Vehicle Speed Estimation by License Plate Detection and Tracking

Diogo Carbonera Luvizon [1],
Bogdan T. Nassu (*Co-advisor*), Rodrigo Minetto (*Advisor*)
Department of Informatics
Federal University of Technology - Paraná (UTFPR) - Curitiba, Brazil
Email: diogo@luvizon.com, {nassu, rminetto}@dainf.ct.utfpr.edu.br

***Abstract*—We propose a non-intrusive methodology, based on digital videos, for measuring vehicle speeds by detecting and tracking license plates — a technology that does not exist in Brazil. We proposed novel algorithms for detecting regions containing moving vehicles, as well as their license plates. Each vehicle's speed is then estimated by tracking the license plates and applying geometric transformations. Experiments were performed on videos containing 8,000 vehicles, associated to speeds measured by a precise inductive loop detector. The average speed measurement error was -0.5 km/h, staying inside the +2/-3 km/h margin determined by international regulatory agencies, in more than 96% of the cases.**

***Keywords*-vehicle speed; speed measurement; license plate detection; feature tracking;**

I. Introduction and Motivation

Systems for vehicle detection and speed measurement play an important role in enforcing speed limits. They also provide relevant data for traffic control. Those systems are divided in intrusive and non-intrusive [1]. Intrusive sensors, usually based on inductive loop detectors, are widely used, but have complex installation and maintenance, accelerate asphalt deterioration, and can be damaged by wear and tear. Non-intrusive sensors, which include laser meters and Doppler radars, avoid these problems, but are usually more expensive and require frequent maintenance. As digital cameras become cheaper and able to produce images with higher quality, video-based systems can become a lower cost alternative for non-intrusive speed measurement. In fact, existing systems are often connected to video cameras [2] that record the license plates of vehicles that exceed the speed limit — thus, the infrastructure for such systems is already available in most cases.

In the M.Sc. thesis presented in [3], briefly summarized in this paper, we describe the pipeline for a non-intrusive video-based system for vehicle speed measurement in urban roadways. Our goal is measuring vehicle speeds with accuracy comparable to that obtained by a system based on inductive loop detectors. The input video is captured by a single fixed overhead camera, positioned so that the rear license plate of vehicles in three adjacent lanes are clearly visible, as shown in Fig. 1. A sample image from this setup is shown in Fig. 2. This setup allows the same images to be used for both speed measurement and license plate identification (e.g. for locating stolen vehicles, or in the case of a speed limit violation).

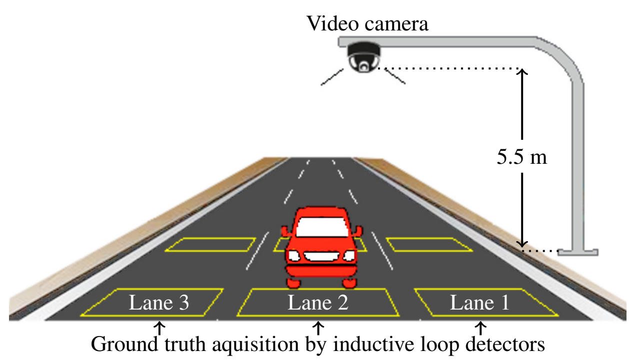

Fig. 1. System setup.

Fig. 2. Sample image captured by our system.

We make some assumptions about the scene and the problem domain: video frames are equally spaced in time; each lane lies on a plane; the vehicles move at a constant speed and with a straight trajectory from the lower to the upper part of the image; and the license plates are at approximately the same distance from the ground. These assumptions allow us to measure vehicle speeds without modeling the 3-D space, or requiring precise camera calibration or positioning.

