www.revistageintec.net
ISSN: 2237-0722
revista GEINTEC
Gestão, Inovação e Tecnologias

Multiple Object Recognition Using OpenCV

D. Kavitha[1]; B.P. Rishi Kiran[2]; B. Niteesh[3]; S. Praveen[4]

[1]Assistant Professor, Department of Computer Science and Engineering, SRM Institute of Science and Technology, Ramapuram Campus, Chennai, India.

[1]kavithad3@srmist.edu.in

[2]Department of Computer Science and Engineering, SRM Institute of Science and Technology, Ramapuram Campus, Chennai, India.

[2]rishikiran008@gmail.com

[3]Department of Computer Science and Engineering, SRM Institute of Science and Technology, Ramapuram Campus, Chennai, India.

[3]nitheeshnithi39a@gmail.com

[4]Department of Computer Science and Engineering, SRM Institute of Science and Technology, Ramapuram Campus, Chennai, India.

[4]praveen99aug@gmail.com

Abstract

For automatic vision systems used in agriculture, the project presents object characteristics analysis using image processing techniques. In agriculture science, automatic object characteristics identification is important for monitoring vast areas of crops, and it detects signs of object characteristics as soon as it occurs on plant leaves. Image content characterization and supervised classifier type neural network are used in the proposed deciding method. Pre-processing, image segmentation, and detection are some of the image processing methods used in this form of decision making. An image data will be rearranged and, if necessary, a region of interest will be selected during preparation. For network training and classification, colour and texture features are extracted from an input. Colour characteristics such as mean and variance in the HSV colour space, as well as texture characteristics such as energy, contrast, homogeneity, and correlation. The device will be trained to automatically identify test images in order to assess object characteristics. With some training samples of that type, an automated classifier NN could be used for classification supported learning in this method. The tangent sigmoid function is used as the kernel function in this network. Finally, the simulated results show that the used network classifier has a low error rate during training and higher classification accuracy. In the previous researches Object detection has been made possible, but in our current research we have attempted to do live Object Detection using OpenCV and also the techniques involved in it.

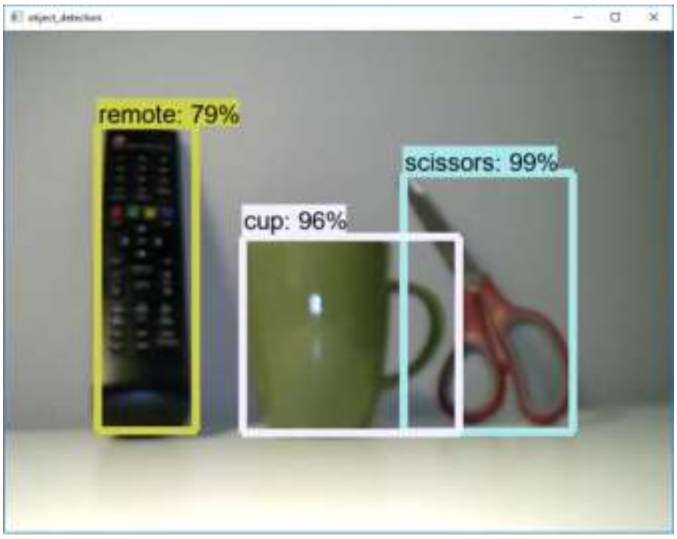
Key-words: OpenCV, HSV, CNN, SVM Classifier.

1. Introduction

Object recognition has seen a rise in popularity in the computer graphics field, thanks to advancements in video monitoring and image analysis. However, in both large-scale applications and embedded platforms, reaching high performance and near-real-time object detection is a major concern. Because of the growing security issues in various fields, a stable and effective near real-time object detection programme running on an embedded device is essential.

The framework can be found on a range of platforms, including a high-performance platform and a web platform. This can be used in surveillance systems with distributed cameras and a backend server that handles the identification. It's all compatible with handheld devices that have a camera and processor.

Fig. 1 - This Figure Explains Object Detection with the Accuracy Percentage.[8]



In terms of detection, such systems need a fast response time. We're especially interested in developing a device that can detect several objects in a scene at the same time. Surveillance cameras can use this tracking information to relay real-time information about observed objects to the back end central device. The data sent to the back-end device may be used to detect the presence of an individual in the protected area or to identify a single entity (blacklisted person) from the identified face.

