Pattern Recognition 48 (2015) 1039–1049

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/pr

Traffic sign detection via interest region extraction

Samuele Salti[*], Alioscia Petrelli, Federico Tombari, Nicola Fioraio, Luigi Di Stefano

Department of Computer Science and Engineering, University of Bologna, Italy

ARTICLE INFO

Article history:
Received 27 December 2013
Received in revised form
29 May 2014
Accepted 30 May 2014
Available online 9 June 2014

Keywords:
Mobile mapping
Traffic sign detection
Local features

ABSTRACT

Mobile mapping systems acquire massive amount of data under uncontrolled conditions and pose new challenges to the development of robust computer vision algorithms. In this work, we show how a combination of solid image analysis and pattern recognition techniques can be used to tackle the problem of traffic sign detection in mobile mapping data. Different from the majority of existing systems, our pipeline is based on interest regions extraction rather than sliding window detection. Thanks to the robustness of local features, the proposed pipeline can withstand great appearance variations, which typically occur in outdoor data, especially dramatic illumination and scale changes. The proposed approach has been specialized and tested in three variants, each aimed at detecting one of the three categories of *mandatory*, *prohibitory* and *danger* traffic signs, according to the experimental setup of the recent German Traffic Sign Detection Benchmark competition. Besides achieving very good performance in the on-line competition, our proposal has been successfully evaluated on a novel, more challenging dataset of Italian signs, thereby proving its robustness and suitability to automatic analysis of real-world mobile mapping data.

© 2014 Elsevier Ltd. All rights reserved.

1. Introduction

Mobile mapping systems gain increasing interest as cost-effective acquisition tools to collect registered geo-referenced data, which in turn enables new applications, such as location-aware mobile applications, emergency response planning, self-driving cars (e.g., the famous Google car, which locates itself thanks to pre-acquired 2D/3D maps of the route), road mapping and facility management. Mobile mapping systems consist of an integrated array of synchronized inertial, positioning and imaging sensors mounted on a mobile platform. Thanks to the presence of GPS and inertial information, collected data are easily registered without complex and costly post-processing operations.

Among the automatic analyses of mobile mapping data, Traffic Sign Recognition (TSR) plays a key role as it is one of the enabling technologies in several of the above-mentioned applications. As such, TSR has been an actively developed area of research during the last years. Although traffic signs present a rigid and simple shape as well as uniform and known colors, the wide appearance variability of traffic signs captured in uncontrolled environments results in challenging working conditions for computer vision algorithms.

Traffic Sign Recognition is usually tackled via a two-step approach: traffic sign detection and traffic sign classification. In detection, the aim is to identify the image regions (bounding boxes) that tightly contain a traffic sign, indeed a similar computer vision problem such as detecting faces [1] or pedestrians [2]. Classification aims at labeling bounding boxes according to the enclosed traffic signs, similarly to the image classification problem.

So far, evaluation on public benchmarks of traffic sign recognition algorithms has focused more on classification than on detection. There are several public datasets to assess the pros and cons of classification algorithms. In particular, the German Traffic Sign Recognition Benchmark (GTSRB) [3] provides a huge dataset and an interesting evaluation of the performance of computer vision algorithms versus the human visual system. Conversely, evaluation of traffic sign detection on publicly available datasets is less explored. Recently, though, the organizers of the GTSRB proposed a follow-up competition targeting traffic sign detection, the German Traffic Sign Detection Benchmark (GTSDB) [4]. The competition addresses detection of three categories of signs: prohibitory (circular red-and-white signs), mandatory (circular blue-and-white signs) and danger signs (upper-triangular red-and-white signs). This paper describes an extension of the traffic sign detection pipeline that we developed and submitted for evaluation to the on–line GTSDB competition [5]. In particular, we present a more detailed description of the ideas and algorithms building up our pipeline, an automatic process to estimate the interest regions rescaling factors, the combination of more HOGs' sources to improve its discriminative power, a thorough analysis on some of the key pipeline steps that provides useful insights into possible design choices, improved figures on the GTSDB dataset as well as results on a novel, more challenging mobile mapping dataset.

[*] Correspondence to: Viale Risorgimento, 2 40135 Rologna – Italy.
Tel.: +39 051 2093545.
E-mail address: samuele.salti@unibo.it (S. Salti).

http://dx.doi.org/10.1016/j.patcog.2014.05.017
0031-3203/© 2014 Elsevier Ltd. All rights reserved.

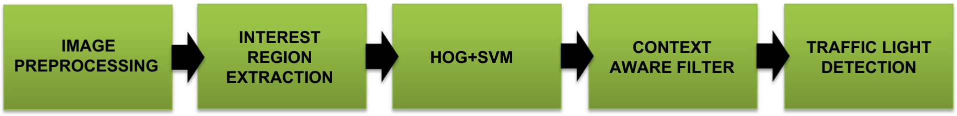

Fig. 1. Proposed traffic sign detection pipeline.

The remainder of the paper is organized as follows: Section 2 presents an overview of recent work on the subject; Section 3 details all the steps of the proposed detection pipeline, i.e. image preprocessing (Section 3.1), interest region detection (Section 3.2), HOGs classification (Section 3.3), context-aware filter (Section 3.4) and traffic light removal (Section 3.5). Section 4 discusses the insights gained by tuning our pipeline on the GTSDB training set (Section 4.1) and provides the experimental results attained on the test sets of three categories of the benchmark (Section 4.2) as well as on a novel mobile mapping dataset (Section 4.3). Section 5 concludes the paper.

2. Related work

Traffic Sign Detection (TSD) usually relies upon two key observations: signs have a well-defined shape and uniform and distinctive colors. The approaches to TSD can then be categorized as geometry-based, segmentation-based or hybrid according to the cue or cues they try to exploit. Geometry-based algorithms employ either Haar-like features in frameworks inspired by the popular Viola–Jones detector [1] or the orientation and intensity of image gradients in frameworks inspired by the Generalized Hough Transform. The first sub-category comprises the works by Bahlmann et al. [6] and by Brkic et al. [7], whereas in the second we find the Regular Polygon Detector [8], the Radial Symmetry Detector [9], the Vertex Bisector Transform [10], the Bilateral Chinese Transform [11] and, alike, the two schemes of Single Target Voting for triangles and circles proposed by Houben [12].

Segmentation-based algorithms tend to follow a common scheme: the image is transformed into a color space that highlights the signs of interest and then thresholded. Several color spaces have been proposed accordingly, such as RGB, normalized RGB, HSV, CIE L*a*b. Gomez-Moreno et al. [13] present a survey of color-spaces used for traffic sign segmentation and an interesting experimental analysis. A variant to this common approach is represented by definition of a specific color transformation that highlights the colors of interest: for example, the transformations defined in [14], which uses the difference between the most characteristic RGB channel of a signal and the other two channels, and that proposed in [15], which uses the maximum between the normalized R and B channels. A more complex approach than simply switching color-space is to develop a saliency-based visual attention model, biologically inspired by selective attention of human vision, such as that proposed in Itti et al. [16]. Such models are based on center-surround differences implemented via operations on Gaussian pyramids and combine into the same saliency map edge magnitude and color. Finally, another variant has been proposed, which pushes further the effort of defining a proper color space for the specific task: in [17], the authors learn from training data via Integer Linear Programming a set of good color transformations together with the optimal threshold for each.

3. Traffic sign detection pipeline

An overview of our detection pipeline is depicted in Fig. 1. The use of interest region detectors to select candidate regions is the main difference between our proposal and the majority of the approaches presented in the previous section, which instead rely either on the Hough Transform or on a sliding window detector trained to learn distinctive cues from data. The only other paper relying on interest regions extraction is the recent traffic sign recognition system described in [15], which employs Maximally Stable Extremal Regions (MSERs) [18] for the detection stage and has been a source of inspiration for our work. Sliding window approaches require a cascade of classifiers [1] to perform detection at acceptable frame-rates, where the first classifiers in the hierarchy quickly discard background regions and let deeper, more specialized but slower, classifiers analyze only a subset of candidate regions. Although the cascade of classifiers works pretty well in practice, its main purpose is to improve the efficiency of the detector, not its effectiveness. Indeed, if a traffic sign gets discarded by higher level classifiers, it cannot be recovered by deeper ones (in other words, recall may only decrease as long as candidate regions move down through the cascade). Given how crucial the first stage turns out in determining the recall of the overall system, we advocate accomplishing candidate regions selection by a robust and fast approach deploying strong prior knowledge on the appearance of the objects of interest.

