Text Recognition on Traffic Panels from Street-level Imagery

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

***Abstract*— Text detection and recognition in images taken in uncontrolled environments still remains a challenge in computer vision. This paper presents a method to extract the text depicted in road panels in street view images as an application to Intelligent Transportation Systems (ITS). It applies a text detection algorithm to the whole image together with a panel detection method to strengthen the detection of text in road panels. Word recognition is based on Hidden Markov Models, and a Web Map Service is used to increase the effectiveness of the recognition. In order to compute the distance from the vehicle to the panels, a function that estimates the distance in meters from the text height in pixels has been obtained. After computing the direction vector of the vehicle, world coordinates are computed for each panel. Experimental results on real images from Google Street View prove the efficiency of our proposal and give way to using street-level images for different applications on ITS such as traffic signs inventory or driver assistance.**

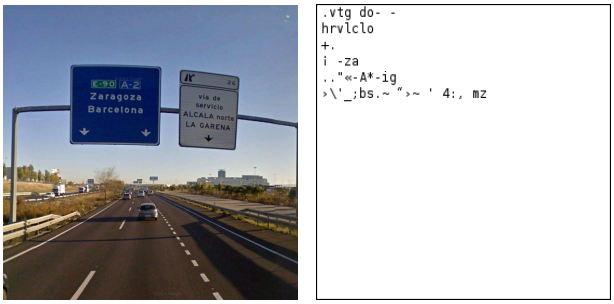
I. INTRODUCTION

Automatic text recognition in complex scenes is a challenging problem in computer vision. Text is usually embedded in any kind of environment, both indoors and outdoors, thus automatic text recognition can have multiple applications in daily life, such as support to visually impaired people, automatic geocoding of businesses, support to robotic navigation in indoor and outdoor environments, driver assistance or touristic services. For instance, some recent applications are:

- a system embedded in a PDA or smartphone to help visually handicapped people [1], in line with the SYPOLE project [2]
- indoor [3] and outdoor [4] mobile robot navigation
- a PDA-based sign translator [5]
- an automatic comic reader on cellular phones [6]
- a translator of signboard images from English to Spanish [7]
- a word translator from English to Spanish and vice versa [8]
- a street sign recognizer for geolocalization [9]

The applications of text recognition in ITS can be also multiple. Automatic text recognition could be useful to support drivers or autonomous vehicles to find a certain place by simply reading and interpreting street signs, road panels or any kind of text present in an environment, when Global Positioning Systems (GPS) suffer from lack of coverage, especially in high-density urban areas. Advanced Driver Assistance Systems (ADAS) could also benefit from text recognition for automatic traffic signs and panels identification. In addition, textual information could be also fused with other data obtained from image or RADAR in order to make more robust systems.

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



(a) Input image (b) OCR (Tesseract)

Fig. 1. Text detection and recognition with a public OCR

In this work, we propose a system to automatically recognized the information contained in traffic panels on street-level images. The motivation of the work is to make an inventory of traffic panels, but it could be also used for driver assistance. Text detection and recognition in street images is a hard task due to size and font variations, complex background, difficult illumination conditions, etc. Applying a publicly available OCR (Tesseract) on street-level images is unsuccessful. Fig. 1 shows an example. The OCR results hardly contain any readable word, but a lot of noise.

We propose a complete system to detect and recognize traffic panels on complex images obtained from the Street View service by Google. Some of our contributions are the use of Hidden Markov Models (HMMs) for word recognition and the use of Web Map Services (WMS) for increasing the effectiveness of the recognition algorithm. A diagram of blocks of the proposed system is shown in Fig. 2.

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 recognition. Section III describes the process of capturing the images. Section IV explains the algorithm to detect and recognize text while Section V describes the panel geolocalization method. Finally, section VI provides the experimental results and section VII concludes the paper.

II. STATE OF THE ART

Because of the wide diversity of the information contained in traffic panels, as well as the usual problems related to outdoor computer vision systems such as occlusions, shadows and non-controlled lighting conditions, to date there has not been much research on automatic visual classification of the

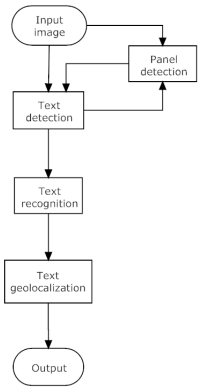


Fig. 2. Diagram of blocks of the proposed algorithm

information contained in road panels. From our knowledge, only three works have been developed. The first one [10] is able to detect candidates to be traffic panels by using an image segmentation for blue and white colours. These candidates are classified by correlating the radial signature of their FFT (Fast Fourier Transform) with the pattern corresponding to an ideal rectangular shape. This algorithm is invariant to traslations, rotations, scaling and projective distortion. However, it is severely affected by changing lighting conditions.