2. Related Works

The following methods are the previous works and methods used by many authors and we have used it as a reference for our paper. The methods are as follows:

2.1. R-CNN

To suggest feature vectors or bounding boxes of possible objects in the scene, an image processing method named as "selective search" is used, though the implementation flexibility allows for other object proposals techniques to be used. The CNN generated a 4,096-element matrix that describes the object's elements, which is then fed into a linear SVM for labelling, with one Classifier trained for each class. It's an effective and easy implementation of CNNs to the issue of object detection and localization.

The technique is slow because each of the region proposals created by the object proposals algorithm requires a CNN-based feature extraction transfer. At test time, the model was running on about 2,000 analysis framework per picture.

2.2 Fast R-CNN

It was proposed in the 2014 article "Spatial Pyramid Pooling in Deep Convolutional Networks for Spatial Identification" that the process referred to as visual sustainable framework networks, or SPPnets, be accelerated. This did accelerate the feature extraction process, but it was basically forward and pass caching technique. Rather than a pipeline, Quick R-CNN is suggested as a single model that learns and outputs areas and categories explicitly.

The model's analysis brings an image as input and returns a series of neural network, which are then processed by a deep neural network. A deep neural network processes the results of the neural network. A well before CNN, such as a Pre - trained models, is used to retrieve functionality. At the end of the deep CNN, a customised system called a Place of Importance Convolutional Layers, or RoI Accumulating, is used to extract the features specific to a given input target area. Although the model is considerably easier to train and forecast, each source picture also needs a number of feature vectors to be proposed.

2.3 YOLO

The method utilizes a specific neural network that is trained from concept to execution, taking a photograph as data and predicting feature vectors and group points for each feature vector. Despite running at 45 frames per second and up to 155 frames per second for a pace version of the algorithm, the methodology has lower predictive accuracy (e.g., more localization errors).

The model works by splitting into a range of units, with each unit liable for determining a boundary if its centre falls inside the cell. Based on the x, y dimensions, as well as the size, height, and trust, each grid cell generates a bounding box. About any cell is used to predict a class.

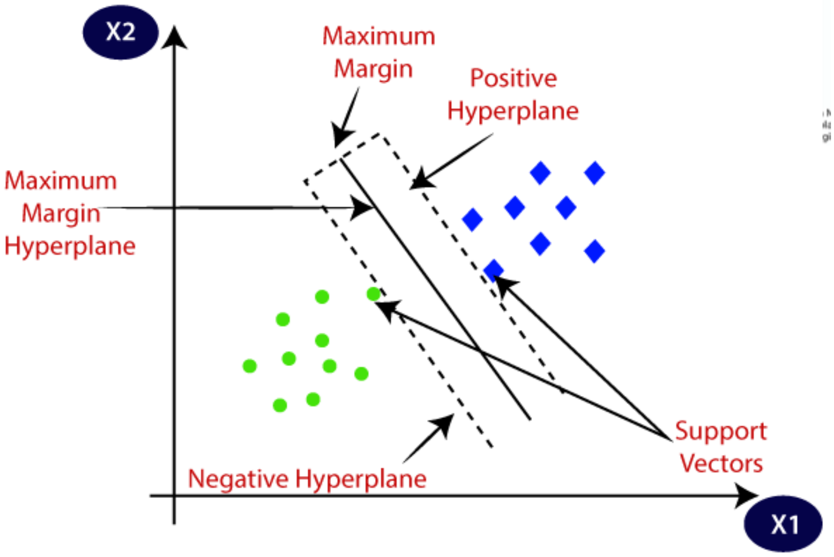
3. Methods

The framework takes live fed image and converts the RGB image into HSV format and uses OpenCV for image detection, the methodologies include 1.SVM Algorithms,2. **C**NN,3. OpenCV.

