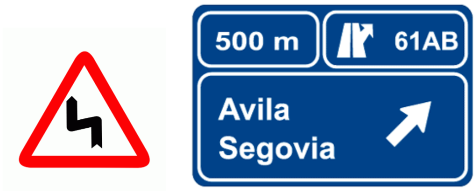
Traffic Panels Detection Using Visual Appearance

Á. González, L.M. Bergasa, J. Javier Yebes and J. Almazán

Abstract— **Traffic signs detection has been thoroughly studied for a long time. However, road panels detection still remains a challenge in computer vision due to the huge variability of types of traffic panels, as the information depicted in them is not restricted. This paper presents a method to detect traffic panels in street-level images as an application to Intelligent Transportation Systems (ITS), since the main purpose can be to make an automatic inventory of the traffic panels located in a road to support maintenance and to assist drivers in order to improve human quality of life. The proposed method extracts local descriptors at some interest points after applying a color detection method for blue and white pixels. Then, the images are modeled using a Bag of Visual Words technique and classified using Naïve Bayes theory and SVM. Experimental results on real images from Google Street View prove the efficiency of the proposed method and give way to using street-level images for different applications on robotics and ITS.**

I. INTRODUCTION

This paper presents an approach for detecting the presence or absence of traffic panels on street-level images using computer vision techniques. Traffic panels are a special case of traffic signs. They are typically rectangular big signs that are located above the road or at the side of the road. They are aimed at depicting some kind of information to the road users, typically information related to the road itself, distance to the next town, direction of the next exit, etc. Therefore, the information depicted in road panels is not restricted, unlike traffic signs which represent certain information (see Fig. 1 to understand the differences between traffic signs and road panels). Most of the organisations responsible for managing the road networks are interested in having up-to-date inventories of the road furniture to support maintenance and cost control. Traffic signs and panels are of especial interest due to the fact that sign visibility degrades due to aging and other reasons such as vandalism, accidents, pollution or vegetation coverage. In addition, during the recent years several private companies and public organisations have started to record street-level panoramic images. The most well-known service is Street View provided by Google. Computer vision techniques on these images simplify the automatic creation of traffic signs inventories, minimizing the human interaction. These inventories can be useful for ITS applications, such as road maintenance and driver assistance, and even for robotic applications to help visually impaired people.



(a) Traffic sign (b) Traffic panel

Fig. 1. An example of a traffic sign and a traffic panel

However, traffic panels detection still remains a very challenging problem due to several reasons. Firstly and above all, there is a huge variability of traffic panels as each of them depicts different information. Therefore, traffic panels vary in size, color and shape. Moreover, there are large viewpoint deviations due to the fact that the images are captured from a driving vehicle. There may also be occlusions due to vegetation or other road users. In addition, weather and illumination conditions are a key problem in any kind of vision-based system. Apart from this, there are many elements in the roads or close to the roads that can be easily confused with traffic panels, such as advertisement panels or trucks.

In this paper we focus on detecting the presence of traffic panels in street-level images. We simply constrain to blue and white background color panels, as our dataset has been obtained from the Spanish road network and most of the panels there have a blue or a white background. The main contribution of this paper is that we use visual appearance techniques to detect the traffic panels. In other words, we model the panels using local descriptors and classify the new samples using panel appearance, instead of using other features such as edges or geometrical characteristics. In addition, we focus on detecting traffic panels as opposed to most of the works of the state of the art, which have concentrated their efforts only on detecting traffic signs, which have a higher intra-class correlation, *i.e.* the variability of traffic signs of the same class is lower. They are typically of the same size, the same shape, the same color and they always depict the same information. This is not the case of traffic panels.

The remainder of the paper is organized as follows. In section II, we make an overview of the state of the art on traffic panels detection. Section III describes the process of capturing the images. Section IV explains the implemented approach for training the system and classifying new input images. Section V provides the experimental results and section VI concludes the paper.

Authors are with the Department of Electronics, Escuela Politécnica, Universidad de Alcalá, 28871 Alcalá de Henares, Madrid, SPAIN. Email: {alvaro.g.arroyo;bergasa;javier.yebes;javier.almazan}@depeca.uah.es

II. STATE OF THE ART

Detection and recognition of traffic signs has been studied for a long time. However, there has not been much research on detection of traffic panels. Some of the reasons have been stated in the previous section. The main cause may be the fact that the variability of traffic panels is immense. In other words, there are not two identical traffic panels. This fact makes the training of automatic traffic panels detection and recognition systems very difficult. From our knowledge, only four works have been developed in this subject.

