TRAFFIC SIGN RECOGNITION USING MSER AND RANDOM FORESTS

Jack Greenhalgh and Majid Mirmehdi

Visual Information Laboratory,
University of Bristol
{greenhalgh,majid}@cs.bris.ac.uk

ABSTRACT

We present a novel system for the real-time detection and recognition of traffic symbols. Candidate regions are detected as Maximally Stable Extremal Regions (MSER) from which Histogram of Oriented Gradients (HOG) features are derived, and recognition is then performed using Random Forests. The training data comprises a set of synthetically generated images, created by applying randomised distortions to graphical template images taken from an on-line database. This approach eliminates the need for real training images and makes it easy to include all possible signs. Our proposed method can operate under a range of weather conditions at an average speed of 20 fps and is accurate even at high vehicle speeds. Comprehensive comparative results are provided to illustrate the performance of the system.

***Index Terms*—** traffic sign recognition, MSER, HOG features, intelligent transportation systems

1. INTRODUCTION

Automatic traffic sign detection and recognition is an important task for an Advanced Driver Assistance System (ADAS). Traffic symbols possess several distinguishing features which can be used for their detection and recognition, including contrast, colour, and shape, which can be used to place symbols into specific semantic groups. Rotational and geometric distortion is fairly limited, as road signs are usually found facing the camera and oriented approximately upright. However, there are several issues which can negatively affect the accurate detection and recognition of traffic signs, including motion blur, variations in illumination, fog, occlusion, and deterioration. Furthermore, road scenes often contain strong geometric shapes which closely resemble road signs, making it harder to avoid the misclassification of background information.

Our proposed method is composed of two main stages: (a) detection, which is performed using a novel application of maximally stable extremal regions (MSERs) [1], and (b) recognition, which is performed with histogram of oriented gradients (HOG) features [2], classified using Random Forests [3].

An additional novel aspect of the work is the use of a road sign database provided by the UK Department for Transport consisting of simple graphical representations of UK road signs. A large training dataset is generated by applying a number of randomised distortions to these synthetic template images, including geometric distortion, blurring, and illumination variations. Use of this database allowed our system to be trained on the entire range of road signs in operation, whereas previous (some very recent) works, such as [4, 5], used only a targeted subset of classes. Training the classifiers on all possible road signs is essential in order to avoid the false detection of excluded road signs, which may be similar in appearance to included road signs. Synthetically generating data also avoids the tedious, time consuming process of hand-labelling a large dataset of real videos, with no loss of accuracy as shown later.

In Section 2, we review past road sign recognition systems and state the improvements we make against them. In Section 3 we describe the approach used for the detection of candidate regions. In Section 4, we outline the generation of synthetic training data and the approach to classification of the candidate regions, and in Section 5 we provide comparative results. We conclude the paper in Section 6.

2. RELATED WORK

The problem of recognition of ideogram based road signs in real road scenes has been dealt with in numerous works, such as [7, 8, 9, 5]. The nature of the problem demands a two stage system commonly practiced by most works in this area: detection and recognition. The detection stage identifies the regions of interest, and is mostly performed using colour segmentation, followed by some form of shape recognition. Detected candidates are then either identified or rejected during the recognition stage using, for example, template matching [10], or some form of classifier, such as SVM [9, 5], or neural networks [4].

Colour information is the cornerstone of many methods that segment traffic sign candidate regions, e.g. as in [9, 11, 12]. The performance of such systems is highly dependent on illumination and weather conditions (e.g. when there is fog). Various colour models have been used to attempt to overcome

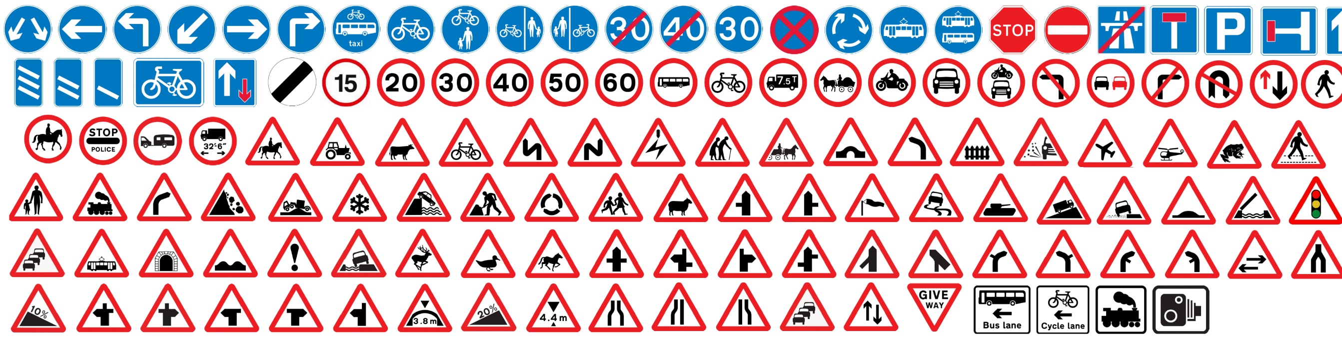

Fig. 1. The full set of graphical road signs used in training our system (obtained from [6]).