3.1. SVM Algorithm

A support vector machine (SVM) is a classification and regression artificial intelligence method for data processing. SVM is a supervised learning technique for categorizing data into different classes. The output of an SVM is a map of the segmented image with the largest possible margin between the two. A supervised learning algorithm called a help vector machine separates data into two groups. It's trained on data that's already been divided into two categories, and it builds the prototype as it moves. An SVM algorithm is responsible for determining which group a specific data set refers to. As a result of this, SVM is a non - linear system classification algorithm. For an SVM algorithm to place objects into groups, the gaps between them on a graph should be as wide as possible. The extreme vectors that help construct the hyperplane are chosen by SVM. Help vectors are the extreme cases, and the algorithm is considered a Support Vector Machine. The example we used in the KNN classifier will help you understand SVM. If we see a curious cat with any puppy-like features, we'll use the Svm classifier to construct a model that can reliably determine if it's a cat or a dog. We'll initially prepare our prototype with a huge number of photos of cats and dogs so it can learn about their various traits, and then bring it to the check with this strange beast. Face recognition, image processing, text categorization, and other tasks will all benefit from the SVM algorithm.

Fig. 2 - It Explains the Hyperplanes which Separate the Values in a SVM.[11]



There are two forms of SVM:

Linear SVM is a classifier for linearly separable data, which is defined as a sample that can be divided into two classes by a single plane. Linear SVM is the classification algorithm for linearly separable data.

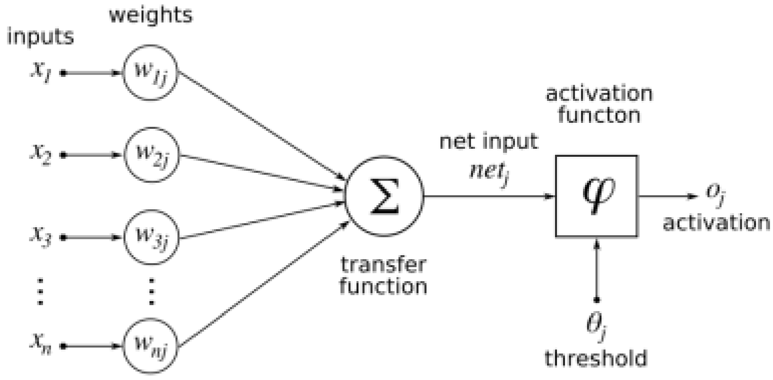
A non-linear SVM is a classifier that is used to classify data that is not isolated linearly. This implies that a sample is non-linear if this could be classified using a straight line, and the classifier used is Non-linear SVM.

Support vectors are pieces of information that are near to the vector space and have an impact on the hyperplane's direction and orientation. These class labels are used to increase the margin of the classifier. If the support vectors are removed, the hyperplane's position would change.

3.2. CNN

The Convolutional Neural Network is the most well-known image detection and classification algorithm (CNN), also known as a ConvNet, the revival of deep neural networks in computer vision, a branch of machine learning was sparked by CNNs, which were one of the primary developments. Multiple layers of artificial neurons make up convolutional neural networks.

Fig. 3 - The Basic Component of Artificial Neural Networks, The Artificial Neuron, has the Following Structure



Artificial neurons, like their biological counterparts, are mathematical functions that calculate the weighted number of several inputs and provide an activation value. Weights decide how each neuron behaves. When fed with pixel values, the artificial neurons in a CNN identify various visual information.

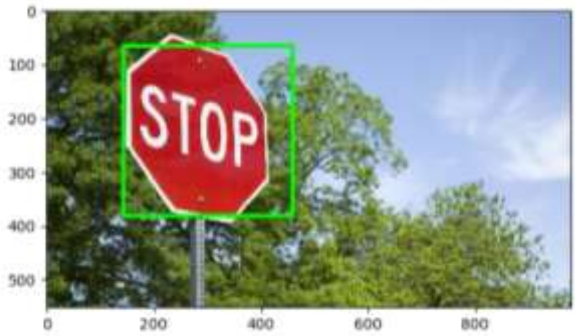
3.3 OpenCV

OpenCV is a fairly wide resource for image recognition, deep learning, and image analysis that is becoming increasingly important in contemporary networks. OpenCV can recognize objects, faces, and even human handwriting from photos and videos. Object detection is a subset of machine learning, signal processing, and big data that deals with identifying features in photos and images. The first step entails using a large number of negative and positive labelled images to train a cascade function. After the classifier has been trained, the training images are used to extract identifying features, known as "HAAR Features." HAAR features are basically rectangular features with bright and dark pixels in different areas.