The first one [1] detects candidates to be traffic panels using a segmentation method that detects blue and white colors from the hue and saturation components of the HSI space. Resulting connected components are analyzed and those which do not fulfill certain geometrical constraints in aspect ratio or size are discarded. Then, the resulting candidates are classified into panel or not by correlating the radial signature of their Fast Fourier Transform with the pattern corresponding to an ideal rectangular shape. This algorithm is invariant to rotations, deformations and camera projection distortions, but it is very sensitive to changing lighting conditions.

On the other hand, Chen et al. [2] extract regions of the same color using a k-means algorithm. Road signs candidates are extracted by searching for flat regions perpendicular to the camera axis and considering some a priori knowledge of the geometry of the panels. The main advantage of this technique is its high computational capacity. In addition, it provides good results under different lighting conditions and it is not affected by rotations and projective distortion. However, the segmentation method based on Gaussian Mixture Models depends highly on the contrast between foreground and background, which is affected at the same time by lighting conditions.

An edge image is firstly obtained in [3] using the Canny edge detector. Then, the authors look for contours in the edge image and they are analysed using some aspect constraints. Finally, the Hough transform is applied over the contours to select those which belong to certain shapes (rectangular, circular and triangular) in order to extract the traffic signs.

Finally, traffic panels depict information in terms of text and symbols. The same authors in [4] propose to use a text detection algorithm in first place in order to detect the text present in the image. Then, the regions in the image where there is a high density of text are classified as road panels. However, it is complemented with a panel detection algorithm that is applied after the text detection method. This technique uses both color and edge information. This method achieves very good performance and it is not affected by rotations, scaling or distortions. However, the main disadvantage of this technique is its high computational time due to the fact that it applies the text detection method over all the images, independently if there is not any panel in the image.

The aim of the method here proposed is to continue the work developed in [4]. The idea is to detect only the images where there are road panels with the method presented in this paper and to apply the text detection and recognition method proposed in [4] in order to reduce the computational time and to increase the efficiency of the text detection method.

III. IMAGE CAPTURE

The images used in this work have been obtained from the Street View service by Google. It provides high-resolution 360^{o} panoramic views from various positions along many streets and roads in the world. It is possible to zoom in on each panoramic image. Fig. 2 shows the first 5 zoom levels (0-4) for a certain view. At each zoom level, the image is given in 512-pixel square tiles. For our purpose of detecting traffic panels, we have chosen a zoom level of 4 and we have cropped the panoramic view to the region shown in red in Fig. 2(e), that is, the tiles $(x = 6, y = 2)$, $(x = 6, y = 3)$, the right half side of the tiles $(x = 5, y = 2)$ and $(x = 5, y = 3)$ and the left half side of the tiles $(x = 7, y = 2)$ and $(x = 7, y = 3)$. This region is optimum to detect the panels located above and on the right margin of the road.

