

AUTOMATIC RECOGNITION OF ROAD SIGNS IN DIGITAL IMAGES FOR GIS UPDATE

André R. S. Marçal

Faculdade de Ciências, Universidade do Porto, DMA, Rua do Campo Alegre, 687, Porto, Portugal

Isabel R. Gonçalves

*Escola Superior de Tecnologia e Gestão, Instituto Politécnico de Viana do Castelo
Av. do Atlântico, Ap. 574, Viana do Castelo, Portugal*

Keywords: Image Processing, Road Sign Recognition, Mobile Mapping Systems, Geographic Information System.

Abstract: A method for automatic recognition of road signs identified in digital video images is proposed. The method is based on features extracted from cumulative histograms and supervised classification. The training of the classifier is done with a small number of images (1 to 6) from each sign type. A practical experiment with 260 images and 26 different road sign was carried out. The average classification accuracy of the method with the standard settings was found to be 93.6%. The classification accuracy is improved to 96.2% by accepting the sign types ranked 1st and 2nd by the classifier, and to 97.4% by also accepting the sign type ranked 3rd. These results indicate that this can be a valuable tool to assist Geographic Information System (GIS) updating process based on Mobile Mapping System (MMS) data.

1 INTRODUCTION

There is a growing interest in having a detailed geo-referenced representation of our environment. Digital mobile mapping integrates digital imaging with direct geo-referencing, providing an ideal tool for the acquisition of large amounts of geo-referenced data. In the last 10-15 years there has been considerable technological developments, allowing for vehicle-based Mobile Mapping Systems (MMS) to be available at a reasonably low cost. These systems are usually based on Global Positioning System (GPS) and Inertial Navigation System (INS) for the navigation component, and two or more imaging cameras for the image data component. The final goal of a MMS is usually to create or update a Geographic Information System (GIS) with interest objects, such as postal boxes, bus stops or road signs.

The standard approach to data extraction from MMS based image videos is to have an operator viewing the video to identify interest objects. Once a relevant object is encountered, the video is stopped and an image pixel from the object is selected. The system then identifies the conjugate point and, using the stereoscopic image pair together with the position and attitude recorded by GPS and INS, computes the ge-

ographic coordinates of the object identified. The operator will then have to provide the attributes of the object to be inserted in the GIS database (e.g. the type of road sign).

Automatic object recognition can provide valuable assistance to this process in two ways: (1) the identification of an interest object in an image, (2) the recognition of the object type or other relevant attributes. For the case of road signs, several attempts to address the issue of automatic identification and recognition are reported in the literature. The system proposed in (Piccioli et al., 1996) uses a normalized cross correlation approach for the recognition component, reporting different values for the detection and classification rates (21% to 98%), depending on the type of images. A combined detection and classification system is described in (Escalera et al., 1997), where the classification, based on neural networks, was tested with 18 sign types, but the experimental details are unclear. The system proposed in (Hsu and Huang, 2001) uses matching pursuit filters for sign recognition. A total of 40 sign types were tested, 30 circular and 10 triangular, with a recognition rate reported of 94% for triangular signs and 91% for circular signs. The automatic road sign recognition system described in (Fang et al., 2004) reports very high classification

results (99%), but the experiment was mostly centred on the detection of signs in video sequences. The recognition rate reported in (Kim et al., 2006) is also 99%, tested with 107 images, but only 10 sign types were considered. Another system combining detection and classification, based on template matching, is described in (Vavilin and Kang, 2006), with an average detection rate of 97.7% and a recognition rate of 91.3% reported (for 172 signs), but the number of sign types used is unclear.

The purpose of this work is to present an alternative method for the automatic recognition of road signs identified in digital images, assuming that the approximate location of the sign in the image is known. The manuscript is organized as follows: in section 2 the proposed methodology for road sign recognition is presented, in section 3 the experimental evaluation strategy is described, section 4 presents the results, and section 5 the conclusions.

2 METHODS

The road sign recognition method developed works in three stages: (1) pre-processing, (2) feature extraction, (3) classification. The system accepts as an input a RGB image of any size, and returns the sign type, from a pre-defined set of types. Although the input image can be of any size, it is expected that the margins are not too large.

