ROAD-SIGN DETECTION AND RECOGNITION BASED ON SUPPORT VECTOR MACHINES

MALDONADO-BASCON ET AL.

Presenter: **Rojin Aslani**

ECG 782

Spring 2021

OUTLINE

- Introduction
- Proposed Method
- Results
- Conclusions

PROBLEM DEFINITION AND SIGNIFICANCE

- **Problem**:
  - Automatic road-sign detection and recognition system based on support vector machines (SVMs)

- **Importance**:
  - Traffic signs are important because they regulate the traffic and, indicate the state of the road, guiding and warning drivers and pedestrians
  - Road signs provide drivers important information
  - They help drivers to drive more safely and more easily by guiding and warning them and thus regulating their actions

- **Application**:
  - Automatic traffic-sign maintenance
  - Visual driver assistance system

ROAD SIGN DETECTION

- **System**:

Input

Video

Bmp frames → Road-Sign Detection and Recognition → display Road Signs

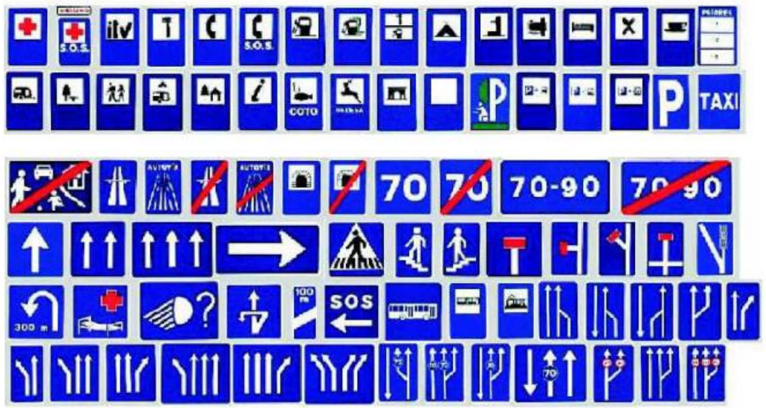
Output

Based on:
Color
Shape
Content

Using:
SVM

DIFFICULTIES IN SIGN DETECTION

- **Difficulties**:
  - The variable lighting conditions of the scene in a natural environment
  - The possible rotation of the signs
  - Different sign dimensions
  - Occlusions because of reducing invisibility by objects such as trees, other signs, or vehicles
  - The large number of different road signs
    - Different shapes
    - Different colors
    - Different contents
    - Color combination



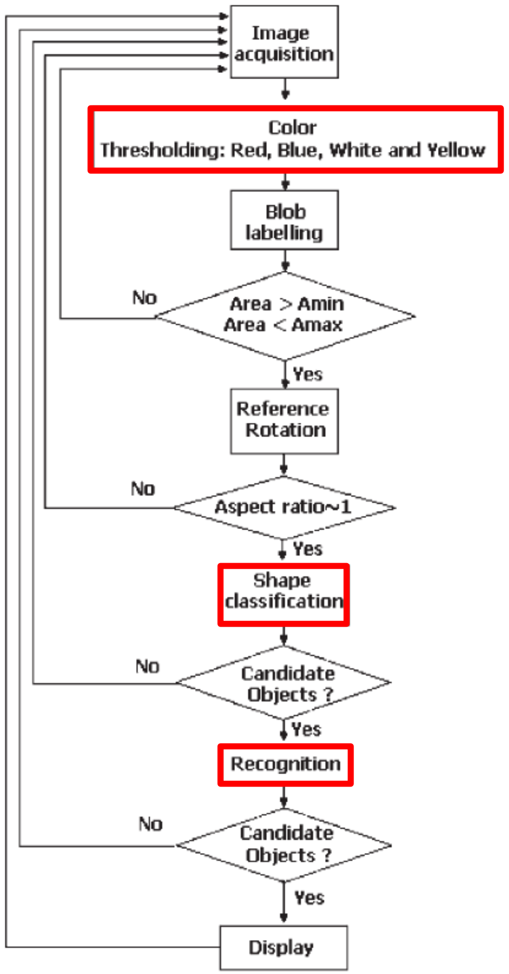
PROPOSED SYSTEM

- **Approach**:
  - The proposed detection and recognition system consists of three stages:
    - **Segmentation** according to the color of the pixel;
    - Traffic-Sign detection by **Shape classification** using linear SVM
    - **Content Recognition** based on Gaussian-kernel SVMs

