Using Fourier Descriptors and Spatial Models for Traffic Sign Recognition

Fredrik Larsson and Michael Felsberg

Linköping University Post Print

N.B.: When citing this work, cite the original article.

Using Fourier Descriptors and Spatial Models for Traffic Sign Recognition

Fredrik Larsson and Michael Felsberg

Computer Vision Laboratory, Linköping University,
SE-581 83 Linköping, Sweden,
{larsson,mfe}@isy.liu.se

Abstract. Traffic sign recognition is important for the development of driver assistance systems and fully autonomous vehicles. Even though GPS navigator systems works well for most of the time, there will always be situations when they fail. In these cases, robust vision based systems are required. Traffic signs are designed to have distinct colored fields separated by sharp boundaries. We propose to use locally segmented contours combined with an implicit star-shaped object model as prototypes for the different sign classes. The contours are described by Fourier descriptors. Matching of a query image to the sign prototype database is done by exhaustive search. This is done efficiently by using the correlation based matching scheme for Fourier descriptors and a fast cascaded matching scheme for enforcing the spatial requirements. We demonstrated on a publicly available database state of the art performance.

Keywords: Traffic sign recognition, Fourier descriptors, spatial models, traffic sign dataset

1 Introduction

Traffic sign recognition is important for the development of driver assistance systems and fully autonomous vehicles. Even though GPS navigator systems work well for most of the time, there will always be situations when no GPS-signal is available or when the map is invalid, temporary sign installations near road works just to mention one example. In these cases, robust vision based systems are required, preferably making use of monochromatic images.

1.1 Related Work

Many different approaches to traffic sign recognition have been proposed and there are commercial vision based systems available, for example in Volkswagen Phaeton [4], Saab 9-5 [3] and in the BMW 5 and 7 series [1].

One common approach is to threshold in a carefully chosen color space, e.g. HSV [5], HSI [14] or CBH [18], in order to obtain a (set of) binary image(s). This is then followed by detection and classification using the authors favorite choice

of classifier, e.g. support vector machines [12] or Neural networks [5,14]. Another common approach is to consider an edge map using e.g. Fourier descriptors [10], Hough transform [6] or distance transforms [7]. For an excellent overview of existing approaches see [14].

One thing that most published methods have in common is that they report excellent results on their own data sets. Typically they achieve more than 95% recognition rate with very few false positives. However, there are unfortunately no publicly available database for comparing different road sign recognition systems. Meaning that most authors report results on their own dataset and do not provide any means of comparing against their method. One of the main contributions of this paper is to provide a labeled database of more than 20 000 frames captured while driving 350 km on highways and in city environment.

It was recently announced that a dataset of patches containing German traffic signs will be released [2]. The patches contains a traffic sign and an additional border of 10 % around the actual sign. The dataset used in this paper contains the entire image, meaning that both detection and recognition have to be solved. Not only recognition.

1.2 Main Contribution

The main contributions of this paper are:

1. Extending the work [10] with an implicit star-shaped object model, leading to improved performance.
2. Removing the need for a region-of-interests detector used in [10], leading to a fully automatic system.
3. Releasing a database with more than 20 000 frames, 20% being hand labeled, containing a total of 3488 traffic signs.

2 Methods

The proposed method consists of three steps: extraction of Fourier descriptors (FDs), matching of FDs, and matching of previously acquired prototypes with spatial models.

2.1 Fourier Descriptors

The Fourier descriptor (FD) of a shape/contour is created by applying the Fourier transform to a periodic representation of the contour, which results in a shape descriptor in the frequency domain.

Adopting the notation in [8], the closed contour c with coordinates x and y is parameterized as a complex valued periodic function

$$c(l) = c(l + L) = x(l) + iy(l), \tag{1}$$

where L is the contour length.[1] The Fourier coefficients C are obtained by taking the 1D Fourier transform of c,

$$C(n) = \frac{1}{L} \int_{l=0}^{L} c(l) \exp(-\frac{i2\pi nl}{L})\, dl \ \ n = 0, ..., N, \qquad (2)$$

where $N \leq L$ is the descriptor length.

