Frequency Maps as Expert Instructions to lessen Data Dependency on Real-time Traffic Light Recognition.

Thiago Almeida  
Programa de Pós-Graduação em Ciência da Computação (PROCC)  
Universidade Federal de Sergipe (UFS)  
Email: thiagosa@dcomp.ufs.br

Hendrik Macedo, Leonardo Matos,  
Bruno Prado and Kalil Bispo  
Departamento de Computação (DCOMP)  
Universidade Federal de Sergipe (UFS)  
Email: (hendrik, leonardo, bruno, kalil)dcomp.ufs.br

Abstract—Research on Traffic Light Recognition (TLR) has grown in recent years, primarily driven by the growing interest in autonomous vehicles development. Machine Learning algorithms have been widely used to that purpose. Mainstream approaches, however, require large amount of data and a lot of computational resources. In this paper we propose the use of Expert Instruction (EI) to reduce the amount of data required to provide accurate ML models for TLR. Given an image of the exterior scene taken from the inside of a vehicle, we stand the hypothesis that the picture of a traffic light is more likely to appear in the central and upper regions of the image. Frequency Maps of traffic light locations were thus constructed to confirm this hypothesis. Results show increased accuracies for two different benchmarks, by at least 15%. The inclusion of EI in the PCANet achieved a precision of 83% and recall of 73% against 75.3% and 51.1% of its counterpart. We finally presents a prototype of a TLR Device with such EI model to assist drivers. To show the feasibility of the apparatus, a dataset was obtained in real time usage and tested in an AdaBSF and SVM algorithms to detect and recognize traffic lights. Results show precision of 100% and recall of 90.9%.

Type / Symposium: Regular Paper / CSCI-ISCI.

I. INTRODUCTION

Although traffic lights are worldwide used and known as a very intuitive traffic regulator, a study from the city of São Paulo/Brazil [1] points out that in August 2016 the advance of the red light was the 2nd most frequent violation associated with fatal accidents. A Traffic Light Recognition (TLR) device could help to lower these numbers. According to [2], the main task of a TLR is to prevent accidents by warning the driver of the existence of a traffic light ahead in a non-intrusive way. In addition, the TLR can detect the main traffic light for that road when there is more than one and how far the traffic light is.

There are, though, several issues to be overcome in order to reliably detect and recognize the state of a traffic light and implementing a robust TLR is not a trivial task [2]. Among the challenges, are the weather condition that alters the lighting of the environment, different types of traffic lights, and traffic lights that are partially hidden due to exterior aspects.

Current work with Image Processing techniques and Machine Learning (ML) from traffic lights image collections shows accuracy around 80%, except for the times when limitation occurs. A way to get around the limitations can be the use of knowledge about the problem and the environment that cannot be obtained directly from the image. We call such type of knowledge of Expert Instruction (EI), since knowledge may be implicit in the form the model is built and not in the data itself, a theory that this paper proposes to validate. Embedding expert instructions can also favor a smaller amount of data needed by ML algorithms to perform well, which is highly desired considering that collecting and categorizing traffic light data is indeed a extreme laborious task.

We are particularly interested in what is the impact of knowing the traffic light object location in a photograph obtained from a camera in the interior of the vehicle to the ML-based recognition process. The hypothesis is that there should be regions where the traffic light object image appear more frequently within the photograph (on the top? on the bottom? right? left?) and, for that reason, these regions may have higher likelihood. We believe that frequency maps properly sum up the result of a deep analysis performed by human experts over traffic light image datasets towards the object location discovering.

This work thus intends to assess whether the use of a traffic light detection and recognition model that explicitly embeds frequency maps as expert instructions achieves better results than a purely ML-based solution. Scientific contribution is two-fold: (1) provide a human-tailored mechanism to produce frequency maps as expert instructions from traffic light image datasets and (2) design an hybrid model to the automatic recognition of traffic lights that demands lower data.