OUTLINE

- Introduction
- Proposed Method
- Results
- Conclusions

ALGORITHM



For each video frame, the process is triggered by **color segmentation** of the frame where the system will search objects with similar colors as traffic signs, i.e., red, blue, white, and yellow.

Once all objects representing possible candidates have been extracted, some features such as size or aspect ratio are analyzed, and some of noisy objects are discarded.

Each blob of interest is rotated until they are all aligned in the same way before the classification process begins.

As the number of different traffic signs is quite large, **shape classification** is performed prior to the **recognition**

For each frame, if no suitable objects that correlate in either color and geometric properties are found, another frame is analyzed.

Noisy objects (with similar colors as traffic signs, like cars and buildings) are rejected in one of these three stages:
1) geometric feature selection
2) shape classification
3) recognition of the inner area

REVIEW OF THREE STAGES

- **Segmentation**
  - Candidate blobs are extracted from the input image by thresholding using HSI color space for chromatic signs. At the same time, white signs are detected with the help of an achromatic decomposition.
- **Shape classification**
  - Blobs that are obtained from segmentation are classified in this stage using linear SVMs. According to the color that has been used in the segmentation, only some given shapes are possible. For example, signs that are segmented using the red clues can be circular, triangular, or octagonal.
- **Content Recognition**
  - The recognition process is based on SVMs with Gaussian kernels. Different SVMs are used for each color and shape classification.

SEGMENTATION

- **Difficulties** in image segmentation: <u>illumination changes</u> and <u>deterioration of signs</u>
  - **Solution**: The hue and saturation components of the **HSI** color space are sufficient for color signs
    - **Reason**: Because hue and saturation components present low variations for objects with a similar color
  - **Approach**: Histograms of hue and saturation for red, blue, and yellow signs are built.
- The hue and saturation components do not contain enough information to segment white signs
  - The image's achromatic decomposition then helps to detect white signs

$$f(R,G,B) = \frac{(|R-G| + |G-B| + |B-R|)}{3D} \qquad (1)$$

  - where *R*, *G*, and *B* represent the brightness of the respective color, and *D* is the degree of extraction of an achromatic color (D=20)
  - f(R,G,B) < 1 represents achromatic colors
  - f(R,G,B) > 1 represents chromatic colors

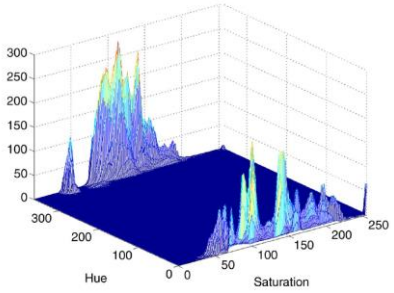


Fig. 3. Hue–saturation histogram for red signs.

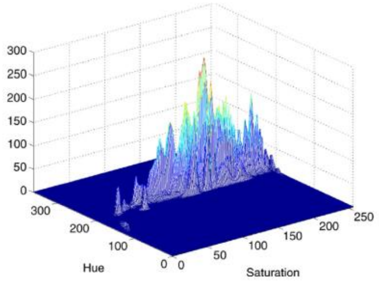


Fig. 4. Hue–saturation histogram for blue signs.