Each coefficient has a clear physical meaning making FDs easy to interpret. Using only a few of the low frequency coefficients is equivalent to using a smoothed version of the contour, see fig. 1 where a pedestrian outline is reconstructed starting with two low frequency coefficients and gradually using more and more high frequency components.
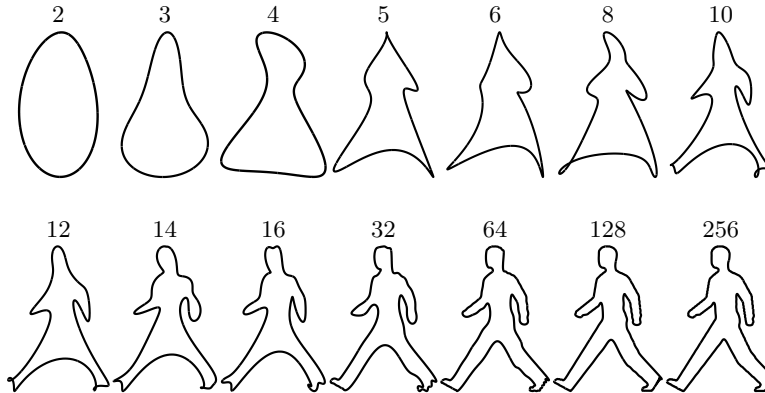

Fig. 1. Reconstruction of a detail from a Swedish pedestrian crossing sign using increasing number (shown above respective contour) of Fourier coefficients.

The main reason for the popularity of FDs is their behavior under common geometric transformations, such as translation, scaling and rotation. The DC component $C(0)$ is the only one that is affected by translations c_0 of the curve $c(l) \mapsto c(l) + c_0$. By disregarding this coefficient, the remaining $N-1$ coefficients are invariant under translation. Scaling of the contour, i.e. $c(l) \mapsto ac(l)$, affects the magnitude of the coefficients and the FD can thus be made scale invariant by normalizing with the energy (after $C(0)$ has been removed). Without loss of generality, we assume that $\|C\|^2 = 1$ ($\|\cdot\|^2$ denotes the quadratic norm) and $C(0) = 0$ in what follows.

Rotating the contour c with ϕ radians counter clockwise corresponds to multiplication of (1) with $\exp(i\phi)$, which adds a constant offset to the phase of the Fourier coefficients

$$c(l) \mapsto \exp(i\phi)c(l) \qquad \Rightarrow \qquad C(n) \mapsto \exp(i\phi)C(n) \ . \qquad (3)$$

---

[1] We treat contours as continuous functions here, where the contour samples can be thought as of impulses with appropriate weights.

Furthermore, if the index l of the contour is shifted by Δl, a linear offset is added to the Fourier phase, i.e. the spectrum is modulated

$$c(l) \mapsto c(l - \Delta l) \qquad \Rightarrow \qquad C(n) \mapsto C(n) \exp(-\frac{i2\pi n \Delta l}{L}) \ . \qquad (4)$$

Note that in what follows, the term translation refers to spatial translation of the entire contour while shift refers to a shift of the start point for sampling the contour.

2.2 Matching of FDs

Rotation and index-shift result in modulations of the FD and it has been suggested to neglect the phase in order to be invariant to these transformations. However, most of the information is contained in the phase and simply neglecting it means to throw away potentially useful information [15].

According to (3) and (4), the phase of each FD component is modified by a rotation of the corresponding trigonometric basis function, either by a constant offset or by a linear offset. Considering the magnitudes only can be seen as finding the optimal rotation of all different components of the FD independently. Given a FD of length $N - 1$, matching the magnitudes only corresponds to finding $N - 1$ different rotations instead of estimating two degrees of freedom (constant and slope). Due to the removal of $N - 3$ degrees of freedom, two contours can be very different even though the magnitude in each FD component is the same.