The proposed system works by tracking sets of distinctive features extracted from image regions around each vehicle's license plate, and is divided into five main parts, as shown in Fig. 3. Initially, an optimized motion detection algorithm identifies image regions containing moving vehicles. These

[1]This work summarizes the M.Sc. thesis of the first author.

regions are fed to a novel license plate detector, which returns a set of axis-aligned rectangular sub-images around the vehicles' license plates. Features are then extracted from each sub-image [4], and tracked using the Kanade-Lucas-Tomasi (KLT) algorithm [5]. To cope with large displacements, from vehicles moving at high speeds, an initial motion estimation is performed by matching features extracted by the Scale-Invariant Feature Transform [6]. Finally, vehicle speeds are measured by comparing the trajectories of the tracked features to known real world measures.

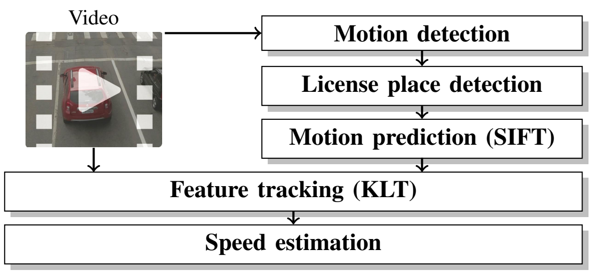

Fig. 3. Overview of the proposed system.

A proof-of-concept of our system was evaluated on approximately five hours of videos in different weather and recording conditions. The videos have an associated ground truth dataset containing vehicle speeds measured by a high precision system based on inductive loop detectors, properly calibrated and approved by the Brazilian national metrology agency (Inmetro). This data set is itself a contribution of our work, and can be freely obtained for research purposes[1]. Our system was able to measure speeds with an average error of -0.5 km/h, staying inside the [-3,+2] km/h limit determined by regulatory authorities in several countries, in over 96.0% of the cases. We also show that our license plate detector outperforms other two published state-of-the-art text detectors, as well as a well-known license plate detector. A preliminary version of our system was published at the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP – Qualis A1) [7].

The main contributions of this work are:

1) A novel video-based system for vehicles speed estimation with precision comparable to intrusive systems.
2) An optimized algorithm for detecting moving vehicles.
3) A new text-based licence plate detector.
4) A combination of the KLT feature tracker and the initial motion prediction by SIFT descriptors.
5) Application of a homograph transformation to convert pixels per frame to meters per second.
6) A new dataset containing more than five hours of videos in high quality, captured in different weather conditions, associated with the manual selection of all license plates, as well as the ground truth speed of vehicles measured by high precision inductive loop sensors, calibrated and certified by the Brazilian metrology agency.

[1]The full dataset will be made available at the time of publication of our article submitted to ITS. A sample is available at www.dainf.ct.utfpr.edu.br/%7erminetto/projects/vehicle-speed.

The rest of this paper is divided as follows. In Section II, we present the proposed method, divided in motion detection, license plate detection, feature selection and tracking, and speed measurement methods. Experimental evaluation and results are shown in Section III. Finally, in Section IV we state the conclusions of our work.

II. Proposed Method

A. Motion detection

The first step in our system's pipeline is detecting moving vehicles, limiting further processing to a set of regions of interest. Ideally, each region of interest must contain the entire license plate from a single vehicle. An overview of the motion detection approach that we develop is shown in Fig. 4.

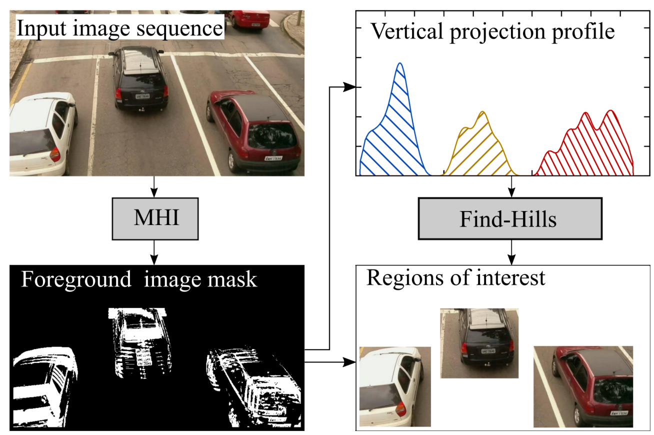

Fig. 4. Motion detection: regions of interest are delimited based on a foreground image mask and a vertical projection profile.