In section 2 we list some of the main attempts to solve the problem. Two of them were chosen as the basis for our investigation and so we highlight enrolled techniques to expose the theoretical foundations of our approach and ease the understanding. Section 3 presents our approach to embed frequency maps as expert instruction in the ML model training. Results of model training as also shown. In section 4 we present the TLR device we developed to show the feasibility of our proposal. We conclude the work in section 5.
II. CURRENT APPROACHES FOR SMART TLR DEVICE

An object recognition mechanism in an image consists of two parts: (1) a detection phase to identify target objects and (2) a posterior classification phase. A sort of different techniques has been tried to deal with the detection phase, ranging from the use of Convolutional Neural Network (CNN) [3] and PCAnet Neural Network [4] to Salience maps [5], [6], [7]. In the recognition phase, most work employs Support Vector Machines (SVM) [8], [9]. A non ML approach can be seen in the work of [6], where Fuzzy Logic has been successfully applied. Other techniques have also been applied in this phase as an alternative to ML to improve false positive detection rates or to soften the connection between detection and recognition phases; [10], [11] and [7] use Histograms whereas [12] and [13] considers the use of Transforms.

[14] proposes a new method that combines Computer Vision and ML to recognize traffic lights. Color extraction and Blob detection are used to perform detection. Color extraction is performed in the HSV color space. Blob detection is implemented by combining image processing algorithms such as flooding, contour tracking and closing. The combination of such techniques allows both circular traffic lights and traffic lights with arrows to be identified. After detecting the regions of interest, a Principal Component Analysis (PCA) classifier is applied. The proposed PCA classifier consists of a PCA network (PCAnet), a type of neural network that simulates the functioning of a CNN with fewer layers and a linear SVM. Five PCA networks (number of groups of identified traffic lights) are created to estimate weights or features capable of identifying the types of traffic light. The weights then feed the SVM that perform the final classification. Tests are performed using a dataset produced and made publicly available by the authors.

[15] proposes an adaptive background suppression algorithm in order to highlight regions of interest, the AdaBSF. Normalized gradients of the input image are calculated. Combining the normalized gradients with the values of the R, G, and B layers of the original image gives a simple 4-layer feature map. Each \( W_i \) detection window in this feature map can be represented as a basic feature vector \( x_i \in IR^{D \times 1} \). In the recognition phase, each candidate region is checked and classified into different traffic light classes using an SVM. The local color histogram and the gradient-oriented histogram of each candidate region are used as descriptors of linear SVM characteristics and are used to train the model and to classify the regions found by AdaBSF. The tests are performed with a specific dataset, provided by the author.

III. EMBEDDING EXPERT INSTRUCTIONS

One way to reduce the amount of data needed to train the algorithms is to introduce some sort of previously known information about the data (or the domain), a prior knowledge [16]. We implemented a model of traffic light detection and recognition that uses Frequency Maps as expert instructions and compared it with the works of [14] and [15] as baselines.

A. Experimentation Scenarios

In the first scenario, data from two datasets were used; the first provided by the work [15] and the second from [14]. To elaborate the EI, test data for each dataset were firstly divided into 2 groups, one for testing and the other to generate the EI. From this second group, 650 images were randomly selected. A total of 1300 images divided into 2 sets of EI were thus provided. Classification algorithm of [14] was implemented and used to compare the influence of EI on training results. Each experiment was carried out with and without the inclusion of EI and has the following tasks: (1) Model training (for different dataset lengths:1000, 4000, 7000, 10000 and 13000 items), (2) Model testing, (3) Usage of EI, (4) Accuracy evaluation. All experiments were performed twice with data selected at random. Accuracy was then measured. Data used in the experiment are images of traffic lights taken from the inside a vehicle.

In the second scenario, the EI was considered on the complete traffic light detection and recognition flow. The traffic light detection algorithm of [14] was implemented and included in the experiment. The regions selected by the detection algorithm were submitted to classification by the algorithm trained with and without EI. The regions identified as traffic lights were marked in the original test image for further evaluation by a human expert. Precision and recall were then calculated. The 17 test sequences of the dataset were used in this scenario. The tasks performed on the test are: (1) Model training with 7000 data, (2) Use of the detection algorithm from [14] to find ROIs in traffic scenes obtained from the inside of a vehicle, (3) Model testing with 17 traffic scene sequences, (4) Use of the original EI, (5) Precision and recall analysis.

All experiments were performed on a Dell Inspiron 15 7000, i7 processor and 16 Gb RAM.

B. Frequency Maps as Expert Instructions

If we consider an image taken from the inside of a vehicle towards a traffic light, it is more likely to note the traffic light in the central and upper regions of the image, since they are usually suspended on poles along the road. This knowledge is an important clue. We, thus, produced traffic light frequency maps to test this hypothesis.