As highlighted in the previous section, the two main cues for traffic sign detection are color and shape: in our pipeline we exploit both, relying purposely on two complementary interest regions' detectors. In particular, we extract as candidate patches of the image those regions that exhibit either a uniform value of the main colors of a sign, as found by the MSER algorithm, or show a strong symmetry, as found by a recently proposed detector based on the wave equation (WaDe) [19]. Color is further exploited by preprocessing the image to enhance the main color of each sign, likewise segmentation-based TSD algorithms. Preprocessing also tries to correct over and under exposure of signs that are common in outdoor pictures and set forth serious challenges for automatic detection.

The candidate regions provided by the interest region detectors are then classified as either traffic sign or background. We deploy Support Vector Machines (SVMs) to carry out classification; as for features, we rely on HOGs [2], as it is widely recognized as an effective general purpose descriptor in a variety of scenarios, including traffic sign detection. Moreover, quite peculiarly with respect to e.g. SIFT [20] and SURF [21], which are paired with a dedicated detector, it turns out quite straightforward to compute HOG on multi-scale interest regions provided by any kind of detector.

Finally, two filters help further pruning out false positives: a generative context-aware filter discards regions that are unlikely to correspond to traffic signs given the relationship between their size and their position in the image; a traffic light detector checks if regions corresponding to traffic light lamps have been erroneously classified as traffic signs by previous filters, so as to discard them. Next subsections present all the steps in detail.

3.1. Image preprocessing

Road scenes are affected by significant changes in lighting conditions, so that an initial preprocessing step is necessary to enhance traffic sign regions and fade background. We experimented with the following preprocessing steps. First, linear *contrast stretching* is applied separately to the three RGB channels (3*CH-CS*) so as to deal with under or over exposed images. Then, a single channel image to be fed to the interesting region extractor is

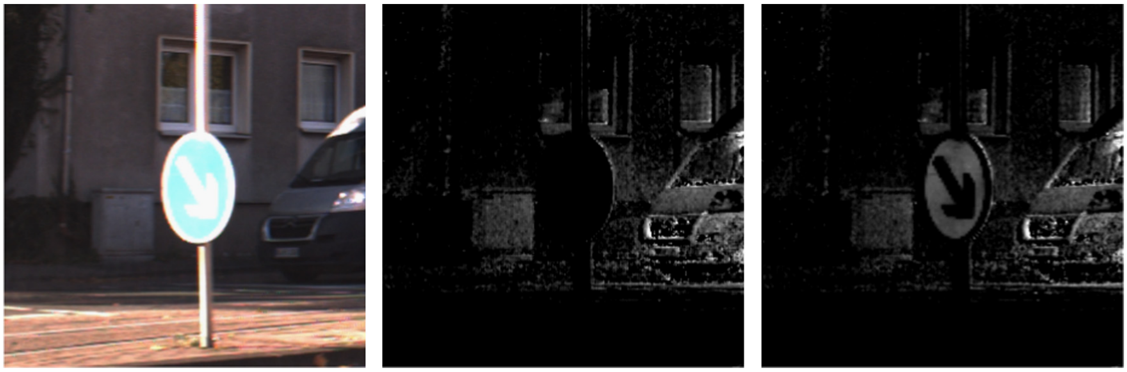

Fig. 2. Enhancement of blue channel. Left, original image (a detail): a very bright mandatory sign; center, enhanced image with blue/green comparison: the sign is not visible; right, enhanced image without blue/green comparison: visibility improves significantly. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

obtained by enhancing the channel C that characterizes the sought signal (blue in the case of mandatory signs, red otherwise). We experimented with two enhancing methods. The first is RGB normalization:

$$C' = \frac{C}{R+G+B}, \quad C \in \{R, B\} \tag{1}$$

The second method has been suggested in [14]. In the case of *danger* and *prohibitory* signs, red is enhanced according to

$$R' = \max\left(0, \frac{\min(R-B, R-G)}{R+G+B}\right) \tag{2}$$

whereas, for mandatory signs, the blue channel is enhanced according to

$$B' = \max\left(0, \frac{B-R}{R+G+B}\right) \tag{3}$$

In turn, the enhancement in (3) differs from the original proposal in [14], which is the exact dual of (2). Indeed, we found that with very dark or bright mandatory signs the blue and green channels tend to have similar values and cancel each other: we do not consider thus the strength of the blue with respect to the green (see Fig. 2 for an example). Finally, a further *contrast stretching* step is applied to the resulting one channel image (1*CH-CS*).

3.2. Interest regions extraction

We substantiate the intuition of relying on multiple interest region detectors by investigating on the use of two complementary algorithms: the well-known Maximally Stable Extremal Regions (MSER) detector [18] and the recently proposed Wave-based Detector (WaDe) [19].

MSER detects high-contrast regions of approximately uniform gray tone and arbitrary shape, and is therefore likely to fire at evenly colored regions within traffic signs. MSERs are found by binarizing each frame at all the possible threshold levels, and then analyzing the connected components at each level: the connected components that maintain their area through several threshold values are selected as MSERs. We can distinguish between dark connected components on a brighter background (MSER+) and bright ones surrounded by a darker background (MSER−).

WaDe detects symmetric regions at various scales by sharp spatio-temporal extrema of a novel scale-space-like family of images, which are obtained as solutions of the wave partial differential equation at consecutive time-steps. The initial condition is given by the gray-scale image under analysis, with approximately absorbing boundary conditions enforced to limit spurious interferences between traveling waves. By letting the image evolve according to the wave equation, contributions from edges of a symmetric structure, such as a circle or square, would reach the center of the symmetry in phase, thereby creating a sharp extremum. WaDe is therefore peculiarly suited to detection of round traffic signs. Similarly to MSER, minima highlight bright symmetric structures on a darker background (WaDe−), maxima the opposite polarity (WaDe+).

For danger and prohibitory signs we consider MSER+ and WaDe+ regions extracted from the red channel of the preprocessed image, so as to detect the red border of the signs, as well as MSER− and WaDe− regions extracted from the preprocessed image to detect the white inner part. As the white part of the signs is the hardest to segment as a uniform color because of its achromaticity [13], we also include MSER− regions from the original image converted to gray-scale to detect the white inner part, as proposed in [15]. For mandatory signs, instead, we consider MSER− and WaDe− regions extracted from the blue channel of the preprocessed image, so as to detect the uniform blue region within the signs.

All regions extracted by MSER− and WaDe− when looking for the inner part of prohibitory or danger signs are rescaled to include the whole traffic sign. We also rescale MSER+ and WaDe+ regions as they tend to be too conservative in estimating the actual sign size. In [5] rescaling factors were chosen manually. In this work, instead, suitable rescaling factors are learned from data to avoid mis-detecting some signs due to imprecise tuning of trivial geometric constraints. The estimation procedure is as follows:

- regions are extracted without rescaling;
- only regions whose overlap with a ground-truth bounding box is higher than a threshold and that are not bigger than the corresponding ground-truth bounding box are kept (we used 0.8 overlap for MSER+ and WaDE+ regions, as they are associated with the whole sign, and 0.4 for regions corresponding to the white inner part of signs);
- we robustly regress an affine rescaling model for the new height of the bounding box h' according to

  $$h' = mh + q \tag{4}$$

  by using the detected bounding boxes and the corresponding ground-truth bounding boxes with Iteratively Re-weighted Least Squares [22];
- at run-time, we rescale both the bounding box width and the height according to the regressed mapping, i.e.

  $$h' = mh + q \quad \text{and} \quad w' = mw + q. \tag{5}$$

Finally, we pass down to the next stage of the pipeline only those bounding boxes that satisfy the constraints reported in Table 1, obtained from ground-truth statistics.

Table 1
Constraints on bounding box size and aspect ratio.

Min. area	225
Max. area	27,300
Min. aspect ratio	0.6
Max. aspect ratio	1.3

3.3. HOG and SVM classification

Given a set of candidate regions detected in the previous step, Histogram of Oriented Gradient (HOG) [2] features are computed to exploit their well-known ability to robustly capturing key shape traits in spite of large intra-class variance.

