translation and rotation invariant<sup>[9]</sup>.

*Farthest distance:*

Farthest distance is the distance between each edge pixel. So, farthest pixel distance for each pixel is found out which will be useful for the object recognition. It is translation and rotation invariant<sup>[9]</sup>.

*Structural methods:*

In structural methods of contour based shape representation and description technique the boundary information is broken into segments i.e. shapes are broken into boundary segments called primitives. The result is encoded into a string form like:

$$S = s_1, s_2, \dots, s_n;$$

Where,  $s_i$  may be an element of a chain code.  $s_i$  may contain an attributes like length, orientation, etc. The string can be directly used to represent the shape or can be used as an input to recognition system for image retrieval process<sup>[8]</sup>.

*Chain code representation:*

The chain code method was introduced in 1961 by Freeman<sup>[8]</sup>. This method represents an object by a sequence of unit-size line segments with a given orientation i.e. an object is represented by a sequence of small vectors of unit length and a limited set possible directions, so, this method is also known as the unit vector method. The choice of the first boundary pixel in the sequence should be independent. One method to choose the first boundary pixel is find the pixel in the border sequence which results in the minimum integer number and

that pixel is then used as the starting pixel i.e. first boundary pixel. But this chain code is not rotation, scale invariant. The chain code is sensitive to the noise<sup>[8]</sup>.

*Polygon decomposition:*

Polygon decomposition is introduced in, in which shape boundary is broken into line segments by polygon approximation<sup>[8]</sup>. The polygon vertices are used as primitives and the feature for each primitive is expressed as a four element string. Four elements of the strings are internal angle, distance from the next vertex, and its x and y co-ordinates. These features are not rotation, translation and scale invariant. Its application for natural objects is impractical.

*Syntactic analysis:*

Syntactic analysis method is inspired by the method, that is, sentences are built up from phrases, phrases are built up from words and words are built up from alphabets, etc.<sup>[8]</sup>. In syntactic methods, object shape is represented by the set of predefined primitives. The set of predefined primitives are referred as the codebook and the primitives are referred as the code words. So, this method is based on the theory of formal language<sup>[8]</sup>. It attempts to simulate nature of the human vision system. But this method is impractical for applications, because it is not possible to generate a pattern of grammar which can generate the valid patterns only.

*Shape invariants:*

Shape invariant is a technique which attempts to represent the properties of the boundary configuration which remain unchanged even if the transformations are performed on the object<sup>[8]</sup>.

Invariant is usually named according to the number of features used to define it. An invariant defined on the single feature is called unary invariant or order one invariant. An invariant defined between two features is known as the binary invariant or order two invariant, similarly ternary invariant, quaternary invariant and so on. Common invariants includes geometric invariants like area, angle, length, triangle; algebraic invariants like eigen-values; differential invariants like curvature<sup>[8]</sup>. Geometric invariants and algebraic invariants are suitable for man-made object recognition. Shape representation using invariants has several problems like invariants are generally obtained from the pure geometric transformation of the object shape, but in reality shape rarely changes according to strict geometric transformations. Another problem of the shape invariant is, they are very sensitive to boundary noise and errors.

*Drawbacks of structural approach:*

Generation of primitives and features is the main drawback of the structural approach. Because the number of primitives required for each shape is not known and so there is no formal definition for an object or shape. The other drawback is its computation complexity and this method does not guaranteed for the best match. A variation of object boundaries causes changes to the primitives, so, it is not less reliable than global methods<sup>[8]</sup>.

**2.9.2 Region based shape representation and description techniques:**

In the region based technique all the pixels within the shape region, i.e. whole region is taken into account for the shape representation and description. Similar to the contour based methods, region based methods can also be divided into the

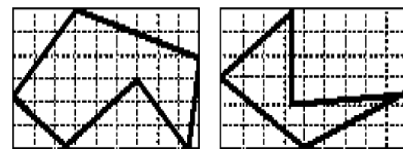
global methods and structural methods, depending on whether they divide the shapes into sub parts or not<sup>[8]</sup>.

*Global methods:*

Global methods take the whole shape region into consideration for shape representation and description<sup>[8]</sup>.

*Grid based method:*

The grid shape descriptor is introduced by Lu and Sajjanhar<sup>[8]</sup>. The shape image is divided into pixels(cells) and then scanned from left to right and top to bottom. The cells covered by the shape are assigned 1 and the cells which are not covered by the shape are assigned 0. So, the shape can be represented by the binary feature vector. The binary Hamming distance is used to measure the similarity between two shapes. For example, the grid descriptors for the two shapes in figure 2 are 001111000 011111111 111111111 111111111 111110011 001100011 and 001100000 011100000 111100000 111100000 011111100 000111000, respectively, and the distance between the two shapes will be 27 by the XOR operation on the two sets.