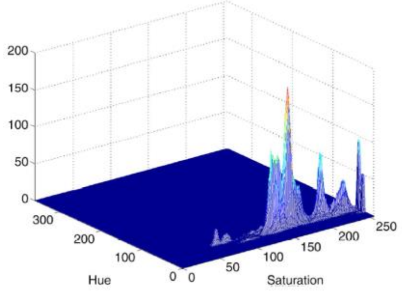
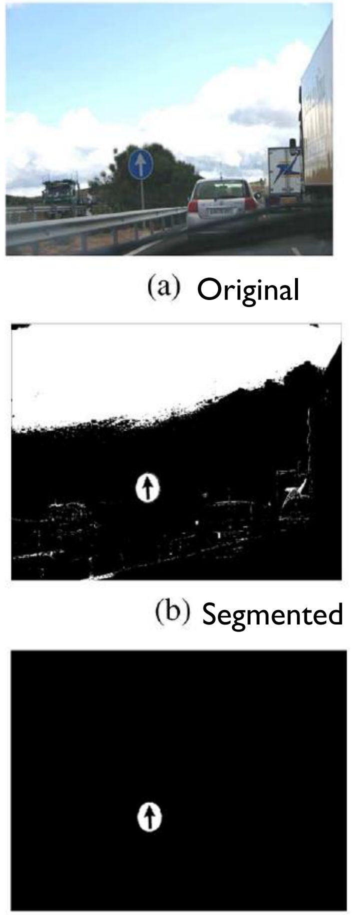


Fig. 5. Hue–saturation histogram for yellow signs.

SEGMENTATION 2

- All candidate blobs are analyzed in a **selection process**
- Some of them are discarded according to their size or aspect ratio
  - Small blobs and big blobs are rejected as noise and noninterest objects, respectively
  - Thresholds for the size criterion are fixed and are derived based on road images
- Signs do not always appear in the ideal position (perpendicular to the direction of driving)
  - **Solution**: each candidate blob in the image is rotated to a reference position before the shape classification in order to obtain a rotation-invariant method.

results of:
segmentation process → **blobs of interest (BoI)** possible traffic signs → will be fed to:
shape classification



(a) Original (b) Segmented (c) BoI

Fig. 7. Segmentation results.

SHAPE CLASSIFICATION

- The BoIs obtained from the segmentation stage are classified in this stage according to their shape using linear Support Vector Machine (SVM).
- Although SVMs are often used to solve binary classification problems, they can also be applied to regression.
- For shape classification, linear SVMs are used.
- Note: The segmentation color in previous stage determines the possible geometric shapes