In [5], we already evaluated the discriminative power of HOGs extracted from different image kinds such as the input gray-scale image, the preprocessed gray-scale image (which better highlights the shape of evenly colored regions within signs), the input color image (as proposed in [2], in such a case the gradient used at each pixel being the gradient with the largest magnitude between those computed on each channel). The best source turned out to be the input color image, which simultaneously considers both shape and color cues. In this work we also evaluate the juxtaposition of the HOGs extracted from the color image and the preprocessed images, as these two input images may be robust to different nuisances. Unlike the HOG parameters used in the GTSRB benchmark reported in [3], we found that the best parameters are those suggested by Dalal and Triggs in their original work. We only increased from 9 to 16 the binning of the “unsigned” gradient orientations over 0–180° and we considered a square region instead of a vertically elongated one. Summarizing, we rescale each region to a size of 64×64 pixels and describe it by 16×16 blocks of 8×8 cells with a spacing stride of 8 pixels, obtaining a 3136-dimensional feature vector (6272-dimensional in case the HOGs extracted from the color and preprocessed images are juxtaposed), that is fed to the classifier.

Despite the robustness of HOG to different illumination conditions, some traffic signs in the GTSDB dataset undergo huge photometric distortions (Fig. 3, central row). Therefore, before extracting HOGs, we carry out an illumination compensation step on detected patches: first we compute the Value histogram, $V = \max(R, G, B)$, then transform each channel by a piecewise linear function obtained by mapping the median of the histogram, V^*, to the midpoint of the range (i.e. 128) and joining it to (0,0) and (255,255) through two linear mappings (Fig. 3, top row). This function has been devised to ensure that underexposed or overexposed patches redistribute their mass more evenly around the midpoint of the range (Fig. 3, top row), thereby attaining new patches which are more amenable to classification by subsequent stages (Fig. 3, bottom row).

Finally, HOGs extracted from the compensated patches are classified as either the traffic sign of interest or background by a binary SVM with RBF kernel.

3.4. Context-aware filter

Bounding boxes classified as traffic signs are then passed down to an additional stage of the pipeline aimed at filtering out wrong detections (i.e. reducing the number of false positives). The idea is to enforce spatial constraints based on context information: indeed, context has not been exploited by previous pipeline stages for the detection of traffic signs, although it could represent an important source of discriminative information.

The proposed filtering stage exploits the typical position of traffic signs in urban environments, with roadsigns appearing not too close to the ground floor or too above the horizon line. More specifically, the devised feature relies on two main cues: the size and height of a traffic sign in the image. It is indeed intuitive to observe that there is typically a strong correlation between the two cues: the smaller the size of the sign, the more its height will tend to approximate the horizon in the current image frame, due to roadsigns of the same category having approximately the same size. This observation relies also on the assumption that the camera is mounted in a fixed position on the roaming vehicle, which is reasonable in mobile mapping applications and allows simplifying the formulation by considering a constant horizon position in each frame. Hence, the limited range that the traffic sign height tends to assume when the size gets smaller is due to the fact that a smaller projection onto the image plane means that the sign is physically far away from the vehicle, while when its size increases the possible positions along the vertical image dimension tend to be more varied and thus less predictable.

Let $B = \{b_1, \ldots, b_n\}$ be a given set of n bounding boxes, the four corners of a generic element $b_i \in B$ denoted as $b_i^{tl}, b_i^{tr}, b_i^{bl}, b_i^{br}$, where t, b, r, l indicate *top*, *bottom*, *right*, *left* respectively. To estimate the traffic sign height $h(b_i)$ associated with a bounding box b_i, we have selected the vertical coordinate of the top-left corner of its enclosing bounding box normalized by the number of image rows H:

$$h(b_i) = \frac{y(b_i^{tl})}{H}, \tag{6}$$

while, as for its size, r, we have chosen the number of rows in its enclosing bounding box normalized by r_{min} and r_{max}, i.e. the minimum and maximum sizes assumable by a traffic sign, which can be computed from the geometric constraints listed in Table 1:

$$r(b_i) = \frac{y(b_i^{bl}) - y(b_i^{tl})}{r_{max} - r_{min}}. \tag{7}$$

These two normalization steps let the two parameters r and h span the range [0, 1]. A graphic illustration of the two values h and r extracted from a real bounding box is shown in Fig. 4.

As our goal is defining a classification problem whereby true traffic signs and false positives are told apart based on context, it turns out hard to model the negative class due to the dependence of negative samples on the actual classification method and parameters employed throughout the previous stages of the pipeline. Due to this difficulty of defining – and thus learning – the negative class, we rely on a generative classifier, with a parametric model aimed at explaining positive samples (i.e. true traffic signs) learned from training data.

The idea is to *learn* the relationship between r and h by means of a parametric model, represented by a family of one-dimensional Gaussian distributions parametrized by the possible normalized size values r, each Gaussian modeling the probability distribution of the normalized height value h given the size r and the event $\mathcal{S}$ that the bounding box represents a traffic sign:

$$p(h|r, \mathcal{S}) \sim \mathcal{N}(\mu(r), \sigma(r)). \tag{8}$$

In turn, we also propose to explicitly model the parameter set $\Theta = \{\mu(r), \sigma(r)\}$ and to estimate this model from training data. To perform robust parameter estimation, we assume that both the means and the standard deviations of the estimated distributions follow a parametric law of the independent variable r:

$$\mu(r) = a_\mu \cdot r + b_\mu, \tag{9}$$

$$\sigma(r) = a_\sigma \cdot r^2 + b_\sigma \cdot r + c_\sigma. \tag{10}$$

We have thus assumed linear dependency between r and $\mu(r)$, and a quadratic relationship for $\sigma(r)$. Nonetheless, different parametric models could be employed, as the whole approach is general enough to be easily customized (in turn, using a linear model for both parameters would not change significantly the

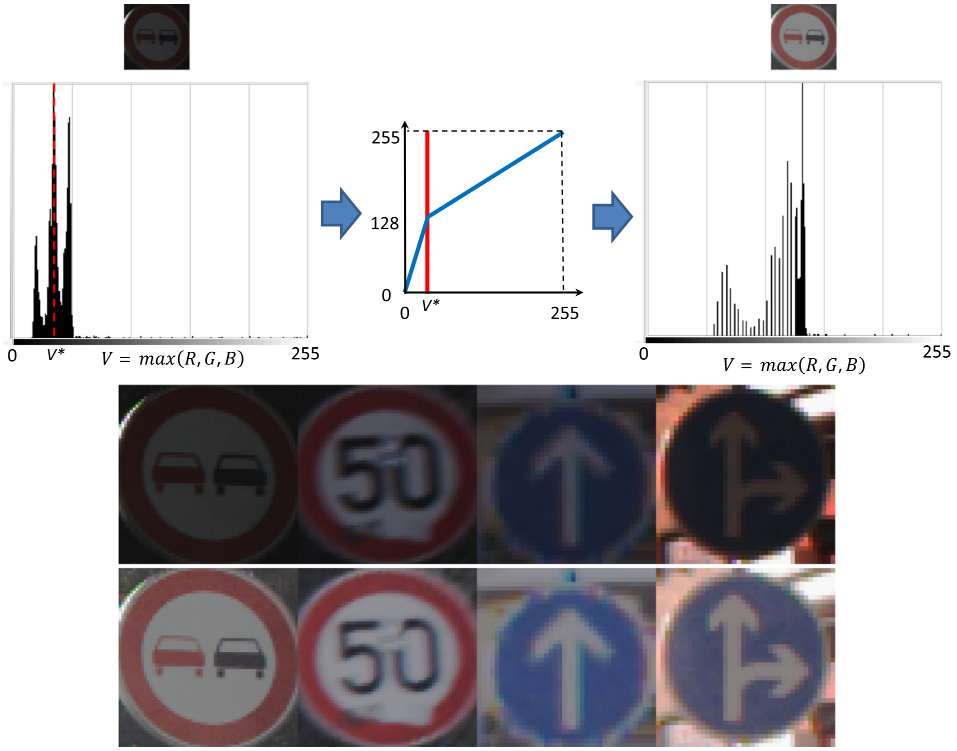

Fig. 3. Illumination compensation. Top row: input patch and histogram of value V (left); piecewise linear mapping (center); output patch and new histogram of value V (right). Central row: original signs. Bottom row: patches obtained by the proposed illumination compensation.