The value of each function is determined by subtracting the amount of pixel intensity in the bright region from the pixel intensity in the dark region. These attributes are calculated using all of the image's potential sizes and locations. Many irrelevant features can be present in an image, while only a few ways define the object, a few related features may be used. The classifier is learned to extract useful features from the pre-labeled dataset and add sufficient weights to each feature to obtain the lowest possible errors. Poor function refers to a single feature. The weighted sum of the weak features is the final classifier.

ISSN: 2237-0722
Vol. 11 No. 2 (2021)
Received: 28.03.2021 – Accepted: 28.04.2021

Fig. 4 - Image Recognition Using OpenCV.[10]



The context takes up a large portion of the image; The item to be viewed is only a small portion of the picture. Cascaded classifiers are used to speed up the detection process. If even a single negative feature is detected in a region of an image during this step, the algorithm goes on to the next region after ignoring the region for further processing. The requisite object in the image is the only area that contains all of the identifying features. The requisite object in the image is the only area that contains all of the identifying features.

4. Experiment and Result

4.1 Pre-processing

Pre-processing is a technique for improving image data by removing unnecessary distortions and optimizing image features that are needed for further processing. a photograph Pre-processing is a term that describes operations on images at the most fundamental degree of abstraction. It accepts and outputs power files.

Image restoration is the method of estimating the clean original image from a corrupted/noisy image. Motion blur, noise, and camera misfocus are all examples of corruption. Image enhancement differs from image restoration in that the latter is intended to highlight aspects of the image that make it more attractive to the viewer, rather than actually providing data that is practical from a science standpoint. "Imaging kits" provide image enhancement approaches that do not rely on an a priori model of the process that created the image. Image enhancement can effectively remove noise by sacrificing any resolution, but this is insufficient in many applications. Resolution in the z-direction of a Fluorescence Microscope is still poor. To recover the picture, more sophisticated image processing techniques must be used. Image restoration methods such as de-convolution are an example. It has the ability to improve resolution, particularly in the axial direction, and also remove noise and boost contrast.

ISSN: 2237-0722
Vol. 11 No. 2 (2021)
Received: 28.03.2021 – Accepted: 28.04.2021

4.2 Colour Space Conversion

Colour space is a method of arranging colours in computer vision and image processing. A colour space consists of two components: a colour model and a mapping function. Colour models are useful because they allow us to represent pixel values using tuples. The mapping function converts the colour model into a list of all possible colour combinations.

Objects in images typically have distinct colours (hues) and luminosities, which can be used to differentiate different areas of the image. The hue and luminosity of an RGB image are represented as a linear combination of the R, G, B channels, while they are single channels in an HSV image (the Hue and the Value channels). By simply thresholding the HSV channels, a simple segmentation of the image can be accomplished.

Fig. 5 - The Image Represents the Hue and Value Channel which is Separated from the Original RGB Image

4.3 Feature Extraction

One of the best and most interesting domains is image processing. Feature extraction is a step in the dimensionality reduction process, which separates and reduces a wide collection of raw data into smaller classes. Features are the pieces or designs of an entity in a picture that aid in its recognition. A cube, for e.g., has four corners and four edges, which are known as square features and aid in human recognition. Corners, margins, regions of interest points, ridges, and other properties are examples of features. As a result, processing would be simpler. In this domain, you can essentially begin to play with your images in order to comprehend them. To process a digital image or video, we use a variety of techniques, including feature extraction and algorithms, to detect features such as shapes, edges, and motion. Convolutional neural networks (CNNs) can replace existing feature extraction because they can extract complex features that articulate the image in greater depth, learn task-specific features, and are more accurate.

