A Machine Learning Based Method for Pedestrian and Vehicle Collision Detection

Rodrigo Meira Andrade*, Marcelo Rudek**, Osiris Canciglieri Junior**

* Pontifical Catholic University of Parana - PUCPR - Control and Automation Eng. Dept., Curitiba, Brazil

** Graduate Program in Production Engineering and Systems - PPGEPS, Curitiba, Brazil

***Abstract*—The automatic processes operated by machine learning algorithms that use information extracted from the images require input of data that can be used as parameters for object classification. The preprocessing step for image segmentation is often a non-trivial task, especially for outdoor environments, where acquisition conditions depend on environmental factors whose characteristics cannot be controlled. The paper addresses all necessary steps to segmentation and classification of vehicles and pedestrians to a traffic control system in order to capture vehicles' infractions during pedestrian crossing the streets. A set of images from potential collision situation were detected from real traffic conditions. The text presents the performance evaluation of machine learning based algorithm combined with feature extraction methods applied in the identification of collision situations.**

I. INTRODUCTION

The traffic of pedestrians and vehicles in the most of cities, requires a responsible conduct from both, following a set of rules and protocols determined by the laws. Always, according to traffic legislation, vehicle drivers must give preference to pedestrians and wait for the pedestrian's complete crossing to resume the movement of the vehicle. Many accidents involving pedestrians occur due to factors such as haste, lack of attention, deficient signaling and irresponsible driving among others.

In order to minimize fatal accidents and injuries, commonly it is employed surveillance equipment, (as radars) normally fixed at street corners as a way to force the correct behavior from the vehicle's drivers. The equipment can also contain several types of sensors such as cameras and lasers used to measure the speed of vehicles, and record the infraction through an image to prove the event, including the capture and the vehicle license plates recording. Traffic monitoring environments, can use inductive loop sensors installed on the track to detect the presence of a vehicle through the generated magnetic profile, or alternatively use Doppler Effect sensors that can detect several vehicles simultaneously. Thus, in general, the detection is done by sensors that serve as the trigger to activate the camera to capture the images of the license plate. Thus, the detection of the infraction is carried out by a sensor and only its registration is done by images [1].

The problem is that these sensors only detect vehicles, and this becomes a limitation in situations where the driver does not respect the preference of people crossing in the pedestrian lane. In addition to this problem, other technical and financial issues can be decisive for obtaining authorization for installation on public roads, because they require intervention on the road, generating interruption of passage of vehicles for installation and maintenance, and it is still possible to cause damage to the pavement due to the necessary cuts to the installation of the inductive loops.

The only use of images in these situations is a feasible alternative to be considered, because it is possible to carry out the differentiated identification of two classes of objects, *ie*, vehicles and pedestrians, and it would be possible to link the vehicle tracking with its invasion on pedestrian lanes during the crossing of pedestrian [1-2].

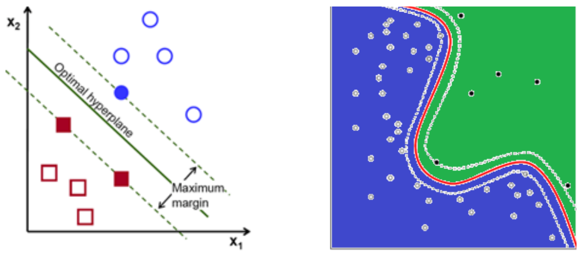
From this presented problem, the aim of the work is to study and to evaluate methods to identify the differentiated displacement of pedestrians and vehicles, and to detect and record a pedestrian invasion infraction only from images. The research questions are related to understand how to apply the computer vision and machine learning algorithms for detection and classification of objects, as well as to define the necessary hardware resources to define the data gathering environment. Thus, the proposal of this research is to make the study and the analysis of the viability of using only cameras to identify infractions of vehicles over pedestrian.

II. BACKGROUND

The problem addressed in this paper is concerned with the selection of the best classification algorithm which solve the problem of identification of the collisions. Thus, an initial step was the literature review about the theme. After filtering a large set of papers, as presented in [1], some of them were selected from at the most cited in correlated researches in order to give the theory support for the proposition of this study. Thus, the selected papers presents the methods' concepts supported by the machine learning context, compound by support vector machines (SVM) applied together with extraction feature tools, such as Histogram Oriented Gradient (HOG) and 3D-Histogram.