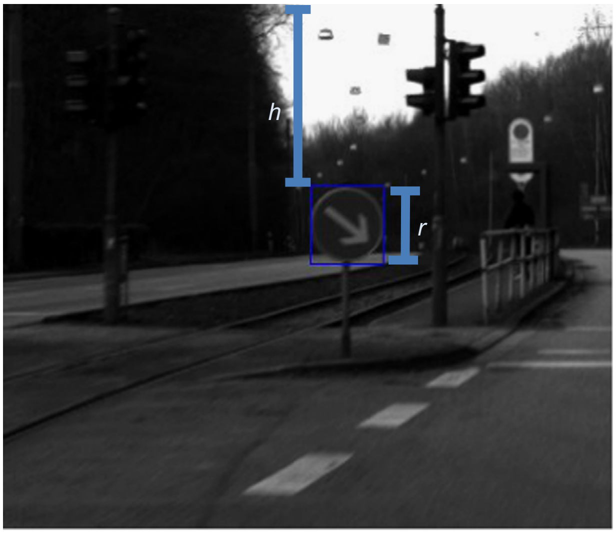

Fig. 4. Features exploited by the context-aware filter: h is the height of the sign in the image, i.e. the vertical coordinate of the top side of the sign bounding box; r is the height of the sign bounding box.

results on the GTSDB dataset). Accordingly, the set of parameters Θ includes the coefficients of the parametric models for the mean and the standard deviation of the Gaussian family we are trying to estimate: $\Theta = \{a_\mu, b_\mu, a_\sigma, b_\sigma, c_\sigma\}$. Given the whole training set of (r,h) pairs, and once $\mu(r)$ and $\sigma(r)$ have been computed for all r values spanned by the training data, an over-determined linear system can be defined so as to estimate in closed-form via least-squares the five parameters. In particular, given the non-uniform distribution of (r,h) pairs over the domain spanned by r, we employ weighted least-square estimation, to render the coefficient estimation process less biased by outliers. The weights in this case are proportional to the available training population for each value of the variable r.

At run-time, the posterior probability that a test sample (r,h) represents a traffic sign according to our model is

$$p(\mathcal{S}|r,h) \propto p(h|r,\mathcal{S})p(\mathcal{S}|r) \simeq p(h|r,\mathcal{S}) \tag{11}$$

where we have used the Bayes rule and assumed an uninformative prior $p(\mathcal{S}|r)$. Therefore, to decide if a bounding box represents a sign we can threshold the likelihood

$$p(h|r,\mathcal{S}) = \frac{1}{\sqrt{2\pi}\sigma(r)} \exp\left(-\frac{(r-\mu(r))^2}{2\sigma^2(r)}\right). \tag{12}$$

3.5. Traffic light filter

The presence of traffic lights in urban images represents a nuisance for automatic traffic sign detection due to shape, position and luminosity of traffic light lamps being often similar to that of certain traffic signs (in particular round-shaped ones, such as *mandatory* and *prohibitory*). This problem is worsened by the widespread presence of traffic lights within road scenery, which may induce a high number of false positives in a traffic sign detection pipeline. For this reason, we have designed a specific stage aimed at rejecting potential false positives due to traffic lights wrongly recognized as traffic signs. This module can also be seen as a traffic light detector, so that it may be deployed within similar application scenarios for other purposes than pruning falsely detected traffic sign.

The proposed approach follows a two-step procedure based on some working hypotheses in order to examine and potentially discard each previously detected Region Of Interest (ROI). First of all, each ROI is assumed centered at one of the lamps of a traffic light. Secondly, we assume that all traffic lights consist of three lamps, with at most one of them switched on at a certain time

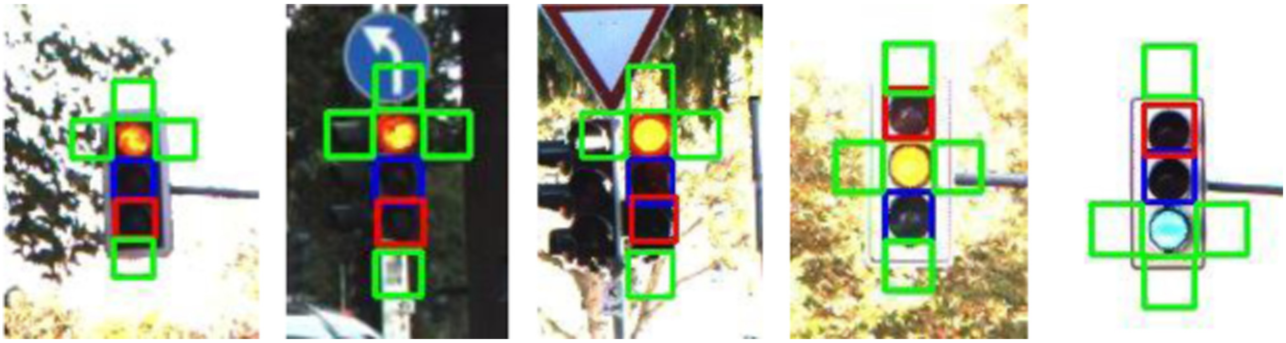

Fig. 5. Examples of traffic lights detected by the proposed filter. For each example, auxiliary ROIs are depicted in green ($ROI_l, ROI_r, ROI_u, ROI_d$), blue ($ROI_{ref}$) and red ($ROI_{in}$). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

instant. This obviously do not consider other configurations, such as e.g. two-lamp traffic lights, which are anyway much less frequent, and two-lights-on configuration. In any case, we wish to point out that the heuristics deployed by the proposed traffic light filter could be adapted to deal also with these two cases.

The goal of the first stage is to detect the most probable status of the lamp on which the current ROI is centered, choosing between *Red*, *Yellow*, *Green* and *Off*. This is done simply by thresholding the average red and green values of the pixels within the circle inscribed in the ROI. The status of the lamp as well as its size (represented by the size of the current ROI) allows us to define the relative position of the traffic light and of the remaining lamps with respect to the current ROI.

Successively, six additional auxiliary ROIs are determined. As shown in Fig. 5, two of them, referred to as ROI_{ref}, ROI_{in}, are centered on the two switched-off lamps of the traffic light (given our working hypotheses, there are always at least two lamps switched-off) and are localized based on the status of the traffic light determined during the first stage (i.e. a green light will have both off lamps along its upward vertical direction, and so on). Since the traffic light size and tilt angle can be only approximately determined from the current ROI dimension - imagine e.g. the common case of traffic lights slightly tilted toward the ground plane-a template matching refinement stage is deployed. Within this stage, the first switched-off lamp (i.e. ROI_{ref}) is positioned according to the size and status of the central lamp. Then, a small search area centered at the expected position of ROI_{in} is defined, and ROI_{ref} is used as a template: the refined ROI_{in} position is found by minimizing the Euclidean distance δ between all equally sized patches within this search area and the template.

Four auxiliary ROIs (see again Fig. 5) are extracted in four locations outside the hypothesized traffic light, along the left, right, up and down directions (referred as $ROI_l, ROI_r, ROI_u, ROI_d$ respectively). The condition applied to test whether a ROI includes or not a traffic light is thus:

$$\frac{\delta(ROI_{ref}, ROI_{in})}{\min\limits_{i \in \{l,r,u,d\}} \delta(ROI_{ref}, ROI_i)} < \tau_\delta \qquad (13)$$

The role of the four external ROIs is twofold. On the one hand, they provide an adaptive term of comparison to rescale the distance between internal ROIs under different lighting conditions, thus allowing for using a fixed threshold τ_δ. On the other hand, in case the input ROI lays on an actual sign, it is likely that ROI_{ref} and ROI_{in} lay on the same background and would report, in this case, a small relative distance. By using four different external ROIs, each placed along a different direction, the chance of having one external ROI laying on the same background part as that of ROI_{in} is increased, this allowing one to robustly discriminate real traffic signs from traffic lights even in presence of disparate background conditions. By thresholding the relative distances between inner and outer ROIs, the whole traffic light detection is normalized with respect to the photometric conditions of the image.

