International Journal of Computing and Applications | **ISSN: 0973-5704**
Volume 13, Number 2, (July-December 2018), pp 241-246

A SURVEY ON OBJECT RECOGNITION METHODS

Nurzamal Islam[1], Senjam Hemba[2], Vijay Prasad[3] and Yumnam Jayanta Singh[4]

[1-3]Dept. of Computer Science & Engineering and IT, School of Technology, Assam Don Bosco University, Guwahati, Assam, India.
Email: [1]n.zaman98[@]gmail.com, [2]hembacoc[@]gmail.com, [3]vpd.vijay82[@]gmail.com
[4]Adjunct at Dept. of CSE and IT, Assam Don Bosco University, Currently with NIELIT, Kolkata, West Bengal.
Email: y_jayanta[@]yahoo.com

Abstract: Object Detection and Recognition is an important task in the area of computer vision and artificial intelligence. Object recognition is the task of identifying or classifying a particular object into one of the predefined categories. Object Recognition includes Image pre-processing, segmentation, feature extraction, object detection, and classification. Although there are different classification technique have been developed, SVM, ANN Decision Tree are the most common classifier in computer vision application. This study attempt to address some segmentation technique, feature extraction, and classification technique.

Keywords: Image pre-processing, Segmentation, feature extraction, classification.

1. INTRODUCTION

Object recognition is the process of identifying the category of the objects. Recognition of object is an easy task for humans, but it is one of the most challenging tasks in computer vision application because of various circumstances like different lighting conditions, the position of the object in the image, orientation, occlusion, scale, etc. Object detection and recognition are the two parts in object identification. Classification of the object in category wise according to its several predefined categories based on its features is called object recognition while the main objective of detection is to distinguish the object from its background. Object recognition finds its application in various domains, including Medical imaging, video surveillance, vehicle navigation, robot navigation, industrial inspection, remote sensing.

2. BASIC STEPS FOR OBJECT RECOGNITION

In this section, the basic steps for object recognition are shown in Figure 1.

A. Image Acquisition

Image acquisition is the first stage of any object recognition system. The images are acquired with a digital camera from the environment or already available dataset for processing.

B. Image Pre-processing

In this stage, the acquired images are pre- processed to remove the undesired distortions present in the image. Pre-processing enhances the quality of the object in the image. Some filtering techniques: Low pass filter, high pass filter, mean filter, median filter are used to remove noise in the image.

C. Image Segmentation

The goal of segmentation is to partition the image into parts with similar properties or subdividing the image into its homogeneous parts. Some of the commonly used segmentation methods are Thresholding Method, Region Growing Method, Region Splitting and Merging, Clustering Method, etc. [1]

1. *Thresholding Method*

Thresholding method is the simplest and most commonly used image segmentation method [2]. It converts the given image into a binary image by selecting an appropriate threshold T. The values of all the pixel below the threshold T is set to zero and all the pixels having a value higher than the threshold T is set to one [2].

$$g(x, y) = \begin{cases} 1, \text{ if } f(x, y) > \mathrm{T} \\ 0, \text{ if } f(x, y) \leq \mathrm{T} \end{cases} \quad (1)$$

2. *Region Growing Method*

Region growing is a method that groups the pixels into a larger region of homogeneous nature [1]. It is started from an initial seed pixel and merge the adjacent pixels according to similarity constraints. The regions are grown iteratively by adding the homogeneous neighboring pixels and increased the region size from the seed pixel. When the seed pixel does not find any similar pixel in its neighbor, it simply stops growing and select another seed pixel that yet to belong a particular region and starts growing again till all the pixels are merged into some regions [3]

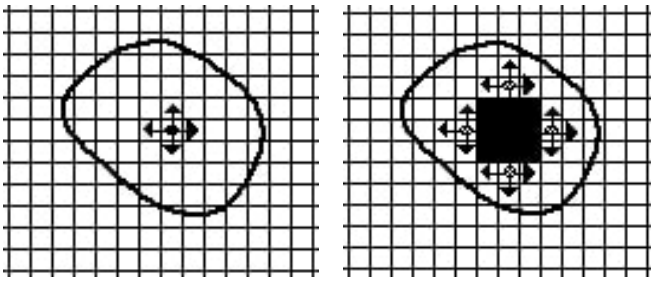

Figure 2: (a) Start of Region Growing. (b) Growing Process

3. *Region Splitting and Merging*

In this split and merge technique, initially the whole image is considered a single region and split the image into four equal regions which are coherent with themselves. Then the each region is again subdivided into four equal parts. After each splitting, the adjacent regions are compared and merge the similar region according to its grayscale, variance, etc. This splitting process is continued until all the regions are coherent (i.e.,- if all the regions have similar properties), and merging is done after each splitting. This splitting phase forms a quadtree [4].

