A Traffic Sign Detection pipeline based on interest region extraction

Samuele Salti, Alioscia Petrelli, Federico Tombari, Nicola Fioraio and Luigi Di Stefano

Abstract— **In this paper we present a pipeline for automatic detection of traffic signs in images. The proposed system can deal with high appearance variations , which typically occur in traffic sign recognition applications, especially with strong illumination changes and dramatic scale changes. Unlike most existing systems, our pipeline is based on interest regions extraction rather than a sliding window detection scheme. 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. Our proposal has been evaluated experimentally within the German Traffic Sign Detection Benchmark competition.**

I. Introduction

TRAFFIC SIGN RECOGNITION (TSR) is an actively developed area of research with applications in Autonomous Driving Systems and Driving Assistance Systems. The wide appearance variability of traffic signs in real-world environments represents a challenging test-bench to assess the performance of computer vision algorithms in uncontrolled environments.

Traffic Sign Recognition is usually solved via a two-step approach: traffic sign detection and traffic sign classification. In detection, the aim is to identify the regions of an image (bounding boxes) that tightly contains a traffic sign: this is indeed the kind of problem solved by traditional detectors developed by the computer vision community for face [1] and pedestrian detection [2]. In classification, given a detected bounding box, the aim is to assign it a label corresponding to the traffic sign contained in it. However, the boundary between detection and classification may be more blurred than implied by previous descriptions. For example, traffic sign detection may produce an initial coarse classification of the bounding box into sign classes, based on color (blue versus white/red signs) and/or shape (triangular versus round signs); classification may discard false positive in the detection result by considering also the not-sign class.

So far benchmarking of traffic sign recognition algorithms has focused on classification rather than detection. There are several publicly available datasets to assess the relative merits of several classification algorithms. In particular, the German Traffic Sign Recognition Benchmark (GTSRB) [3] provided a massive amount of data and an interesting evaluation of the performance of computer vision algorithm versus the human visual system. State-of-the-art algorithms obtain recognition rates on par or even slightly better than humans. Conversely, evaluation of traffic sign detection on publicly available datasets is far more difficult and unexplored. Recently, though, the organizer of GTSRB proposed a follow-up competition targeting traffic sign detection, the German Traffic Sign Detection Benchmark (GTSDB) [4]. The competition addresses separate detection of three categories: prohibitory (circular red-and-white signs), mandatory (circular blue-and-white signs) and danger signs (upper-triangular red-and-white signs). This paper describes the pipeline that we submitted for evaluation to the on–line competition.

All the authors are with the Department of Computer Science and Engineering, University of Bologna, Italy (email: { name.surname}@unibo.it).

The remainder of this paper is organized as follows: sec. II presents an overview of recent works on the subject; sec. III details all the steps of the proposed detection pipeline, i.e. image preprocessing (Sec. III-A), interest region detection (Sec. III-B), HOGs classification (sec. III-C), context-aware filter (sec. III-D) and traffic light detection and removal (sec. III-E). Sec. IV briefly presents the German Traffic Sign Detection Benchmark, so as to make the paper self-contained, and Sec. V reports the results attained by our pipeline on the three categories of the benchmark. Sec. VI concludes the paper.

II. Related Work

Traffic Sign Detection (TSD) usually relies upon two observations: that 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 usually 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 [5]. The first sub-category comprises the works by Bahlmann *et al*. [6] and by Brkic *et al*. [7], whereas in the second one 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 usually follow a common scheme: they transform the image in a color space that highlights the signs of interest and then threshold it. Several color spaces have been proposed, 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 of this common approach is represented by custom definition of a specific color transformation that highlights colors of interest: for example, the transformations defined by [14], which uses the difference of the most characteristic RGB channel of a signal and the other two channels, and the transformation defined by [15], which uses the maximum between the normalized R and B channels. Finally, another variant has been proposed, which pushes further the effort of defining a proper color

978-1-4673-6129-3/13/$31.00 ©2013 IEEE

Fig. 1
PROPOSED TRAFFIC SIGN DETECTION PIPELINE.

space for the specific task: in [16], the authors learn from training data via Integer Linear Programming a set of good color transformations and also the best threshold for each transformation.