The second work on this topic [11] considers a priori knowledge of the geometry and other features of the panels to detect them in the image. Text detection is carried out by applying a technique that incorporates edge detection, a segmentation method based on GMM (Gaussian Mixture Models) theory and search for lines through a geometrical analysis. 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. On the other hand, this algorithm uses geometrical restrictions to put the objects into lines and words, but it does not take into account other features such as size or colour, which can be vital in some contexts. As well as this, the segmentation method based on GMM depends highly on the contrast between foreground and background, which is affected at the same time by lighting conditions.

Another work on this topic is described in [12], by the same authors. The motivation of this work is to complement the functionality of a traffic signposting inspection system based on computer vision. The vehicle used to capture the images has an active infrared illuminator. As traffic signs are made of retroreflective materials, panels detection is carried out by using shape detection algorithms over the frame difference of two consecutive illuminated and non-illuminated single frames. Then, textual information is searched only over the detected panels instead of doing it over the whole image. Text detection is achieved by applying a technique that fuses local edge detection, perspective distortion correction and geometrical analysis.

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. 3 shows the first 4 zoom levels 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 3 and we have cropped the panoramic view to the region shown in red in Fig. 4, that is, the tile $(x = 3, y = 1)$, the right half side of the tile $(x = 2, y = 1)$ and the left half side of the tile $(x = 4, y = 1)$. This region is optimum to detect the panels located above and on the right margin of the road.

IV. TEXT DETECTION AND RECOGNITION

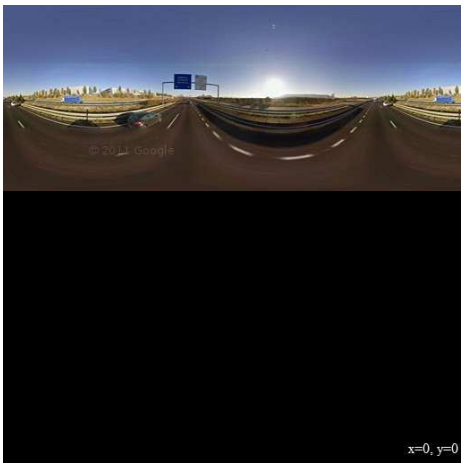
It has been applied a text detection algorithm developed by the authors. This method is aimed at locating and recognizing text on any kind of images, including complex background images. It has been seen that it is able to detect text on street-level images with high precision and producing a very low number of false positives. However, as it can be seen in Fig. 5(a), sometimes the text detection method does not locate all the text contained in traffic panels due to different reasons, such as small size of the text in the image or image distortion that produces non-horizontally alligned text. Therefore, it has been complemented with a panel detection algorithm that is applied after the text detection method. Where panels are found in the image, we apply a character restoration step to locate all the characters that have not been detected during the first global text detection stage. An example is shown in Fig. 5(b). The panel detection algorithm uses both color and edge information. It works as follows.

The text localization algorithm uses MSER to detect characters. It has been seen that the images obtained when applying MSER over the street-level images allow to detect the traffic panels easily, as the panels appear as big stable rectangular regions with high text density inside them. Fig. 6 shows an example. In addition, as many panels have blue background, it has been developed a blue-color segmentation technique to detect big blue rectangular regions in the image. We compute the normalized RGB color coordinates defined by (1)-(3). Then, we compute a single-channel image by forming a weighted sum of the rgb components, using (4). If we compute MSER algorithm on this image, blue panels can be easily extracted, as it is shown in Fig. 7. This technique does not allow us to detect all the panels in an image, but it allows us to know that a detected panel is a true panel with a very high probability, in order to search for previously undetected characters in the extracted panel regions.