IV. IMPLEMENTATION

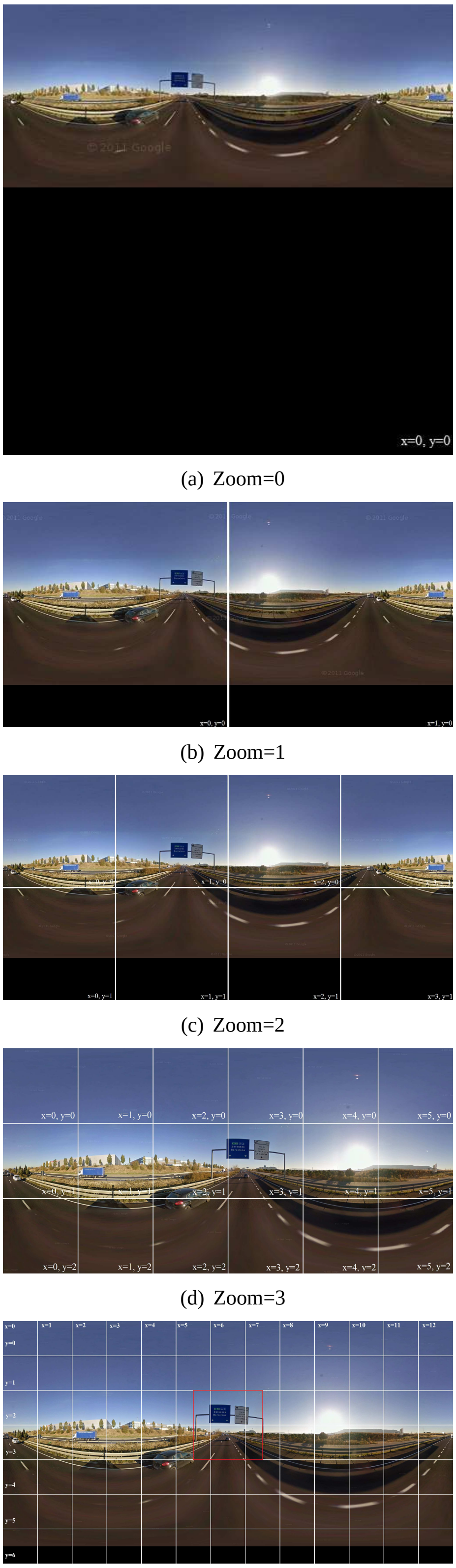
The main objective of this system is to detect the presence of blue-background and white-background traffic panels in the images, either located on the right side of the road or above the road. In first place, a total of 16277 images has been extracted and three independent subsets of images have been obtained, one for training the system (approximately 50% of the images), one for validating it (around 25% of the images) and the last one for testing it (25% of the images). All the images have been obtained from street-level images of the Spanish road network, specifically from the roads shown in Fig. 3 (the training and validation sets from the roads shown in red and the test set from the roads shown in blue).

Since the traffic panels are located above the road or on the right side of it, two independent regions of interest have been applied on the images. These regions are shown in Fig. 4. In addition, as there are blue-background and white-background traffic panels, four independent training subsets have been created: one for blue-background panels located above the road, another one for blue-background panels on the right side of the road, the third one for white-background panels above the road and the last one for white-background panels on the right side of the road.

A method that detects blue pixels and white pixels in the images has been developed. The goal is to compute the features in the image only where it is likely to be a traffic panel in order to minimize the number of false positives. We propose to detect the blue regions in the image as a combination of three independent methods using a logical AND operation as in (1).

$$BlueMask = g_1(x,y)\ AND\ g_2(x,y)\ AND\ g_3(x,y) \quad (1)$$

$g_1(x,y)$ is computed using (2) as it was proposed in [5]. $R(x,y)$ is the red channel of the image and $T_r = 90$ has been found to be the optimum value using genetic algorithms.



(a) Zoom=0

(b) Zoom=1

(c) Zoom=2

(d) Zoom=3

(e) Zoom=4 and Region of Interest in the panoramic view (red)

Fig. 2. [Best viewed in color] Different zoom levels for a panoramic view

$$g_1(x,y) = \begin{cases} 255 & \text{if } R(x,y) \leq T_r \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

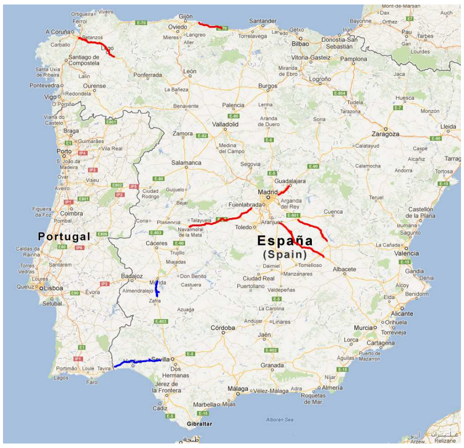
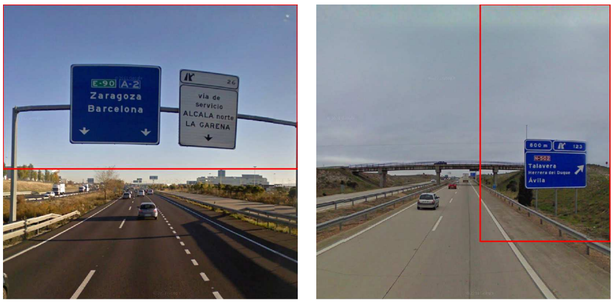


Fig. 3. [Best viewed in color] Roads from which the images have been obtained: training and validation sets (red) and test set (blue)



(a) Upper region of interest

(b) Lateral region of interest

Fig. 4. Regions of interest on the images

On the other hand, $g_2(x,y)$ is computed using (3) as it was proposed in [6]. $H(x,y)$ is the Hue component of the image and $T_1 = 200°$ and $T_2 = 280°$ are the optimum values of the thresholds. These values have been optimized using genetic algorithms again.