Recently, a new efficient correlation based matching method for FDs was proposed by Larsson et al. [10]. This approach is partly similar to established methods such as [9,17], but differs in some respects: Complex FDs are directly correlated to find the relative rotation between two FDs without numerically solving equation systems. Let $\mathcal{T}$ denote a transformation corresponding to rotation and index-shift. Let c_1 and c_2 denote two normalized contours, then

$$\min_{\mathcal{T}} \|c_1 - \mathcal{T}c_2\|^2 = 2 - 2 \max_l |r_{12}(l)| \qquad (5)$$

where $|\cdot|$ denotes the complex modulus and the cross correlation r_{12} is computed between the Fourier descriptors C_1 and C_2 according to [16], p. 244–245,

$$r_{12}(k) = (c_1 \star c_2)(k) \doteq \int_0^L \bar{c}_1(l) c_2(k + l) \, dl = \mathcal{F}^{-1}\{\bar{C}_1 \cdot C_2\}(k) \ . \qquad (6)$$

In particular, if c_1' and c_2' denote two contours so that $c_2' = \mathcal{T}'c_1'$, where $\mathcal{T}'$ denotes a transformation covering scaling, translation, rotation and index-shift, then

$$\min_{\mathcal{T}} \|c_1' - \mathcal{T}c_2'\|^2 = 2 - 2 \max_l |r_{12}(l)| = 0 \qquad (7)$$

where the correlation r_{12} is computed after the FDs have been normalized with respect to scale and translation. The parameters of the transformation $\mathcal{T}$ that

minimize (7) are given as

$$\Delta l = \arg \max_{l} |r_{12}(l)| \qquad \phi = \arg r_{12}(\Delta l) \tag{8}$$

$$s = \frac{(\sum_{n=1}^{\infty} |C_1'(n)|^2)^{\frac{1}{2}}}{(\sum_{n=1}^{\infty} |C_2'(n)|^2)^{\frac{1}{2}}} \qquad t = C_1'(0) - C_2'(0) \ . \tag{9}$$

It is also shown in [10] that considering the maximum of the real part instead of the absolute value in (7), corresponds to not compensating for the rotation, i.e. rotation variant matching is given according to

$$\min_{\mathcal{T}/R} \|c_1' - c_2'(\Delta l)\|^2 = 2 - 2 \max_{l} \text{Re}\{r_{12}(l)\} \ . \tag{10}$$

2.3 Sign Prototypes

A traffic sign prototype is created from a synthetic image of the traffic sign, see first row in fig. 5. The synthetic image is low-pass filtered before local contours are extracted using Maximally Stable Extremal Regions (MSER)[13]. Each extracted contour c_k is described by its Fourier descriptor C_k.

In order to describe the spatial relationships between the local features an extra component v_k is added, creating a pair (C_k, v_k) where the first component captures the local geometry (contour) and the second component the global geometry of the sign. This second component v_k is simply the vector from the center of the local feature to the center of the object, see fig. 2. This can be seen as a simple implicit star-shaped object model [11] where each local feature is connected to the center of the object. The combination of FDs and corresponding spatial vectors gives the final traffic sign prototype as

$$P = \{(C_k, v_k)\} \quad k = 1..K \tag{11}$$

where K is the total number of contours for the sign prototype.

These spatial components effectively remove the need for a region-of-interests detector as a first step. Even though each C_k might give matches not corresponding to the actual sign, it is unlikely that multiple matches vote for the same position if they not belong to the actual traffic sign.

2.4 Matching Sign Prototypes

J contours q_j are extracted from a query image and represented by their FDs Q_j, see top row in fig. 3. For each sign prototype, all prototype contours C_k are compared to all extracted contours Q_j. Since traffic signs have a well defined orientation, we use the rotation variant matching score:

$$e_{jk} = 2 - 2 \max_{l} \text{Re}\{\mathcal{F}^{-1}\{\bar{Q}_j \cdot C_k\}(l)\} \ . \tag{12}$$

Fig. 2. Extracted local features (green contours) and corresponding vectors (red arrows) pointing towards the center of the traffic sign.

This results in the binary matrix $\mathbf{M} = (m)_{jk}$ of matched contours with

$$m_{jk} = \begin{cases} 1 & e_{jk} \leq \theta_k \\ 0 & e_{jk} > \theta_k \end{cases} \tag{13}$$

where θ_k is a manually selected threshold for each prototype contour k, see fig. 3 lower left for an example of matched contours.

The next step is to verify which combinations of matched contours $\mathbf{Q}_j$ fit to the spatial configuration of the sign prototype. This is done by a cascaded matching scheme. For each individual match m_{jk}, we obtain by means of (9) the parameters s_k and t_k and compute an estimate $\mathbf{v}'_{jk} = s_{jk}\mathbf{v}_k + t_{jk}$.