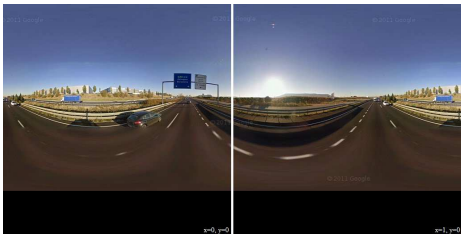
$$r = \frac{R}{R+G+B} \tag{1}$$

$$g = \frac{G}{R+G+B} \tag{2}$$

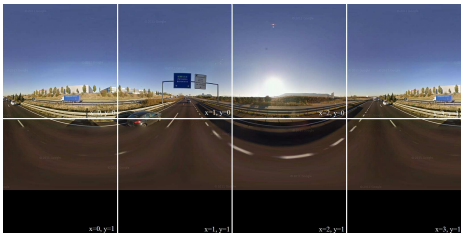
$$b = \frac{B}{R+G+B} \tag{3}$$



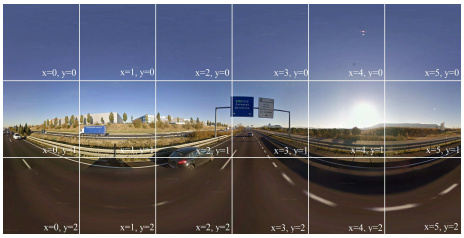
(a) Zoom=0



(b) Zoom=1



(c) Zoom=2



(d) Zoom=3

Fig. 3. Different zoom levels for a panoramic view

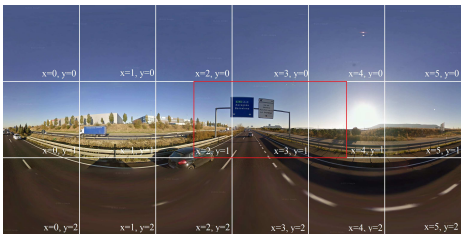
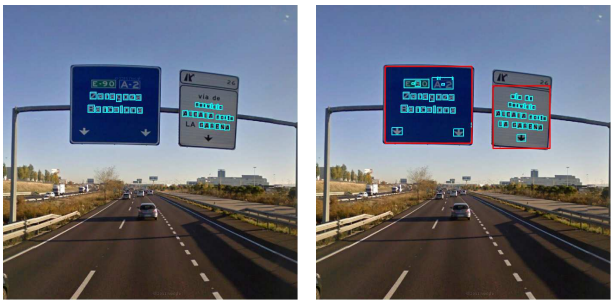


Fig. 4. [Best viewed in color] Region of Interest in the panoramic view (red)

$$I = 0.2989 \cdot r + 0.5870 \cdot g + 0.1140 \cdot b \quad (4)$$



(a) Text detection before panel detection (b) Panels detected and text detection after panel extraction

Fig. 5. [Best viewed in color] Text detection

(a) Dark-on-bright MSER image (b) Bright-on-dark MSER image

Fig. 6. Panels detection

(a) Normalized RGB image (b) MSER on the normalized RGB image

Fig. 7. [Best viewed in color] Blue panels segmentation

Finally, character and word recognition is applied. Character recognition is based on computing histogram of gradient direction over the edge points of the binarised objects. It fails when panels are far, as text is small and difficult to segmentate. However, it is not necessary to recognize all the characters perfectly. They are just an estimation, because a word recognizer is applied later. The word recognizer used was proposed earlier by the authors in [12]. It is a HMM-based approach that computes the most probable model that has generated the set of observations to be recognized. It uses a dictionary of words, which includes all the words that the system is able to recognize, that is, name of cities and other common words that typically appear in road panels. In order to increase the effectiveness of the recognition algorithm, we make use of a Web Map Service (WMS) to reduce the size of

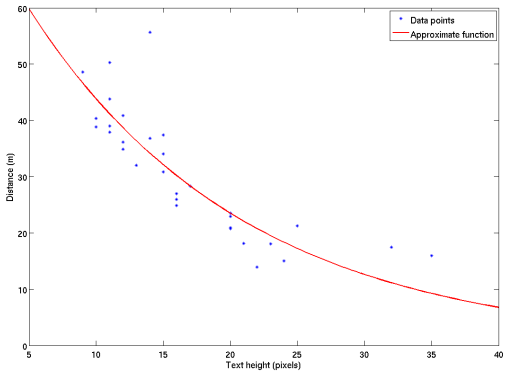


Fig. 8. Function to convert from height (pixels) to distance (meters)

the dictionary to a limited geographic area, *i.e.* to the nearest places. The service used is Cartociudad [13], an official database of the Spanish road network, supported by the Spanish government. As the GPS coordinates (latitude and longitude) where every image has been taken are known, a request to the Cartociudad server giving the position as input returns a document parsed in XML format where different features are stated for the point under request, such as the name of the road or the street, the milestone and the province among others. We make a request just for the images where text has been found.