Motion detection begins with a rough foreground / background segmentation. We use the *Motion History Image* (MHI) concept from Bobick and Davis [8]. An example is shown in Fig. 4. To reduce the processing time when generating the MHI, the images are subsampled — i.e. the values of x and y in the input image are restricted to a regular sparse grid, with pixels outside the grid being skipped. The segmentation may become less precise, but this is acceptable as long as the vehicle license plate remains entirely within its corresponding region. In tests involving the complete system, we observed that a subsampling factor of 4 in both axes (i.e. processing 1 of each 16 pixels) reduced the processing time to 16.47% of the original time, without any loss in detection performance. The segmentation mask is generated by setting all non zeros pixels from the MHI.

After the sub-sampled binary segmentation mask is obtained, we perform a vertical *projection profile* analysis to separate vehicles horizontally. We take the lower part of the segmentation mask, which shows the region closer to the camera, and count the foreground pixels in each column, generating a histogram with n bins (for n image columns). This histogram is smoothed, to reduce noise, and interpreted as an array Ψ. Figure 4 shows an example of vertical projection profile. It can be seen that the interval containing a vehicle is delimited by an ascending and a descending slope, corresponding respectively to the left and right boundaries of the vehicle. Due to lack of space, the algorithms used to determine these boundaries for each vehicle is described in details in the Master's thesis [3], although the method is exemplified in Fig. 5. To guarantee a vehicle's license plate is inside its corresponding region of interest, we discard regions with lower boundaries close to the image bottom — these cases may correspond to a vehicle that is entering the frame, so its license plate is not visible yet.

By pairing the ascending and descending boundaries from arrays A and D, we can determine the left and right boudaries of each region of interest, as exemplified in Fig. 5(c). We assume each region of interest corresponds to a vehicle in the image. Upper and lower boundaries for each region are obtained directly from the binary segmentation mask.

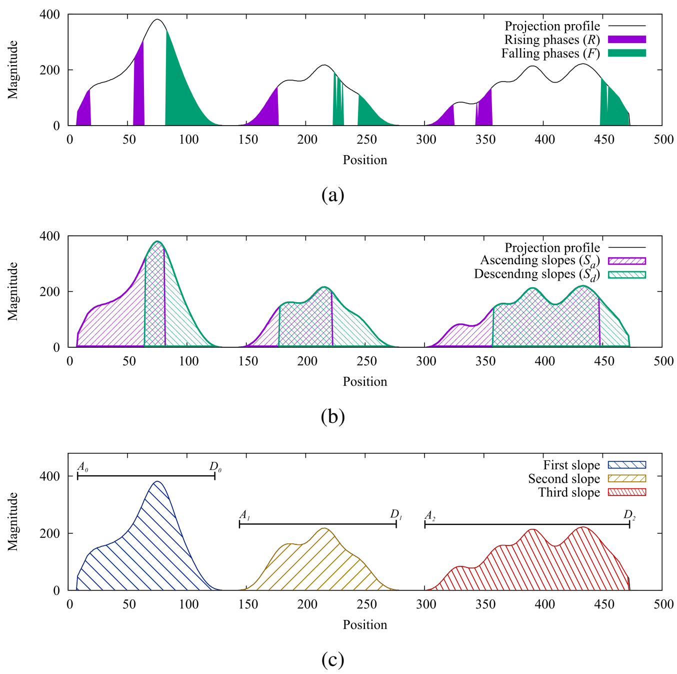

Fig. 5. Representation of the algorithms described in [3] to find the vehicle's boundaries. Rising and falling phases (a), ascending and descending slopes (b), and three slope regions delimited by the rising edge of ascending slopes and the falling edge of descending slopes (c).