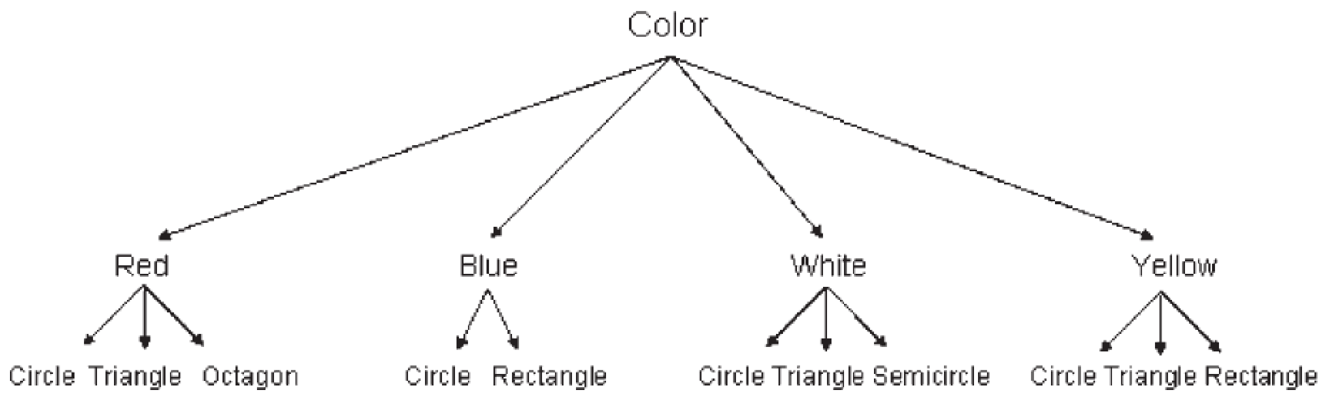


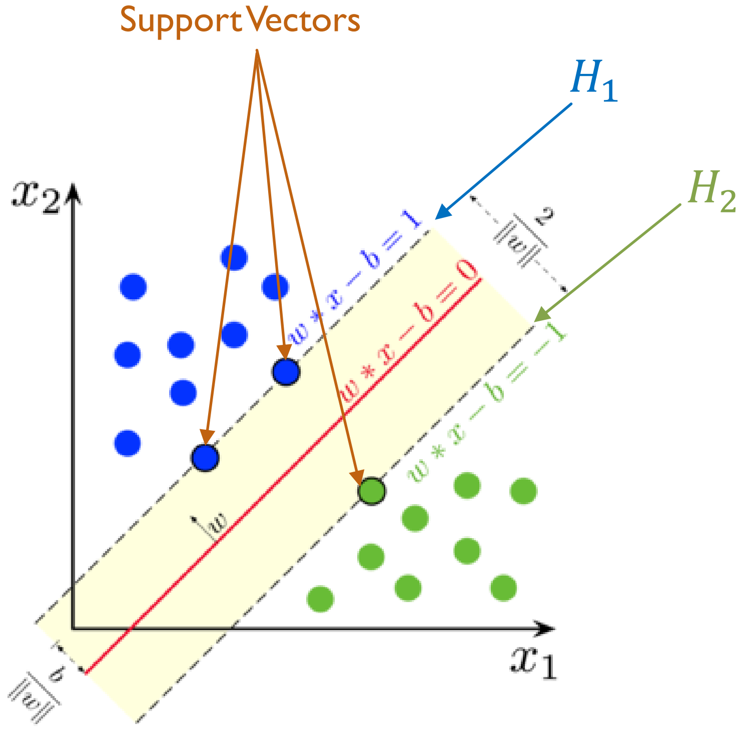
Fig. 10. Tree structure of the classification. The structure shows how the segmentation color determines the possible shapes of the BoI.

SVM

- In the case of two separable classes:
  - The training data are labeled $\{\boldsymbol{x_i}, y_i\}$ where $i = 1, \ldots, l$, $y_i \in \{-1,1\}$, $\boldsymbol{x_i} \in R^d$
- Vectors $\boldsymbol{x_i}$ are the DtBs (distances from the external edge of the blob to its bounding box)
  - Values y_i are "1" for one class and "-1" for the others
  - d is the dimension of the vector
  - l is the number of training vectors
- If a hyperplane $\{\boldsymbol{w}, b\}$ separates the two classes, the points that lie on it satisfy $\boldsymbol{x}.\boldsymbol{w}^T + b = 0$
  - $\boldsymbol{w}$ is normal to the hyperplane and $||w||$ is the Euclidian norm of $\boldsymbol{w}$
- In the separable case, the following constraints hold:

$$y_i(\mathbf{x}_i \cdot \mathbf{w}^T + b) - 1 \geq 0 \quad \forall i. \qquad (2)$$

  - The points that satisfy (2) give us the scale factor for **w** and b
  - These points (support vectors) lie on both hyperplanes $H_1 = \boldsymbol{x}_i.\boldsymbol{w}^T = 1$ and $H_2 = \boldsymbol{x}_i.\boldsymbol{w}^T = -1$



Source: Wikipedia

- The goal is to maximize the margin between both sets by minimizing $||\boldsymbol{w}||^2/2$ subject to (2), leading to the following optimization problem:

$$Lp = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_{i=1}^{l} \alpha_i y_i \left(\mathbf{x}_i \cdot \mathbf{w}^T + b\right) + \sum_{i=1}^{l} \alpha_i. \qquad (3)$$

- Once the optimization is completed, we can classify **x** to class "1" or "-1" using the decision function: $f(\mathbf{x}) = \mathrm{sgn}(\mathbf{x} \cdot \mathbf{w}^T + b). \qquad (4)$

DTBS

- DtBs are used as feature vectors for the inputs of the linear SVMs
- DtBs are distances from the external edge of the blob to its bounding box (D_1, D_2, D_3, D_4)
- Four DtB vectors of 20 components are obtained, and they feed specific SVMs depending on the previous color extraction
- **Example**: An extracted blob by Blue color feeds four DtB SVMs to classify the shape as a Circle "1" or not "−1". Another four SVMs to classify the shape as a Rectangle "1" or not "−1"
- Thus, four votes are possible for each shape. A <u>majority voting method</u> is applied to get the classification with a threshold; so, if the total number of votes is lower than this value, the analyzed BoI is rejected as a noisy shape.