III. PROPOSED PIPELINE

As highlighted in the previous section, the two main cues for traffic sign detection are color and shape: in our pipeline, we exploit both of them. In particular, we define candidate patches of an image those regions that exhibit a uniform value of intensity of the main color of a sign (red and white for prohibitory and danger signs, blue for mandatory signs) or show a strong symmetry. In order to extract these candidate regions we employ two interest regions detector: the well known Maximally Stable Extremal Regions (MSERs) [17] and a recently proposed symmetry detector based on the wave equation (WaDe) [18]. Color is further exploited by preprocessing the image in order to enhance the colors present in each sign of interest. Preprocessing also tries to correct over and under exposure of images that are common in outdoor pictures and set forth serious challenges to automatic detection.

The use of interest region detectors to select candidate regions is the main difference of our proposal with respect to the majority of the approaches presented in the previous section, that 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 by [15], which employs MSERs 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 are able to quickly discard background regions and let deeper, more specialized but slower classifiers analyze only a subset of regions. Although the cascade of classifiers works pretty well in practice, its main purpose is to improve the efficiency of the detector, not its performance: indeed, if a sign is discarded by higher level classifiers, it cannot be recovered by deeper classifiers (in other words, recall may only decrease as long as candidate regions move down in the cascade). Given how crucial the first stage is in determining the recall of the overall system, we advocate the use as first stage of the pipeline of an algorithm which deploys a strong prior knowledge on the appearance of the object of interest, but is robust and not computationally expensive. The Hough Transform and its variants could be used purposely, but they suffer from two shortcomings: they only exploit pixels at the boundaries of the sign, which may be occluded or camouflaged by a similar background; they are very sensitive to out-of-plane rotations of the signs.

The candidate regions extracted by the chosen interest region detector are then refined by a classifier trained on shape and color. We experimented with both Random Forests (RFs) [19] and Support Vector Machines (SVMs) [20], and found out that the latter provide better performance. For the sake of space, in this paper we report only the results obtained by using SVMs. As for features, we used HOGs [2] on color images to capture both shape and color cues within one feature.

Finally, two filters help further pruning out false positives: a generative filter discards regions that are unlikely to be traffic signs given the relationship between their size and their position in the image; a traffic light detector check if the selected region corresponds to a lamp of a traffic light, that may be erroneously considered a traffic sign by previous filters, and discards it.

The next subsections presents all the steps in details. An overview of our detection pipeline is depicted in Fig. 1.

A. Image preprocessing

Road scenes are affected by significant variations in lighting conditions so an initial preprocessing may be necessary to enhance traffic sign regions and fade background. We experimented with the following preprocessing steps. First, *contrast stretching* [21] is applied separately to the three RGB channels (*3CH-CS*) in order to deal with under or over exposed images. Then, a single channel image to be processed by the interesting region extractor is obtained by a suitable enhancing of the channel C that characterize the sought signal (blue in the case of mandatory signs, red otherwise). We experimented with two enhancing steps. The first is the simple RGB normalization:

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

The second is the image enhancing suggested in [14]. In the case of *Danger* and *Prohibitory* signals, red is enhanced as follows:

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

TABLE I

RESCALING FACTORS OF INTEREST REGIONS BOUNDING BOXES

Prohibitory MSER+	1.35
Prohibitory MSER-	1
Prohibitory MSER- Greylevel	1.35
Danger MSER+	1.6
Danger MSER-	1
Danger MSER- Greylevel	1.6
Mandatory MSER+	1
Mandatory MSER- Greylevel	1
Prohibitory WaDe-	1
Prohibitory WaDe+	1.3
Danger WaDe-	1.75
Danger WaDe+	1.75
Mandatory WaDe-	1.2

TABLE II

MSER ADN WADE PARAMETERS

MSER		WaDe	
δ	2	ρ	0.1
max. variation	0.5	T	120
min. diversity	0.2	first scale	8

TABLE III

CONSTRAINTS ON BOUNDING BOX SIZE AND ASPECT RATIO

Min. area	225
Max area	27300
Min aspect ratio	0.6
Max. aspect ratio	1.3

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

$$C' = \max(0, \frac{B - R}{R + G + B}) \quad (3)$$

The last enhancement differs from the one originally proposed in [14], which is the exact dual of (2): we found that on dark mandatory signs blue and green channels tend to have a similar value and cancel each other and, therefore we do not consider the strength of the blue channel with respect to green (as illustrated in Fig. 2). Finally, a further step of *contrast stretching* is applied to the resulting one channel image (*1CH-CS*).