2.1 Pre-processing

The aim of the pre-processing stage is to select the area of the input image that actually contains the road sign. The RGB (Red Green Blue) input image (I_{in}) is converted to the HSI (Hue Saturation Intensity) color model. A thresholding segmentation is performed to identify the areas of red and blue in the image. A binary image for red (B_{red}) is produced from pixels with $H \in ([0.00, 0.10] \cup [0.80, 1.00]) \cap S \in [0.30, 1.00]$, and a binary image for blue (B_{blue}) is produced from pixels with $H \in]0.57, 0.70[\cap S \in]0.25, 0.65[\cap I \in [0.13, 0.60[$. Both binary images are subjected to a filtering process to remove small objects and irregularities due to mixed pixels and noise. First a 3 by 3 median filter is applied, which removes all small isolated objects in the binary images. Then two morphological filters are used to further smooth the object edges and remove non-isolated small objects: an erosion with a 2x2 square structuring element, and a dilation with a diamond shaped structuring element (Gonzalez and Woods, 2008). All remaining small objects (less than 40 pixels) are removed from the

binary images.

After this processing step, only the largest object and all other objects that are at least 60% of its size are retained from each binary image (B_{red} and B_{blue}). The interior of the remaining objects are then filled, and the two binary images combined. The binary image (B_s) with only the object of interest (the road sign) is obtained by selecting the largest object of the two processed binary images B_{red} and B_{blue} . A sub-section of the RGB image is then obtained using the minimum enclosing rectangle of the object in B_s . The binary image B_s is used to mask out the pixels that do not belong to the road sign, resulting in a RGB image I_s where only the pixels belonging to the road sign have non zero values. Examples of such images are presented in grey scale in figure 1.

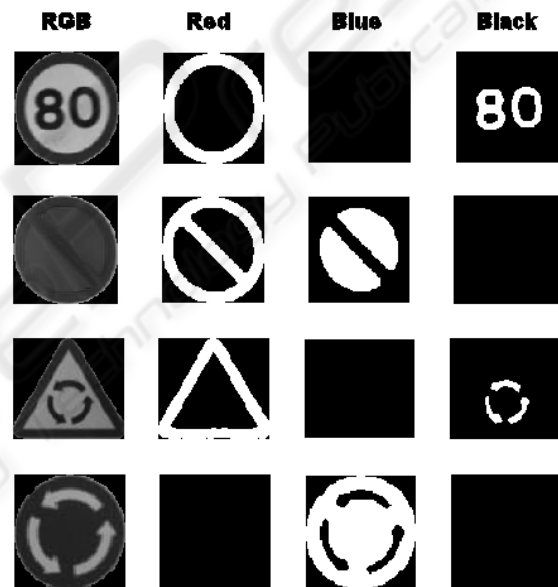


Figure 1: Examples of binary component extraction for red, blue and black, from the RGB color images (here in greyscale) obtained after the pre-processing stage.

2.2 Feature Extraction

The features that characterize the observed object (road sign) are obtained from the red, blue and black components of the RGB color image I_s . The same criteria described in section 2.1 is used for the extraction of the red and blue binary image components. The binary image for black (B_{black}) is obtained from pixels with $H \in ([0.00, 0.10] \cup [0.69, 0.90]) \cap I \in [0, 0.25] \cap S \in [0, 0.35[$. The implementation was done in a way that each pixel can only belong to a binary image, with priority for red, then blue, and black last. Examples of the red, blue and black binary component extraction are presented in figure 1.

Let $B(x,y)$ be a binary image, with X_0 by Y_0 pixels (or X_0 columns by Y_0 lines). There are two possible values for each pixel (x,y) : 0 and 1. The operators F and G applied to a binary image produce cumulative histograms for columns (f) and lines (g), according to (1) and (2), where x and y are integers between 1 and X_0 and Y_0 .

$$F\{B(x,y)\} = f(x) = \sum_{i=1}^{Y_0} B(x,i) \quad (1)$$

$$G\{B(x,y)\} = g(y) = \sum_{i=1}^{X_0} B(i,y) \quad (2)$$

The application of operators F and G to $B(x,y)$ produces two vectors: f with X_0 elements and g with Y_0 elements. As the binary images used have different sizes, these vectors need to be normalized. The normalization is done in two ways: in terms of the values of f and g , and in terms of their number of elements. The normalization of the vector values is done by dividing its values by the number of lines / columns of the interest image, so that the range of values used is 0 to 1. The reason for normalizing the number of vector elements is to obtain a constant (relatively small) number of elements, independently of the binary image size. Let n be the number of elements of the normalized vectors. Modified versions of vectors f and g are initially created, where each element is repeated n times. These new vectors (f' and g') have nX_0 and nY_0 elements. The normalized vectors f^n and g^n are computed by (3) and (4).