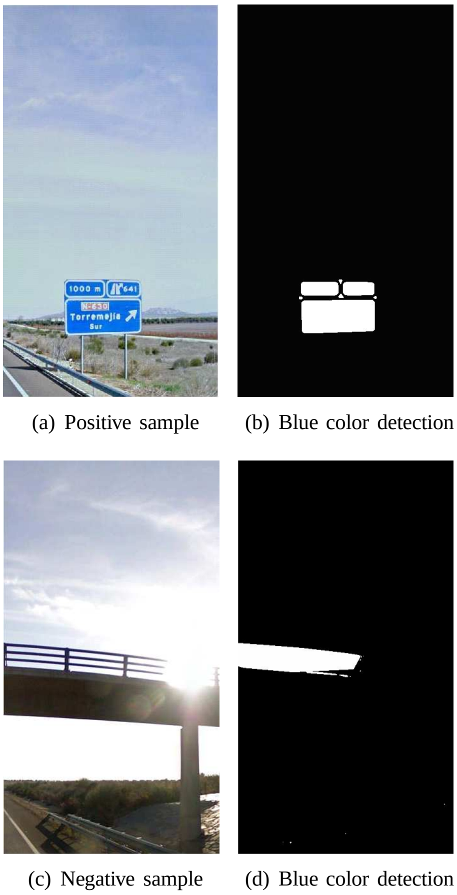
$$g_2(x,y) = \begin{cases} 255 & \text{if } H(x,y) \geq T_1 \ and \ H(x,y) \leq T_2 \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

Finally, we propose to compute $g_3(x,y)$ using (4), which consists of applying the Otsu segmentation method [7] on the image obtained by subtracting the blue color component from the red color one.

$$g_3(x,y) = Otsu(|R(x,y) - B(x,y)|) \tag{4}$$

Figure 5 shows the result of applying this blue color detection method on two images (a positive and a negative sample).

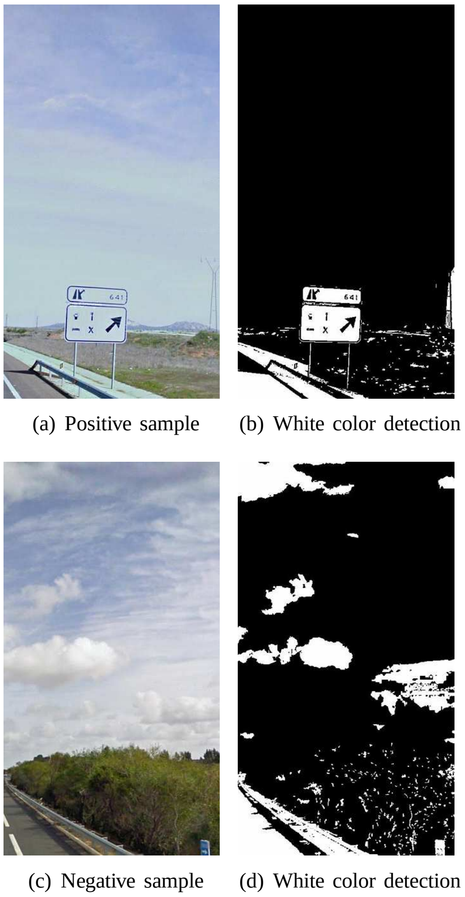
On the other hand, the method to detect white regions in the image is based on the Maximally Stable Extremal Regions method (MSER) [8], which is a region detector that allows to detect bright-on-dark regions in the image. Figure 6 shows an example of applying this white color detection method on an image with a traffic panel and on an image without any panel.



(a) Positive sample (b) Blue color detection

(c) Negative sample (d) Blue color detection

Fig. 5. Blue color detection



(a) Positive sample (b) White color detection

(c) Negative sample (d) White color detection

Fig. 6. White color detection

Then, feature descriptors are computed on the masks obtained after applying the blue and white color detection methods. The descriptors are extracted at some interest points, which are obtained using the Harris-Laplace salient point detector [9]. It uses a Harris corner detector and subsequently the Laplacian for scale selection.