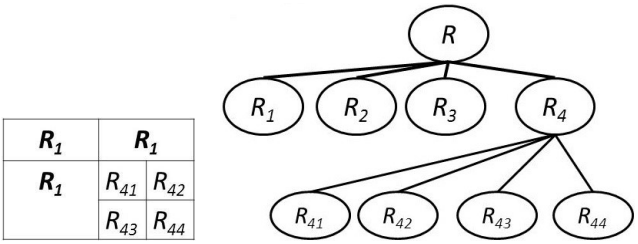

Figure 3: (a) Splitting of an image. (b) Quadtree

4. *K-Means Clustering*

The K-means clustering algorithm partitions all the pixels into k clusters where each pixel belongs to the clusters with minimum distance. The value of K is chosen as random or using some heuristic at the beginning. Later it can be updated as per the requirement based on the data set. The algorithmic steps of K mean clustering are as [5]:

(a) Select K cluster randomly or based on some heuristic at the beginning.

(b) Place each pixel in the cluster whose current centroid in nearest.

(c) Update the location of the centroid by averaging all the pixels in the cluster.

Repeat 2 and 3 until convergence.

D. Feature Extraction

Feature extraction is the process of extracting meaningful information from the images. Feature extraction is required to reduce the dimensionality that represents the most important part of the image as a feature vector. This process is very useful when the size of the images are large enough [6]. There are several techniques to extract features from images like color feature, shape feature, texture feature, etc.

1. *Color Feature*

In the case of visual perception information, the most useful feature is the color features. The Red, Green, and Blue are the three primary color component in an image that explored by human beings. By mixing these three individual color component with appropriate portion, we can generate any required color. This

primary color model is specified by RGB color model [7]. This RGB color model is represented by a Cartesian coordinate system. Are shown in the Figure 4. Where the three primary color RGB are represented along three Cartesian coordinate axes x, y and z.

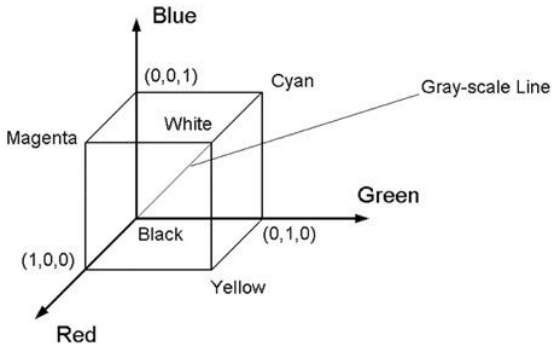

Figure 4: RGB Color Model

Another color model consisting of three components: Hue, Saturation and Value (HSV) can also be derived from the RGB color model.

When we see any color image, it cannot distinguish how much of a color component are present in that image. In order to discriminate the color component present in a particular image, we use three characteristics of a color image called Brightness, Hue, and Saturation. Brightness is the chromatic notion of the intensity of an image, Hue represents the dominant wavelength in a mixture of color and Saturation indicates the purity of that particular color. The three attributes Hue, saturation and Brightness are used to distinguish the color component from one color to another color. The amount of RGB is required to form any particular color is called Tristimulus. A color is normally specified by its chromatic coefficient given in the equation (1), (2) and (3) as the coefficient of red, the coefficient of green and coefficient of blue respectively.

$$r = \frac{R}{R+G+B} \tag{2}$$

$$g = \frac{G}{R+G+B} \tag{3}$$

$$b = \frac{B}{R+G+B} \tag{4}$$

The normalization is represented as $r + g + b = 1$. The color feature is best suited for recognition of object invariant of rotation and size of the object [8].