B. Interest region detection

We investigated on the use of two complementary interest region extractors. To detect regions exhibiting a uniform value of intensity of the distinctive color of a sign we employ MSER [17], whereas we use the recently proposed WaDe algorithm [18] to identify highly symmetric regions. MSER detects high-contrast regions of approximately uniform gray tone of arbitrary shape by thresholding at different level and extracting the connected components that exhibit a stable area across a range of thresholds. We can distinguish between dark connected components on a brighter background (MSERs+) and bright ones surrounded by a darker background (MSERs-). WaDe detects symmetric regions at various scales as sharp spatio-temporal extrema of the one-parameter family of signals built by numerical approximation of the solution of the initial boundary value problem associated with the wave Partial Differential Equation (PDE). In particular the initial condition is given by the graylevel image under analysis and approximately absorbing boundary conditions are imposed. Likewise MSER, minima highlight bright symmetries on a dark background (WaDe-) and maxima the opposite polarity (WaDe+).

In particular, for danger and prohibitory signs we consider: MSERs+ and MSERs- from the preprocessed image; MSERs- from the original color image converted to greylevel; WaDe+ and WaDe- from the preprocessed image. For mandatory signs, we consider MSER- and WaDe- from the preprocessed image. As dark blobs are typically found in the inner white part of prohibitory or danger signs, we rescale the bounding box so as to include the whole traffic sign. We also rescale WaDe- regions as they tend to be too conservative in estimating the actual sign size. Suitable rescaling factors were found experimentally and are reported in Tab. I.

We tuned the algorithms so that they are able to detect almost all the traffic signs in the training set, regardless of the number of false positives they also feed to the further stages of the pipeline. MSER and WaDe parameters are reported in Tab. II. We pass down to the next stage of the pipeline only those bounding boxes that satisfies some loose constraints on their area and aspect ratio (see Tab. III).

C. Histogram of Oriented Gradients

Once a set of candidate regions are detected in the previous step, Histogram of Oriented Gradients (HOGs) [2] are computed in order to exploit their ability to robustly capture shape despite high intra-class variance. To capture both color and shape in one feature, we extract HOGs directly from color patches: as proposed in the original paper, in such a case the gradient at each pixel is the gradient with the greatest magnitude among the gradients computed on each of the channels. Unlike the HOGs parameters used in the GTSRB benchmark reported in [3], we found that the proper parameters are those suggested by Dalal and Triggs in they 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, that is fed to the classifier.

Despite the robustness of HOGs to different illumination conditions, some sign appears with extreme photometric distortions in the GTSDB dataset (top row of Fig. 3). Therefore, before extracting HOGs, we apply an illumination normalization to the patch: we compute the histogram of the value, $V = max(R, G, B)$, of image pixels, and then we transform each channel by mapping the median of the histogram to the midpoint of the range, i.e. 128, and by joining this point with the points $(0, 0)$ and $(255, 255)$ through two linear mappings. This piece-wise linear mapping ensures that underexposed or overexposed patches redistribute their mass more evenly

Fig. 2

ENHANCEMENT OF BLUE CHANNEL. LEFT, ORIGINAL IMAGE (A DETAIL): A VERY DARK MANDATORY SIGN IS HIGHLIGHTED; CENTER, ENHANCED IMAGE WITH BLUE/GREEN COMPARISON: THE VISIBILITY OF THE SIGN IS WORSE; RIGHT, ENHANCED IMAGE WITHOUT BLUE/GREEN COMPARISON: THE VISIBILITY IS IMPROVED.

Fig. 3

EFFECT OF THE PROPOSED ILLUMINATION NORMALIZATION.

around the range midpoint (bottom row of Fig. 3), thereby easing the classification task for subsequent stages.