The method consisted on the analysis of a series of images by a human expert. In each image, the region where the traffic light is located is demarcated. A matrix of the same dimensionality of the analyzed images is then created. The points of the matrix related to the demarcated regions receive an increment. In the end, the entire matrix is divided by the total number of images analyzed, thus generating the frequency of traffic lights for each pixel in the image.

The datasets from [15] and [14] were used to generate two frequency maps as shown in Figure 1; redder hues represent higher frequencies. A random sample of 650 images was generated from the test data for each dataset, totaling 1300 images. The frequency data was smoothed using a masked averaging algorithm.
To properly compare the results of two different ML models, it is necessary to train both models with the same data. However, the datasets do not provide the coordinates from where the traffic light events had been obtained from the original images, making it impossible to directly combine the previously calculated frequency with the training samples. To solve this problem, it was assumed that all training samples were found in regions with a non-zero value on the frequency map, so the combination could be made with random frequency values (in respect to the beta distribution). Figure 2 shows the frequency distribution histograms of the frequency maps. The values of the parameters of the beta distribution found for the 1st map were $\alpha = 0.01655$, $\beta = 1$, and for the 2nd map were $\alpha = 0.03420$, $\beta = 12.8985$.

C. Model training

The method used to evaluate the insertion of the EI was that proposed by [14]. The authors proposed a PCANet to find feature vectors given the set of traffic lights for training. The feature vectors are then used to train SVMs that will classify pixels as belonging (or not) to a traffic light. The addition of EI in this process can take several forms. The most convenient form is to treat the EI as the first layer of recognition. Soon after the detection phase, the Regions of Interest (ROI) are added or multiplied by the value of the frequency obtained for that region in the frequency map. Since the frequency map and the original scene have the same dimensionality, and knowing the ROI coordinates of the original scene, this combination will tend to highlight the ROI found for the next algorithm. In Figure 3, we see the structure of a PCANet in the form of a process model. It is possible to notice the inclusion of the expert instruction soon after obtaining the input image for classification, in this case, the ROIs.

An important concern is to ensure that expert instruction does not make the method deterministic, as it would be if it was applied at the end of the method. It is possible that some part of the frequency map has a value of 0 (zero), which means that in the analyzed sample, no traffic lights appeared in that region. However, it is also possible for a traffic light to appear in an unusual region of the scene and the TLR should be able to find it. It was necessary to define an increment $inc$ that is added to every multiplication factor $fat$, which consists of the frequency value found in the frequency map. All the experiments depicted in this section assumed $inc = 0.1$.

The [15] training dataset is composed of 2 main classes: GREEN and RED. The dataset has 9,977 units of green traffic lights and 10,975 units of red traffic lights, summing up 20,952 units of training. The recognition method proposed by [14]
Fig. 3. Structure of a two-phase PCANet (atages) including EI.

was trained with different amounts of data obtained from the training dataset from [15]. Firstly, no EI was considered. Next, we embedded the frequency maps. After each training episode, an accuracy test was performed. Final accuracy values can be seen in Figure 4. It is possible to notice the great positive effect of EI on the accuracy for a smaller amount of data (orange line vs. blue line in the figure). For greater data, behavior becomes similar until over-fitting occurs.

Similar experiment was carried out with the dataset from [14], using the same amounts of training data. As in the previous experiment, data is distributed along 4 classes according to the type of traffic light: round, with an up arrow, left arrow, and right arrow. However, in this case, data are divided into more specific groups depending on how the traffic light appears: GreenRoi1, GreenRoi3 and GreenRoi4 for variations of green traffic lights; and RedRoi1 and RedRoi3 for red traffic light variations. Final accuracy values can be seen in Figure 5. As before, it is possible to notice the great positive effect of expert instruction on the accuracy for smaller amount of data. Also, results seem to indicate that over-fitting for EI version happens earlier.

Fig. 4. Test accuracy levels with EI (orange line) and without EI (blue line) for the dataset from [15].

It is worth noting that the behavior of applying EI to different data may vary, even within the same problem domain. Figure 6 further detail the accuracy for each traffic light set. Although the EI has improved the rates for all sets, the groups concerning the red traffic lights got minor improvements if compared to the green ones. This suggests a lower sensitivity of the red traffic lights to the addition of EI. This hypothesis is further supported by the sharper fall in the accuracy rate for the green groups GREENROI3 and GREENROI4 (over-fitting), which not happens with that intensity in the red groups.

Fig. 5. Test accuracy levels with EI (orange line) and without EI (blue line) for the dataset from [14].