2. *Shape Feature*

The use of object's shape feature is one of the most challenging tasks in object recognition system. Shape feature plays an important role in searching similar objects in content-based image retrieval system. For finding the shape feature, the main step is to calculate the boundary (also called perimeter) and the area of the object using the equation (4) given below [9].

$$\text{Shape} = 4\pi\left\{\frac{\text{Area}}{\text{Perimeter}^2}\right\} \tag{5}$$

3. *Texture Feature*

Gabor filters are commonly used for representation and discrimination the texture feature. This is a Gaussian Kernel function which is modulated by a sinusoidal plane wave. It can be defined in 2D as [7]:

$$G2D(x, y; \sigma) = \frac{1}{2\pi\sigma^2} \exp^{-\left(\frac{x^2+y^2}{2\sigma^2}\right)} \tag{6}$$

The width of the Gaussian kernel is satisfied by σ.

4. *Local Binary Pattern*

Local Binary Pattern (LBP) is a very efficient and highly successful feature. It looks at nine pixels at a time, usually 3x3 pixels. It particularly interested at the center pixel. A Local Binary pattern is then, in turn, the set of the 3x3 pixel into a single value, and it is done by comparing every neighboring pixel with the center pixel. If the neighboring value is greater than or equal to the center pixel, then the neighbor pixel is set to 1 else it is set to 0. Then this string of number is used to train the system. The interesting thing about LBP is that it is luminous invariant. LBP improves its performance in object recognition when it is combined with HOG features. The computation of LBP and C in a 3x3 neighborhood [10]:

6	5	2
7	6	1
9	8	7

1	0	0
1		0
1	1	1

Figure 5: (a) 3x3 Image, (b) Thresholded with center pixel

1	2	4
128		8
64	32	16

Figure 5: (c) Power of 2 for LBP calculation

$$LBP = 1 + 16 + 32 + 64 + 128 = 241$$

$$C = (6 + 7 + 8 + 9 + 7)/5 - (5 + 2 + 1)/3 = 4.7$$

5. *Histogram Oriented Gradient*

The main concept of Histogram of Oriented Gradient is the local object presence and the shape of the image that can be represented by the distribution of intensity gradient or edge direction [11]. The HOG algorithm includes four main steps: Gradient computation, orientation binning, descriptor blocks and block normalization [12].

The implementation of HOG is achieved by decomposing the image into a small region called cell, for each cell collecting a local 1-dimensional histogram of gradient directions or orientation of the edges over the pixels of the cell. The resulting combined histograms represent the descriptor. For better performance and invariance of illumination and shadowing, the local histogram is normalized by collecting a measure of local histogram across a larger region called blocks [12].

(i) *Gradient Computation:* The gradient of an image $f(x, y)$ is actually a vector that can be calculated by using the partial derivative of that image along x- and y-direction.

$$\nabla F = \left(\frac{\partial f}{\partial x}, \frac{\partial f}{\partial y} \right) \tag{7}$$

For gradient computation, the first step is to calculate the gradient values. The most popular method is to apply the 1D centered point discrete derivative mask in both x and y directions. This method requires filtering the gray scale image with following filter kernels.

$$D_x = [-1 \quad 0 \quad 1] \quad \text{and} \quad D_y = \begin{matrix} 1 \\ 0 \\ -1 \end{matrix} \tag{8}$$

So, for any given image I, we obtain the derivative along x- and y-direction using the kernels

$$I_x = I \times D_x \quad \text{and} \quad I_y = I \times D_y \tag{9}$$

The Magnitude of the gradient is given by

$$|G| = \sqrt{I_x^2 + I_y^2} \tag{10}$$

The Orientation of the gradient is given by

$$\theta = \tan^{-1}\left(\frac{I_x}{I_y} \right) \tag{11}$$

(ii) *Orientation Binning:* The second step is to calculate the cell histograms. The 8x8 pixel size cells are computed with 9 orientation bins for [0°, 180°] interval. For each pixel's orientation, the equivalent orientation bin is found and the orientation's magnitude $|G|$ is voted to this bin [12].

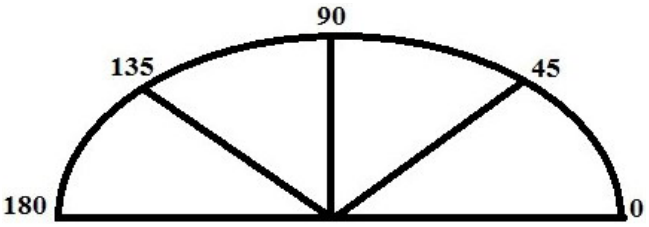

Figure 6: Histogram oriented Bins

(iii) *Descriptor Blocks:* In order to normalize the cells' orientation histograms, they should be grouped into larger cells, especially connected blocks. Usually, 2x2 cell forms a block with 50% overlap, so each cell contributes more than once to the final descriptor.