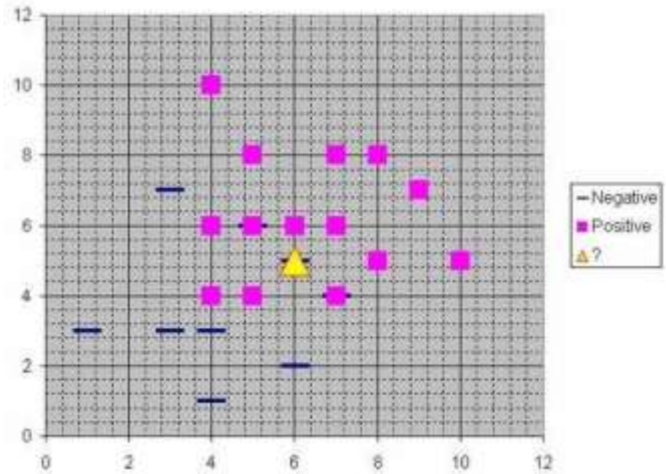
4.4 NN Classifier

Neural networks are statistical structures that are built on the behaviour of biological neurons. The best method for detecting and distinguishing between various sets of signals is a neural network. It is important to choose a suitable architecture and learning algorithm when using a neural network to achieve the best performance. Unfortunately, there is no sure-fire way to do so. One of the big PR victories of the twentieth century was the choice of the word "neural network." The phrase "A system of graded, proportional values with key derivation features" sounds a lot more fascinating than "A system of graded, proportional values with key derivation features." Given their name, neural networks are not "conscious computers" or "synthetic minds." A hundred neurons make up a standard artificial neural network. The human nervous system, on the other hand, is thought to have about $3x10^{10}$neurons.

Although the implementation is somewhat different, neural networks and K-Nearest Neighbour (k-NN) models are conceptually similar. The most important theory though is an item's expected target value is likely to be similar to other items with similar predictor variable values. Find the following diagram:

Assume that each case in the training set has two predictors, x and y. The instances are mapped with their x and y axes, as shown in diagram. Assume the target variable is split into two groups: positive (represented by a block) and negative (represented by a line). Assume we're attempting to forecast the result of a new situation, as described by the triangle's position, with attribute values of x as 6 and y as 5.1.

Fig. 6. This Figure Represents an Example of How Neural Network Works



ISSN: 2237-0722
Vol. 11 No. 2 (2021)
Received: 28.03.2021 – Accepted: 28.04.2021

A dash denotes a negative value, and the triangle is almost perfectly placed on top of it. The dash, though, is in an unusual position as opposed to the other dashes, which are grouped below the squares and to the left of centre. As a result, the resulting negative value may be considered as an outlier.

The nearest neighbour classification is affected by the number of neighbouring points considered in this case. Since it is located on top of a proven negative point, the latest point can be labelled as negative if 1-NN is used and only the nearest point is taken into account. However, When the 9-NN model is being used, and only the nearest 9 points are considered, the effect of the nearest 8 positive elements outweighs the near negative one.

4.5. Digital Image Processing

Image processing techniques such as noise removal will most likely be used to classify artefacts in an image, after that, (limited) attribute extraction is used to find lines, regions, and likely areas with distinct shapes.

These tasks are often performed unintentionally by the human visual system, but to approach human performance, a computer needs professional programming and a lot of processing power. Various methods are used to interpret information in the form of an image. A image is usually depicted as a two-dimensional series of brightness values, similar to those seen on a photographic print, slideshow, tv screen, or movie screen. A picture can be processed optically or digitally by a computer.

A picture must first be condensed to a sequence of blocks that the computer can manipulate before it can be digitally manipulated. A picture component, also known as a pixel, is a numerical representation of an image's brightness value at a given spot. Although much larger images are becoming more common, a typical digitalised image could have 512 512 pixels, or around 250,000 pixels. There are three simple operations that can be performed on the image in the machine after it has been digitised. For a point operation, a data point in the output image is reliant on a particular image pixel in the source images. The quality of an output image pixel is calculated by multiple neighbouring pixels in the image for local operations. In a global operation, all of the source image pixels add to the output image pixel value.

These operations, taken individually or in combination, are used to enhance, restore, or compress the image. A picture is enhanced when it is changed to make the details it contains more visible, but it can also be enhanced by making it more visually appealing.

