Object Detection

By Yiwen Jiang, ECE '21

Introduction

Computer vision is one of the "hottest" areas nowadays with its prominent application in fields including facial recognition, autonomous vehicle, and robotics systems. The problem that this area is trying to solve is to decipher how to let computers see and identify things like we humans do. While computers have the power to solve extremely complicated problems and do thousands of iterations of command in seconds, they actually cannot perform the tasks that we see as the simplest. Training the computer to do things that even babies do, such as distinguishing a tree from a human face, a banana from a chair is what encompasses the field of computer vision. A variety of techniques and areas constitute the overall realm of computer vision, this tech note focuses on the field of object detection.

Basic Object Detection Methods and Concepts

Object detection focuses on the problem of locating or defining objects of desired labels and categories within an image. There are several popular methods for object detection including interest points with Hough voting, sliding windows, and region proposal.

Sliding Windows

To start, the concept of sliding windows and region proposal are all rather easy to understand.

The sliding window method uses a square box and shifts it around the image. For each instance, compute the confidence value for the current image that is inside the box. Hoping that for one of the instances the object would be within the border of the box and be detected. An example is shown below in Figure 1, a red box representing the window is being slid across the photo and the goal is to detect the walking man. As seen on the leftmost subfigure, the red box included the man that is intended to be detected.

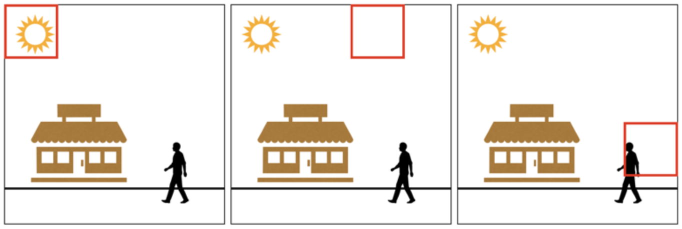

Figure 1. Sliding Windows Approach

Region Proposal

The region proposal method starts with grouping the pixels in an image into predicted image-like regions using traits such as contours and patterns. The confidence score for each of the regions is then being computed. This method is used with the hope that one of the regions proposed would contain the actual object. The same picture as before is shown in Figure 2, still, with the goal of detecting the waking man, a variety of different pixel groups represented by the different color boxes are being proposed. The confidence score of these boxes containing the walking man is being calculated, and as shown in Figure 2, one of the boxes proposed, the purple box, actually included the object intended to be detected.

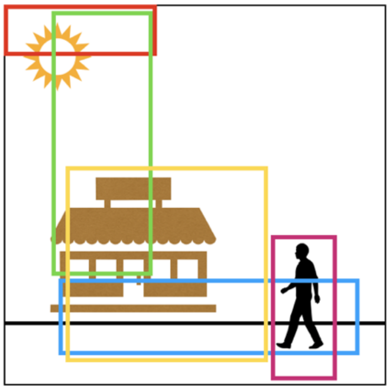

Figure 2. Region Proposal

Interest Point Based

The interest point based method is slightly different. The concept of Hough transform is used here. With a line in the image space, $y = mx + b$, Hough space takes coordinates as b and m, which are the slope and the vertical shift. Thus, a line in the image space will correspond to a point in the Hough space, as (m, b). Similarly, a line in the Hough space, $b = -mx + y$, will correspond to a point (x, y) in the image space.

The use of interest points is followed by Hough voting. The idea of voting is such that it is not possible to fit a model for all combinations of features, thus voting is used to let the features vote for models that are compatible with it. Interest point is first found in an image. Interest points are usually identified by a change in one of the characteristics of the picture, such as color, shape, that in turn make these points distinct and relatively easy to detect. There are many methods to do that, one of the popular ones is using Harris corner detector. For each interest point, an area around the point is used as a training patch. Voting is done to decide where the center of the object could be. Points with multiple votes would be determined to be the center of an image. The patches that voted for this point will be used in determining the overall area of the object.

An example of this concept is shown in Figure 3. Three interest points are being detected and calculated as indicated by the three red circles. With Hough voting, the three interest points voted for the most likely position of the center of the object to be, represented here with the red triangles. This provides great confidence that there is likely to be an object at the location.