We propose to represent the images with a Bag of Visual Words (BOVW) technique [10]. This method models an image as a sparse vector of occurrence counts of visual words. In other words, it translates a very large set of high-dimensional local descriptors into a single sparse vector of fixed dimensionality across all images. To do so, the feature space of the local image descriptors is quantized into a discrete number of visual words using k-means clustering. In this case, we have found that the optimum size of the vocabulary is $k = 300$, as it will be shown in section V. The visual words are the cluster centers. The image is represented as a histogram which counts how many times each of the visual words occurs in the image and the classes or categories are learned by a Naïve Bayes classifier [11] using this vector representation. Therefore, given a new image, the nearest visual word is identified for each of its features using the Euclidean distance between the cluster centers and the input descriptor and the classification decision is made by the Naïve Bayes classifier previously trained.

A comparison of different grey-based and color-based descriptors has been carried out. Specifically, the following descriptors have been used: SIFT [12], C-SIFT [13], Hue-SIFT [14], RGB-SIFT [15], Hue Histogram [14] and Transformed Color Histogram (TCH) [15]. Only with SIFT, Hue Histogram and TCH it has been possible to successfully cluster the descriptors and train the classifier, as the classification error rate with the other descriptors was higher than 70%. The dimensionality of the SIFT descriptor used is 64 elements and it is computed from the grey-level image. On the other hand, the dimensionalities of the Hue Histogram descriptor and TCH are 37 and 45 elements respectively. Hue Histogram is computed from the hue and saturation color models, while TCH is computed from the red, green and blue color components after normalizing each channel independently.

V. Experimental Results

The test images are completely independent to the training set in order to assure the reliability of the results. Tables I-IV show the results for each defined class: blue-background lateral panels, blue-background panels above the road, white-background lateral panels and white-background panels above the road. Table V shows the results for all the panels on the right of the road regardless of their color, while

Table VI shows the results for the panels above the road regardless of the color. The results are shown in terms of detection rate, sensitivity and specificity. The detection rate is the percentage of correctly detected panels. Usually a panel appears in several images at different distances. In case the algorithm detects a panel in at least one of the images where it appears, we count it as a correct detection. Therefore, the detection rate is computed in multi-frame. On the other hand, sensitivity and specificity are computed in single-frame. They are defined as in (5) and (6).

$$Sensitivity = \frac{TP}{TP\ +\ FN} \tag{5}$$

$$Specificity = \frac{TN}{TN\ +\ FP} \tag{6}$$

TP stands for the number of true positives, FN stands for the number of false negatives, TN is the number of true negatives and FP is the number of false positives. The sensitivity measure relates to the system's ability to identify positive samples, while the specificity relates to the system's ability to identify negative samples. In order to join both measures into one, the f-measure is defined in (7).

$$f = \frac{Sensitivity + Specificity}{2} \tag{7}$$

It can be seen that the best results are obtained for the color descriptors, being TCH the best one. The detection rate is above 95% for the four situations under study and the value of the f-measure is the highest in all cases except for blue panels located on the side of the road, although it is very close to the highest value which is obtained with the Hue Histogram descriptor. However, the highest value of the specificity measure is achieved in most cases for the SIFT descriptor. It means that the number of false positives for this descriptor is very low. Nevertheless, the sensitivity is much lower for SIFT respect to the other descriptors. It means that the number of false negatives is very high respect to the number of true positives. In other words, the classifier trained with the SIFT descriptor categorizes most of the images as if there is not any panel present in the image. That is the reason why the detection rate for SIFT is so low respect to the other descriptors.

