A Study of Feature Based Techniques in Visual Object Recognition

K. Pani Sekhar[1], Dr. Subhashish Bose[2]

[1]Research Scholar, Department of Electronics and Communication Engineering, Sri Satya Sai University of Technology & Medical Sciences, Sehore, M.P., India.
[2]Research Supervisor, Department of Electronics and Communication Engineering, Sri Satya Sai University of Technology &Medical Sciences, Sehore, M.P., 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. 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. 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.

Keywords: *Feature Extraction, Object Recognition, Shape, Shape Descriptor, Survey*.

---

1. INTRODUCTION

The problem of object recognition by computers has been active for more than two decades, and as yet has not been solved. To solve this would create many new opportunities for the automation industry, and increase efficiency in many other industries. Many current vision systems can distinguish between objects if the image is clear and without occlusions. The system described in this paper proposes a solution that is designed to identify objects that are partially occluded or viewed from different angles. This paper forms part of a generalised object recognition system that is described further elsewhere [1,2]. The function of the method described here is to extract, from the image contours, features and relationships that can be represented in a database. A number of database objects can be selected that have a high probability of being contained within a captured image. The method must select all of the object images contained within the captured image (no false negatives) while including only a modicum of the others (false positives).

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 [2]. 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 clearer for feature extraction process. Image processing is the technique that would enhance image quality for preparing images for measurement of the present features. In feature extraction stage from preprocessed 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.

Recognition of general objects within a scene is a complex problem with vast potential. People have been trying to develop methods to deal with this for many years, with varying degrees of success. If an artificial vision system is created to a high enough standard, automation of many repetitive tasks will be possible. The most common problems encountered are occlusions over part of the image, scaling and rotation of the image, and distortion of the image. Systems have been designed for recognition of objects from very specific sets, such as skeletal object recognition, which works very well for objects with complex skeletal structures [3]. Template matching [4] can identify many different object types, but only with small database sizes due to the complex calculations required. Many object recognition systems have been attempted that use contours to characterise each object, with many different ways to store and search through the data. One example [5] is similar to the system described this paper, by splitting a contour and analysing the components; however the system compares all the contour parts, which is slow, and still susceptible to occlusions. The field lies open for a system efficient enough to be scaled up to larger databases. One such candidate [1,2] is based on the use of image contours and the extraction of the contour fingerprint.

Fingerprints and Contour Generation

The image is translated into a set of contours; these are sets of adjacent points that have the same grey level. They will occur around edges or outlines in the image as can be seen in figures 1 and 2.

Figure 1. Original image of the mug

Figure 2. Image with contours superimposed

The fingerprint for each contour is given by the first derivative of the contour local tangent angle, which is the local rate of curvature. This means sharp corners on the contour show up as spikes on the fingerprint (see figure 3), straight lines are successive points close to the x-axis, and arcs are successive points of equal value away from the x-axis. The use of fingerprints makes the recognition of significant features far easier, and the endpoints are found very accurately.

Figure 3. Fingerprint of a contour on the outer edge of the mug.

There will be many contours for each image. Depending on the complexity of the image this can range from 200 to over 1000. Because a human can still easily recognise the object from the contours, it can be assumed that these contours contain all the data necessary to identify the object within the image.

2. FEATURE EXTRACTION TECHNIQUES

The most of the image retrieval systems are based on the shape, color, texture and object layout[2]. 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 Texel's. 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. Visual Seek also uses the object layout as an image feature [2]. 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[2] .

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[2] .

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[2] .

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:

1. Color Histograms

Swain and Ballard introduces the histogram based recognition methods[3][4]. 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[3][5]. In the color histogram method color histogram of the sample object is used for the retrieval of the similar object image[3]. 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. 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[2]. This feature vector is used for matching process for the image retrieval. This method is scale and translation invariant, but not rotation invariant.

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 50 each [2]. HED is also useful for shape representation.

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[2]. The distribution of edges is useful for image to image matching. But this descriptor is not effective for the rotation invariance.

5. Angular radial partitioning (ARP)

In ARP method, the images in the stored database are converted to grayscale and edge detection is performed[2]. 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 makes 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.

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 Eigen space is calculated under consideration of all possible change.

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[4][7]. 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.

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[7]. 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.

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[8] . 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.

3. FEATURE-BASED IMAGE RECOGNITION

We come now to the extraction of features from the contours of the image. The features must be extracted one by one from the multitude of contours, which is a time-intensive process. Currently there are three different types of feature:

• **Line:** The line is defined by its centre point, length, and tangent angle. It is found by isolating the flat spots that occur at zero amplitude on the fingerprint plot. The line itself is fitted to the two endpoints on the part of the contour associated with the flat spot.

• **Arc:** The arc is defined by its middle point, length, radius and tangent angle. It is found similar to the line, but the flat spots must occur at higher amplitudes. The arc is then fitted to the two endpoints and the middle point along the part of the contour associated with the flat spot.

• Lobe: The lobe is defined by its middle point, two endpoints, length, tangent angle, and the angle the contour changes from start to finish. The exact contour cannot be recreated from this definition but the most important parts are covered, and it provides a very good approximation. The lobe is fitted to each spike on the fingerprint plot using the corresponding contour endpoints and tangents.

Each contour's fingerprint has to be examined to extract the relevant features. The features are stored with the associated contour in the program. The number of total features usually ranges from one thousand to ten thousand. All the features collected must be combined and reduced to find only the most important ones. This is done by checking for redundancy, then intelligently combining smaller features into larger ones.

• All types of features go through a grouping algorithm which finds matches using the various attributes for each feature type. A set of average feature attributes is created for each group.

• All the groups are checked for sufficient redundancy to ensure the feature occurs on multiple contours, meaning it is part of a significant outline. This means the number of features is now vastly reduced so a more intelligent grouping algorithm can be performed.

• Lines of different length are combined provided they are parallel and in close proximity. Calculations are done to make a new line that encompasses both lines that created it.

• Arcs are combined similarly to lines but the radius, intersection point, and local tangent angles are also taken into account. Again calculations are done to ensure it takes both smaller arcs into account.

• Lobes are combined if they are offset from one another; this is checked by finding the vectors between each of the three sets of points. The new lobe is calculated by moving the points a weighted distance along the vectors.

4. 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.

5. 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", 3rd 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 ScaleInvariant Keypoints”, International Journal of Computer Vision 60(2), 91-110, 2004
8. Zhang Dengsheng, Lu Guojun, “Review of Shape Representation and Description Techniques”, ElsevierComputerScience, 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
10. R.C. Flemmer & H.H.C. Bakker, "Generalised Object Recognition", submitted to ICARA 2009.
11. H.H.C. Bakker & R.C. Flemmer, "Data Mining for Generalised Object Recognition", submitted to ICARA 2009.