these issues. For example, Shadeed et al. [13] segmented red road signs by exploiting the U and V chrominance channels of the YUV space, with U being positive and V being negative for red colours, in combination with the hue channel of the HSV colour space. For another example, Gao et al. [8] applied a quad-tree histogram method to segment the image, based on the hue and chroma values of the CIECAM97 colour model. In contrast, there are several approaches that ignore colour information entirely, instead using only shape information from gray-scale images, such as [7, 14]. For example, Loy and Zelinksy [14] proposed a system which used local radial symmetry to highlight points of interest in each image and detect octagonal, square, and triangular road signs.

The vast majority of existing systems consist of classifiers trained using hand-labelled real images, for example [4, 5], which is a repetitive, time-consuming, and error-prone process. Our method avoids collecting and manually labelling training data since it only uses synthetic, graphical representations of signs obtained from an online road sign database [6] (see Fig. 1). Furthermore, while many existing systems report high classification rates, the total number of traffic sign classes they recognise is generally very limited, e.g. 7 in [5], or 20 in [10], and hence are less likely to suffer mismatches against similar signs missing from their databases. Our proposed system uses all instances of ideogram based (non-text) traffic symbols used in the UK and hence performs its matching in this full set. We expect our approach to be equally functional if applied to other countries' traffic sign databases obtained in a similar fashion.

It is also worth noting many proposed systems suffer from slow speeds, making them inappropriate for application to real-time problems. Some methods report processing times of several seconds for a single frame, e.g. [9, 15, 5]. Our system runs at an average speed of 20 fps.

Several commercial traffic sign recognition systems exist, including [16, 17]. Such systems also recognise a very limited set of traffic signs, for example the system developed by Mobileye [17] only detects speed limit signs and 'no-overtaking' signs. Comparison with these commercial systems is difficult, as little information on their performance is available.

In Section 5, we compare our proposed method against a similar road sign detection system proposed by Gómez-Moreno et al [9] (whose method will be described later).

3. DETECTION OF ROAD SIGNS AS MSERS

MSERs are regions which maintain their shape when the image is thresholded at several levels. This method of detection was selected due to its robustness to variations in contrast and lighting conditions. Rather than detecting candidates for road signs by border colour, the algorithm detects candidates based on the background colour of the sign, since these backgrounds persist within the MSER process.

For the detection of traffic symbols with white backgrounds, MSERs are found in grayscale. Each frame is binarised at a number of different threshold levels, and the connected components (CC) at each level are found. The CCs which maintain their shape through several threshold levels are selected as MSERs. Fig. 2 shows different thresholds for an example image with the CCs coloured. It can be seen that the CC representing the circular road symbol maintains its shape through several threshold levels. CCs are found across many threshold levels, rather than at a single level. This ensures robustness to both variations in lighting and contrast. Several features of the detected CCs are used to reduce the number of candidates, such as aspect ratio, CC perimeter, and CC area. The parameters for these were determined empirically.

We approach the detection of traffic symbols with red or blue backgrounds slightly differently. Rather than detecting MSERs in a grayscale image, the frame is transformed from RGB into a 'normalised red/blue' image, Ω_{RB}. This is such that for each pixel of the original image, values are found for the ratios of its red and blue channels to the sum of all channels respectively, and the greater of these two values is used as the pixel value of the normalised red/blue image, i.e.

$$\Omega_{RB} = max\left(\frac{R}{R+G+B}, \frac{B}{R+G+B}\right). \quad (1)$$

The values in Ω_{RB} are higher for red and blue pixels, and

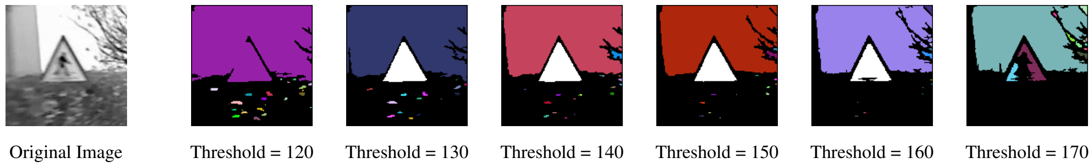

Fig. 2. Original image, and connected components at several threshold levels.