A. *support vector machines (SVM)*

The SVM (Support Vector Machine) is one of the main machine learning classifiers used in computer vision to classify images [2-3]. The principle of operation of the method is based on the mapping of the input information of the problem function, creating a dimensional space with k dimensions proportional to the number of independent variables of the problem. By constructing this dimensional plane, this method optimally constructs a hyperplane in space, separating input information into classes. Initially the proposal of this method was to solve binary classification problems, i.e., classifies a given entry into two classes for which it was trained [4]. Currently it has been extended also to be applied to for problems of regressions and multi-class classification [5]. The SVMs for binary classification use the unsupervised training method, that is, the algorithm is trained with only input information, without the need to use output (response) information. During training, the model to be generated will create an optimized hyperplane separating the information between two classes, and from this, all input information will be based on the separation of the hyperplane created during the training. Figure 1 demonstrates the principle of separation of two classes by the characteristics of the objects.



(a) (b)

Figure 1 - Hyperplane examples, (a) Linear and (b) Non-linear, adapted from [1].

For the application in the measurement of traffic of people and vehicles in public roads, the SVM can be used as classifier to separate two classes of very specific objects, such as class "person" and class "vehicles".

Thus, the purpose of the algorithm is to find a separation line that will divide the data by separating them by classes, this line is known as hyperplane. The formation of the hyperplane always seeks to find the greatest distance between two points in relation to two distinct classes.

B. *Histogram Oriented Gradient (HOG)*

The HOG is a feature descriptor used in digital image processing, based on counting the number of occurrences of the gradient orientation of small areas (subdivisions) evenly distributed in the image [6]. The basic information for the HOG are the geometric characteristics of the shape of objects, neglecting information such as colors.

HSV (Hue, Saturation, Value) is a representation of an RGB color system (Red, Green, Blue) in cylindrical coordinates [7-8]. The image in the HSV pattern is represented by a matrix mxnx3, where each H, S and V channel occupies one of the dimensions of the matrix. The first channel (H) is the dominant color or hue, the second channel (S) is the saturation, and the last channel (V) contains the brightness / color intensity information of each pixel in the image [1].

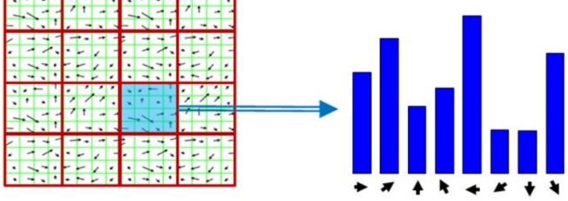


Figure 2 - HOG example. a) Number of configured quadrants over original image. b) Histogram calculation from orientations for the segmented image. Adapted from (LEVICV, 2013).

$$Gx = \begin{bmatrix} -1 & 0 & 1 \end{bmatrix} \times Im \tag{1}$$

$$Gy = \begin{bmatrix} -1 \\ 0 \\ 1 \end{bmatrix} \times Im \tag{2}$$

$$G = \sqrt{Gx^2 + Gy^2} \tag{3}$$

$$\theta = tan^{-1}\frac{Gy}{Gx} \tag{4}$$

Thus the features are extracted from his 3-dimensional histogram. Within this three-dimensional space, each possible position of the histogram contains the number of times that respective color appeared in the image.

Opposite to HOG, the 3D histogram of the HSV format uses only the colors of the object, neglecting its shape. The use of this method in a data set for vehicle or pedestrian recognition tends to have worse results in relation to the HOG, since vehicles and pedestrians have a great variety of colors, and thus more susceptible to noise as the variation of luminosity. What does not happen with HOG since the object maintains its shape. In contrast HSV is faster to be processed compared to HOG.

C. *3D Histogram*

The 3D histogram represents the set of individual histograms from each RGB plane, given by a tri-dimensional matrix, which contains respective axes (X, Y and Z). Through 2 axes (X and Y) is formed a two-dimensional matrix, whose values represent the palette of visible colors [9]. The length of the axes is calculated by the square root of the maximum number of colors used, an RGB color palette based on the 8-bit RGB color has 65536 colors. The Z axis represents the amount of pixels in the image which has that color represented by the Im(x, y) coordinates to some image Im. Those all values are applied with the SVM as extra parameter classification [10].

III. Proposed Method

The proposed method, comprehends the following tasks: image acquirement and enhancement; segmentation & background removing; objects detection; objects classification; objects tracking; infraction detecting; infraction recording and transmission & storage, grouped in three main layers as presented in figure 3.