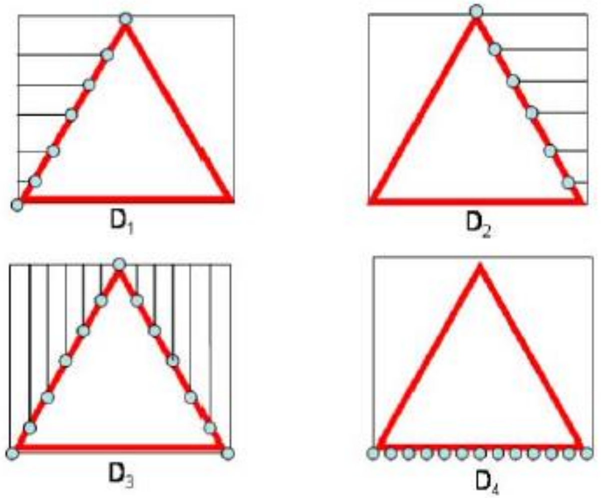


Fig. 9. DtBs for a triangular shape.

SHAPE CLASSIFICATION PROPERTIES

- The proposed method is:
  - invariant to translation, because it does not matter where candidate blob is.
  - invariant to rotation, because, before obtaining DtB vectors, all blobs are oriented in a reference position.
  - invariant to scale, due to the normalization of the DtB vectors to the bounding-box dimensions.
- The method is robust to occlusions, because four feature vectors are obtained to characterize every blob.

results of: shape classification → **Classified Candidates blobs** → will be fed to: content recognition

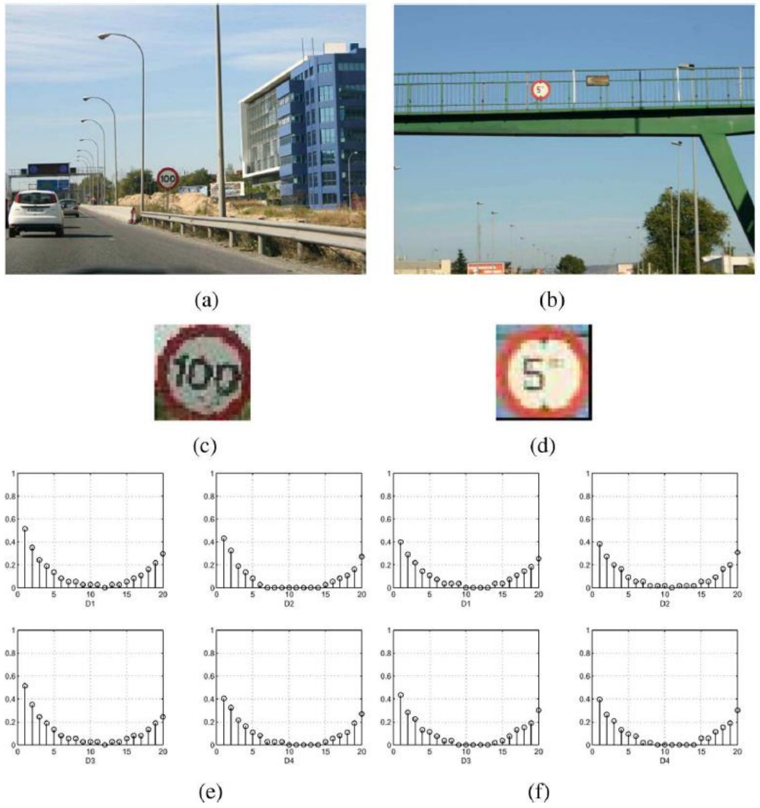


Fig. 11. Translation invariance. (a) and (b) Original images. (c) and (d) BoIs of (a) and (b). (e) and (f) DtB vectors of (c) and (d).