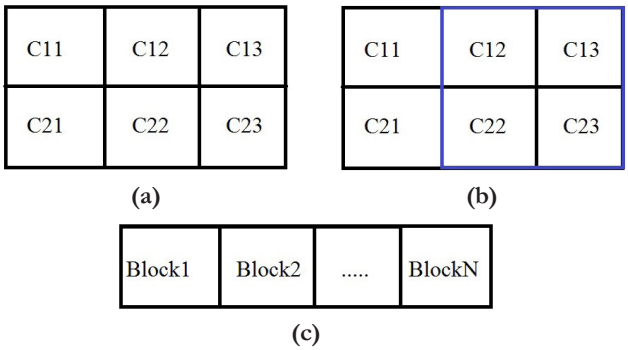

Figure 7: (a) an Image with six cells (b) HOG block 2-by-2 cells (c) HOG feature vector is arranged by HOG block

(iv) *Block Normalization:* Rather than the individual normalization of each histogram, the cells are grouped first into blocks then normalize the block based on the histograms present in the block [12].

E. Classification

There are a different kind of classification technique present for object recognition. Support Vector Machine, Artificial Neural Network, Multilayer Perceptron Neural Network, Decision Tree, Rule Based classifier, Naive Bayes classifier, etc. for classification. Some of them are discussed below:

1. *Artificial Neural Network*

Artificial Neural Network is a computational paradigm used to solve various problems in Artificial intelligence and pattern recognition purpose depending on the structure and functions of biological neurons of the human brain. Artificial Neural Network is approved to have better classification ability in many different areas. ANN has multiple interconnected layers: The first layer is called input layer, and there can be multiple hidden layers in between input and the final output layer. The first input layer sends data to the hidden layers which in turn passes the calculated result to the output layer.

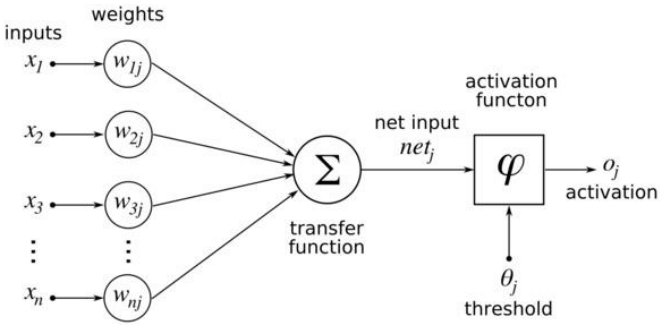

Figure 8: Artificial Neural Network Model.

To train ANN classifier, choosing an appropriate model is involved for which there present the corresponding algorithms. Training any dataset using neural network model is challenging because of the adjustment of weights and bias value. The output of the network is mainly associated with the adjustment of the input weights, and the bias value of the network and the output need to be matched with existing predefined data. In the Back Propagation model of neural network where multiple layers is present, and an error value for each layer is calculated including the final output layer. The error value is then passed on to the input layer, according to that error value the input weight and bias is adjusted in order to get the desired output, where the output is known, and that is why it is a supervised learning method. It is used in different applications like pattern recognition, optimization, speech recognition, etc. [13].

2. *Support Vector Machine*

A support vector machine (SVM) is a binary classifier for solving two class problem by their nature. However, it can be adjusted to manage multiclass problem using multiple svm called multiclass SVM. A support vector machine thus constructs a hyperplane for two class problem or builds $(n-1)$ hyperplane for n class problem of classification [14]. A support vector machine actually tries to find a hyperplane between any two classes that increases the boundary maximally of the separated class.

