ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 11, November 2018

Intelligent Systems that Emerges 'Visual Generic Object Recognition'

Mansi Patel[1]

Graduate Student, Department of Computer Science, California State University-Sacramento, California, United States[1]

ABSTRACT: Visual Object Recognition is the base of many Machine Learning Models and Artificial Intelligence Models. Object Recognition deals with identifying the object given and learn to identify the new same object with the knowledge of the previously given object by grouping the same kind of objects to one category. Generic Object Recognition tends to detect the general objects from the given image i.e. person. while Specific Object Recognition detects the more specific objects i.e. Nelson Mandela. The Visual Generic Object Recognition requires the object representation, the training technique, and the object detection technique. This paper describes an introduction to the object recognition, representation of the generic objects, detection of the generic object, and one intelligent system - face detector.

KEYWORDS: Artificial Intelligence, Machine Learning, Generic Object Recognition, Feature Extraction, Face Detector, Intelligent Systems, Object Classification, Object Detection, Haar Feature

I. INTRODUCTION

Object Recognition gives the machine the understanding the difference between different objects and categorizes the same object in the same category. It is related to the intelligent system which captures the image and differentiates the objects in the given image based on the trained data. Recognition has two variant with respect to object recognition, the specific case recognition, and the generic case recognition. The specific case recognizes a particular person's face, a particular place, and a particular object, for instance, Michael Jackson's Face, Statue Of Liberty, and Neighbour's car while generic case recognition more focuses on the two cars parked in the parking lot and it identifies both. Overall the specific case recognition is vastly used for the systems for information retrieval where a system has been given the knowledge of every famous person living or dead, it has to be more specific to get the correct details and deliver to the people while generic case recognition is vastly used for the self-driving car project where a system has to recognize cars and trucks moving on the roads, it has to be resulting in to identification of all pedestrians, cars, trucks, and bicycles where the system does not have to identify exactly who is the owner of the car or which specific pedestrian is passing.

As stated in Visual Object Recognition by Kristen Grauman and Bastian Leibe, Series Editors - Ronald J. Brachman and Thomas G. Dietterich, For specific object recognition in computer vision relies on a matching and geometric verification paradigm while for generic object recognition it also depends on training data of statistical model of appearance and shape for the same object from different angles even for different illumination. Paper illustrates the different representations for generic object categories and detection of the same object with the various pose, texture, etc. The paper illustrates the process of Visual Object Recognition for Generic Object involves representation of the Generic Objects, Generic Object Detection, Generic Object Training Models, The Viola-Jones Face Detector - An example of such system which includes several phases of the training process, recognition process, and observation.

Face Recognition among Object Recognition is most promising; unlocking the smartphone with our face is no more imagination. This kind of the Face Recognition falls into the specific object recognition because it unlocks the smartphone detects the face of the owner of the device. Although the generic object recognition is very trending.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 11, November 2018

II. RELATED WORK

Recent prominent work on Visual Object Recognition can be seen in the project of the self-driving car made by Google-Waymo. TensorFlow is one of the promising API which supports Object Recognition, a better way computationally for Deep Learning and Machine Learning. Intel is currently working on the project of the autonomous vehicles and for such intelligent system which uses the concept of object recognition, their team is working on their own GPU and GPU frameworks. Inspired by the work of Viola and Jones, Mehul K Dabhi and Bhavna K Pancholi improved and tested the face detector implemented Haar features for the feature extraction. Highly motivated by interesting work of the authors Ronald J. Brachman and Thomas G. Dietterich for the better understanding of concepts and contents. The main goal of this paper is to illuminate the theory and concepts of how generic object recognition works in the machine-learning establishment.

III. REPRESENTATIONS OF THE GENERIC OBJECTS

A. *Window Based Object Representation*

This particular representation refers to a single descriptor for the same region of interest or image object with respect to appearance. Given a window descriptor to find out which category the object should be is exactly the window based object representation. There are several ways to represent an object from an image in single descriptor window, it could be represented and detected by its pixel, resolution or its textures. However, the perfect representation for given system varies from its purpose of the application, for instance, *The Viola-Jones Face Detector* [Viola and Jones,2001] where the object category well evaluated by window based representation by 2D texture consisting rectangular patch centred of different face instances.