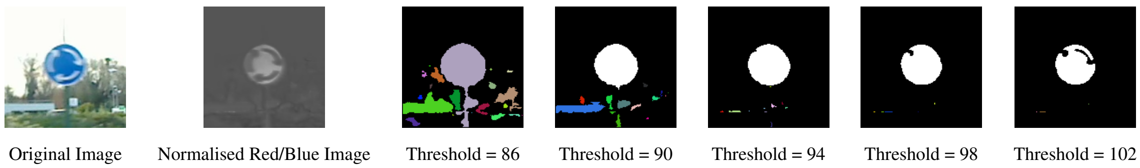

Fig. 3. Original image, normalised red/blue image, and connected components at several thresholds

lower for other colours. MSER regions are then found in this new image. Fig. 3 shows an example image, the result of the normalised red/blue transform, and the image at several different thresholds with their CCs. Again it can be seen that the red and blue road signs maintain their shape at several threshold levels, making them candidates for classification.

To increase the speed of the algorithm, we threshold only at an appropriate range of values, rather than at every possible value that is the norm in the original MSER [1].

4. ROAD SIGN CLASSIFICATION

The recognition stage is used to confirm a candidate region as a traffic sign, and classify the exact type of sign. For classification of candidate regions their HOG features are computed to extract the occurrence of gradient orientations in the image, since HOG features are expected to capture the strong, high contrast edges of the traffic signs well. This is performed on a dense grid of cells, using local contrast normalisation on overlapping blocks. While this intensive normalisation produces large feature vectors (1764 dimensions for a 64×64 image), it provides higher accuracy results. Traffic signs are generally found to be approximately upright and facing the camera, which limits rotational and geometric distortion, removing the need for rotation invariance.

In order to choose the best possible classifier for our HOG features, a comparison was made between SVM, MLP (Multi-Layer Perceptron), and Random Forests, the results of which are shown in Table 1. Random Forests gave both the highest accuracy and fastest classification time. While SVM provides high accuracy and fast classification for a small number of classes, it suffers when dealing with large multi-class problems, given that it is primarily a binary classifier. Multi-class classification can be achieved by training N binary *one-versus-one* classifiers, where $N = n \times \frac{n-1}{2}$ and n is the total number of classes. However, this would severely affect the speed of classification when dealing with a very large number of classes, e.g. 100 classes would require 4950 binary

Classifier	Accuracy	Average Time for Single Classification (ms)
Support Vector Machine	87.8 %	115.87
Multilayer Perceptron	89.2 %	1.45
Random Forest	94.2 %	0.15

Table 1. Comparison of different classification methods

classifiers. A Random Forests classifier requires training with large datasets, which in our case is readily available due to the nature and extent of our synthetically generated data. It can also handle feature vectors with thousands of variables, and produces a classifier which is highly accurate.

Regardless of the classification machinary used, each region is classified using a cascade of classifiers, as illustrated in Fig. 4. First the candidate region is resized to 64×64 pixels. A HOG feature vector with 1764 dimensions is then calculated and this feature vector is used to classify the shape of the region as a circle, triangle, inverted triangle, rectangle, or background. Octagonal ‘stop’ signs are considered to be circles. If the region is found to be background, it is rejected. If the region is found to be a shape, it is then passed on to a (symbol) sub-classifier for that specific shape.

For the Random Forest approach, different classifier trees are used for candidates with white backgrounds (MSERs for grayscale images), and candidates with red or blue backgrounds (MSERs for Ω_{RB}). Therefore, each sub-classifier is specific to symbols with a certain background colour and shape. Colour background triangles, and colour background inverted triangles are rejected as background, as no signs of this type exist in the UK road sign database [6].