The vector $\mathbf{v}'_{j1}$ defines a hypothesized prototype center. We then go through all prototype contours $k = 2 \dots K$ and verify for all $m_{ik} \neq 0$, $i \neq j$, that s_{ik}/s_{j1} is sufficiently close to 1 and that $\mathbf{v}'_{ik}$ is sufficiently close to the hypothesized prototype center. These contours are consistent with respect to scale and location and if only if sufficiently many contours are consistent, a detection of the corresponding sign is flagged, see fig. 3 lower right.

2.5 Dataset

A dataset has been created by recording sequences from over 350 km of Swedish highways and city roads. A 1.3 mega-pixel color camera, a Point-Grey Chameleon, was placed inside a car on the dashboard looking out of the front window. The camera was pointing slightly to the right, in order to cover as many relevant signs as possible. The lens had a focal length of 6.5mm, resulting in approximately 41 degrees field of view. Typical speed signs on motorways are about 90 cm wide, which corresponds to a size of about 50 pixel if they are to be detected at a distance of about 30 m.

A human operator started the recording whenever a traffic sign was visible and stopped the recording when no more signs were visible. In total, in over 20 000 frames have been recorded of which every fifth frame has then been

Fig. 3. Upper left: Query image. Upper right: Extracted contours. Lower left: Contours that matched any of the contours in the pedestrian crossing prototype are shown in a non-yellow color. Lower right: The final result after matching against all sign prototypes.

manually labeled. The label for each sign contains sign type (pedestrian crossing, designated lane right, no standing or parking, priority road, give way, 50 kph, or 30 kph), visibility status (occluded, blurred, or visible) and road status (whether the signs is on the road being traveled or on a side road), see fig 4 for examples. The entire database including ground truth is available at http://www.cvl.isy.liu.se/research/traffic-signs-dataset.

3   Experiments

Synthetic images of Swedish road signs, see bottom row of fig. 5, were used for creating models according to the methodology described in Sec. 2.3. The sign models were then matched against real images from two datasets. The first dataset, denoted *Manually ROIs dataset*, is the one used in [10] which is using patches from bounding boxes of 200x200 pixels, see fig. 5. The second evaluation was done on the the newly collected dataset, denoted *Summer dataset*, see Sec. 2.5. All processing is done frame wise not using temporal clues.
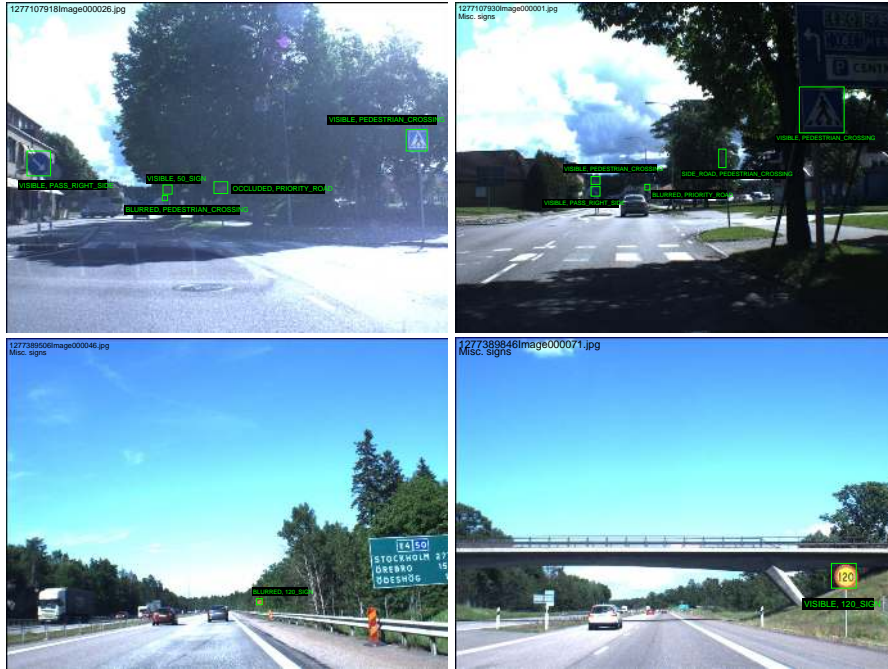
Fig. 4. Examples from the database.