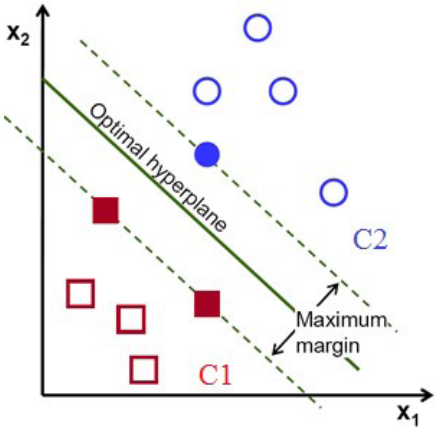

Figure 9: Maximum margin hyperplane of SVM trained for two class problem

Therefore, the ways the support vector machine design the classifier is that it tries to maximize the distance of the dividing boundary between the two classes by maximizing the dividing plane from each of the feature vectors belongs to class C1 and C2. We can compute a linear discriminant function as [15]:

$$g(x) = w^t x + b \tag{12}$$

The equation (12) represents a straight line in 2D, a plane in 3D and a hyperplane in multidimensional space. Where w is the vector which is perpendicular to that hyperplane. So the vector w represents the orientation of the hyperplane, and the constant b represents the position of the hyperplane in the d dimensional space, it is usually known as bias term. And for the classification problem, every feature vector x, it is required to compute the linear equation (equation 12). If the specific vector x_i lies on the positive side of the hyperplane, then the equation becomes:

$$g(x) = w^t x_i + b > 0 \quad (13)$$

If the specific vector x_i belongs to the negative side of the hyperplane, the linear equation becomes:

$$g(x) = w^t x_i + b < 0 \quad (14)$$

And when the vector x_i lies on the hyperplane then the equation:

$$g(x) = w^t x_i + b = 0 \quad (15)$$

An SVM classifier is trained using supervised learning method used for classification and regression analysis. SVM has been approved by a number of current studies to have better classification accuracy comparing another classification algorithm [16].

3. CONCLUSION

This paper presents a study and provides a brief knowledge about various feature extraction technique and some classification technique. This paper basically gives the idea of object recognition methods such as K-Means Clustering, Local Binary Pattern, Histogram Oriented Gradient with theoretical and some calculation knowledge.

References

[1] A Kaur, "A Review Paper on Image Segmentation and its Various Techniques in Image Processing." International Journal of Science and Research (IJSR) 3.12 (2014).

[2] A.B Patil, J.A Shaikh-OTSU Thresholding Method for Flower Image Segmentation, International Journal of Computational Engineering Research, Volume 6 (5), 2016.

[3] A.D. Marshall, "Vision systems." Tinklalapis http:// www. cs. cf. ac. uk/Dave/Vision lecture (1997).

[4] S. Gupta, Y.J. Singh, "Region and Shape Features Based Image Retrieval.", International Journal of Advanced Research in Computer Science and Software Engineering, Volume 5(11), (2015).

[5] Z.R. Barot, N. Limbad, "An Approach for Detection and Classification of Fruit Disease: A Survey." *International Journal of Science and Research (IJSR), (2013).*

[6] Feature extraction (https://in.mathworks.com/ discovery/feature-extraction.html), retrived on 2dec2016.

[7] S. Naskar, T. Bhattacharya, "A Fruit Recognition Technique using Multiple Features and Artificial Neural Network." International Journal of Computer Applications 116.20 (2015).

[8] Y. Zhang, L. Wu, "Classification of fruits using computer vision and a multiclass support vector machine." Sensors 12.9 (2012).

[9] W.C. Seng, S.H. Mirisaee, "A new method for fruits recognition system." *Electrical Engineering and Informatics, 2009. ICEEI'09. International Conference on.* Vol. 1. IEEE, (2009).

[10] T. Ahonen, A. Hadid, "Face description with local binary patterns: Application to face recognition." IEEE transactions on pattern analysis and machine intelligence 28.12 (2006).

[11] N. Dalal, B. Triggs, "Histograms of oriented gradients for human detection." Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on. Vol. 1. IEEE, (2005).

[12] C. Yıldız, "An Implementation on Histogram of Oriented Gradients for Human Detection."

[13] G. Kaur, "Satellite Image Classification using Back Propagation Neural Network." *Indian Journal of Science and Technology* 9.45 (2016).

[14] G. Anthony, H. Greg, M. Tshilidzi, "Classification of images using support vector machines." arXiv preprint arXiv:0709.3967 (2007).

[15] D.K. Srivastava, L. Bhambhu, "Data Classification using Support Vector Machine." Journal of Theoretical and Applied Information Technology 12.1 (2010)

[16] S.R. Suralkar, A.H. Karode, "Texture image classification using support vector machine." International Journal of Computer Applications in Technology 3.1 (2012)