Generating Synthetic Training Data - Training the classifiers on all possible road signs is essential to avoid misclassification of unknown signs. However, gathering a sufficient amount of real data on which to train the classifiers is difficult and time consuming given the sheer number of different existing signs, and scarcity of particular signs. Our proposed solution to this problem is to use easily available graphical

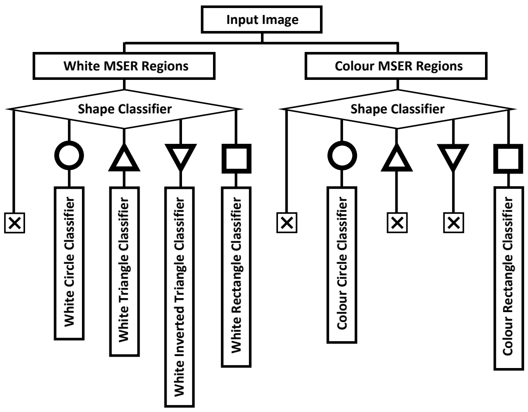

Fig. 4. Cascaded approach to classification.

data and synthetically generate variations and distortions of them to create training data for the classifiers. This allows us to use the entire range of road signs, avoid tedious manual hand labelling for training purposes, and report more reliable classification results that are a true reflection of a complete search. We believe using only a subset of signs (as in [5] or [10]) simplifies the problem of classification within that subset, for example by avoiding misclassification against other similar but excluded signs, therefore in many cases the quality of the reported results are unreliable and not worth comparing against.

The graphical base images we use were obtained from a free online database provided by the UK's Department for Transport [6]. Randomised geometric distortions were then applied in order to replicate the range of distortions likely to be seen in real data, and the type of regions likely to be found during the detection stage. Each distorted example image is superimposed over a random section of background. Randomised brightness, contrast, noise, blurring, and pixelation is also applied to each image. For each sign in the complete set of 132 road-sign images (Fig. 1) used for training, 1200 synthetic distorted images were generated. For comparison, Fig. 5 shows a number of real road-sign images next to a number of our synthetic training images.

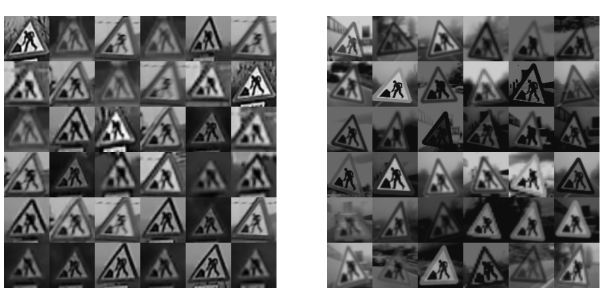

Fig. 5. Comparison of real (left) and synthesised (right) data

5. EXPERIMENTAL RESULTS

The proposed system is accurate at high vehicle speeds, and under a range of lighting and weather conditions. A considerable increase in speed was gained by running the algorithm in parallel as a pipeline to around 20 fps running on a 3.33 GHz Intel Core i5 CPU under OpenCV, where the frame dimensions were 640×480. However, the system retained a latency of around 200 ms.

We compare our proposed method against a similar road sign detection system proposed by Gómez-Moreno et al. [9]. Their system detects candidate regions using colour information, and performs recognition using SVM based on a training set of between 20 and 100 images per class, on an unspecified number of classes. Each frame is segmented using the hue and saturation components of an HSI image for which histograms are built for red, blue, and yellow sign colours, created using images with a range of weather and lighting conditions. For the segmentation of white road signs, the image is binarised based on the achromaticity of each pixel and then each candidate blob is classified by shape. The distance from the side of the candidate blob to its bounding box is measured at each side (left, right, top, bottom), at several points. Binary SVMs for each shape are then used to vote for each side of the blob (circle or triangle). If the blob receives four votes for a particular shape, that shape is chosen. An SVM with a Gaussian kernel is then used to classify each sign type, based on shape and colour. This is trained using pixel values from the candidate region which falls into a template representing the shape (circle or triangle). For our comparison, we used between 20 and 100 real training images per class to train SVM classifers for recognition, as suggested in [9].

For test data, we used several videos, filmed under a range of weather conditions, at a variety of different vehicle speeds. Video 1 was filmed in clear weather conditions, at low speeds of around 20 mph. Video 2 was filmed in thick fog, at high vehicle speeds, e.g. above 50 mph. Video 3 was filmed in clear weather conditions, at a variety of vehicle speeds, ranging from 20 to 60mph. Examples of these scenes, with results from our system, are shown in Figure 6.

In the results provided in Table 2, Precision represents the percentage of recognised road signs which were correct, Recall represents the percentage of total road signs successfully recognised, and the F-Measure, a combination of Precision and Recall, represents overall accuracy. The results show that our method outperforms the method of Gómez-Moreno et al [9]. While their detection method classified reasonably in clear weather conditions, scenes suffering from poor lighting conditions and strong illumination caused it to fail. Our MSER detection system provides robustness by searching for candidate regions at a range of thresholds, rather than using a single fixed value. Gómez-Moreno et al.'s [9] recognition method also produced a large number of false positives.