B. License plate detection

The motion detector produces one region of interest for each moving vehicle present in the scene at a given time. The license plate detector finds, for each region of interest, an axis-aligned rectangle, which is an approximate bounding box of the vehicle's license plate region. This procedure is performed for each region of interest only until a license plate is detected — afterwards, features extracted from the license plate region are tracked across frames, as explained in Section II-C.

Our detector follows the *hypothesis generation and validation paradigm* [9]. Namely, in the hypothesis generation phase we use *edge extraction*, *edge filtering*, and *region grouping* modules to provide coarse candidate regions based on the edge attribute that makes up the license plate. At this phase, we aim to isolate the license plate region and prevent false negatives, even at the cost of several false positives. In the *hypothesis validation* phase, we use a *region classification* module to refine the candidates. For this classification we use the *Text HOG* (T-HOG) descriptor [10], which is based on the observation that the license plate textual information can often be characterized by the distribution of the directions of the image gradients. An overview of the license plate detector is outlined in Fig. 6.

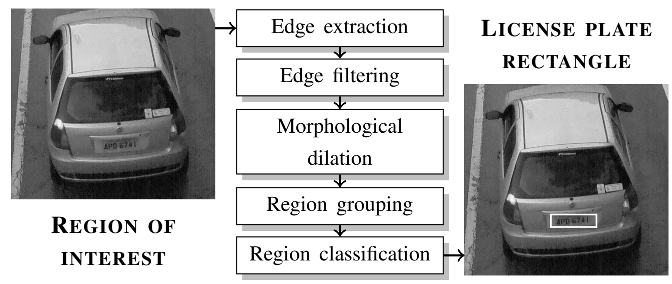

Fig. 6. License plate detection scheme

C. Feature Selection and Tracking

Once a license plate region is detected, our system selects a set of distinctive features and tracks it across multiple video frames. Our aim is producing a list containing the trajectories of the tracked features. The structure of our tracking scheme is outlined in Fig. 7. Feature selection is performed only once for each vehicle, immediately after its license plate is detected. Following the approach from Shi and Tomasi [4], a "*good feature*" is a region with high intensity variation in more than one direction, such as textured regions or corners.

The traditional Lucas-Kanade algorithm only works for small displacements (in the order of one pixel). To overcome this limitation, we consider the pyramidal version of KLT, described by Bouguet [11]. The algorithm builds, for each frame, a multi-scale image pyramid by using the original image at the pyramid base, and putting at each subsequent level a version of the image in the previous level with width and height reduced by half. The pyramidal KLT algorithm starts by finding the displacement vector $\vec{d}$ at the last pyramid level, using the result as the initial estimate for $\vec{d}$ in the next level, repeating the process until the pyramid base (i.e. the original image) is reached.

The number of levels ℓ in the pyramid determines the maximum allowed displacement for a feature in pixels, given by $2^{\ell+1}-1$ (*e.g.* for $\ell = 3$, the maximum displacement is of 15

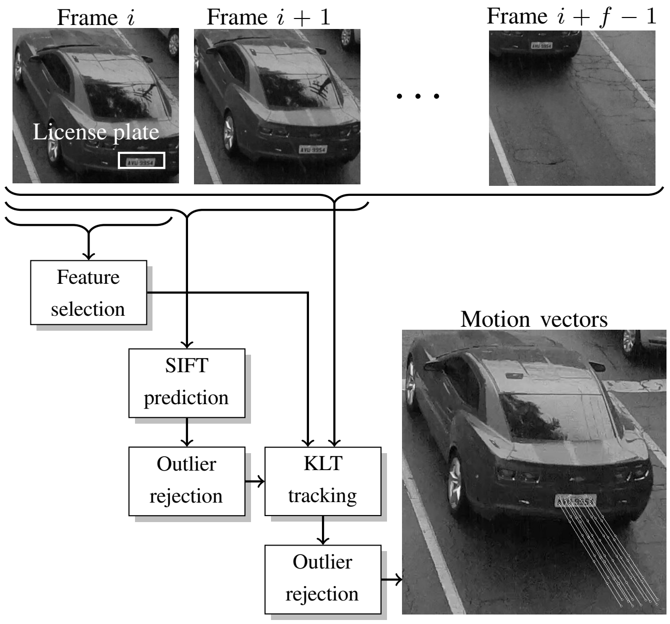

Fig. 7. Overview of the proposed feature tracking method.