Several ways for Window Based Object Representation are Pixel Intensities and Colors, Window Descriptor: Global Gradients and Texture, Patch Descriptor: Local Gradients and Texture, A Hybrid Representation: Bags of Visual Words, Contour and Shape Feature, Feature Selection. One most promising example of the window based object representation is pixel-based detection of an image, to achieve this the most general technique is parallel processing which uses GPU (Graphical Processing Unit) that divides every image to small individual parts that work simultaneously on GPU to get best parallel processing throughput. This technique combines machine learning to the parallel processing for its best outcome.

B. *Part-based Object Representation*

This introduces more spatial relations in the recognition process in addition to the window based object representation. This representation involves models such as the Bag of words, Constellation, Start shape, Tree, k-fan, Hierarchy, and Sparse Flexible Model. From Fischler and Elschlager [1973], the basic idea is that assembly parts represent the objects and such parts have flexible spatial relations among them. Recent models learn the part appearance from training data given and try to separate the local features of the parts to one category. The major task is the selection method and grouping algorithms, most popular models use local features to select the parts the grouping choice depends on the tendency to assume of that part locations have mutual independence - different models deal with it differently, need to choose the best fit model for the targeted data with given numbers of parameters. A Bag of words model represents the completeness while Constellation Model is fully connected model defines the pairwise relationship of selected parts. A drawback with such a large number of parameters is that when it runs on a large number of the parts it limits the applicability for complex visual data.

C. *Mix Representation*

We can get the advantage of both the representation. Marcus de AssisAngeloniand HelioPedrini explore that Part-based representation enhances the performance of the face recognition evaluating every cell partition's

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 11, November 2018

performance rather than the bounding box, on the other hand, Dalaland Triggs shows the accuracy of the human detector using a histogram of oriented gradients descriptors rather than the shape contexts.

'Deformable Part-based Model' introduced by Pedro Felzenszwalb, David McAllesterand DevaRamanan[2008] for PASCAL person detection challenge efficiently performed the challenge with having two-fold improvement. For the selection, they used bounding box representation, defined the object part by latent variable and run the classifier which selects a window covering a larger area with the labeled bounding box.

In contrast, Chum and Zisserman[2007] did Star Model part representation that used to generate an initial hypothesis that compares the image to the set of premeasured model. The reverse methodology from the 'Deformable Part-based Model'.

IV. GENERIC OBJECT DETECTION

A. *Detection as Classification*

The classification is responsible for the generic object detection with one of the window based representation and trained classifier. The sliding window approach achieves the result of the classification. This detection technique tries to fit all objects of the image in one square or rectangle box and shows the generic object model name from the knowledge of the training data. Thus given an image to recognize the generic object detection, the square acts as a sliding box on the screen and tries to detect at each position and scale in the image. An image is resampled into the pyramid to achieve a multiscale run of the image and then runs through each level and scale for probable object detection, storing these outputs while checking for true detection by performing non-maximum suppression which is responsible for the post-processing step that results in true detections. The Figure illustrates the main components for classification model used for object recognition.

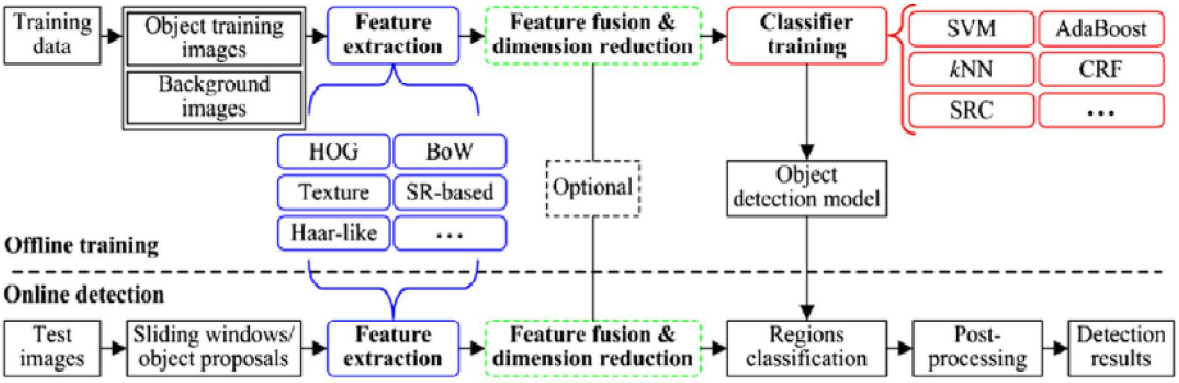

Fig.1. Flow Diagram of The Sliding Window based Detection

The flow shows that for the given test image and applied sliding windows representation, first it extracts the features. Feature extraction has an important role, given an image features extraction run extracts sub-windows of each level thus multiple scales are extracted and tested classifier on them. The feature fusion and dimension reduction layer is optional this omits some processing overhead by filtering the features, which are extracted from the feature extraction. The model than has object detection models on the basis of the trained classifiers that can be SVM, AdaBoost etc. followed by post-processing that prunes out the nearby true detection.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 11, November 2018

Detector here detects with the threshold given to it for false positive and negative detections. Generally, the threshold chosen between 0 to 1, 0 means the false detection and 1 means that there is an overlap. The threshold depends on the application and the precision and recall curves of the application at the testing phase.