HOGs extracted from the normalized patch are classified as either sign of interest or background by a binary SVM with RBF kernel. Our 10-fold cross-validation tuning on the GTSDB training set indicates proper values for C and γ to be 1 and 0.01, respectively.

D. Context-aware filter

An additional stage in our pipeline has the goal to discard wrongly detected traffic signs (i.e. reducing the number of false positives) by enforcing spatial constraints based on context information. The enforced feature exploits the typical position of traffic signs in urban environments, in particular it forces signs to appear not too close to the ground floor nor too above the horizon line. More specifically, the feature relies on two main cues: the size and height of a traffic sign in the image. It is indeed easy 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. This assumption relies on the fact that the camera is mounted in a fixed position on the vehicle throughout processed frames, this allowing to simplify the problem by assuming a rather constant horizon position in each frame. The limited range that the sign heights tends to assume when its 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.

To estimate the sign height h, we have selected the vertical coordinate of the top-left corner of its enclosing bounding box normalized by the number of image rows, while as for its size, r, we have chosen the number of rows in its enclosing bounding box normalized by the minimum and maximum size assumable by roadsigns. These two normalization steps allows the two parameters r and h to span the range $[0, 1]$.

The approach we propose is a generative classifier, with a parametric model explaining positives samples (i.e. true traffic signs) learned based on training data. As a parametric model for the relationship between r and h, we propose to use a family of one-dimensional Gaussian distributions parametrized by the possible normalized radius values r, each modeling the probability distribution of the normalized height value h:

$$p\left(y|r\right) \sim \mathcal{N}\left(\mu\left(r\right), \sigma\left(r\right)\right) \tag{4}$$

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

$$\mu(r) = a_\mu \cdot r + b_\mu \tag{5}$$

$$\sigma(r) = a_\sigma \cdot r^2 + b_\sigma \cdot r + c_\sigma \tag{6}$$

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 sensibly the results with the GTSDB dataset). Accordingly, the set of parameters Θ includes the coefficients of the

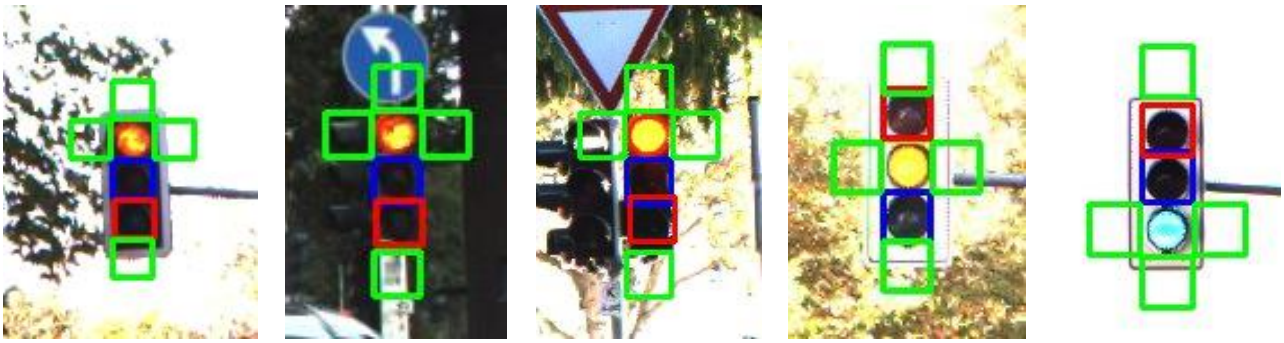

Fig. 4

EXAMPLES OF TRAFFIC LIGHTS DETECTED BY THE PROPOSED FILTER. FOR EACH EXAMPLE, AUXILIARY ROIS ARE DEPICTED RESPECTIVELY IN GREEN ($ROI_l, ROI_r, ROI_u, ROI_d$), BLUE ($ROI_{ref}$) AND RED ($ROI_{in}$). [FIGURE BEST VIEWED IN COLORS]

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 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 have employed a weighted least-square estimation, to render the coefficient estimation stage less biased by outliers. The weights in this case are represented by the available training population for each value of variable r.