CONTENT RECOGNITION

- Once the candidate blobs are classified into a shape class, the recognition process is initiated to classify candidate blobs based on their content.
- Recognition is implemented by SVMs with Gaussian kernels.
  - **Reason**: in many cases, the data cannot be separated by a linear function.
  - **Solution**: mapping the input data into a different space $\Phi(\mathbf{x})$.
- The training data are used through a dot product.
  - To avoid computing $\Phi(\mathbf{x})$, we use kernel $K(x_i, x_j) = <\Phi(x_i), \Phi(x_j)>$
- A Gaussian kernel is used: $$K(\mathbf{x}_i, \mathbf{x}_j) = e^{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|^2}{2\sigma^2}} \quad (6)$$
- The decision function is:

$$f(\mathbf{x}) = \operatorname{sgn}\left(\sum_{i=1}^{N_s} \alpha_i y_i K(\mathbf{s}_i, \mathbf{x}) + b\right) \quad (7)$$

  - N_s is the number of support vectors
  - S_i are the support vectors.

CONTENT RECOGNITION 2

- The input of the recognition stage is a block of 31 × 31 pixels in grayscale image for every candidate blob
  - Therefore, the interior of the bounding box is normalized to these dimensions.
- To reduce the feature vectors, only those pixels that must be part of the sign (pixel of interest, **PoI**) are used.
- To reduce the problem complexity, every candidate blob is only compared to those signs that have the same color and shape as the blob.
- Amount of training samples per class: [20 ,100] (Average=50)
- To search for the decision region, all feature vectors of a specific class are grouped together against all vectors corresponding to the rest of classes (including noisy objects), following the one-versus-all classification algorithm.
  - Different one-versus-all SVMs classifiers with a Gaussian kernel are used, so that the system can recognize every sign.
  - One-versus-all SVM is a method for using binary classification algorithms for multi-class classification

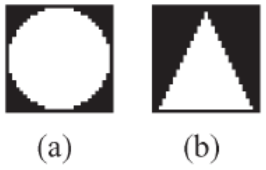


Fig. 16. PoIs in white. (a) Circular and (b) triangular.

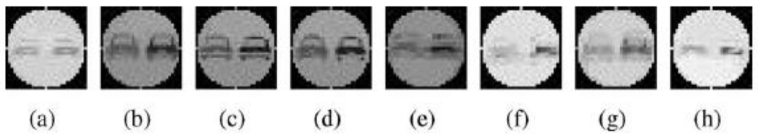


Fig. 17. Positive support vectors for the "No overtake" traffic sign by achromatic segmentation.

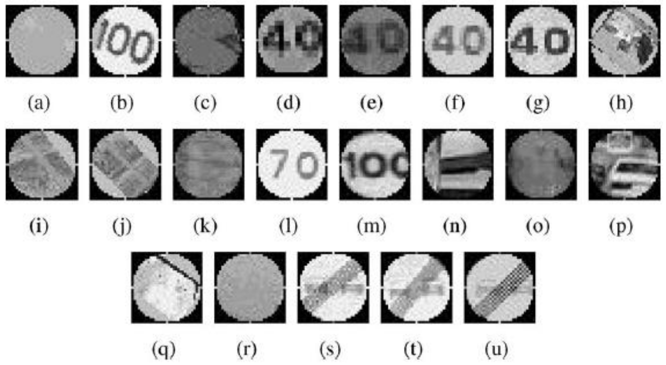


Fig. 18. Negative support vectors for the "No overtake" traffic sign by achromatic segmentation.

OUTLINE

- Introduction
- Proposed Method
- Results
- Conclusions

RESULTS

- Test sequences have been recorded with a video camcorder in a vehicle while driving at usual speed for 4 km, during both the day and at night, in different weather conditions of sunny, cloudy, and rainy weather.
- The video sequences are converted into “.bmp” images
- The size of each image is 720 × 576 pixels
- <u>All signs have been correctly detected at least twice.</u>
- **Suggestion**: For future tracking performance, we can establish the criterion that a candidate traffic sign is dismissed if it appears in only one frame of the sequence.