1466
D. Detection and Recognition

To validate the inclusion of EI in the complete flow of a TLR, an experiment was carried out performing the ROIs detection phase followed by the submission of the ROIs found to the trained model. The detection algorithm of [14] was used in conjunction with the PCANet for classification. Unlike the previous experiment, the frequency value obtained from the original frequency map gets an increment of $inc = 0.1$, since the ROIs are now identified by the detection algorithm and keep the region’s origin coordinates. In the previous experiment, a frequency estimate was generated based on the beta distribution of the original frequencies.

Two tests were performed with the images test dataset provided by [14]: with and without EI. This test dataset consists of 17 sequences of 2142 images (in which there is a traffic light), the shortest with 56 frames, and the largest with 518 frames. A sample of length 7,000 for the training set of both tests was chosen. This value is the average of the minimum and maximum amounts of data used in the previous experiment. Accuracy measure is manually done. Indeed, automating the counting of hits and errors in such kind of problem is quite complicated because it is not possible to associate an image to a specific class, since it is not a matter of classifying images, but embedded objects. In addition, the number of objects that can be found in the image may also vary in each image of the dataset, and could even be 0.

Table I shows the precision and recall rates obtained with and without EI. The use of EI presents significantly better results, achieving accuracy of 83% and recall of 73%, in contrast to 75.3% and 51.1% for the version without EI.

IV. TLR Prototype Deployment

We developed a car device to interact with the driver and continuously capture images of the environment from the inside of the vehicle. A Galaxy S8+ was configured to capture video with HD resolution. The images were subjected to classification using the method of [15]. The final dataset contains 929 traffic images, being 638 images with green or red traffic lights and 291 images without traffic lights, the negative group. The images were extracted from videos at 5 frames per second (fps) rate. Figure 7 shows an example of the images obtained with the device.

Fig. 7. Image obtained using the TLR device with a Galaxy S8+

The samples of traffic lights used in training are very different from some of the examples found in the test set, as shown in Figure 8. Indeed, when the training set do not fully represent the real world, some traffic lights are not recognized. Lighting conditions in the training dataset also vary from those under which the test dataset was created; indeed, the geography, weather, and the device used to take the images have a major influence on lighting.
It is worth noting that the distance between the TLR and the traffic light is crucial for recognition. The greater the distance from which the TLR is able to correctly classify the traffic light, the greater its robustness. A threshold has been empirically defined in respect to the size of the traffic light in the image: only traffic lights with at least 20 pixels on their diagonal were taken into account.

Despite all these issues, a precision of 100% and a recall of 90.9% were achieved. Such results, if compared to original’s [15] (precision of 92.2% and recall of 94.6%), validate the TLR and justify its use in future research.

V. CONCLUSION

The main goal of this work was to evaluate whether the inclusion of Expert Instruction (EI) could decrease the amount of data needed to train a ML algorithm in the Traffic Light Recognition (TLR) problem. For this, a PCANet with and without EI was trained. The EI consisted of the frequency of occurrence of semaphores per pixel of a sequence of images. This was chosen based on the driver’s implicit knowledge of what is the usual location of a traffic light in his/her field of view from the inside of a vehicle: central and upper regions of the scene have higher likelihood. Frequency maps were built to confirm such hypothesis.

Results have shown that the use of EI helps to decrease the amount of data for model training for both datasets of [15] and [14], used as baseline. Test results with EI model training performed better than the conservative approach. Even for little data (around 1,000 elements) EI-based approach achieved 75% of accuracy. To evaluate the use of this training in a complete TLR process, an experiment was carried out in which an ROI detection algorithm was applied and then the ROIs were subjected to classification by PCANet using EI and without EI. EI-based TLR achieved 83% of precision and 73% of recall against 75.3% of accuracy and 51.1% of recall of its counterpart.

A TLR device was also deployed. The goal was to show the actual feasibility of such a device for driver assistance. The images obtained with the TLR device were submitted to the TLR algorithm. Precision of 100.0% and recall of 90.9% validate the use of this TLR device layout.

Further research includes fine tunning of parameters and values. One of these parameters is the number of images used to generate the frequency map, which can be expanded to be even more representative. Other important parameters are: the mask used for smoothing the frequency map and the increment size added to every frequency value f[i] for the combination with the algorithm’s input image.