pixels). Larger values of ℓ may allow for larger displacements, but they may also produce very small images at the upper pyramid levels, which may lead to confusion and large errors in the initial estimates. Another limitation of the algorithm is that for the first frame in a sequence (i.e. the moment a license plate is detected), there is no known motion. In these cases, it is common to start with zero. If there are large feature displacements, e.g. from a vehicle moving at a high speed, this initial estimate will be too far from the correct displacement, preventing the feature from being properly tracked.

To overcome the limitations described above, an initial value for the displacement is estimated by using a different method for matching features extracted from two sequential images. We have used the SIFT (Scale-Invariant Feature Transform) features proposed by Lowe [6]. SIFT is a popular method for detecting and describing image keypoints, being robust to illumination variations and a number of geometric distortions. Using the standard parameters proposed by Lowe [6], SIFT features are extracted from an expanded window around the license plate region (frame i), and matched to other features extracted from frame $i+1$, using the *nearest neighbor distance ratio* matching strategy described by Mikolajczyk [12]. The average feature displacement is then used as the initial value for the motion vectors. Note that this process occurs only once for each detected license plate — after the motion is roughly predicted, the system relies on the KLT algorithm, which allows for faster and more accurate estimates.

D. Speed Measurement

Our system measures vehicle speeds based on the motion vectors obtained by the feature selection and tracking method. Each motion vector can be associated with a measurement of a vehicle's instantaneous speed at a particular time, given in pixels per frame, in the image plane. The purpose of the speed measurement module is converting these measurements to kilometers per hour (km/h) in the real world.

An important assumption of our system is that each street lane lies on a plane. That assumption makes it possible for us to map the motion vectors $\vec{d_i}$, which are given in pixels in the image plane, to displacements $\vec{v}_i$, given in meters in the ground plane (see Fig. 8).

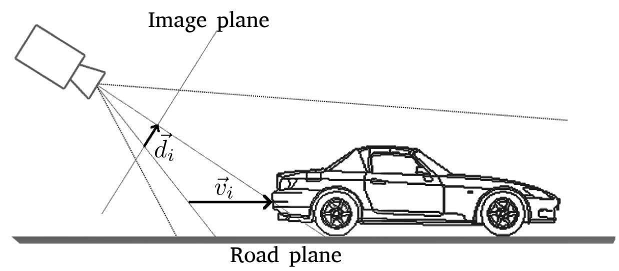

Fig. 8. Vehicle speed measurement scheme.

The complete description of the homography process and the method to convert the displacement vectors computed by the KLT tracking algorithm to speed in km/h is shown in the M.Sc. thesis [3], due to lack of space in this paper.

III. Experimental Evaluation

A proof-of-concept system was built for evaluating the proposed approach. Besides the cameras and physical infrastructure, we used a 2.2 GHz Intel Core i7 machine with 8 GB of RAM running Linux, with the algorithms implemented in C++. In the next sections we briefly describe our dataset, and evaluate our system's performance regarding motion detection, license plate detection, and speed measurement.

A. Dataset

Our dataset, summarized in Table I, contains 20 videos captured by a low-cost 5-megapixel CMOS image sensor, with frame resolution of 1920×1080 pixels, at 30.15 frames per second. The videos are divided in 5 sets according to weather and recording conditions. Each video has an associated *ground truth file*, in a simple XML format, containing bounding boxes for the first license plate occurrence of each vehicle, as well as each vehicle's actual speed. The ground truth for the license plates was obtained by human inspection. The ground truth speeds were obtained from a high precision speed meter based on inductive loop detector, properly calibrated and approved by the Brazilian national metrology agency (Inmetro). Note that the videos contain some vehicles with no visible license plate, and that the ground truth speed meter sometimes fails to properly assign a speed to a vehicle. The "No. valid" column in Table I indicates the number of vehicles which have both a visible license plate and an assigned speed.

