IEEE TRANSACTIONS ON INTELLIGENT TRANSPORTATION SYSTEMS, VOL. 16, NO. 5, OCTOBER 2015

Neha M

Reg. No: 14CO127

8095299474

nehamohan1996@gmail.com

Vehicle Color Recognition with Spatial Pyramid Deep Learning

Authors: Chuanping Hu, Xiang Bai, Senior Member, IEEE, Li Qi, Pan Chen, Gengjian Xu, Lin Mei

- It has been a practice from a long time, to monitor license plates for the purpose of video monitoring, detection of criminal activities, etc. But the major problem is that it is not visible always.
- On the other hand, a vehicle's body color occupies most of the visible area and could be used as one of the traits to be monitored.
- In this paper, a method to extract color of a vehicle using deep learning has been proposed.
- A Convolution Neural Network (CNN) and Spatial Pyramid techniques are used as feature extractors and Support Vector Machine (SVM) as classifier.
- Advantage: No preprocessing required. Performs better than other models in every challenging condition. Limitations: Cannot expect 100% accuracy.

Work Described

- Feature Learning with CNN: Train the deep CNN architecture having 5 convolution layers and 3 fixed layers. The trained CNN learns features from each of the images and feed it to the classifier.
- Color Recognition with SP and SVM: The features from CNN are scanned region by region and features are extracted. Hence every pixel is considered. The images of features extracted by SP are then classified by SVM.

Why is this system important?

In urban cities, the roads are filled with huge number of vehicles, of different/same sizes, colours and shapes. But since the colour of a vehicle is the most prominent attribute, it will be of big use in the field of criminal detection, etc. But this involves huge data input. To classify this image data manually requires lot of man power. Hence having a system which can classify the images based on color, accurately, would be convenient. CNN combined with SP and SVM techniques gives us most accurate results.

Methodology

- The proposed system uses CNN architecture use it to learn features from each of the input images. The responses of the filters are regarded as our features for vehicle color recognition. To learn a more discriminative representation, the network is trained.
- The network consists of 5 convolution layers C1, C2, C3, C4, C5 and three fully-connected layers {fc6, fc7, fc8}.
- Following each convolution layer, the contrast normalization, pooling and nonlinear function are connected to it one after the other. The output of each fully-connected layer is computed as

  $$Y(t) = W(t-1)Y(t-1) + B(t-1),$$

  where W and B are the parameters learnt in training phase.

  The result obtained after all the processes will be the score of each category in classification.

- After training the deep CNN, we choose the outputs of the last three layers, since these outputs of the first four layers are not discriminative enough, and the dimensions of the features are too high.
- Now, the Spatial Pyramid extractor is used on the results of above step.
- SVM is then used on the extracted features to classify different colors.

Results

- Convolutional layers execute convolutions with a group of filters to generate the response maps, which are also used as the features to train a classifier.
- In the feature learning stage, network with five convolution layers and three fully connected layers is constructed. The images are resized to $227 \times 227 \times 3$.
- Then, the images are fed to the convolutional layer with 96 filters of size $11 \times 11 \times 3$ and stride of four pixels.
- The outputs of C5, fc6 and fc7 will be the vehicle recognition features for our system.
- As a result, the vectors of sizes 46080, 20480, and 20480 are generated.
- To encode spatial information, the features of different subregions a vehicle image are also generated. The images are divided into 2×2 subregions. Then, the features of each subregion and the whole image are concatenated to form a long vector. The aggregated features and the corresponding labels are used to train the linear SVM classifier.

Fig 1. Vehicle color recognition examples by the proposed algorithm.

Pros

- The proposed feature is more informative than conventional features, even at an extremely small scale.
- The features from the training data are learnt automatically, and hence doesn't adopt manually designed features.
- The algorithm is run on raw pixels, and hence pre-processing of images to remove noise, hue, etc., is not required.
- Using spatial method for recognition increases the accuracy further.

Cons

- The system is not 100% accurate.
- Severe illumination and several indistinguishable colours results in wrong predictions.

Future work

- The method proposed is general and can be applied to other AI systems as well.
- Since the system is highly accurate and robust, the authors are planning to integrate it in real world applications in future.

Implementation as a part of mini project

As a part of this project, a system which takes an input a picture of a vehicle and gives color of the vehicle as output using deep learning will be built. The SP and SVM strategy will be used instead of fully connected layers as the authors claims that it is more effective. This will be implemented in python.