Note that the evaluation was done using grey scale images and thus do not use the distinct colors occurring in the signs as a descriptor. The images used correspond to the red channel of a normal color camera. This is easily achieved by placing a red-pass filter in front of an ordinary monochromatic camera. Using normal grey-scale conversion would be problematic since some of the signs are isoluminant, e.g. sign (c) in fig. 5. The reason for not using colors is that color cameras have lower frame rates given a fixed bandwidth and resolution. High frame rates are crucial for cameras to be used within the automotive industry. Higher frame rates mean for example higher accuracy when estimating the velocity of approaching cars.

3.1 Results Manually ROIs dataset

The first dataset is used in order to compare to the reported results in [10] and contains 316 regions-of-interests (ROIs) of 200x200 pixels, see fig. 5. The ROIs were manually extracted around 216 signs and 100 non-signs. The result is summarized in table 1. This dataset is fairly simple and the proposed method increases the already good performance of [10]. Using constraints on the spatial arrangement of contours removes some of the false positives (FPs) while keeping the same recall level or increasing it by allowing for less strict thresholds on the individual contours. The classes *Priority road* and *Give way* are unaffected since
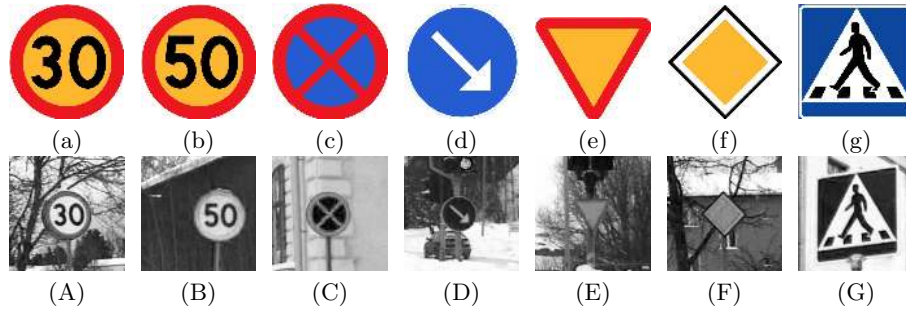
Fig. 5. First row: Synthetic signs used to create models. Second row: Corresponding real world examples.

they consist of a single contour each, thus not benefiting from the added spatial constraints.

Sign type	Proposed method		[10]	
	Recall%	#FP	Recall%	#FP
Pedestrian crossing	98.0	0	98.0	1
Designated lane right	95.8	0	95.8	2
No standing or parking	100.0	0	96.6	1
50 kph	91.7	2	91.7	2
30 kph	95.8	1	95.8	1
Priority road	95.7	0	95.7	1
Give way	94.7	0	94.7	2

Table 1. Performance on the Manually ROIs dataset for the method presented in [10] and the proposed algorithm.

3.2 Results Summer dataset

The second evaluation is done on the new Summer dataset, see Sec. 2.5 for details regarding the dataset. The evaluation was limited to include signs for the road being traveled on with a bounding box of at least 50x50 pixels, corresponding to a sign more than 30 m from the camera. Table 2 contains the results for the same sign classes that was used in the Manually ROIs dataset, with one exception. The class 30kph was removed since only 11 instances of the sign were visible, not giving sufficient statistics. The entire image was giving as query without any ROIs.

All classes except from *Give Way* show excellent precision. The recall rates for the *Pedestrian crossing* and *Designated lane right* classes are above 90% while the *50 kph, Priority road* and *No standing or parking* classes show recall over 70%. The *Giveway* class shows less impressive performace, recall rate under

Sign type	Total Signs	Precision%	Recall%
Pedestrian crossing	158	96.03	91.77
Designated lane right	107	100.00	95.33
No standing or parking	44	97.14	77.27
50 kph	67	100.0	76.12
Priority road	198	98.66	74.24
Give way	67	59.26	47.76

Table 2. Results on the Summer dataset for the proposed method.

50% combined with poor precision. The *Giveway* class contains only one single contour and thus can not benefit from the spatial model. Increasing the recall rate for the *Giveway* class is possible but would also lead to a dramatic increase in the number of FPs. Using [10] on this dataset resluts in a high recall rate but in the order of 2000 FPs for all classes giving precision rates of less than 10%, showing the dependency of ROIs. The three top rows in fig. 6 show examples of correctly detected and recognized traffic signs.