TABLE I

EXPERIMENTAL RESULTS FOR BLUE LATERAL PANELS

Descriptor	Detection rate	Sensitivity	Specificity	f
SIFT	67.86%	0.2500	0.9192	0.5846
Hue Histogram	94.05%	0.6625	0.8782	0.7704
Transformed Color Histogram	98.81%	0.6042	0.9253	0.7674

TABLE II

EXPERIMENTAL RESULTS FOR BLUE UPPER PANELS

Descriptor	Detection rate	Sensitivity	Specificity	f
SIFT	86.66%	0.5366	0.9789	0.7577
Hue Histogram	100%	0.9512	0.8438	0.8438
Transformed Color Histogram	100%	0.8963	0.9536	0.9300

TABLE III

EXPERIMENTAL RESULTS FOR WHITE LATERAL PANELS

Descriptor	Detection rate	Sensitivity	Specificity	f
SIFT	45.83%	0.1724	0.9264	0.5494
Hue Histogram	58.33%	0.3563	0.6107	0.4835
Transformed Color Histogram	95.83%	0.6552	0.5079	0.5815

TABLE IV

EXPERIMENTAL RESULTS FOR WHITE UPPER PANELS

Descriptor	Detection rate	Sensitivity	Specificity	f
SIFT	75%	0.3740	0.9542	0.6641
Hue Histogram	93.75%	0.8293	0.6827	0.7560
Transformed Color Histogram	96.88%	0.7480	0.8998	0.8238

TABLE V

EXPERIMENTAL RESULTS FOR ALL THE LATERAL PANELS

Descriptor	Detection rate	Sensitivity	Specificity	f
SIFT	62.96%	0.3304	0.8511	0.5907
Hue Histogram	86.11%	0.7625	0.5385	0.6505
Transformed Color Histogram	98.15%	0.7464	0.4772	0.6118

TABLE VI

EXPERIMENTAL RESULTS FOR ALL THE UPPER PANELS

Descriptor	Detection rate	Sensitivity	Specificity	f
SIFT	81.82%	0.5708	0.9394	0.7551
Hue Histogram	97.40%	0.9292	0.5975	0.7634
Transformed Color Histogram	98.70%	0.8821	0.8817	0.8819

It has been found that the optimum number of visual words is $k = 300$. Figure 7 shows how the f-measure for blue lateral panels using the TCH descriptor varies as a function of the size of the vocabulary. It can be seen that the value of f increases rapidly from 25 to 300 visual words and then it tends to be asymptotic from 300 onwards. Therefore, we have chosen $k = 300$ as the size of the vocabulary, because with a higher number of visual words the training is slower and the testing is higher but the results obtained do not change drastically.

Finally, a different classifier apart from Naïve Bayes has been tested. This classifier is based on Support Vector Machine (SVM) [16] with linear kernel. We have found

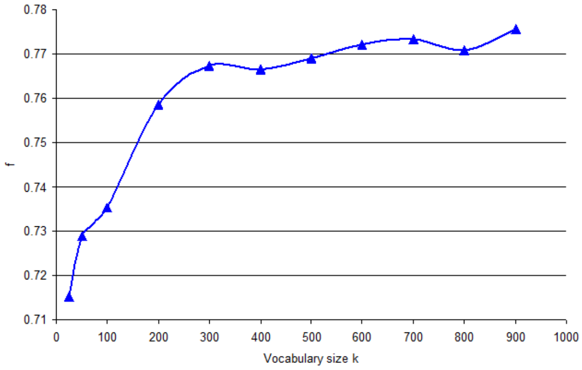


Fig. 7. f-measure as a function of the size of the vocabulary for TCH descriptor

that, in general, the number of false positives using SVM is much lower than using Naïve Bayes and, therefore, the specificity is higher. However, the number of false negatives (when the algorithm does not detect a panel but there is one in reality) is higher and consequently the sensitivity is lower than if a Naïve Bayes classifier is used. The panel detection rate is also lower and, in addition, we have seen that the computational time using SVM is much higher than using the original Naïve Bayes classifier. Therefore, in this application it is preferred to use Naïve Bayes against SVM. As an example, the comparison between Naïve Bayes and SVM using TCH for blue-background panels located on the side of the road is shown in Table VII.

TABLE VII

EXPERIMENTAL RESULTS FOR BLUE LATERAL PANELS

Classifier	Detection rate	Sensitivity	Specificity	f
Naïve Bayes	98.81%	0.6042	0.9253	0.7674
SVM	90.48%	0.4167	0.9794	0.6980

VI. CONCLUSIONS AND FUTURE WORKS

This paper has presented an approach for detecting road panels in street-level images. The main contribution of this work is the modelling of traffic panels using a BOVW technique from local descriptors detected at interest keypoints, instead of using other features such as edges or geometrical characteristics as it has been done up to now in the literature. This is not an easy task due to the immense variability of the traffic panels. However, the experimental results show the effectiveness of the proposed method. Using a color descriptor like TCH, a panel detection rate higher than 95% is achieved. In addition, as the dimensionality of this descriptor is small (only 45 elements), the training time is lower than using other descriptors. A comparison of different descriptors has been carried out and the best results are obtained for TCH.