B. *Detection with Part-based Models*

Part-based Models seeks search efforts during the detector's runtime although its efficiency depends on the application, generally, it is easy to deal with parts and spatial occultation where fewer training examples are required. Combinational Classifiers have the simplest approach that combines information from multiple object parts during the classification. This was first approached by Mohan et al. [2001] and Heisele et al.[2001]. The approach has a direct connection between the part's location and its spatial dependencies. This technique has some drawback if the object parts can move independently in such a case the more training data is needed and even the results cannot be assured. Taking it to the Generalized Hough Transform and RANSAC detectors for the part-based detection model, both the approach have the advantage with higher dimensional transformation spaces when including various aspects of an image i.e. aspect ratio, image-plane rotation etc. The difference is that the Generalized Hough Transform detection relies on the integration of votes in a certain tolerance window while RANSAC detects the same with inlier features by tolerance threshold.

The below figure shows the stages of the GHT(Generalized Hough Transform) for part-based detection.

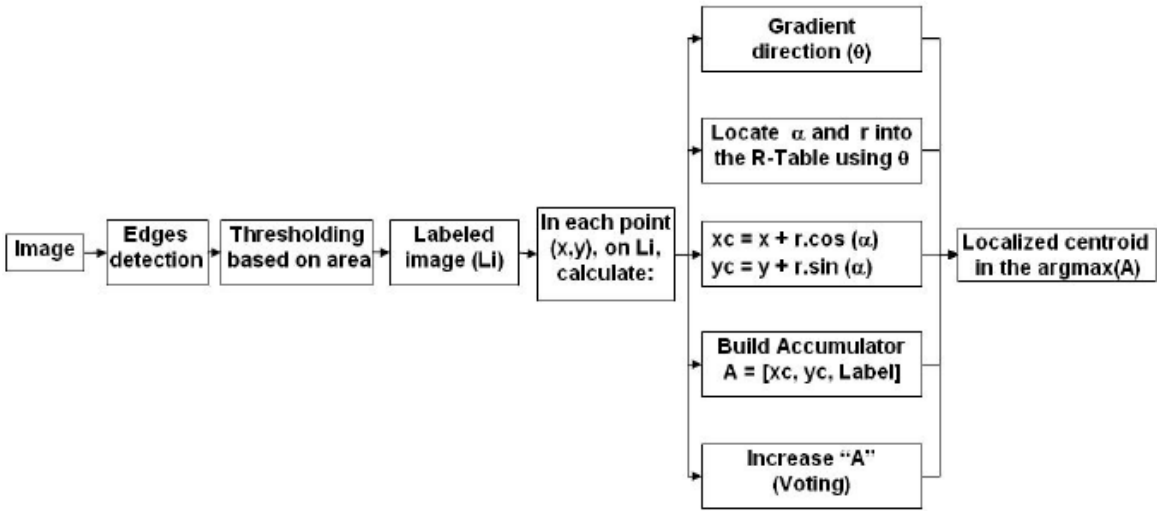

Fig. 2. The flow diagram of the Generalized Hough Transform for part-based detection

The figure illustrates how the GHT based on area threshold and part's voting can establish the parts and spatial relationship among those parts and returns the build accumulator with the location measures for the applied label with it. The flow shows that for a given image it first detects the edges then based on the area the threshold is selected that follows by the labelling of an image. For each location (x, y), define the gradient direction, generalise all parameters, build accumulator and vote for the accumulator by increasing it and then get the local centroid of 'A'.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 11, November 2018

V. THE VIOLA-JONES DETECTOR

Let's look at one Generic Object Recognition System - The Viola-Jones Detector that is the face detector. This model is considered to be a milestone in the field of generic object detection that used discriminative learning to detect the regularities among faces. This technique developed by an ensemble of weak classifiers, these classifiers implemented to extract some simple contrast-based feature from the sliding window face. The detection algorithm scans an image to detect if it has similar features as the human face. Training of this technique is relatively expensive and expects the cropped images of frontal face instances that also in a large number of images while the detection on this trained model is very quick because of internal images and attentional classifier cascade used. Further discussed the main aspect of this detector which is the training process and the recognition process.