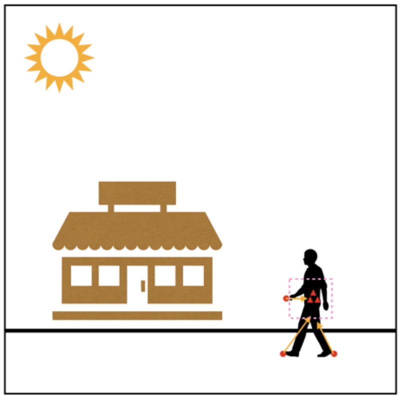

Figure 3. Interest Point with Hough Voting

Deep Learning Object Detection

Haar Feature-Based Cascade Classifier

The cascaded classifier is first proposed by Paul Viola and Michael Jones in a 2001 conference paper "Rapid Object Detection using a Boosted cascade of Simple Features". This method utilizes both positive and negative images in the training process. While it is first proposed with the purpose of facial recognition, with sufficient training data this model can be trained for other objects as well. This classifier employs Haar feature, where three types are being used, edge, line and four rectangle features as shown below in Figure 4. The value for each feature is the difference between the sum of pixels under the white and black rectangles. While it is highly likely that each image contains an enormous number of features, it is also true that some of them are very irrelevant. The features that classify the positive and negative images with the minimum error rates are being selected. The final classifier consists of a weighted sum of weaker classifiers that themselves are not sufficient to detect the image.

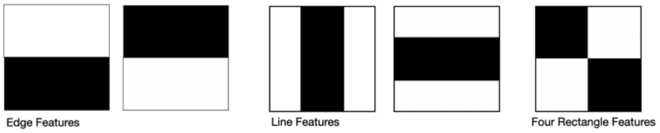

Figure 4. Haar Features

The cascaded concept is introduced based on the idea such that there are areas in the picture where no positive instances exist. Thus, after the rough check, if the region doesn't contain any desired instances, then the region should not be focused on in later processing. To achieve this, the paper divides up the total features into smaller groups such that for each stage only a few features are being applied. When a region fails on any of the stages, it will be discarded and not considered for the remaining testing process.

Aggregated Channel Features Detector

The Aggregated Channel Features detector is first being introduced in the paper "Fast Feature Pyramids for Object Detection" written by Dollár et al. in 2014. The paper proposed a new and efficient method of computing feature pyramid, which is a scheme commonly used for object detection. The traditional method of computing Feature Pyramid includes computing feature channels at different scales of the original picture. The convention is to calculate the value with 4 to 12 scales per octave, however, this requires a lot of calculation, thus extremely costly. The paper suggested instead of calculating the channel for the 4-12 scales per octave, only the channels on the octave would be calculated, while the intermediate scales will be computed with approximation, an illustration is shown in Figure 5 to represent this pipeline. Moreover, to improve the robustness and accuracy of the approximation, downsampling is being used so that the pixels being utilized in the resulting channel are weighted sums of the ones from the original channel.

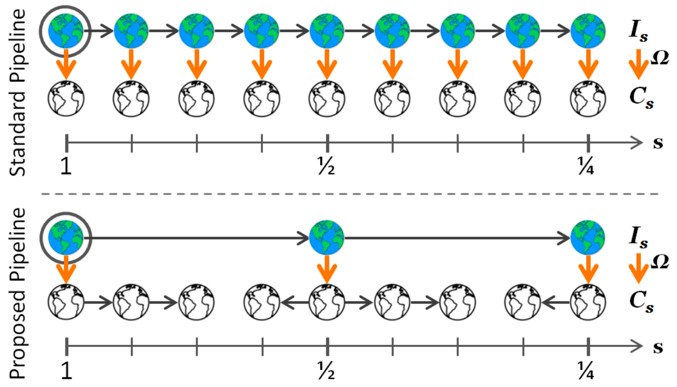

Figure 5. Fast Feature Pyramids. Reprinted from Dollar, Piotr, et al. "Fast Feature Pyramids for Object Detection." IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, no. 8, 2014, pp. 1532–1545., doi:10.1109/tpami.2014.2300479.