Computer vision has long sought to recognise object groups in real-world images. This is conceptually challenging due to wide appearance differences within object instances belonging to the same class. Furthermore, distortions caused by background clutter, size, and perspective variations can cause even the same object instance to appear quite differently.

5. Conclusion and Future Enhancement

Detecting instances of a particular class, such as within an image or video, is what object detection is all about. It can do so by studying the unique characteristics of each individual. The Algorithm developed successfully detected multiple objects live as well as in a saved video. The details in the image are accurately identified by this proposed method. The proposed system is to test real time object/image detection using OpenCV and we have successfully implemented it. Since this method doesn't work at the speed we want and also our accuracy ranges about 98%. We can further enhance it by using these models in real-time applications like cctv and other means to observe anything we need. Since we have added only a limited number of sample objects to our model, our model can only detect those predefined objects.

References

Kushal, M., BV, K. K., MJ, C. K., & Pappa, M. (2020). ID Card Detection with Facial Recognition using Tensorflow and OpenCV. *In Second International Conference on Inventive Research in Computing Applications (ICIRCA)*, 742-746. https://doi.org/10.1109/ICIRCA48905.2020.9183342.

Othman, N.A., Salur, M.U., Karakose, M., & Aydin, I. (2018). An embedded real-time object detection and measurement of its size. *In International Conference on Artificial Intelligence and Data Processing (IDAP),* 1-4. https://doi.org/10.1109/IDAP.2018.8620812.

Cameron, K., & Islam, M. S. (2018). Multiple Objects Detection using HSV. *In NAECON 2018-IEEE National Aerospace and Electronics Conference*, 270-273.

https://doi.org/10.1109/NAECON.2018.8556711

Chandan, G., Jain, A., & Jain, H. (2018). Real time object detection and tracking using Deep Learning and OpenCV. *In International Conference on Inventive Research in Computing Applications (ICIRCA)*, 1305-1308. https://doi.org/10.1109/ICIRCA.2018.8597266

Ahmed, S., Khan, M.F.R., Labib, M.F.A., & Chowdhury, A.E. (2020). An Observation of Vision Based Underwater Object Detection and Tracking. *In 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE)*, 117-122. https://doi.org/10.1109/ICETCE48199.2020.9091752

Lei, Z., Xue-fei, Z., & Yin-ping, L. (2008). Research of the real-time detection of traffic flow based on OpenCV. *In International Conference on Computer Science and Software Engineering*, *2*, 870-873. https://doi.org/10.1109/CSSE.2008.872

Yu, L., Sun, W., Wang, H., Wang, Q., & Liu, C. (2018). The design of single moving object detection and recognition system based on OpenCV. *In IEEE International Conference on Mechatronics and Automation (ICMA)*, 1163-1168. https://doi.org/10.1109/ICMA.2018.8484437.

towardsdatascience.com, 'Live Object Detection', 2019. https://towardsdatascience.com/live-object-detection-26cd50cceffd.

towardsdatascience.com, 'How to Detect Objects in Real-Time Using OpenCV and Python', 2020. https://towardsdatascience.com/how-to-detect-objects-in-real-time-using-opencv-and-python-c1ba0c2c69c0.

Geeksforgeeks.com, 'Detect an object with OpenCV-Python', 2020.

https://www.geeksforgeeks.org/detect-an-object-with-opencv-python/.

medium.com, 'Support Vector Machine', 2020.

https://medium.com/@skilltohire/support-vector-machines-4d28a427ebd.

Pactpub.com, 'AI and Deep Learning', 2018.

https://www.packtpub.com/product/deep-learning-essentials/9781785880360.

scikit-image.org, 'RGB to HSV Conversion',

https://scikit-image.org/docs/dev/auto_examples/color_exposure/plot_rgb_to_hsv.html.

Dtreg.com, 'Probabilistic and General Regression Neural Networks',

https://www.dtreg.com/solution/probabilistic-and-general-regression-neural-networks

machinelearningmastery.com, 'A Gentle Introduction to Object Recognition with Deep Learning', 2019. https://machinelearningmastery.com/object-recognition-with-deep-learning/.