To increase the robustness of the filter, additional constraints that need to be satisfied in order for a ROI to be classified as a traffic light are enforced. In particular, the mean and variance of the pixels within the circles inscribed in ROI_{ref}, ROI_{in} must be smaller than a certain threshold, so as to check that they correspond to switched-off lamps:

$$\begin{aligned} &\mu(ROI_{ref}) < \tau_\mu \quad \wedge \quad \mu(ROI_{in}) < \tau_\mu \\ &\sigma(ROI_{ref}) < \tau_\sigma \quad \wedge \quad \sigma(ROI_{in}) < \tau_\sigma \end{aligned} \qquad (14)$$

4. Results

The proposed pipeline has been tuned and tested on the dataset made publicly available for the German Traffic Sign Detection Benchmark [4]. It has also been evaluated on data from a real mobile mapping acquisition campaign carried out in Verona, Italy.

4.1. Pipeline tuning on the GTSDB training dataset

The German Traffic Sign Detection Benchmark [4] is a publicly available dataset targeting traffic sign detection. An on-line competition was also launched when the dataset was released, at the beginning of December 2012, and the submission phase was closed at the end of February 2013. The dataset made available to participants to download features 900 images (split into 600 training images, released at the start of the competition, and 300 evaluation images without ground-truth, released when the submission phase started) with very tough size and illumination condition variations. In this subsection, we present the tuning of the pipeline that was carried out on the training images during the first part of the competition. The next subsection presents and discusses the results achieved on the test set by the tuned pipeline.

As mentioned in the Introduction, participants were required to detect signs belonging to three categories characterized by differences in shape and color: prohibitory (circular red-and-white signs), mandatory (circular blue-and-white signs) and danger signs (upper-triangular red-and-white signs). When tuning our pipeline on the GTSDB dataset, for each category we first selected the best parameters for the image preprocessing and interest region detection steps. As the best preprocessing can only be defined in terms of the quality of the regions found by the detectors, the two stages were tuned jointly. Interest region detection is the only stage where the bounding boxes associated with traffic signs are injected into the pipeline: subsequent stages

Table 2
MSER and WaDe performance for different sign categories.

Category	Detector	3CH-CS	Enhance/Norm.	1CH-CS	FN	FP
Prohibitory	MSER	No	Norm.	No	0	1127K
Prohibitory	WaDe	Yes	Norm.	Yes	2	947K
Mandatory	MSER	Yes	Norm.	No	6	365K
Mandatory	WaDe	Yes	Enhance	Yes	1	1498K
Danger	MSER	No	Norm.	No	2	1129K
Danger	WaDe	No	Norm.	Yes	12	1979K

Table 3
Rescaling factors of interest regions bounding boxes.

Category	m	q
Prohibitory MSER− normalized red	1.09	−0.3
Prohibitory MSER+ normalized red	1.43	0.83
Prohibitory MSER− gray-scale	1.43	0.83
Danger MSER− normalized red	1.1	−0.75
Danger MSER+ normalized red	1.47	0.73
Danger MSER− gray-scale	1.47	0.73
Mandatory MSER+ normalized blue	1.09	−0.54
Prohibitory WaDe− normalized red	1.21	−3.65
Prohibitory WaDe+ normalized red	1.53	−1.11
Danger WaDe− normalized red	0.84	6.65
Danger WaDe+ normalized red	0.95	6.06
Mandatory WaDe− enhanced blue	1.18	−2.56

can thus only prune false detections. Therefore, the main aim in tuning jointly preprocessing and interest region detection was to minimize the number of missed traffic sign detections (false negatives, if we consider detection as a binary classification of all the possible bounding boxes). When the number of false negatives is on par, we prefer a tuning that yields less false positives.

We simultaneously tuned the rescaling factor according to the procedure described in Section 3.2 as well as whether to use or not the three preprocessing steps. The best combinations of preprocessing and rescaling are reported in Tables 2 and 3, along with the number of false negatives and false positives over the whole training set. To avoid combinatorial explosion of the parameter space, we instead manually selected the detector parameters, according to the typical size of signs in the GTSDB dataset and the idea of using loose thresholds (such as ρ for WaDe and δ for MSER) so to try not to miss even the most challenging, poorly contrasted signs. The selected parameter tunings are for MSER, $\delta=2$, max variation $=0.5$, and min diversity $=0.2$; for WaDe, $\rho=0.1$, $T=120$, and first scale $=8$.

What is interesting about the results in Table 2 is that the use of complementary region detectors turns out very important to let the first stage of the pipeline achieve the highest recall, thus corroborating our reasoning of Section 3.2: should only MSER or WaDe be used, sub-optimal results would be achieved. This is because different signs call for different detector strengths: for instance, in the case of mandatory signs, the white arrows can create separated blue regions out of a single mandatory sign, especially if the scale of the sign is small, which in turn leads to detection of several MSERs with wrong scales. On the other hand, the circular symmetry of the sign remains evident and is captured effectively by WaDe (Fig. 6).

Next, we tuned on the training set the context-aware filter and the traffic light detector. Given the three categories of traffic signs that need to be detected, we have decided to separately learn two different context parameter sets, one for the *danger* category, and the other, jointly, for both *prohibitory* and *mandatory* signs. This is due to the size of *prohibitory* and *mandatory* signs being extremely similar, so that merging the two classes allows for a more robust parameter estimation. Fig. 7 shows the functions learned on the GTSDB training dataset to represent the mean and standard deviation of the family of Gaussian distributions modeling the likelihood $p(h|r,\mathcal{S})$ for the *danger* and *prohibitory+mandatory* traffic signs. The blue and red dots show the training population of, respectively, mean and standard deviation values for each value of r (note the different sizes of each dot being proportional to the amount of training samples associated with each value). The green and yellow lines denote, instead, the learned functions: the distribution of the training data clearly support our hypothesis of a predictable relationship between the size of a traffic sign and its position in the image: only a small subset of signs, especially belonging to the mandatory class, deviates from the estimated mean value (rightmost figure). This can be ascribed to the presence of mandatory signs close to the ground in the proximity of pedestrian crossings, junctions and roundabouts. By estimating a higher standard deviation value for closer, and therefore bigger, signs, the proposed model avoids to filter such traffic signs, although reducing its ability to reject false positives. Table 4 reports the learned coefficients for the functions associated with the two considered categories.

As far as the traffic light filter is concerned, we manually estimated the thresholds $\tau_\delta=0.05$, $\tau_\mu=65$, and $\tau_\sigma=40$ based on the GTSDB training dataset. Clearly, the danger signs pipeline can avoid the traffic light removal step, as the HOGs+SVM classifier trained on triangular signs already discards circular patterns. Examples of false positive detections that can be pruned thanks to the proposed filters are reported in Fig. 8.

Given the best preprocessing and rescaling parameters for each traffic sign category and the tuned final filters, we then estimated the impact of using as feature for the SVM classifier the justapoxition of HOGs extracted from the preprocessed and the color image, instead of using only HOGs from the color image, as done in [5]. To estimate the performance of different input HOGs+SVM classifier, we perform 10-fold cross-validation on the training set and run the overall pipeline, so as to evaluate the effect of the different features on performance. Results are reported in Table 5 in terms of Areas Under the Curves (AUCs) of the precision–recall curves obtained by varying the SVM decision threshold. We can see that, although HOGs from the preprocessed image is not always better than HOGs from the color image, their combination achieves always a better AUC. This is especially true for mandatory signs, which is indeed the most challenging category. From the cross-validation we also obtained as best parameters for the SVM classifier the values $C=1$ and $\gamma=0.01$, i.e. the same values as already used in our online submission [5].