Taken from video 1 | Taken from video 2 | Taken from video 3.

Fig. 6. Example frames of system run on test videos

Method	Gómez-Moreno et al [9] method				Proposed method			
	video 1	**video 2**	**video 3**	**total**	**video 1**	**video 2**	**video 3**	**total**
Signs correctly detected as candidate regions	12	1	23	36	14	5	35	54
Detected candidate regions correctly classified	8	1	10	19	13	4	33	50
Signs misclassified as other sign	4	0	4	8	1	1	2	4
Background misclassified as sign	7	2	8	17	3	1	2	6
Precision	**42.11 %**	**33.33 %**	**45.45 %**	**43.18 %**	**76.47 %**	**66.67 %**	**89.19 %**	**83.33 %**
Recall	**57.14 %**	**20.00 %**	**26.32 %**	**33.33 %**	**92.86 %**	**80.00 %**	**86.84 %**	**87.72 %**
F-Measure	**0.48**	**0.25**	**0.33**	**0.38**	**0.8387**	**0.7273**	**0.88**	**0.85**

Table 2. Comparative results for Gómez-Moreno et al.'s [9] and the proposed method. The total number of signs were 14 in video 1, 5 in video 2, and 38 in video 3.

6. CONCLUSION

We proposed a novel, real-time system for the automatic detection and recognition of traffic symbols. The proposed system detected candidate regions as MSERs, a method which is robust to variations in lighting and illumination in the scene. These candidate regions were then classified using HOG features with a cascade of Random Forests. All training data was synthetically generated by applying various randomised distortions to graphical template images. The approach allows the system to recognise all classes of ideogram based traffic symbols, and eliminates the need for a hand-labelled database of real images. The system operates with a high accuracy and performs better than other reported results.

7. REFERENCES

[1] J. Matas, "Robust wide-baseline stereo from maximally stable extremal regions," *IVC*, vol. 22, no. 10, pp. 761–767, 2004.

[2] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *CVPR*, 2005, pp. 886–893.

[3] L. Breiman, "Random forests," *ML*, vol. 45, no. 1, pp. 5–32, 2001.

[4] M. Hossain, M. Hasan, M. Ali, M. Kabir, and A. Ali, "Automatic detection and recognition of traffic signs," in *RAM*, 2001, pp. 286–291.

[5] K. Ohgushi and N. Hamada, "Traffic sign recognition by bags of features," in *TENCON*, 2009, pp. 1–6.

[6] D. for Transport, "Traffic Signs Image Database," http://www.dft.gov.uk/trafficsignsimages/, 2011.

[7] P. Paclik and J. Novovicová, "Road sign classification without color information," in *ASIC*, 2000.

[8] X. Gao, L. Podladchikova, D. Shaposhnikov, K. Hong, and N. Shevtsova, "Recognition of traffic signs based on their colour and shape features extracted using human vision models," *JVCIR*, vol. 17, no. 4, pp. 675–685, 2006.

[9] S. Maldonado-Bascón, S. Lafuente-Arroyo, P. Gil-Jimenez, H. Gomez-Moreno, and F. Lopez-Ferreras, "Road-sign detection and recognition based on support vector machines," *IEEE-ITS*, vol. 8, no. 2, pp. 264–278, 2007.

[10] Y. Gu, T. Yendo, M. Tehrani, T. Fujii, and M. Tanimoto, "A new vision system for traffic sign recognition," in *IV Symposium*, 2010, pp. 7–12.

[11] P. Wanitchai and S. Phiphobmongkol, "Traffic Warning Signs Detection and Recognition Based on Fuzzy Logic and Chain Code Analysis," *IITA*, pp. 508–512, 2008.

[12] K. C.G., L. V. Prabhu, A. R. V., and R. K., "Traffic Sign Detection and Pattern Recognition Using Support Vector Machine," *ICAPR*, pp. 87–90, 2009.

[13] W. Shadeed, D. Abu-Al-Nadi, and M. Mismar, "Road traffic sign detection in color images," in *ICECS*, vol. 2, 2003, pp. 890–893.

[14] G. Loy, "Fast shape-based road sign detection for a driver assistance system," *IROS, 2004*, pp. 70–75.

[15] A. Reiterer, "Automated Traffic Sign Detection for Modern Driver Assistance Systems," in *Facing the Challenges Building the Capacity*, 2010, pp. 11–16.

[16] V. M. Services, "Phaeton debuts with new design and new technologies," https://www.volkswagen-media-services.com/, 2011.

[17] Mobileye, "Traffic Sign Detection," http://mobileye.com/, 2011.