In the first layer, in the initial step we apply the image smoothing method by convolution [11-14] with a 5x5 size kernel array. An additional division scale factor is also added to the new pixel created, this division causes the image to become darker, causing the effect that darker pixels saturate to 0 value, thus reducing image information.

Following to the second step, the gamma of the image is adjusted to a fixed value, for instance 0.5, keeping this pattern for all frames, thus the small variations of luminosity in the scene are removed, this helps to improve the non-detection of false positives due to fact that during the first stage the image is darkened purposely. By the adjustment of the gamma to 0.5, causes the image to be grayed, and losing image information it become an advantage for background detection, since the variation of brightness and shadows of the scene is largely removed.

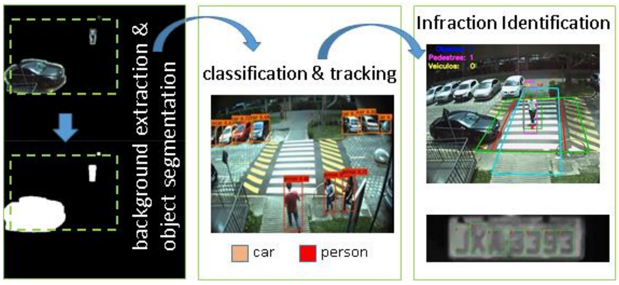


Figure 3 – the three main layers of infraction detection proposition.

The third step on first layer is the background removal. When the background of the scene is removed, consequently the objects that are in motion, are found and have the position detected. During this step, a binary image is generated that distinguishes the background of moving objects (foreground), and then from this binary image it is possible to create the masks for each moving object, by background detection [15].

The second layer is the classification [16] and tracking [17] of the objects. In this step the SVM and HOG algorithms were tested as presented in discussion section.

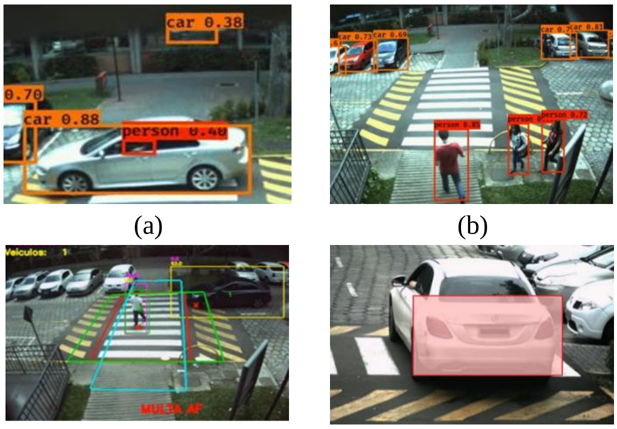
The third layer represents the infraction detection. Despite this step is not covered in this paper, it is fully influenced by classification. Thus, we discuss the combination of SVM, HOG and 3D Histogram in order to evaluate which combination causes a better pre-processing to the next steps.

IV. Results and Discussion

For experimental analysis of the method, the environment for data collecting was arranged with two cameras, A “tracking camera” was positioned with its field of view over the pedestrian lane, and a “registering camera” was positioned over street pointed to vehicles flow. In Figure 4, the images (a) and (b) show the scenery captured by registering camera with its respective objects identification, car and person with its respective assertiveness percentage. The image (c) presents an infringement situation example. In this case a car in yellow bounding box is in movement during a pedestrian crossing. The figure (d) presents the license plate acquisition during a registered infraction.

The validation of the proposed method was performed by infraction simulations, by calculation of percentage accuracy when correctly classifying objects in the image. A video file with several irregular situations among regular situations was recorded and submitted to the algorithm. The simulated situations in this video are:

- The vehicle does not stop near the white strip while there is a pedestrian crossing.
- The vehicle does not stop near the track while there is a pedestrian near the strip waiting for a vehicle to stop to make the safe crossing.
- The vehicle does not completely wait for the pedestrian to cross over the track, causing to advance in the track.
- Vehicle crossing while a pedestrian started crossing.
- The same situations tested during the night and different weather conditions.



(a) (b) (c) (d)

Figure 4 – Simulation of the process operated over real images.

The calculation of the accuracy was done through two data sets. The first set of data contains about 20 thousand images containing vehicles and approximately 30 thousand images containing objects that are not vehicles. The second dataset contains approximately 15,000 pedestrian images and 20,000 non-pedestrian object images, totaling approximately 35,000 images. The resulting values are shown in the table 1 whose values demonstrate the precision of combination between SVM and HOG versus SVM and 3D histogram.

Table 1. Algorithm Performance.