4.2. Results on the GTSDB test dataset

This section reports results of the tuned pipeline on the GTSDB test dataset (i.e. the data released in the test phase of the competition, which were not used for training). The precision–recall curves for the three categories are depicted in Fig. 9. The corresponding AUCs are 99.994% for the prohibitory class, 99.79% for the danger class and 98.17% for the mandatory class. By deploying data-driven estimation of the region rescaling factors and HOGs on color and preprocessed images, we improve the AUCs for all the three categories, with respect to the results presented in [5][1]: the corresponding AUCs were 99.98% for the prohibitory class, 98.72% for the danger class and 95.76% for the mandatory class. Fig. 10 shows all the errors of the pipeline at the best working point (i.e. that minimizing the sum of false positive and false negatives): there are three missed detections

[1] Visible on-line at http://www.benchmark.ini.rub.de/index.php?section=gtsdb&subsection=results&subsubsection=prohibitory

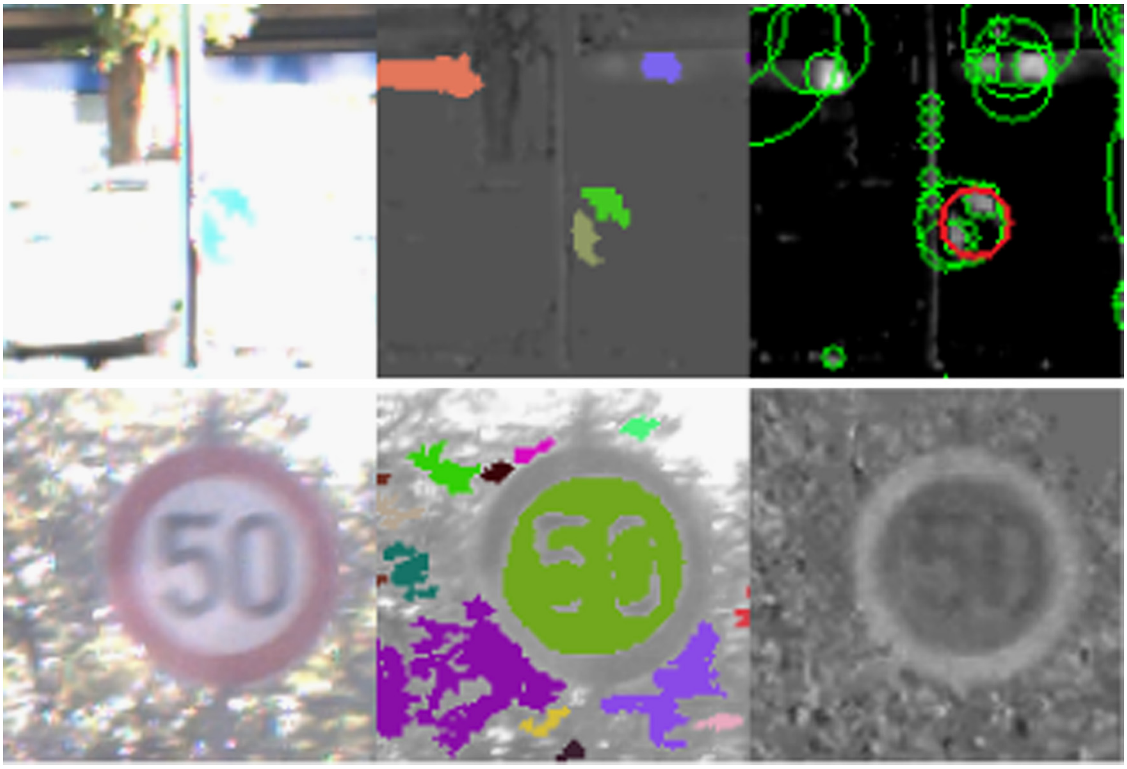

Fig. 6. Complementarity of region detectors. Top row: mandatory sign; bottom row: prohibitory sign. From left to right: input image (a detail); MSER detections (colored regions); WaDe detections (circles, correct detection in red). Note that preprocessing is different for the two detectors and on the last image there are no detections by WaDe due to poor contrast. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

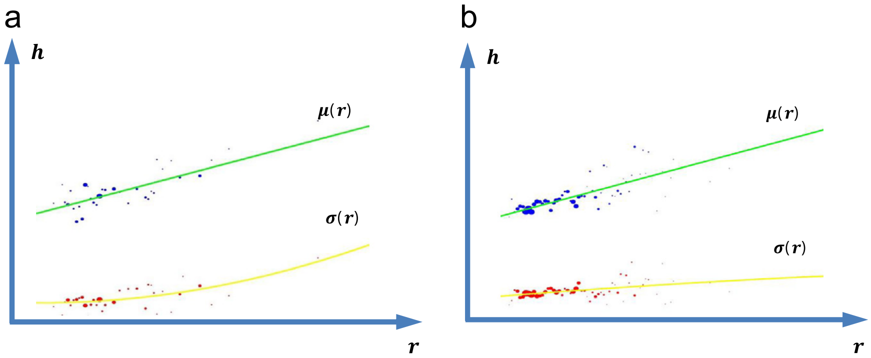

Fig. 7. Context-aware filter: learned functions representing the mean and standard deviation of the family of Gaussian distributions modeling the likelihood $p(h|r,\mathcal{S})$ for danger (a) and mandatory+prohibitory (b) signs. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

Table 4
Context-aware filter: coefficients defining the functions in Fig. 7.

Category	a_μ	b_μ	a_σ	b_σ	c_σ
Mandatory+prohibitory	0.329	0.365	−0.020	0.097	0.061
Danger	0.327	0.383	0.222	−0.006	0.050

because of poorly contrasted signs, four false detections due to signs very similar to mandatory ones but not included in the evaluations and one mandatory sign missed detection due to a very small and bright appearance.

For the sake of completeness, we report here also the AUC scores yielded by the improved pipeline without the context-aware and traffic light filter: 99.990% for the prohibitory class, 99.65% for the danger class and 98.12% for the mandatory class. Although the gap from the ideal AUC is thus already small, the use of filters helps reducing it relatively by 40% for the prohibitory class, 40% for the danger class and 2% for the mandatory class.

With the improved results, our algorithm now ranks 4th out of 67 submissions on prohibitory signs, 6th out of 36 submissions on danger signs and 2nd out of 34 submissions on mandatory signs. The processing time on a standard PC of an unoptimized implementation of our detection pipelines are as follows: 0.6 Frames per second for prohibitory signs, 1.26 Frames per second for danger signs, and 1.75 Frames per second for mandatory signs.

4.3. *Results on mobile mapping data*

Although the images of the GTSDB dataset are hard to classify automatically, they do not fully represent the challenges of real mobile mapping applications. In particular, only one frontal camera is used in the GTSDB dataset, and therefore signs never appear particularly skewed with respect to the camera, whereas this is quite common in a mobile mapping scenario, where more

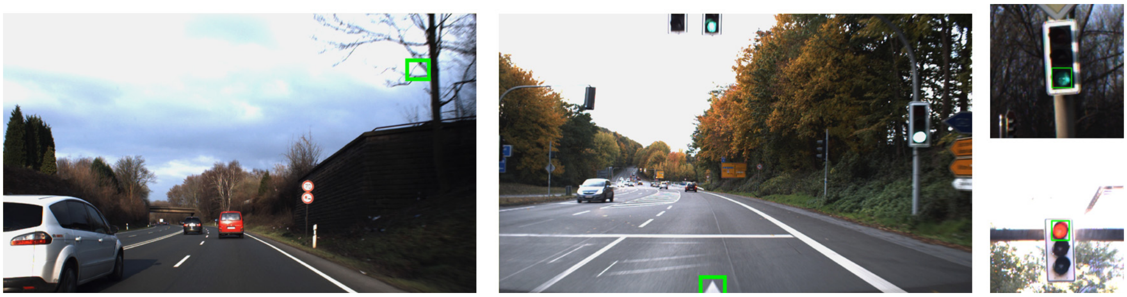

Fig. 8. Examples of false detections pruned by the context-aware (left) and traffic light (right) filters.

Table 5
Comparison between HOGs extracted on color images and the juxtaposition of HOGs from color and preprocessed images.