V. DATA GEOLOCALIZATION

For every image, we know the GPS position of the vehicle where the images were taken. We want to estimate the position of the detected text. The camera parameters are unknown and the system is monocular, so in principle it is not possible to compute the relative position of the objects in the image respect to the camera. However, we know that the size of the text contained in traffic panels is not constant, but it does not differ very much from one panel to another. A function that converts from text height in pixels to depth distance in meters have been computed. The data to compute this function have been obtained from different panels estimating the distance from the vehicle to the panel by using the satellite image in Google Maps. Fig. 8 shows the data points and the approximated function, which is defined by (5), where d is the distance from the panel to the vehicle, h is the text height in pixels, $a = 81.66$ and $b = -0.06225$.

$$d = a \cdot e^{b \cdot h} \tag{5}$$

Once we have detected text in a certain image, we compute the mean height in pixels of the text and apply (5) to obtain the distance d from the panel to the vehicle. Then, in order to estimate the GPS position of the panel, we need to compute the direction vector of the vehicle. We trasform the latitude and longitude values of the vehicle in the current and in the following frame (lat_1, lon_1) and (lat_2, lon_2) respectively to UTM coordinates in meters $P_1 = (x_1, y_1)$ and $P_2 = (x_2, y_2)$. The direction vector $\vec{v}$ and the angle α of the vector respect to the horizontal axis is computed using (6) and (7). α is always referenced to the positive horizontal semiaxis.

$$\vec{v} = (v_x, v_y) = (x_2 - x_1, y_2 - y_1) \tag{6}$$

$$\alpha = \arccos \frac{v_x}{\sqrt{v_x^2 + v_y^2}} \tag{7}$$

Finally, (8) is applied to compute the position of the traffic panel $P_3 = (x_3, y_3)$, using the geometry shown in Fig. 9. These are UTM coordinates, which can be easily transformed into latitude and longitude values applying the correspondent conversion formulas.

$$P_3 = (x_3, y_3) = (x_1 + d \cdot \cos\alpha, y_1 + d \cdot \sin\alpha) \tag{8}$$

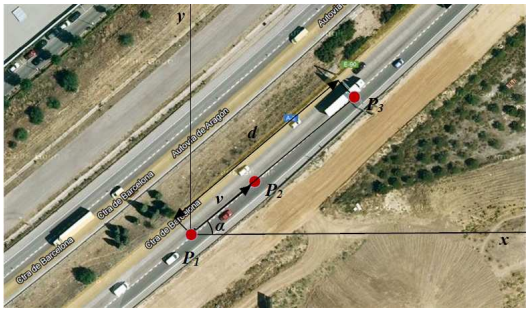


Fig. 9. Geometry to estimate the position of the panel

VI. EXPERIMENTAL RESULTS

It has been analyzed a total of 173 kilometers of two different highways of the Spanish road network. In this stretch, there are 214 panels of which 86 are panels above the road and 128 are traffic panels located at the right side of the road. Traffic panels contain not only words, but also numbers and different symbols such as direction arrows or petrol station indications. In addition, we have several samples for every panel, because each one usually appears in different frames at different distances. Since the detection and recognition have been carried out for every frame independently, we show the detection and recognition rates in Tables I-III for different ranges of distances in order to see how the distance to the panel affects. We have defined three ranges: short distance when the panel is less than 30 meters far, medium distance when it is in the range 30-50 meters and long distance when the panel is further than 50 meters. The nearer the panel is, the better the performance of the algorithm is. However, the recognition rate for symbols remains above 70% no matter the distance of the panel is, due to the fact that symbols are typically bigger than characters and numbers, being easier to segmentate. Fig. 10 and Fig. 11 show some examples of the obtained results.

TABLE I

EXPERIMENTAL RESULTS AT SHORT DISTANCE

Data	Detection rate	Recognition rate
Words	92.00%	67.21%
Numbers	73.35%	64.13%
Symbols	63.08%	80.00%

TABLE II

EXPERIMENTAL RESULTS AT MEDIUM DISTANCE

Data	Detection rate	Recognition rate
Words	53.87%	41.94%
Numbers	20.88%	40.00%
Symbols	47.00%	75.53%

TABLE III

EXPERIMENTAL RESULTS AT LONG DISTANCE

Data	Detection rate	Recognition rate
Words	13.73%	15.29%
Numbers	3.77%	9.68%
Symbols	20.76%	84.15%

VII. CONCLUSIONS AND FUTURE WORKS

This work uses a text detection algorithm for complex background images, which was developed by the same authors, to detect and recognize the visual information depicted in traffic panels. It may have a wide range of applications on ITS, such as driver assistance, traffic panels inventory or support to panels design. Unlike other methods that detect traffic panels in first place and text inside the panel regions secondly, a different approach has been implemented. Firstly, we detect text on the image, while traffic panels are supposed to be high-density text areas in the image. It has been showed that the performance of the proposed method is quite good, as we detect most of the words in almost all the panels at short distance. Numbers and symbols detection needs to be improved, as well as character recognition at long distance.