At run-time, the likelihood that a test sample (r, h) represents a traffic sign according to our model can be calculated as:

$$p\left(y|r\right) = \frac{1}{\sqrt{2\pi}\sigma\left(r\right)} \exp\left(-\frac{\left(r-\mu\left(r\right)\right)^2}{2\sigma^2\left(r\right)}\right) \qquad (7)$$

The likelihood is then thresholded to discard possibly wrong traffic sign detection hypotheses. A major advantage of the proposed generative approach is that the learning stage does not require negative samples, which indeed would be hard to deploy in our scenario given their dependence on the actual classification method and parameters employed in the previous stage of the pipeline.

Given the three categories of traffic sign that need to be detected, we have decided to separately learn two different parameter sets, one for the *Danger* category, and one, jointly, for both *Prohibitory* and *Mandatory* signs, given the similar size of the signs falling belonging to the last two categories.

E. Traffic light filter

The presence of traffic lights in urban images represents a major nuisance for automatic traffic sign detection, due to shape, position and luminosity of the traffic light lamps being often similar to that of certain traffic signs (in particular concerning round-shaped categories, such as *Mandatory* and *Prohibitory*). This problem is worsened by traffic lights being frequent in road images. 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 also for other tasks within similar computer vision applications.

The proposed approach follows a two step procedure to examine and potentially discard each previously detected Region Of Interest (ROI), based on certain working assumptions. First of all, each evaluated RoI is assumed as being centered on one of the lamps of the traffic light. Secondly, we assume that all traffic lights are characterized by three lamps, with at most one of them being switched on at a certain time frame. This obviously do not consider specific situations such as two-lamp traffic lights, which are anyway not particularly frequent sources of errors.

Given these premises, 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, 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 hypothesizing the complete 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. 4, two of them, referred to as ROI_{ref}, ROI_{in}, are centered on the two switched-off lamps of the traffic light (given the current assumptions, there are always at least two switched-off lamps) 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 based on the minimization of the Euclidean distance δ is deployed to refine the position and scale of ROI_{in} with respect to that

Fig. 5
EXAMPLAR SCENES FROM THE GTSDB DATASET.

of ROI_{ref}. The remaining four auxiliary ROIs (see again Fig. 4) are extracted in four locations outside the hypothesized traffic light, respectively along the left, right, up and down directions (referred as $ROI_l, ROI_r, ROI_u, ROI_d$). The condition applied to test whether a ROI includes or not a traffic light is thus:

$$\frac{\delta\left(ROI_{ref}, ROI_{in}\right)}{\min\limits_{i\in\{l,r,u,d\}} \delta\left(ROI_{ref}, ROI_i\right)} < \tau_{tl} \tag{8}$$

By thresholding the relative distances between inner and outer ROIs, the whole traffic light detection is normalized with respect to the current photometric conditions related to image frame acquisition. 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 that the mean and variance of the pixels within the circles inscribed in ROI_{ref}, ROI_{in} must be smaller than a certain threshold. Fig. 4 reports some examples of traffic lights detected with the proposed approach, also depicting the auxiliary ROIs deployed by the method.

IV. THE GERMAN TRAFFIC SIGN DETECTION BENCHMARK

The German Traffic Sign Detection Benchmark [4] is an on-line traffic sign detection competition. The dataset made available for download to the participants features 900 images (split into 600 training images and 300 evaluation images) with very tough signs' size and illumination conditions variations (some exemplar images are reported in Fig. 5). As mentioned in the Introduction, the partecipants are not only required to detect signs but also to classify them in three coarse categories, based on shape and color: prohibitory (circular red-and-white signs), mandatory (circular blue-and-white signs) and danger signs (upper-triangular red-and-white signs).