$$f^n(j) = \frac{\sum_{i=1}^{X_0} f'(i + (j-1)X_0)}{X_0}; j = 1, \dots, n \quad (3)$$

$$g^n(j) = \frac{\sum_{i=1}^{Y_0} g'(i + (j-1)Y_0)}{Y_0}; j = 1, \dots, n \quad (4)$$

The normalized vectors f^n and g^n are computed for the binary components B_{red} , B_{blue} and B_{black} . The features are thus 6 vectors (f_{red}^n , f_{blue}^n , f_{black}^n , g_{red}^n , g_{blue}^n , g_{black}^n), each with n elements. As an illustration, figure 2 shows four examples of the binary component for red and the corresponding vectors f , g , f_{red}^n and g_{red}^n (with $n = 10$), presented as line and bar plots. In this example the two feature vectors extracted from the red binary image components (f_{red}^n and g_{red}^n) are clearly capable of distinguishing the signs.

2.3 Classification

A supervised classification process is used. Initially, reference vectors (f_{red}^n , f_{blue}^n , f_{black}^n , g_{red}^n , g_{blue}^n , g_{black}^n) are obtained for each road sign type, from a number of training images (t). For each road sign type (or class),

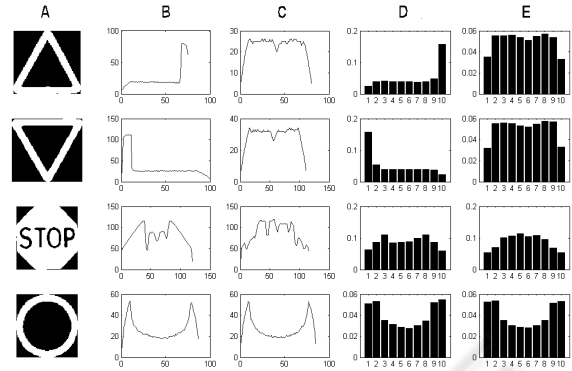


Figure 2: Example of feature extraction from binary images. Red binary image components (A), cumulative histograms for lines (B) and columns (C), normalised feature vectors f_{red}^n (D) and g_{red}^n (E), with $n = 10$.

the distance between the observed and reference vector is computed for all six features. The sum of the six distances (d) is the discriminative criteria used to classify a sign. The class with the lowest value of d is assigned to the observed road sign. The distances between two vectors were computed using the absolute and the Euclidean distances.

3 EXPERIMENTAL SETUP

A practical application was carried out to evaluate the performance of the proposed method with real MMS image data.

3.1 Test Images

A video dataset acquired by a MMS was made available. The dataset has over 14000 images, of 640 by 480 pixels, acquired by a AVT Marlin camera (Madeira, 2007). These images were inspected and sub-sections with road signs extracted. Although there are over 150 different road signs (DGV, 2003), most of these signs are rarely used. The requirement imposed to the experiment was to have at least 10 different images of the same road sign type. This limited the number of different road sign types to only 26, which are presented in figure 3 as standard references (DGV, 2003). The road sign types #13 to #17 have all the same standard reference, as they are all speed limit signs: of 40 (#13), 50 (#14), 60 (#15), 80 (#16) and 100 (#17). The sign type #25 used was in fact for a speed of 50 (instead of 30, as presented in the standard image of figure 3). The original version of the images in figure 3 are in color, with the light gray corresponding to red and darker gray to blue (except for

sign #4, where the traffic signal also uses red, green and yellow).

Real road signs are often different from the official standards, such as those presented in figure 3. This can be confirmed by an inspection of the test images used for signs #4 and #22, presented in figure 4. The most noticeable differences are the shape of the arrow in image *RS22_8*, which is thicker than the standard shape, and the absence of the black background on the traffic signal in image *RS4_1*. As for the other signs, there are occasional variations from the standard shape. Furthermore, in real images there are differences in terms of illumination, size, orientation (although very oblique views were not used) and noise (including blurred images and damaged signs). Some of these situations can be observed in the examples presented in figure 4. There are images with oblique view (*RS4_2* and *RS22_7*), with a large margin (*RS22_2*), blurred (*RS4_5*) and signs damaged by graffiti (*RS22_9*). There is also a large variety of illumination conditions, both in terms of background and on the sign itself, and in sizes (between 55x62 pixels to 212x195 pixels in the examples of figure 4).

A total of 260 images (sub sections of the full MMS video frames) were thus selected for the experiment (10 images of each type), with a variety of sizes, illumination and viewing conditions, margin sizes and noise.