	Vehicle		Pedestrian	
	accuracy	velocity	accuracy	velocity
SVM + HOG	83.4%	18ms	85.8%	18ms
SVM+3D histogram	87.1%	8ms	86.9%	8ms

According table 1, the performance of SVM with 3D histogram is a little bit upper from SVM+HOG in terms of accuracy. The most important gain is in velocity, that it is a necessary condition to real time applications. The tests show that when using the SVM classifier with the 3D histogram, this classifier becomes quite susceptible to the overfitting process, where a small change in the camera configuration as for instance the brightness. Thus, it demonstrates an example of some difficulty in maintaining its performance during variation of ambient illumination.

The research demonstrates that is feasible to operate a system with a machine learning based classifier in order to classify vehicles and pedestrians in own respective classes. Thus, a separable tracking can analyses the collision situation between vehicles and pedestrians.

ACKNOWLEDGMENT

The authors would like to thank the Pontifical Catholic University of Parana by the financial support.

REFERENCES

[1] R. M. Andrade, Proposal for a Method Based on Computational Vision for Identification and Registration of Vehicle's Advance on the Pedestrian Crossing Line, Master Thesis (*in portuguese*), Department of Mechatronics, PUCPR, Brazil, 2018.

[2] Tatsuya Hanyu; Qiangfu Zhao.Incremental Training of SVM-Based Human Detector.IEEE 11th International Symposium on Embedded Multicore/Many-core Systems-on-Chip (MCSoC). p. 181 – 185, 2017.

[3] Keiichi Shirasuka. Mosaic block detection based on HOG with SVM classifier and template matching. IEEE International Conference on Consumer Electronics (ICCE). 2018 p: 1 – 2.

[4] S. Bougharriou; F. Hamdaoui; A. Mtibaa. Linear SVM classifier based HOG car detection. 18th International Conference on Sciences and Techniques of Automatic Control and Computer Engineering (STA). 241 – 245, 2017.

[5] WEISEN, V. Deep Learning Methods for Object Detection @ vincentweisen.wordpress.com. 2016.

[6] DALAL, N.; TRIGGS, B. Histograms of Oriented Gradients for Human Detection. IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2010.

[7] SONKA, Milan; HLAVAC, Vaclav; BOYLE, Roger. Image processing, analysis, and machine vision. Cengage Learning, 2014.

[8] IMPIOMBATO, D.; et al. Unified, Real-Time Object Detection. Computer Vision and Pattern Recognition, v. 794, p. 185–192, 2015.

[9] Baochang Zhang; Yun Yang; Chen Chen; Linlin Yang; Jungong Han; Ling Shao. Action Recognition Using 3D Histograms of Texture and A Multi-Class Boosting Classifier. IEEE Transactions on Image Processing, Volume: 26, Issue: 10; p. 4648 – 4660; 2017.

[10] YANG, J.; ZHONG, W.; MIAO, Z. On the Image enhancement histogram processing. In: 2016 3rd International Conference on Informative and Cybernetics for Computational Social Systems (ICCSS), 2016.

[11] SZEGEDY, C.; LIU, W.; JIA, Y.; SERMANET, P.; REED, S.; ANGUELOV, D.; ERHAN, D.; VANHOUCKE, V.; RABINOVICH, A. Going deeper with convolutions. Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, v. 07–12–June, p. 1–9, 2015

[12] KAIMING HE GEORGIA GKIOXARI PIOTR DOLL´ AR ROSS GIRSHICK. Mask R-CNN. Computer Vision and Pattern Recognition, 2017.

[13] KRIZHEVSKY, A.; HINTON, G. E. ImageNet Classification with Deep Convolutional Neural Networks. NIPS'12 Proceedings of the 25th International Conference on Neural Information Processing Systems, p. 1097-, 2012.

[14] WU, J. Introduction to Convolutional Neural Networks, 2017.

[15] BENEZETH, Y.; JODOIN, P.-M.; EMILE, B.; ENE LAURENT, H.; ROSEN BERGER, C. Comparative study of background subtraction algorithms. Journal of Electronic Imaging, v. 19, 2010.

[16] UIJLINGS, J. R. R.; VAN DE SANDE, K. E. A.; GEVERS, T.; SMEULDERS, A. W. M. Selective Search for Object Recognition. International Journal of Computer Vision, 2012.

[17] BOCHEM, A.; KENT, K.; HERPERS, R. FPGA Based Real-Time Tracking Approach with Validation of Precision and Performance, 2011.