V. RESULTS

This section reports results of the proposed pipeline at the end of the submission phase of the GTSDB test dataset (i.e. the data released in the test phase of the competition, which were not used for training). The best precision-recall curves for the three categories are depicted in Fig. 6. The

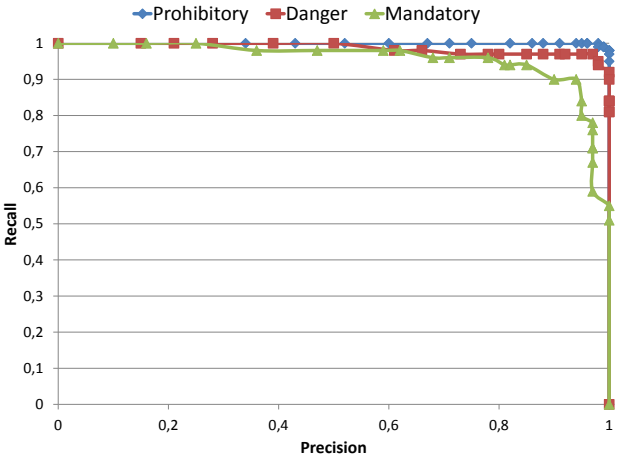

Fig. 6
RESULTS OF THE PROPOSED PIPELINE OVER THE THREE CATEGORIES OF THE GTSDB DATASET. THE CORRESPONDING AUCS ARE 99.98% FOR THE PROHIBITORY CLASS, 98.72% FOR THE DANGER CLASS AND 95.76% FOR THE MANDATORY CLASS.

corresponding Areas Under the Curve (AUCs) are [1] 99.98% for the prohibitory class, 98.72% for the danger class and 95.76% for the mandatory class.

It turned out that the best configuration of the pipeline stages is almost identical for prohibitory and danger signs: the only difference in this case is that the danger pipeline can avoid the traffic light removal step, as the HOGs+SVM classifier trained on triangular signs already discards circular patterns. For this two categories, MSER performs better than WaDe: although both are able to detect all the prohibitory signs and all but two danger signs in the training set, it seems that symmetric regions detected by WaDe are more ambiguous for the HOGs+SVM classifier than less regular regions typically detected by MSER. As red-and-white signs tend to mantain a good constrast between the external red circle and the internal white area even when underexposed, we found

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

that it was not necessary to apply *3CH-CS* and *1CH-CS*; there is also no difference in performing normalization of the red channel (eq. 1) or red enhancing and normalization (eq. 2).

The situation is totally different for mandatory signs. They turned out to be the most challenging class of the GTSDB for our pipeline; they are also the class where there is only one result above 99% AUC in the on-line competition, whereas there are several results above this threshold for the other two classes. We ascribe the inherent difficulty to different behavior of mandatory sign with respect to varying illumination conditions, in particular when they are underexposed: the blue region tends to become particularly dark and therefore its gradient are less strong and it becomes harder to distinguish it from the surrounding background. Indeed, in our experiments it turned out that preprocessing of the input image as well as blue enhancing according to equation (3) is necessary to detect all but one mandatory signs in the training images. For this class, WaDe is definitely more effective than MSER: the white arrows can create separated blue regions out of a single mandatory sign, especially if the sign is small, which in turns lead to the detection of several MSERs with wrong scales. On the other hand, the circular symmetry of the sign remains evident and is captured by WaDe.

Overall, our pipeline ranked 4th out of 42 submissions on prohibitory signs, 8th out of 26 submissions on danger signs and 5th out of 30 submissions on mandatory signs. The processing time on a standard PC of an unoptimized implementation of our detection pipelines are: 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.

VI. Conclusions

The traffic sign detection system we propose demonstrates good performance under challenging conditions such as extremely dark illumination, saturated background, 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 a very good result on the challenging class of mandatory signs. This can be ascribed to the use of interest region detectors, which allows for notably reducing the number of evaluated candidates (thus, possible ambiguities at classification stage) with respect to a sliding window approach, as well as to the use of effective false positive filters based on traffic light detection and context information. Future directions of research concerns experimenting and developing novel representations for interest regions which could be more inherently based on color information, while still proving robust to severe photometric variations found in typical outdoor conditions.

Acknowledgements

We wish to thank QONSULT SPA for funding the research activity leading to the results presented in this paper.

References

[1] P. Viola and M. Jones, "Robust real-time object detection," in *International Journal of Computer Vision*, 2001.