This cost-efficient method of Fast Feature Pyramid lays as the foundation to the Aggregated Channel Features detector. Given an input image, the channels are being computed. The channels used in this detector include normalized gradient magnitude, histogram of oriented gradients, and LUV color channels. The channels are being smoothed through downsampling or lowering the resolution. Next, the aggregated channels are being computed with features being single-pixel lookups. Finally, an augmented boosted tree is being trained using these features to achieve object detection. An overview of the ACF detector is being shown below in Figure 6.

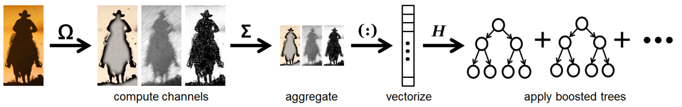

Figure 6. Overview of ACF Detector. Reprinted from Dollar, Piotr, et al. "Fast Feature Pyramids for Object Detection." IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, no. 8, 2014, pp. 1532–1545., doi:10.1109/tpami.2014.2300479.

The method of ACF detection distinguishes itself with its fast speed that is extremely valuable in areas where real-time object detection is needed. As reported in the paper, a pedestrian detector built using this method can operate at over 30 frames per second.

Object Detection Application

Object detection and recognition technology has been applied to a variety of areas nowadays. It is widely used in automated driving systems, where the identification and detection of pedestrians and other cars on the road are crucial to ensure the safety of the driver and others on the road. This has also been employed in areas to alleviate the work of human labor, such as the assisting robot arms in the process of picking up single packages from a pile of packages.

Moreover, this has also been used in the landings of space crafts in outer space where the hazards in potential sites have to be identified and avoided.

Hazard Detection in Outer Space Landing

A method using the Hough transform method and supervised learning decision forests to detect multisize rocks has been proposed by the paper "A Holistic Vision-based Hazard Detection Framework for Asteroid Landings" written by Yan et al.

Similar to the general computer vision algorithm, the process detailed in the paper are divided into training and detection procedures. A supervised machine learning process is used in the process, and the detection is being assumed to originate from pixels as voting elements. The detection problem is formulated based on the following characterizations 1) if the current pixel is from an object 2) if the object is a real object (meaning the area the computer chosen to be an object, might just be a random and meaningless area). The specific algorithm employed in the paper is slightly different from that of the traditional Hough transform concept. Whereas the traditional one sums the votes, this method accumulated the vote in a Hough image, and the maximum value of the hypothesis is considered. A Hough forest is incorporated into the detection process and trained with positive and negative samples. While this method demonstrates high robustness it also incorporates a high level of complexity to implement.

Conclusion

Using the concepts of the basic object detection methods such as interest points with Hough Voting, the incorporation of deep learning enables the possibility of faster and more accurate algorithms. These detection models, such as the ones discussed in this tech note the ACF detector, provide promises in the area of Computer Vision with hope that one day computers will one day be proficient in doing simple tasks of identifying and classifying objects as humans do. This also sets the foundation of the possibility of achieving a heavily automated based society.

References

1. Paul Viola and Michael J. Jones. Robust real-time face detection. International Journal of Computer Vision, 57(2):137–154, 2004.

2. Rainer Lienhart and Jochen Maydt. An extended set of haar-like features for rapid object detection. In Image Processing. 2002. Proceedings. 2002 International Conference on, volume 1, pages I–900. IEEE, 2002.

3. OpenCV: Cascade Classifier. (2020). OpenCV. https://docs.opencv.org/3.4/db/d28/tutorial_cascade_classifier.html

4. Ballard, D. H. (1981). Generalizing the Hough transform to detect arbitrary shapes. Pattern Recognition, 13(2), 111–122. https://doi.org/10.1016/0031-3203(81)90009-1

5. Dollar, Piotr, et al. "Fast Feature Pyramids for Object Detection." IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 36, no. 8, 2014, pp. 1532–1545., doi:10.1109/tpami.2014.2300479.

6. Fidler, S., Object Detection, UTortonto, CSC420-Introduction to Image Understanding,Class Lecture Slides, http://www.cs.toronto.edu/~fidler/slides/2015/CSC420/lecture17.pdf, Retrieved April 27, 2021.

7. Yan, Y., Qi, D., Li, C., Yu, M., & Chen, T. (2016). A Holistic Vision-based Hazard Detection Framework for Asteroid Landings. IFAC-PapersOnLine, 49(17), 218–223. https://doi.org/10.1016/j.ifacol.2016.09.038