Figure 3: Standard road signs used in the experiment (DGV, 2003). Sign types #13-17 correspond to 5 different speed limits: 40, 50, 60, 80 and 100. The speed used in sign #25 is 50 instead the standard value of 30.

3.2 Evaluation Strategy

The evaluation of the proposed road sign recognition system is based on a reference scenario, with 4 images for training ($t = 4$) each sign type, normalization with $n = 10$ and Euclidean distance classifier. Each of these parameters was allowed to vary, within a limited range, with all remaining parameters fixed at the reference values. As the number of images available for each sign type was small (10), the parameter t was only tested for values between 2 to 6, with the remaining images ($10 - t$) used for testing. The normalisation parameter (n) was tested for values 6, 8, 10, 12,

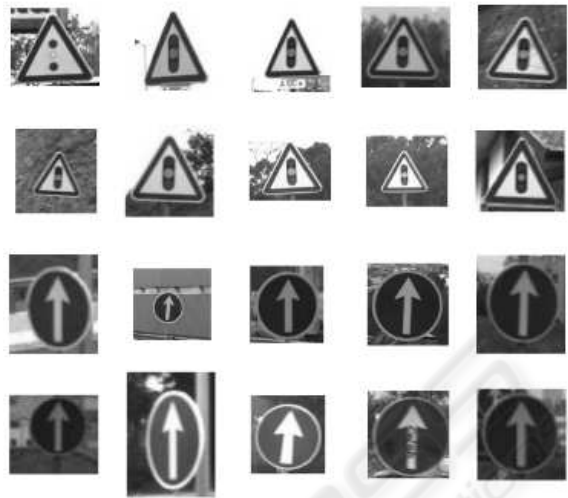


Figure 4: Test images used for road signs #4 and #22.

14, 16, 18 and 20. Two distance metrics were used for the discriminative function used for the classifier: Absolute and Euclidean. The Mahalanobis distance was also tested, but it was not included because the covariance matrix was not always invertible.

The average classification accuracy (A_1) was computed by the ratio between the number of images classified correctly and the total number of test images. Two other classification accuracies were also computed, by accepting the sign types ranked 1st and 2nd (A_{12}) and accepting the sign types ranked 1st, 2nd and 3rd (A_{123}).

4 RESULTS

The classification accuracies for the experiment with the reference parameters ($t = 4$, $n = 10$ and Euclidean distance) were: $A_1=93.6\%$, $A_{12}=96.2\%$ and $A_{123}=97.4\%$. This means that 146 out of the 156 images used for testing were correctly classified. As for the remaining 10 images, 4 had the correct road sign assigned as 2nd option and 2 as 3rd option. In only 4 out of the 156 images (2.6%) the correct road sign was not selected in the top 3 ranking. The difficulties are mostly related to the speed limit signs (#13–#17). Table 1 shows how these images were classified. For the remaining 21 road sign types, there was only one misclassified image (from sign #4 to #7).

The experiment was repeated using different training data sizes (t between 1 and 6) and distance metrics. The results are presented in table 2. Generally, the classification accuracy tends to improve

Table 1: Classification results for the speed limit signs (#13–#17) with the reference scenario ($t = 4$, $n = 10$ and Euclidean classifier).

	#13	#14	#15	#16	#17
#13	5	0	0	0	1
#14	0	3	0	2	1
#15	0	0	3	3	0
#16	1	0	1	4	0
#17	0	0	0	0	6

with the increase in training data size. However, the results are reasonably good even with a single image of each sign type for training ($t = 1$). The results produced using the Euclidean distance were better than those produced by the absolute distance.

The impact of the feature normalization on the classification results was also investigated. The experiment was repeated using normalization values n between 6 and 20. The Euclidean distance and the number of training images ($t=4$) were kept fixed for all cases. The classification accuracies A_1 , A_{12} and A_{123} are presented in table 3. The reference value ($n = 10$) seems to be a good choice, with slightly better values only observed for higher values of n , for A_{12} and A_{123} .

 Table 2: Average classification accuracy for different training data sizes (t between 1 and 6) and distance metrics ($n = 10$ for all cases).