A. *Training Process*

The training set consists the cropped images of faces of the candidate and also non-faces objects with a common resolution which is generally 24 x 24 pixels, thus training set have positive and negative examples. Here Voila & Jones introduces Haar feature. To define some similarities of human faces, they first defined a library of rectangular features shown as the figure below, the sliding window measures each location of the image and try to fit the rectangular feature on it and the result of the rectangular feature is measured by the sum of blue colored filled rectangle minus the sum of white colored rectangle feature from one Haar feature, the figure-3 below shows example for eye feature on the top of it. Voila & Jones defined 180,000 such features for their detection purposes. For each training internal image, it computes the sum of the intensities at each location (x,y) while the response of each feature requires 6-8 array references to store.

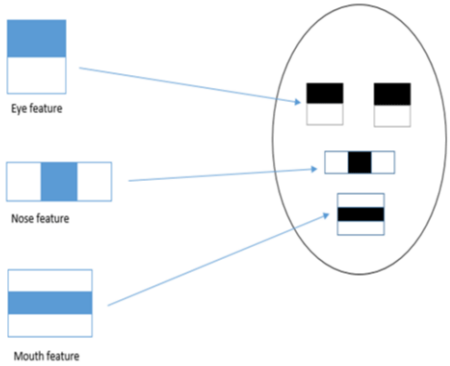

Fig. 3. Haar Features as rectangular representation

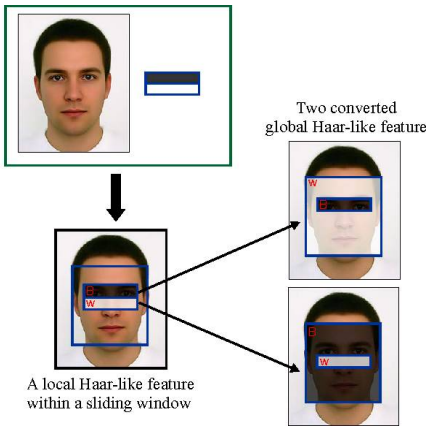

Fig. 4. Identifying rectangle features during the training process

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 11, November 2018

The figure-3 describes the representation of the haar features of the face as rectangular features which will be detected in real time on the basis of given trained model and figure-4 describes detection of the rectangular features on one of the training image given to it. Figure 4 has two haar feature for human eyes and under the eye area which trainer processes and detects on the given image exactly on his eyes and his under eye area. This is just a simple example of two haar feature while in reality there could be many of the haar feature that can be created and detected from human face by the classifier in the training process.

Using the AdaBoost Algorithm to the response of the above process to gain the sequential selection of the discriminative rectangular features, the strong classification over the weak classifier can be achieved. Adaboostclassifier then selects according to weighted error measures on the training data. This turns out to be a strong classifier over the weak classifier and aimed to focus on the remaining errors and build trained classifier with fewer errors.

B. *Recognition Process*

Given an image as test input, classifier use sliding window object presentation for each location having a pyramid for multiple scales run on the image call it a sub-window. The strong classifier runs on each sub-window and computes selected T rectangular feature responses using an internal image. Furthermore, it classifies thesub-window as face object or non-face object as the binary output of the classifier. The below figure shows the overall detection procedure of the Viola-Jones Face Detector.