As future work, we intend to continue the work developed in [4], where a text detection and recognition method for road panels was presented. In that work, the text detection algorithm was applied over the entire image, independently if there is a panel present or not. Therefore, the efficiency of the method is not very high. This could be improved if the text detection algorithm is applied only on the images where there are road panels, which is achieved by the method here proposed.

ACKNOWLEDGMENTS

The authors are grateful for the cooperation rendered by the Community of Madrid in the Robocity2030-II CAM-S2009/DPI-1559 project and by the Spanish Ministerio de Economia y Competitividad through the project Smart Driving Applications (TEC2012-37104). The authors would also like to thank Google for the images, which have been used only for research purposes.

REFERENCES

[1] A.V. Reina, R.L. Sastre, S.L. Arroyo and P.G. Jiménez, “Adaptive traffic road sign panels text extraction”, *Proceedings of 5th WSEAS International Conference on Signal Processing, Robotics and Automation*, pp. 295-300, 2006.

[2] X. Chen, W. Wu and J. Yang, “Detection of text on road signs from video”, *IEEE Transactions on Intelligent Transportation Systems*, vol. 6, pp. 378-390, 2005.

[3] M.A. García-Garrido, M.A. Sotelo and E. Martín, “Fast road sign detection using Hough transform for assisted driving of road vehicles”, *Eurocast 2005*, 2005.

[4] Á. González, L.M. Bergasa, J.J. Yebes and J. Almazán, “Text Recognition on Traffic Panels from Street-level Imagery”, *IEEE Intelligent Vehicles Symposium (IV)*, 2012.

[5] N. Kulkarni, “Color Thresholding Method for Image Segmentation of Natural Images”, *International Journal of Image, Graphics and Signal Processing*, vol.4, no.1, pp.28-34, 2012.

[6] H. Gómez-Moreno, S. Maldonado-Bascón, P. Gil-Jiménez and S. Lafuente-Arroyo, “Goal evaluation of segmentation algorithms for traffic sign recognition”, *IEEE Transactions on Intelligent Transportation Systems*, vol. 11, no. 4, pp. 917-930, 2010.

[7] N. Otsu, “A threshold selection method from gray-level histograms”, *IEEE Transactions on Systems, Man and Cybernetics*, vol. 9, no. 1, pp. 62-66, 1979.

[8] J. Matas, O. Chum, M. Urban, and T. Pajdla, “Robust wide baseline stereo from maximally stable extremal regions”, *Proceedings of British Machine Vision Conference (BMVC)*, pp. 384-396, 2002.

[9] K. Mikolajczyk and C. Schmid, “Indexing based on scale invariant interest points”, *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, vol. 1, pp. 525-531, 2001.

[10] G. Csurka, C.R. Dance, L. Fan, J. Willamowski and C. Bray, “Visual categorization with bags of keypoints”, *Proceedings of European Conference on Computer Vision (ECCV)*, 2004.

[11] D.D. Lewis, “Naïve Bayes at forty: The independence assumption in information retrieval”, *Lecture Notes in Computer Science*, vol. 1398, Machine Learning: ECML-98, pp. 4-15, 1998.

[12] D. Lowe, “Object recognition from local scale-invariant features”, *Proceedings of IEEE International Conference on Computer Vision (ICCV)*, vol. 2, pp. 1150-1157, 1999.

[13] A.E. Abdel-Hakim and A.A. Farag, “CSIFT: A SIFT descriptor with color invariant characteristics”, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1978-1983, 2006.

[14] J. van de Weijer, T. Gevers and A. Bagdanov, “Boosting color saliency in image feature detection”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 1, pp. 150-156, 2006.

[15] K.E.A. van de Sande, T. Gevers and C.G.M Snoek, “Evaluating Color Descriptors for Object and Scene Recognition”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1582-1596, 2010.

[16] C. Cortes and V. Vapnik, “Support-Vector Networks”, *Machine Learning*, vol. 20, no. 3, pp. 273-297, 1995.