The FPs produced are always due to the contour extraction going awry, see bottom left in fig. 6 for an example of a missed sign. The contour extraction algorithm were not able to extract sufficient contours from this image resulting in a missed *Designated lane right* sign. The bottom right image in fig. 6 shows a false positive of the *Give way* class. It is shown in order to illustrate the problems induced when using a single contour.

4 Conclusions

A method for using locally extracted contours in combination with an implicit star-shaped object model was presented. The presented method works fully automatically on query images with no need for regions-of-interests. It is shown that the presented method performs well for traffic signs that contain multiple distinct contours, such as the Swedish pedestrian crossing sign. For traffic signs with few or a single contour, such as the Swedish give way sign, the method still needs improvement.

A major contribution is the release of the first publicly available large traffic sign database not only containing small patches around traffic signs, thus allowing for evaluation of detection and recognition not only recognition. The database contains over 20 000 frames with 20% being labeled. The database will in the future be extended to include different ambient conditions, such as night and rain, and also to include signs from different countries.

5 Acknowledgments

References

1. BMW Automobiles (2010), http://www.bmw.com/
2. German Traffic Sign Recognition Benchmark (2010), http://benchmark.ini.rub.de/
3. Saab (2010), http://www.saab.com/
4. VW Media Services (2010), https://www.volkswagen-media-services.com/
5. Fleyeh, H., Dougherty, M., Aenugula, D., Baddam, S.: Invariant Road Sign Recognition with Fuzzy ARTMAP and Zernike Moments. In: Intelligent Vehicles Symposium, 2007 IEEE. pp. 31 –36 (2007)
6. Garcia-Garrido, M., Sotelo, M., Martin-Gorostiza, E.: Fast Road Sign Detection Using Hough Transform for Assisted Driving of Road Vehicles. In: Computer Aided Systems Theory, EUROCAST 2005, Lecture Notes in Computer Science, vol. 3643, pp. 543–548 (2005)
7. Gavrila, D., Philomin, V.: Real-time object detection for ldquo;smart rdquo; vehicles. In: Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on. vol. 1, pp. 87 –93 vol.1 (1999)
8. Granlund, G.H.: Fourier Preprocessing for Hand Print Character Recognition. IEEE Trans. on Computers C–21(2), 195–201 (1972)
9. Kuhl, F.P., Giardina, C.R.: Elliptic Fourier features of a closed contour. Computer Graphics and Image Processing 18, 236–258 (1982)
10. Larsson, F., Felsberg, M., Per-Erik, F.: Correlating Fourier Descriptors of Local Patches for Road Sign Recognition. IET Computer Vision (2011), accepted for publication
11. Leibe, B., Leonardis, A., Schiele, B.: Robust Object Detection with Interleaved Categorization and Segmentation. International Journal of Computer Vision 77(1), 259–289–289 (May 2008), http://dx.doi.org/10.1007/s11263-007-0095-3
12. Maldonado-Bascon, S., Lafuente-Arroyo, S., Gil-Jimenez, P., Gomez-Moreno, H., Lopez-Ferreras, F.: Road-sign detection and recognition based on support vector machines. Intelligent Transportation Systems, IEEE Trans. on 8(2), 264 –278 (2007)
13. Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust wide baseline stereo from maximally stable extremal regions. In: BMVC. pp. 384–393 (2002)
14. Nguwi, Y.Y., Kouzani, A.Z.: Detection and classification of road signs in natural environments. Neural Comput. Appl. 17, 265–289 (April 2008)
15. Oppenheim, A., Lim, J.: The importance of phase in signals. Proc. of the IEEE 69(5), 529–541 (May 1981)
16. Papoulis, A.: The Fourier Integral and its Applications. McGraw-Hill, New York (1962)
17. Persoon, E., Fu, K.S.: Shape discrimination using fourier descriptors. IEEE Transactions on Systems, Man and Cybernetics 7(3), 170–179 (March 1977)
18. Zhang, Q., Kamata, S.i.: Automatic road sign detection method based on Color Barycenters Hexagon model. In: ICPR 2008 (2008)

Fig. 6. The first three rows show examples of correctly detected signs. Bottom left shows an example of a missed detection and bottom right shows an example of a false positive, classified as a Give way sign.