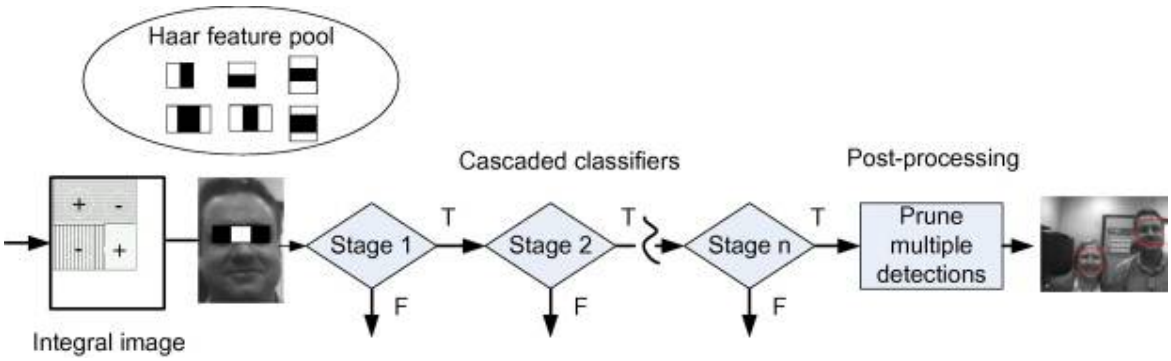

Fig. 5. The Overall detection procedure of the Viola-Jones Face Detector

Figure-5 gives an overall idea of the face recognition combining both training and recognition/detection process of the Viola-Jones Detector. The input is any visible real time image object and the output is the square/rectangle box that implies the detection of the faces on an input image. The process starts with the detection of rectangular haar features of the face that is followed by the cascade classifier and Adaboost algorithm applied on it. For every stage/round of the classifier it selects the responses. Atlast the post-processing prunes the result out of the sliding window.

C. *Observation*

The classifier here plays the vital role for the face detection in the Viola-Jones Face Detector as they use the strong classifier that turns out the best for the given model. The cascade classifier works until T rounds and selects sequentially T responses of the discriminative rectangular feature responses. Thus this implies that the classification process is really expensive and tedious while the detection and recognition process on the trained classifier works pretty fast and efficient.

ISSN(Online): 2320-9801
ISSN (Print) : 2320-9798

International Journal of Innovative Research in Computer and Communication Engineering

(A High Impact Factor, Monthly, Peer Reviewed Journal)

Website: www.ijircce.com

Vol. 6, Issue 11, November 2018

VI. CONCLUSION

Undoubtedly, the 'Visual Generic Object Recognition' is the latest trend for the Machine Learning and Artificial Intelligence whether it is Google's new project of the self-driving car or the Microsoft AI's autonomous vehicle project. The Recognition performance relies on the representation of the objects used and detection technique applied to the representation. From a large number of different representation methods of the window-based representation and the part-based representation, which of them are suitable for the proposed application and among them to decide which performs the best is challenging and even more challenging is to decide the detection technique that benefits the chosen representation. Thus from the numerous combination of the representations and detection techniques, one should try every possible unique thought having a probability of success likewise Viola-Jones tried cascade classifier with the sliding window representation which became milestone model for visual generic object recognition. Thus, every object recognition application needs a thorough knowledge and research for related work and trial-error approach to evaluate the best-fit classifier and detector.

REFERENCES

1. Ronald J. Brachman and Thomas G. Dietterich, 'Synthesis Lectures on Artificial Intelligence And Machine Learning', Morgan & Claypool Publisher, pp. 1-4,61-103, 2011.
2. Burr Settles, 'Active Learning', Morgan & Claypool Publisher, pp. 1-4,61-103, 2012.
3. Wei Zhang, 'Constellation Models for Recognition of Generic Objects',(2015)
4. N. Dalal and B. Triggs, 'Histograms of oriented gradients for human detection', IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), July 2005.
5. Pedro Felzenszwalb, David McAllester and, Deva Ramanan ,'A discriminatively trained, multiscale, deformable part model", IEEE Conference on Computer Vision and Pattern Recognition, August 2008.
6. Marcus de AssisAngeloni and HelioPedrini, 'Part-based representation and classification for face recognition' IEEE International Conference on Systems, Man, and Cybernetics (SMC), February 2017.
7. Mehul K Dabhi and Bhavna K Pancholi, 'Face Detection System Based on Viola - Jones',International journal of Science and Research (IJSR), Vol 5, April 2016.
8. Gong Cheng, Junwei Han, 'A Survey on Object Detection in Optical Remote Sensing Images', ISPRS Journal of Photogrammetry and Remote Sensing, March 2016.
9. Miguel Vera, Antonia Jose bravo, Ruben Medina, 'Improving Ventricle Detection in 3-D Cardiac Multislice Computerized Tomography Images' International Conference on Computer Vision, Imagining and Computer Graphics (VISIGRAPP), 2010.
10. Akash AA, Mollah AS, Akhand MAH, 'Improvement of Haar Feature-Based Face Detection in OpenCV Incorporating Human Skin Color Characteristic', Journal of Computer Science Applications and Information Technology, Nov 2016.
11. Ming Yang, James Crenshaw, Bruce Augustine, Russell Mareachen, and Ying Wu, 'AdaBoost-based face detection for embedded systems', ELSEVIER Computer Vision and Image Understanding, Vol 114, pp 1116-1125, November 2010.