TABLE I
DATASET INFORMATION: TIME (MINUTES); NUMBER OF VIDEOS; NUMBER OF VEHICLES WITH PLATES AND SPEED INFORMATION. THE QUALITY OPTIONS ARE: [H] HIGH-QUALITY, [N] FRAMES AFFECTED BY NATURAL OR ARTIFICIAL NOISE, [L] FRAMES AFFECTED BY SEVERE LIGHTING CONDITIONS, [B] MOTION BLUR, AND [R] RAIN.

Set	Time	No. videos	No. vehicles	No. plates	No. speed	No. valid	Notes
01	34	4	1,146	1,128	1,033	1,019	[H]
02	169	11	4,829	4,713	4,345	4,241	[L]
03	26	2	960	936	876	855	[N]
04	41	2	1,045	1,034	928	917	[N,R]
05	20	1	869	800	795	734	[L,B]
Tot.	291	20	8,849	8,611	7,977	7,766	

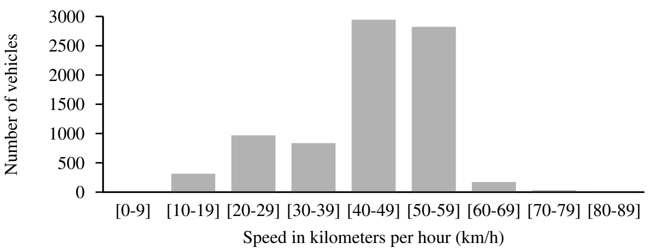

Fig. 9. Vehicles speed distribution.

Figure 9 shows how the vehicle speeds are distributed (the speed limit in this particular roadway is 60 km/h).

The whole dataset used in our experiments will be made available for research purposes, and can be itself considered one of the major contributions of our work.

B. Motion detection evaluation

To evaluate the results from the motion detector, we compare the obtained regions of interest (ROIs) with the license plates in the ground truth. Ideally, all the detected ROIs will contain a license plate, and all license plates will be fully contained within a detected ROI. Objectively, we compute precision and recall metrics as described by Wolf *et al.* [13], with the precision p being given by the proportion of ROIs which contain at least one license plate, and the recall r being given by the proportion of license plates that were inside a ROI. Table II shows the precision and recall, as well as the average processing time, obtained by our motion detector. The different columns show the results obtained with different amounts of subsampling — more sparse grids will reduce the processing time, but can also lead to incorrect results.

C. License plate detection evaluation

To evaluate the performance of the license plate detector, we compare the detected license plates with those in the ground truth. The comparison is based on precision and recall metrics. For ranking purposes, we also consider the F*-measure*, which is the harmonic mean of precision and recall.

We compared our license plate detector with three text and license plate detectors described in the literature (see Chapter 2

TABLE II
MOTION DETECTION PERFORMANCE: THE PRECISION p, RECALL r, AND AVERAGE TIME (IN MS, FOR EACH FRAME) FOR FIVE SUBSAMPLING CONFIGURATIONS AND TWO OVERLAPPING THRESHOLDS.

Subsampling	1×1		2×2		4×4		8×8		32×32	
	p	r	p	r	p	r	p	r	p	r
$\lambda = 1.0$	0.86	0.99	0.86	0.99	0.86	0.99	0.81	0.99	0.58	0.99
$\lambda = 0.5$	0.87	0.99	0.88	0.99	0.87	1.00	0.84	1.00	0.65	1.00
Avg. time (ms)	21.50		7.97		3.54		1.36		0.20	

of [3]): SnooperText [9], the Zheng *et al.* [14] algorithm, and the Stroke Width Transform (SWT) [15]. The results obtained for the entire data set are shown in Table III, divided in 5 subsets according to weather and recording conditions. Our detector significantly outperformed the other approaches in these tests. The average time to process each region of interest was 58 ms for SnooperText; 918 ms for Zheng *et al.*; 402 ms for SWT; and 195 ms for our detector.