ACKNOWLEDGMENT

The authors thank CAPES and FAPITEC-SE for the financial support [Edital CAPES/FAPITEC/SE No 11/2016 - PROEF, Processo 88887.160994/2017-00]. The authors also thank FAPITEC-SE for granting a graduate scholarship to Thiago Almeida and CNPq for granting a productivity scholarship to Hendrik Macedo [DT-II, Processo 306073/2017-00].

REFERENCES

[1] CET - Companhia de Engenharia de Trânsito, “Análise da mortalidade ocorrida por acidentes de trânsito em agosto e setembro de 2016.” http://www.cetsp.com.br/media/523410/parcial2016.pdf, 2016, acesso em: 09 ago. 2017.
[2] M. B. Jensen, M. P. Philipseun, A. Mogelmose, T. B. Moeulund, and M. M. Trivedi, “Vision for looking at traffic lights: Issues, survey, and perspectives,” IEEE Transactions on Intelligent Transportation Systems, vol. 17, no. 7, pp. 1800–1815, July 2016.
[3] M. Weber, P. Wolf, and J. M. Zöllner, “DeepTrl: A single deep convolutional network for detection and classification of traffic lights,” in 2016 IEEE Intelligent Vehicles Symposium (IV), June 2016, pp. 342–348.
[4] Z. Wang, Z. Bi, C. Wang, L. Lin, and H. Wang, “Traffic lights recognition based on pcnn,” in 2015 Chinese Automation Congress (CAC), Nov 2015, pp. 559–564.
[5] S. Kim, E. Jung, S. Hyun, and D. S. Han, “Real time object detection based on saliency map,” in 2016 IEEE International Conference on Consumer Electronics (ICCE), Jan 2016, pp. 534–535.
[6] T. Almeida, N. Vasconcelos, A. Benica, and H. Macedo, “Fuzzy model applied to the recognition of traffic lights signals,” in 2016 8th Euro American Conference on Telematics and Information Systems (EATIS), April 2016, pp. 1–4.
[7] T. Almeida, N. Vasconcelos, and A. Benica, “Framework para detecção de semáforos baseado em atenção visual,” in Conference on Graphics, Patterns and Images, 28(SIBGRAPI), Salvador, 2015.
[8] Y. Zhou, Z. Chen, and X. Huang, “A system-on-chip fpga design for real-time traffic signal recognition system,” in 2016 IEEE International Symposium on Circuits and Systems (ISCAS), May 2016, pp. 1778–1781.
[9] X. Shi, N. Zhao, and Y. Xia, “Detection and classification of traffic lights for automated setup of road surveillance systems,” Multimedia Tools Appl., vol. 75, no. 20, pp. 12547–12562, 2016. [Online]. Available: [url](http://dx.doi.org/10.1007/s11042-014-2343-1)
[10] M. Michael and M. Schlippsing, “Extending traffic light recognition: Efficient classification of phase and pictogram,” in 2015 International Joint Conference on Neural Networks (IJCNN), July 2015, pp. 1–8.
[11] Y. Ji, M. Yang, Z. Lu, and C. Wang, “Integrating visual selective attention model with hog features for traffic light detection and recognition,” in 2015 IEEE Intelligent Vehicles Symposium (IV), June 2015, pp. 280–285.
[12] J. Balcerew, A. Konieczka, T. Marciniak, A. Dabrowski, K. Mańkowski, and K. Piniarski, “Automatic detection of traffic lights changes from red to green and car turn signals in order to improve urban traffic,” in 2014 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA), Sept 2014, pp. 110–115.
[13] Z. Cai, Y. Li, and M. Gu, “Real-time recognition system of traffic light in urban environment,” in 2012 IEEE Symposium on Computational Intelligence for Security and Defence Applications, July 2012, pp. 1–6.
[14] Z. Chen and X. Huang, “Accurate and reliable detection of traffic lights using multiclass learning and multibobject tracking,” IEEE Intelligent Transportation Systems Magazine, vol. 8, no. 4, pp. 28–42, winter 2016.
[15] Z. Shi, Z. Zou, and C. Zhang, “Real-time traffic light detection with adaptive background suppression filter,” IEEE Transactions on Intelligent Transportation Systems, vol. 17, no. 3, pp. 690–700, March 2016.
[16] S. S. Haykin, S. S. Haykin, S. S. Haykin, and S. S. Haykin, Neural networks and learning machines. Pearson Upper Saddle River, 2009, vol. 3.