**Figure 2: Grid representation of two contour shape<sup>[8]</sup>**

To make this method translation, rotation and scale invariant, the shape is scaled into a fixed size rectangle, shifted to the upper left of the rectangle and rotated so that the major axis of the shape is horizontal.



**Figure 3: Two similar contour shapes with different representation<sup>[8]</sup>**



**Figure 4: Two similar region shapes with different grid representation<sup>[8]</sup>**

The advantage of the grid descriptor is it is very simple representation. The main problem of this method is major axis based rotation normalization. The major axis is sensitive to the noise. For example, the two similar shapes in figure 3 have very different grid representations. For the region based shapes, the grid representation is not rotation invariant. For example, the two shapes in figure 4 are similar shapes, but their grid representations are very different.

*Structural methods:*

Region based structural methods partitioned the shape region into sub parts which are used for shape representation and description. Convex hull is the type of structural method. The region structural methods have the similar problems like contour structural methods<sup>[8]</sup>.

### 3. SUMMARY AND CONCLUSION

Color histogram is sensitive to noise. Color histogram method is invariant to translation and rotation. EPNI method is scale and translation invariant, but not rotation invariant. Histograms of edge directions and Edge histogram descriptor are histogram based feature extraction techniques used for image retrieval. ARP is scale and rotation invariant. Eigen vector approach is very sensitive to change in individual pixel value, so, it is not translation, rotation, scale invariant.

Generally, there are two types of shape representation and description techniques: contour based methods and region based methods. Both the methods are further divided into two types: global methods and structural methods, by how they use the contour or region information, whether they use the complete information or partition the contour/region information for representation and description.

Global contour based shape descriptor techniques take the whole shape contour as the shape representation. Global region based method consider the whole region for shape representation, so, it effectively use all the pixel information within the region. Grid methods are not rotation invariant for region based shapes because the major axis is sensitive to the noise. Contour based methods are more popular than the region based methods, because human beings are thought to discriminate the shapes mainly by their contour features. In some shape applications the interior content is not important, only the shape contour is important, so, contour based techniques are widely used for that kind of applications.

There are some limitations of the contour based methods:(i) They are sensitive to noise and variations, because they only use a small part of information, contour information.(ii) In some applications region content is more important than the contour features.

These limitations can be overcome by the region based methods. Region based methods are more robust than the contour based methods, because they use all the available shape information. Region based methods provide more accurate retrieval. The common problem of the contour based method is that they can't cope well with shape deflection, which is overcome by region based methods, because region based methods use all the available information of the shape. Structural approaches are more complex than global approaches to implement.

### 4. REFERENCES

- [1] Raja Tanveer Iqbal, Costin Barbu, Fred Petry, "Fuzzy Component Based Object Detection", Science Direct, International Journal of Approximate Reasoning, 45, 546-563, 2007
- [2] Chalechale A., Mertins A., Naghdy G., "Edge image description using angular radial partitioning", IEE Proc.-Vis. Image Signal Process., Vol. 151, No. 2, April 2004
- [3] Swain M. J., Ballard D. H., "Color Indexing", International Journal of Computer Vision, 7(1):11-32, 1991
- [4] Linde Oskar, Lindeberg Tony, "Composed Complex-Cue Histograms: An Investigation of the Information Content in Receptive Field Based Image Descriptors for Object Recognition", In Computer Vision and Image Understanding 116 (2012) 538{5602.DOI: 10.1016/j.cviu.2011.12.03, March-16, 2012
- [5] Wu Jun, Xiao Zhitao, "Video Surveillance Object Recognition Based on Shape and Color Features", 3<sup>rd</sup> International Congress on Image and Signal Processing (CISP2010), 2010
- [6] Schiele Bernt, Crowley James L., "Recognition without Correspondence using Multidimensional Receptive Field Histograms", International Journal of Computer Vision 36(1), 31–50 (2000)
- [7] Lowe David G., "Distinctive Image Features from Scale-Invariant Keypoints", International Journal of Computer Vision 60(2), 91-110, 2004
- [8] Zhang Dengsheng, Lu Guojun, "Review of Shape Representation and Description Techniques", Elsevier Computer Science, Pattern Recognition, The Journal Of The Pattern Recognition Society, Pattern Recognition 37 (2004) 1-19
- [9] Zare Chahooki Mohammad Ali, Charkari Nasrollah Moghadam, "Learning the Shape Manifold to Improve Object Recognition", Springerlink, Machine Vision and Applications 24:33-46 DOI:10.1007/s00138-011-0400-6, 2013