Category	Color AUC	Preprocessed AUC	Color + preprocessed AUC
Prohibitory	99.11	99.21	99.43
Mandatory	89.66	93.27	95.01
Danger	96.89	96.32	97.22

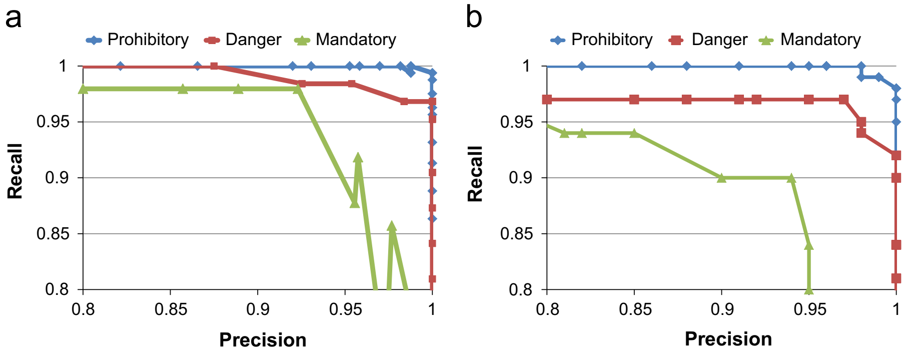

Fig. 9. Left chart: results of the proposed pipeline on the three categories of the test GTSDB dataset. Right chart: our results in the competition [5].

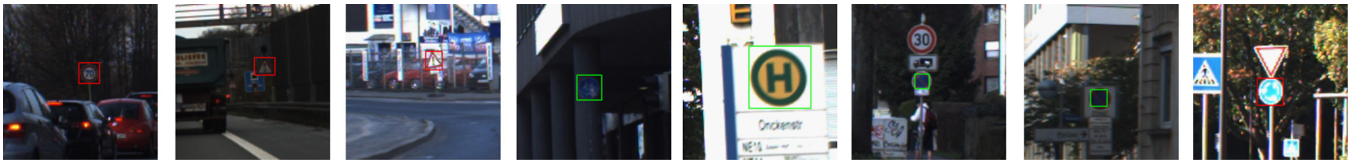

Fig. 10. All errors yielded by the proposed pipeline at its best working point on the GTSDB test set: false positives in green, false negatives in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

cameras are often used and the same sign can, therefore, appear e.g. parallel to the image plane in the frontal view and highly skewed in a lateral camera (or vice versa for signs in lateral roads, on gates, etc.). Moreover, traffic sign sizes in real data span a much greater range than in the GTSDB: in particular, very small signs are more common than in the GTSDB, perhaps because the latter were not included in the dataset by the competition organizers to ease the detection task. On the other hand, as the data are correctly geo-referenced, it is sufficient to detect a sign in one of the camera views to consider it as detected : there is no practical advantage in detecting the same physical traffic sign in more than one view. Therefore, also the evaluation methodology is different: we define a sign as detected (true positive) if it is detected in at least one of the images where it appears, false negative otherwise.

The dataset that we use to test our pipeline was provided by Qonsult S.p.A. and was acquired with a vehicle equipped with the Topcon IP-S2 system, which features a PointGrey Research Ladybug 5 panoramic camera, which in turn is compound by six synchronized traditional pin-hole cameras. The dataset features 6580 images (1316 images for each one of five cameras of the Ladybug, the sixth camera pointing at the sky never includes traffic signs) that contains 213 instances of danger signs (corresponding to 33 physically unique signs), 459 instances of mandatory signs (93 physically unique signs), and 661 prohibitory signs (140 physically unique signs).

We used as training data all the images of the GTSDB dataset. We expect to obtain better performance when a training set of Italian signs will be available, as there are significant formatting differences in traffic signs even between European countries. Nonetheless, we chose not to use as training data a subset of the mobile mapping data so as not to reduce the size of the test data and, hence, the statistical significance of the results. In spite of the differences between the training and the testing data, our pipeline was able to achieve promising results, as vouched by Fig. 11, left. Indeed, we could achieve 78.21% AUC for the prohibitory class, 82.13% for the danger class and 72.78% for the mandatory class.

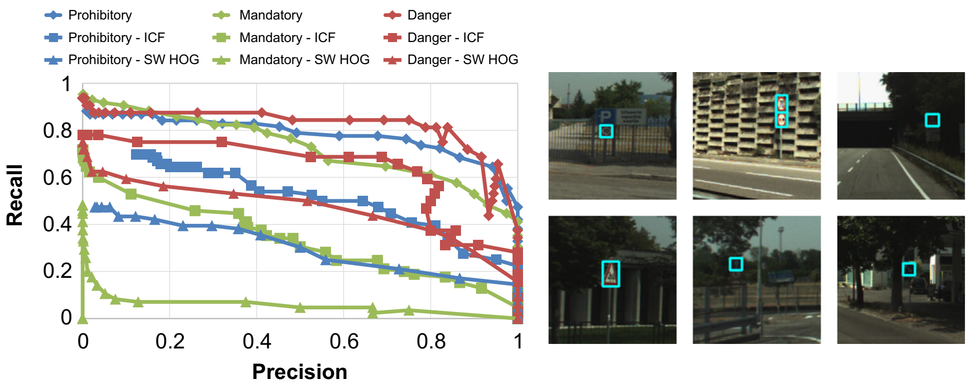

Fig. 11. Left: results of the proposed pipeline vs a sliding window HOG and sliding window ICF on the mobile mapping dataset. Right: some correct detections of the proposed pipeline at the best working point.

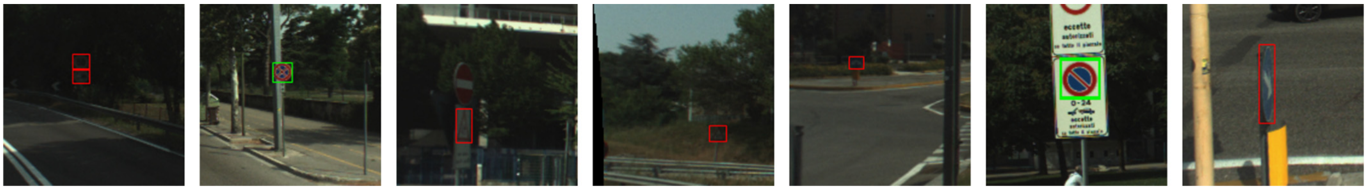

Fig. 12. Some errors on the mobile mapping dataset at the best working point: false positives in green, false negatives in red. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this paper.)

The results also suggest that our pipeline can withstand changes in appearance of traffic signs that are frequent in real-world mobile mapping data, detecting dark, skewed and very small signs, as shown by the results in Fig. 11, right. The main source of errors are signs that were not included in the ground-truth of the testing dataset, but are nevertheless detected by the pipeline, such as the “no parking” signs, together with extremely small or skewed signs. Some detection errors on the mobile mapping dataset are shown in Fig. 12.

Finally, we report a quantitative comparison on this dataset. Due to the difficulty in implementing the other proposals participating in the online competition, and the unavailability of public implementations, we opted for comparing our pipeline against the standard, sliding window approach to detect objects deploying HOG [2] and a state-of-the-art sliding window detector based on Integral Channel Features (ICF) [23], by using the implementation provided by the CCV library (http://www.libccv.org) and setting its parameters according to how the pipeline was used in [24]. Results are reported again in Fig. 11, left. The AUCs for sliding window HOG are 30.98% for the prohibitory class, 46.8% for the danger class and 5.5% for the mandatory class; for ICF are 52.94% for the prohibitory class, 62.92% for the danger class and 33.19% for the mandatory class. The relative ranking among the different sign categories is the same for all detectors. The comparison between sliding window classifiers and our pipeline shows that performance can be dramatically improved by deploying description and classification only at promising locations identified by interest region detectors.

5. Conclusions

The traffic sign detection system we propose demonstrates good performance under challenging conditions such as varying illumination, partial occlusions and large scale variations. Evaluation on the German Traffic Sign Detection Benchmark shows the effectiveness of the proposed approach: our system is able to yield nearly optimal performance on two classes and very good results on the most challenging class of mandatory signs. This can be ascribed to the deployment of complementary interest region detectors, which allows for injecting into the pipeline almost all regions corresponding to traffic signs while notably reducing the number of candidates with respect to a sliding window approach, as well as to the development of effective filters based on traffic light detection and context information. Results on a challenging mobile mapping dataset of Italian signs show the robustness of the proposed approach: our pipeline can successfully be deployed in real and challenging application scenarios.

Future directions of research concerns experimenting with and developing novel representations for interest regions which could be more inherently based on color information, while still proving robust to the severe photometric variations found in typical outdoor conditions. The fusion of information from registered images in mobile mapping data to enhance the pipeline robustness and effectiveness is also worth to be pursued.