	A_1	A_{12}	A_{123}
Absolute			
$t=1$	82.1%	89.7%	92.7%
$t=2$	84.1%	91.4%	94.7%
$t=3$	86.8%	94.0%	96.7%
$t=4$	91.7%	96.2%	97.4%
$t=5$	88.5%	95.4%	97.7%
$t=6$	88.5%	94.2%	98.1%
Euclidean			
$t=1$	85.0%	90.6%	94.0%
$t=2$	85.1%	91.4%	94.2%
$t=3$	88.5%	94.5%	96.7%
$t=4$	93.6%	96.2%	97.4%
$t=5$	90.0%	95.4%	97.7%
$t=6$	91.4%	95.2%	98.1%

The k nearest neighbors method was also tested for the reference scenario ($t = 4$, $n = 10$), using the Euclidean distance. However, the results were not very good. The average classification accuracies were 74.3% for $k=1$, and 72.4% for $k=3$ and for $k=5$.

 Table 3: Average classification accuracy for different feature normalization settings (n), with the other parameters were kept fixed ($t = 4$ and Euclidean classifier).

	A_1	A_{12}	A_{123}
$n=6$	87.8%	93.6%	97.4%
$n=8$	90.4%	96.2%	97.4%
$n=10$	93.6%	96.2%	97.4%
$n=12$	91.0%	96.2%	98.1%
$n=14$	91.7%	96.2%	98.1%
$n=16$	92.3%	96.2%	98.1%
$n=18$	93.0%	96.2%	98.1%
$n=20$	93.0%	96.2%	98.7%

5 CONCLUSIONS

The results of the proposed method for automatic recognition of road signs in digital images are encouraging. The classification accuracies for the experiment with the reference parameters ($t = 4$, $n = 10$ and Euclidean distance) were: $A_1=93.6\%$, $A_{12}=96.2\%$ and $A_{123}=97.4\%$. The features extracted from cumulative histograms of red, black and blue binary components of the RGB image seem to be effective for the discrimination of road signs. The impact of the various classification and feature normalization parameters could not be fully tested due to the limited size of the training dataset (260 images of 26 types). However, one very promising aspect already observed was the small number of training images required to train the classifier. Future work includes the preparation of a more extensive test dataset, as more MMS video data should soon become available. The goal is to have at least 40 sign types with 15 to 20 test images from each. Once this dataset is available, it should be possible to better evaluate the feature normalization parameters and to test other classifiers. The use of more sophisticated classifiers should compensate the likely reduction in accuracy due to the increase in the number of road sign types.

The proposed methodology can be used in a GIS input system from MMS video datasets. The number of road sign types will have to be increased to 50 or more, which should not be a problem as the number of images required for training was found to be small. The classification accuracy will tend to decrease as the number of types considered increases. However, in this type of system the operator will always have to confirm the classification proposed automatically. A useful fixture would be to propose a sign, plus 2 or 3 alternatives (the 2nd and 3rd ranked in the discrimination function). The operator would then only have to confirm the suggestion (1st option), select one of the alternatives or, in the worst case scenario, to identify

manually from the full list of attributes. The successful implementation of such system can improve the working ability of the operator, thus reducing costs and speeding up the GIS updating process based on MMS image data.

ACKNOWLEDGEMENTS

The authors would like to thank Sérgio Madeira and José Alberto Gonçalves for providing the MMS image video dataset.

REFERENCES

- DGV (2003). *Guia de Sinalização Rodoviária*. Ministério da Administração Interna, Lisboa.
- Escalera, A., Moreno, L., Salichs, M., and Armingol, J. (1997). Road traffic sign detection and classification. *IEEE Transactions on Industrial Electronics*, 44:848–859.
- Fang, C., Fuh, C., Yen, P., Cherng, S., and Chen, S. (2004). An automatic road sign recognition system based on a computational model of human recognition processing. *Computer Vision and Image Understanding*, 96:237–268.
- Gonzalez, R. C. and Woods, R. E. (2008). *Digital Image Processing*. Prentice Hall, Upper Saddle River, New Jersey, 3rd edition.
- Hsu, S. and Huang, C. (2001). Road sign detection and recognition using matching pursuit method. *Image and Vision Computing*, 19:119–129.
- Kim, G., Sohn, H., and Song, Y. (2006). Road infrastructure data acquisition using a vehicle-based mobile mapping system. *Computer-Aided Civil and Infrastructure Engineering*, 21:346356.
- Madeira, S. (2007). *Sistema Móvel de Levantamento com Integração em SIG*. PhD thesis, Faculdade de Ciências Universidade do Porto.
- Piccioli, G., De Micheli, E., Parodi, P., and Campani, M. (1996). Robust method for road sign detection and recognition. *Image and Vision Computing*, 14:209–223.
- Vavilin, A. and Kang, H. J. (2006). Automatic detection and recognition of traffic signs using geometric structure analysis. In *SICE-ICASE International Joint Conference*, pages 1451–1456. ICASE.