[2] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *International Conference on Computer Vision & Pattern Recognition*, C. Schmid, S. Soatto, and C. Tomasi, Eds., vol. 2, INRIA Rhône-Alpes, ZIRST-655, av. de l'Europe, Montbonnot-38334, June 2005, pp. 886–893. [Online]. Available: http://lear.inrialpes.fr/pubs/2005/DT05

[3] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel, "Man vs. computer: Benchmarking machine learning algorithms for traffic sign recognition," *Neural Networks*, no. 0, pp. –, 2012. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0893608012000457

[4] S. Houben, J. Stallkamp, J. Salmen, M. Schlipsing, and C. Igel, "Detection of traffic signs in real-world images: The German Traffic Sign Detection Benchmark," in *International Joint Conference on Neural Networks (submitted)*, 2013.

[5] D. H. Ballard, "Generalizing the hough transform to detect arbitrary shapes," *Pattern Recognition*, vol. 13, no. 2, p. 111122, 1981.

[6] C. Bahlmann, Y. Zhu, and V. Ramesh, "A system for traffic sign detection, tracking, and recognition using color, shape and motion information," in *Proc. IEEE Symposium on Intelligent Vehicles*, 2005, p. 255260.

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

[8] G. Loy, "Fast shape-based road sign detection for a driver assistance system," in *Proc. IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2004, p. 7075.

[9] N. Barnes, A. Zelinsky, and L. Fletcher, "Real-Time Speed Sign Detection Using the Radial Symmetry Detector," in *Intelligent Transportation Systems, International IEEE Conference on*, vol. 9, no. 2, jun 2008, pp. 322–332.

[10] R. Belaroussi and J.-P. Tarel, "Angle vertex and bisector geometric model for triangular road sign detection," in *Proc. IEEE Workshop on Applications of Computer Vision (WACV)*, 2009, p. 577583.

[11] ——, "A real-time road sign detection using bilateral chinese transform," in *Proc. IEEE Workshop on Applications of Computer Vision (WACV)*, 2009, p. 11611170.

[12] S. Houben, "A single target voting scheme for traffic sign detection," in *Intelligent Vehicles Symposium*. IEEE, jun 2011, pp. 124–129. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5940429

[13] H. Gomez-Moreno, S. Maldonado-Bascon, P. Gil-Jimenez, and S. Lafuente-Arroyo, "Goal Evaluation of Segmentation Algorithms for Traffic Sign Recognition," *Intelligent Transportation Systems , IEEE Transactions on*, vol. 11, no. 4, pp. 917–930, dec 2010. [Online]. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=5508422

[14] A. Ruta, Y. Li, and X. Liu, "Real-time traffic sign recognition from video by class-specific discriminative features," *Pattern Recognition*, vol. 43, no. 1, pp. 416–430, jan 2010. [Online]. Available: http://dl.acm.org/citation.cfm?id=1595888.1595967

[15] J. Greenhalgh and M. Mirmehdi, "Real-Time Detection and Recognition of Road Traffic Signs," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 4, pp. 1498–1506, dec 2012.

[16] R. Timofte, K. Zimmermann, and L. Van Gool, "Multi-view traffic sign detection, recognition, and 3D localisation," *Machine Vision and Applications*, pp. 1–15, dec 2011. [Online]. Available: http://link.springer.com/article/10.1007/s00138-011-0391-3

[17] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust wide-baseline stereo from maximally stable extremal regions," *Image and Vision Computing*, vol. 22, no. 10, pp. 761 – 767, 2004, ¡ce:title¿British Machine Vision Computing 2002¡/ce:title¿. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0262885604000435

[18] S. Salti, A. Lanza, and L. Di Stefano, "Keypoints from symmetries by wave propagation," in *2013 IEEE International Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE, June 2013.

[19] L. Breiman, "Random forests," in *Machine Learning*, 2001, pp. 5–32.

[20] C. Cortes and V. Vapnik, "Support-vector networks," in *Machine Learning*, 1995, pp. 273–297.

[21] R. C. Gonzalez and R. E. Woods, *Digital Image Processing (3rd Edition)*. Prentice Hall, Aug. 2007.