In the future we propose to make use a priori of the regulation on traffic panels, which describes how they must be designed, in order to make a more effective recognition algorithm. For instance, in the above side of the panels, it is supposed to appear just numbers and symbols. On the other hand, our approach consists of detecting text on the image and then detecting the traffic panels based on high-density text areas together with color segmentation for blue-background panels, in order to make a more exhaustive search of text in the panel regions to restore undetected characters in the first stage of the algorithm. We aim to extend our method to detect also white-background panels using edge extraction. The main problem is that there is image distortion and the panel edges are not straight in the image when the panels are near. It is also necessary to improve the estimation of the distance of the panels in the image in order to geolocalize the traffic panels more

(a) Original image

(b) Text recognized

(c) Original image

(d) Text recognized

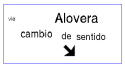
(e) Original image

(f) Text recognized

(g) Original image

(h) Text recognized

(i) Original image



(j) Text recognized

Fig. 10. Image results

(a) Original image

(b) Text recognized

(c) Original image

(d) Text recognized

(e) Original image

(f) Text recognized

(g) Original image

(h) Text recognized

(i) Original image

(j) Text recognized

Fig. 11. Image results

precisely and to make a fusion of the information extracted in consecutive frames, as we detect each panel usually in different consecutive frames. Moreover, it is unnecesary to compute α from latitude and longitude values in two consecutive frames, as we could use directly the heading value given by Google for each image. Finally, we have used a WMS service that is only valid for the Spanish road network. It has been done this way just to validate the proposed algorithm. However, in the future we plan to use a WMS hosted by Yahoo which provides geographical data for many other countries, including province and county information, which is the information that we use in our algorithm to reduce the size of the dictionary of words. For this purpose, we do not use the data provided by Google because it only gives the name of the country, the state and the nearest city, but not the name of the province or the county.

ACKNOWLEDGMENTS

This work has been financed with funds from the Ministerio de Educación y Ciencia through the project DRIVER-ALERT (TRA2008 - 03600), as well as from the Comunidad de Madrid through the project Robocity2030 (CAM - S - 0505 / DPI / 000176). The authors would like to thank Google for the images, which have been used only for research purposes. There is no commercial intention.

REFERENCES

[1] S. Ferreira, V. Garin and B. Gosselin, "A Text Detection Technique Applied in the Framework of a Mobile Camera-based Application", *1st Int. Workshop on Camera-based Document Analysis and Recognition*, 2005.

[2] V. Gaudissart, S. Ferreira, C. Mancas-Thillou and B. Gosselin, "Sypole: a Mobile Assistant for the Blind", *Procs. European Signal Processing Conference (EUSIPCO)*, 2005.

[3] X. Liu and J. Samarabandu, "An edge-based text region extraction algorithm for indoor mobile robot navigation", *Int. Conf. on Mechatronics and Automation*, 2005.

[4] I. Posner, P. Corke and P. Newman, "Using Text-Spotting to Query the World", *Procs. IEEE Int. Conf. on Intelligent Robots and Systems (IROS)*, 2010.

[5] J. Zhang, X. Chen, J. Yang and A. Waibel, "A PDA-based sign translator", *Proc. Int. Conf. on Multimodal Interfaces*, 2002.

[6] M. Yamada, R. Budiarto, M. Endoo and S. Miyazaki, "Comic image decomposition for reading comics on cellular phones", *IEICE Trans. on Information and Systems*, pp. 1370-1376, 2004.

[7] A. Canedo-Rodriguez, K. Soohyung, J.H. Kim and Y. Blanco-Fernandez, "English to Spanish translation of signboard images from mobile phone camera", *IEEE Southeastcon*, pp. 356-361, 2009.

[8] "Word Lens, An Interactive Real-time Spanish/English translator app" (press release): http://www.bestappsite.com/2010/12/20/word-lens-an-interactive-real-time-spanishenglish-translator/, 2010.

[9] S.N. Parizi, A.T. Targhi, O. Aghazadeh and J.O. Eklundh, "Reading Street Signs Using a Generic Structured Object Detection and Signature Recognition Approach", *Int. Conf. on Vision Application*, 2009.

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

[11] 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.

[12] Á. González, L.M. Bergasa, J. Javier Yebes and M.A. Sotelo, Automatic Information Recognition of Traffic Panels using SIFT descriptors and HMMs, *13th International IEEE Conference on Intelligent Transportation Systems*, pp. 1289-1294, 2010.

[13] http://www.cartociudad.es/