TABLE III
LICENSE PLATE DETECTION PERFORMANCE EVALUATION, BASED ON PRECISION (p), RECALL (r), AND THE F-MEASURE. THE BOLDFACE VALUES ARE THE MAXIMA OBTAINED FOR EACH CASE.

	PROPOSED			SNOOPERTEXT			ZHENG *et al.*			SWT		
Set	p	r	F	p	r	F	p	r	F	p	r	F
01	**0.96**	**0.94**	**0.95**	0.81	0.88	0.84	0.92	0.88	0.90	0.76	0.61	0.68
02	**0.92**	**0.84**	**0.88**	0.86	0.81	0.83	0.45	0.29	0.35	0.28	0.23	0.25
03	**0.94**	**0.94**	**0.94**	0.56	0.79	0.66	0.87	0.90	0.88	0.66	0.62	0.64
04	**0.94**	**0.92**	**0.93**	0.44	0.71	0.54	0.91	0.88	0.89	0.79	0.58	0.67
05	**0.88**	**0.82**	**0.85**	0.76	0.72	0.74	0.48	0.48	0.48	0.15	0.15	0.15
Tot.	**0.93**	**0.87**	**0.90**	0.73	0.80	0.76	0.65	0.52	0.58	0.44	0.37	0.40

D. Vehicle speed measurement evaluation

Speed measurement performance was evaluated by comparing the speeds measured by our system with the ground truth speeds obtained by the inductive loop detectors. According to the standards adopted in the USA, an acceptable measurement must be within the [-3 km/h, $+2$ km/h] error interval. Examples of estimated and real speed are shown in Fig. 10.

The first row in Table IV shows the results obtained by our system. Percentages are given regarding the valid vehicles — those with both a license plate and an associated speed in the ground truth — and are divided in 3 classes, depending on whether the measured speed was below, inside, or above the acceptable error interval. The maximum nominal error values for the whole dataset were -4.68 km/h and $+6.00$ km/h, with an average of -0.5 km/h a standard deviation of 1.36 km/h.

We performed testes using an "ideal license plate detector", taking as references the manually annotated license plates from the ground truth, instead of the detected license plates. As shown in the third row of Table IV, the performance in this case was not much different from the performance obtained by our system, indicating that the tracking can be done even with

Estimated: 45.5 km/h | E.: 43.1 km/h | E.: 54.7 km/h | E.: 48.2 km/h | E.: 39.9 km/h

Real: 48.0 km/h | R.: 43.0 km/h | R.: 54.6 km/h | R.: 46.3 km/h | R.: 41.2 km/h

Fig. 10. Examples of vehicle speeds measured by our system and by a high precision meter based on inductive loops

TABLE IV
SPEED MEASUREMENT RESULTS OBTAINED BY OUR SYSTEM AND OTHER APPROACHES: "LOWER", "IDEAL", AND "HIGHER" REPRESENT, RESPECTIVELY, SPEED ERRORS BELOW, ABOVE AND WITHIN THE ACCEPTABLE LIMITS, CONSIDERING THE USA STANDARD [$-3/+2$ KM/H].

	Lower	Ideal	Higher
PROPOSED SYSTEM	1.1%	96.0%	2.8%
IDEAL DETECTOR	0.9%	96.1%	3.0%
FREE FEATURE SELECTION	11.3%	73.4%	15.3%
PARTICLE FILTER	20.3%	22.1%	57.6%

poor license plate regions, hard to be identified even by an human observer. We also verified the "free feature selection", that is, we extracted features from the whole vehicle region. These features are less distinctive, leading to tracking errors. Finally, we compared our system with a blob-based tracker. In this experiment, we used a particle filter algorithm to track the regions of interest found by our motion detection module.