Conflict of interest

None declared.

References

[1] P. Viola, M. Jones, Robust real-time object detection, Int. J. Comput. Vis. 57 (2001) 137–154.
[2] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, vol. 2 (2005) 886–893.
[3] J. Stallkamp, M. Schlipsing, J. Salmen, C. Igel, Man. vs computer: Benchmarking machine learning algorithms for traffic sign recognition, Neural Netw. (2012).
[4] S. Houben, J. Stallkamp, J. Salmen, M. Schlipsing, C. Igel, Detection of traffic signs in real-world images: The German Traffic Sign Detection Benchmark, in: International Joint Conference on Neural Networks (2013).
[5] S. Salti, A. Petrelli, F. Tombari, N. Fioraio, L. Di Stefano, A traffic sign detection pipeline based on interest region extraction, in: Proceedings of International Joint Conference on Neural Networks (2013).
[6] C. Bahlmann, Y. Zhu, V. Ramesh, A system for traffic sign detection, tracking, and recognition using color, shape and motion information, in: Proceedings of IEEE Symposium on Intelligent Vehicles, (2005) 255–260.

[7] K. Brkic, A. Pinz, S. Legvic, Traffic sign detection as a component of an automated traffic infrastructure inventory system, in: Proceedings of Workshop of the Austrian Association for Pattern Recognition, (2009) 1–12.

[8] G. Loy, Fast shape-based road sign detection for a driver assistance system, in: Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems, (2004) 70–75.

[9] N. Barnes, A. Zelinsky, L. Fletcher, Real-Time Speed Sign Detection Using the Radial Symmetry Detector, in: International IEEE Conference on Intelligent Transportation Systems, vol. 2 (2008) 322–332.

[10] R. Belaroussi, J.-P. Tarel, Angle vertex and bisector geometric model for triangular road sign detection, in: Proceedings of the IEEE Workshop on Applications of Computer Vision (2009) 577–583.

[11] R. Belaroussi, J.-P. Tarel, A real-time road sign detection using bilateral chinese transform, in: Proceedings of the IEEE Workshop on Applications of Computer Vision, (2009) 1161–1170.

[12] S. Houben, A single target voting scheme for traffic sign detection, In: Intelligent Vehicles Symposium, IEEE (2011) 124–129.

[13] H. Gomez-Moreno, S. Maldonado-Bascon, P. Gil-Jimenez, S. Lafuente-Arroyo, Goal evaluation of segmentation algorithms for traffic sign recognition, IEEE Trans. Intell. Transp. Syst. 11 (2010) 917–930.

[14] A. Ruta, Y. Li, X. Liu, Real-time traffic sign recognition from video by class-specific discriminative features, Pattern Recognit. 43 (2010) 416–430.

[15] J. Greenhalgh, M. Mirmehdi, Real-Time Detection and Recognition of Road Traffic Signs, IEEE Trans. Intell. Transp. Syst. 13 (2012) 1498–1506.

[16] L. Itti, C. Koch, E. Niebur, A model of saliency-based visual attention for rapid scene analysis, Pattern Anal. Mach. Intell. 20 (1998) 1254–1259.

[17] R. Timofte, K. Zimmermann, L. Van Gool, Multi-view traffic sign detection, recognition, and 3D localisation, in: Mach, Vision Appl. 25 (2011) 633–647.

[18] J. Matas, O. Chum, M. Urban, T. Pajdla, Robust wide-baseline stereo from maximally stable extremal regions, in: Image Vision Comput. 22 (2004) 761–767.

[19] S. Salti, A. Lanza, L. Di Stefano, Keypoints from symmetries by wave propagation, in: Proceedings of the IEEE International Conference on Computer Vision and Pattern Recognition, (2013) 2898–2905.

[20] D.G. Lowe, Distinctive image features from scale-invariant keypoints, Int. J. Comput. Vision 60 (2004) 91–110.

[21] H. Bay, A. Ess, T. Tuytelaars, L. Van Gool, Speeded-up robust features (SURF), Comput. Vis. Image Underst. 110 (2008) 346–359.

[22] P.W. Holland, R.E. Welsch, Robust regression using iteratively reweighted least-squares, Communications in Statistics: Theory and Methods (1977) 813–827.

[23] P. Dollár, Z. Tu, P. Perona, S. Belongie, Integral channel features, in: Proceedings of the British Machine Vision Conference (2009).

[24] M. Mathias, R. Timofte, R. Benenson, L. Van Gool, Traffic sign recognition—how far are we from the solution? in: Proceedings of the International Joint Conference on Neural Networks (2013).

Samuele Salti received the M.Sc. degree in computer science engineering in 2007 and the Ph.D. degree in computer science engineering in 2011, both from the University of Bologna, Italy. Since 2011 he is a Post-Doc at Computer Vision Lab., DISI (Department of Computer Science and Engineering), University of Bologna. In 2007 he visited the Heinrich-Hertz-Institute in Berlin, Germany, working on human–computer interaction. In 2010 he visited the Multimedia and Vision Research Group (MMV) at Queen Mary, University of London to work on adaptive appearance models for video tracking. His research interests are adaptive video tracking, 3D shape matching, Bayesian filtering and object recognition. He has co-authored 19 publications in international conferences and journals. He was awarded the best paper award runner-up at 3DIMPVT, the International Conference on 3D Imaging, Modeling, Processing, Visualization and Transmission in 2011. He serves as a reviewer for IEEE Transactions on Signal Processing, IEEE Transactions on Image Processing and a number of international conferences. He is a member of the IEEE and GIRPR.

Alioscia Petrelli received the degree in computer science engineering from the University of Bologna, Italy, in 2005. He spent four years as research fellow at the Computer Vision Laboratory of the Department of Electronics, Computer Science, and Systems in Bologna. Currently, he is a Ph.D. student with the Department of Computer Science and Engineering, University of Bologna. His research focuses on computer vision, including 3D surface matching and machine learning. He serves as a reviewer for the IEEE International Conference on Computer Vision and is a member of the IEEE Computer Society.

Federico Tombari holds an appointment as an Assistant Professor (RTD) at the University of Bologna, after obtaining from the same institution a Ph.D. in 2009. His current research activity concerns computer and robot vision, and it encompasses co-authoring more than 50 refereed papers on peer-reviewed international conferences and journals, mainly focused on 2D/3D object recognition, stereo vision, video analysis for surveillance and efficient indexing. In 2004 he has been a Visiting Student at the University of Technology, Sydney, while in 2008 he was an intern at Willow Garage, California. He is a Senior Scientist volunteer for the Open Perception foundation and a developer for the Point Cloud Library. In 2012 and 2013 he held a position as an Adjunct Professor at the University of Bologna. He is member of IEEE and IAPR-GIRPR. He is the recipient of the “Best Paper Award Runner-up” of the International Conference on 3D Imaging, Modeling, Processing and Visualization Technologies (3DIMPVT 2011).

Nicola Fioraio received the B.Sc. and M.Sc. degrees in electronic engineering from the University of Bologna, Bologna, Italy, in 2009 and 2011, respectively. In 2011 he visited Willow Garage, Inc., Menlo Park, CA (USA), working on RGB-D real-time multi-constrained registration. Since 2012 he is a Ph.D. student at the Computer Vision Lab., Dept. of Computer Science and Engineering, University of Bologna, Italy, working on semantic SLAM and active object recognition. His current research interests include computer vision, robotics, machine learning and real-time RGB-D SLAM. He is a member of the IEEE Computer Society and the IAPR-IC (GIRPR).

Luigi Di Stefano received the degree in electronic engineering from the University of Bologna, Italy, in 1989 and the Ph.D. degree in electronic engineering and computer science from the Department of Electronics, Computer Science and Systems (DEIS) at the University of Bologna in 1994. In 1995, he was a Postdoctoral Research Fellow at Trinity College, Dublin. He is currently an Associate Professor at the Department of Computer Science and Engineering, University of Bologna, His research interests include computer vision, image processing and computer architecture. He is the author of more than 150 papers and five patents. He is a member of the IEEE Computer Society and the IAPR-IC. From 2012 he is a member of the Scientific Advisory Board of Datalogic Group.