IV. CONCLUSION

This M.Sc. thesis proposed a new non-intrusive video-based system for vehicle speed estimation, which could be used as a low-cost speed enforcement system alternative. The proposed method was evaluated on a new dataset, with almost five hours of full-HD videos, containing more than 8,000 vehicles with associated ground truth speeds obtained by a high precision system, as well as manually labeled ground truth license plate regions. Our license plate detector achieved a precision of 0.93 and a recall of 0.87, outperforming other well-know approaches. In our experiments, our system's error was -0.5 km/h, staying inside the +2/-3 km/h margin in more than 96% of the cases.

PUBLICATIONS AND PATENTS

The scientific publications and patents resulting from this work are presented as follows:

- A conference paper published at the 2014 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) [7], **Qualis A1**.
- An article submitted to the IEEE Transactions on Intelligent Transportation Systems (ITS), **JCR 2.2**, currently with minor changes in the revision process.
- A patent filed at the Brazilian National Institute of Industrial Property (INPI) in 2015, identified as "BR 1020150319150".

ACKNOWLEDGMENT

This work was partially supported by the Brazilian Agency CNPq - Grant 444789/2014-6.

REFERENCES

[1] T. V. Mathew, "Intrusive and non-intrusive technologies," Indian Institute of Technology Bombay, Tech. Rep., 2014.

[2] N. Buch, S. Velastin, and J. Orwell, "A review of computer vision techniques for the analysis of urban traffic," *IEEE Trans. on Intelligent Transportation Systems (ITS)*, vol. 12, no. 3, pp. 920–939, 2011.

[3] D. Luvizon, "Vehicle speed estimation by license plate detection and tracking," Master's thesis, PPGCA - UTFPR, Aug 2015, http://repositorio.utfpr.edu.br/jspui/handle/1/1380.

[4] J. Shi and C. Tomasi, "Good features to track," in *IEEE Int. Conf. on Computer Vision and Pattern Recognition (CVPR)*, 1994, pp. 593–600.

[5] B. D. Lucas and T. Kanade, "An Iterative Image Registration Technique with an Application to Stereo Vision," *Joint Conference on Artificial Intelligence*, pp. 674–679, 1981.

[6] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *Int. Journal of Computer Vision (IJCV)*, vol. 60, no. 2, pp. 91–110, 2004.

[7] D. Luvizon, B. Nassu, and R. Minetto, "Vehicle speed estimation by license plate detection and tracking," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2014, pp. 6563–6567.

[8] A. Bobick and J. Davis, "The recognition of human movement using temporal templates," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 23, no. 3, pp. 257–267, Mar 2001.

[9] R. Minetto, N. Thome, M. Cord, N. J. Leite, and J. Stolfi, "SnooperText: A text detection system for automatic indexing of urban scenes," *Computer Vision and Image Understanding (CVIU) - Elsevier*, vol. 122, pp. 92–104, 2014.

[10] R. Minetto, N. Thome, M. Cord, J. Stolfi, and N. J. Leite, "T-HOG: An effective gradient-based descriptor for single line text regions," *Pattern Recognition (PR), Elsevier,*, vol. 46, no. 3, pp. 1078–1090, 2013.

[11] J.-Y. Bouguet, "Pyramidal Implementation of the Lucas Kanade Feature Tracker," *Intel Corporation, Microprocessor Research Labs*, 2000.

[12] K. Mikolajczyk and C. Schmid, "A Performance Evaluation of Local Descriptors," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 27, no. 10, pp. 1615–1630, 2005.

[13] C. Wolf and J.-M. Jolion, "Object count/area graphs for the evaluation of object detection and segmentation algorithms," *Int. Journal on Document Analysis and Recognition (IJAR)*, vol. 8, no. 4, pp. 280–296, 2006.

[14] D. Zheng, Y. Zhao, and J. Wang, "An efficient method of license plate location," *Pattern Recognition Letters (PRL) - Elsevier*, vol. 26, no. 15, pp. 2431–2438, 2005.

[15] B. Epshtein, E. Ofek, and Y. Wexler, "Detecting text in natural scenes with stroke width transform," in *IEEE Int. Conf on Computer Vision and Pattern Recognition (CVPR)*, 2010, pp. 886–893.