TABLE III
SUMMARY OF RESULTS

Number of sequence	1	2	3	4	5
Number of images	749	1774	860	995	798
Number of traffic signs	21	21	20	25	17
Detections of traffic signs	218	237	227	285	127
Noisy candidate blobs	601	985	728	622	434
False alarms	0	3	4	8	7
Confused recognition	4	4	4	2	7
	Sunny	Sunny	Sunny	Rainy	Night

RESULTS

- The traffic sign is not identified in all frames of the sequence

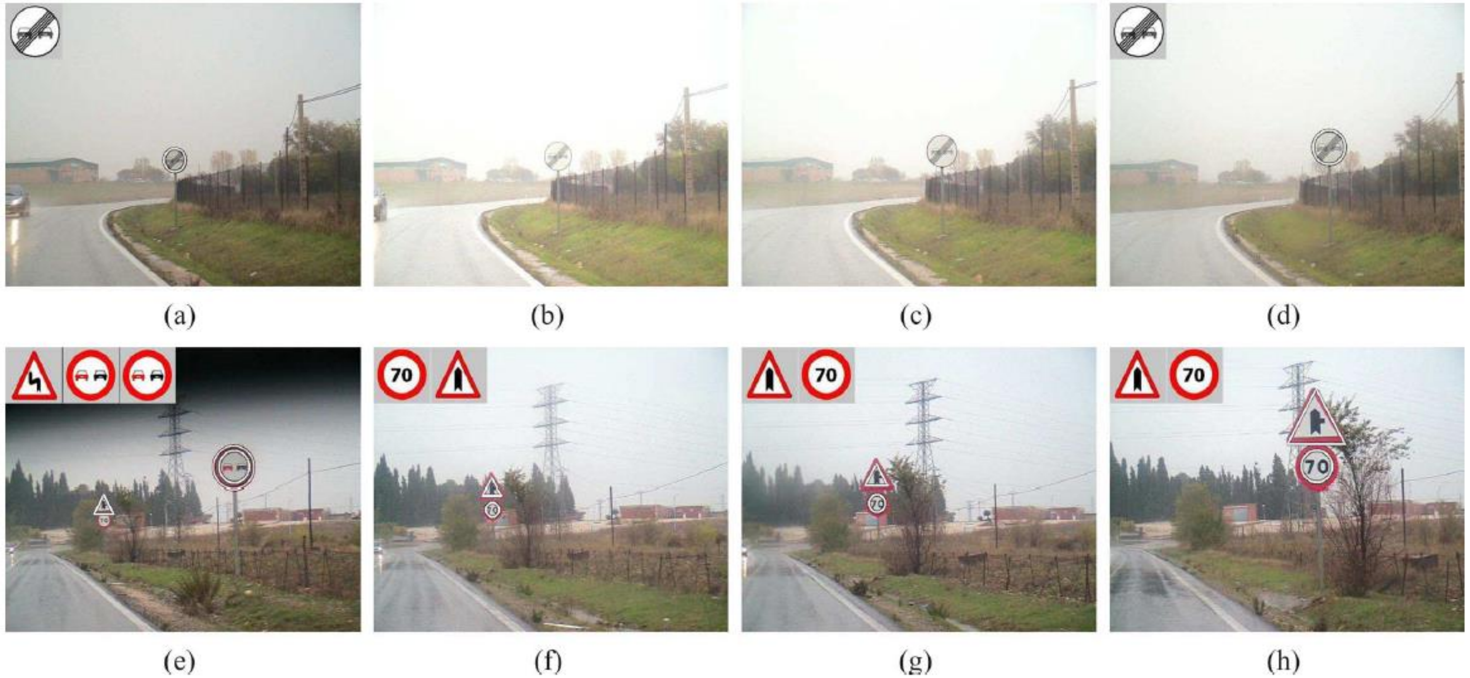
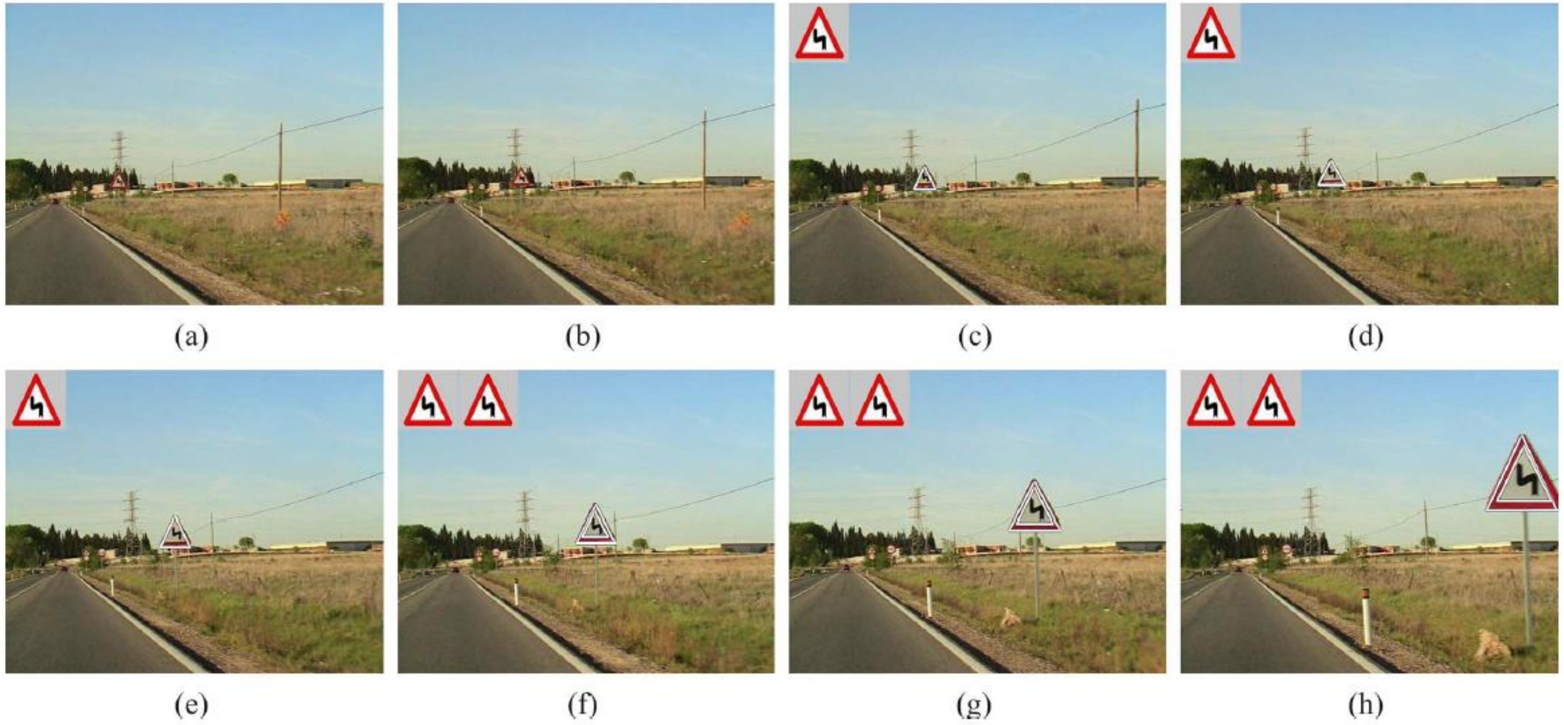


Fig. 20. Examples of recognition process including incorrectly recognized frames. (a)–(d) Recognition with misdetection in (b) and (c). (e)–(h) Recognition with confused classification in (e).

RESULTS

- System is generally able to recognize objects with so many different scales as standard traffic signs



(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 22. Experimental results with eight frames for a triangular road sign.

RESULTS

- The road signs with color combination have been detected successfully.

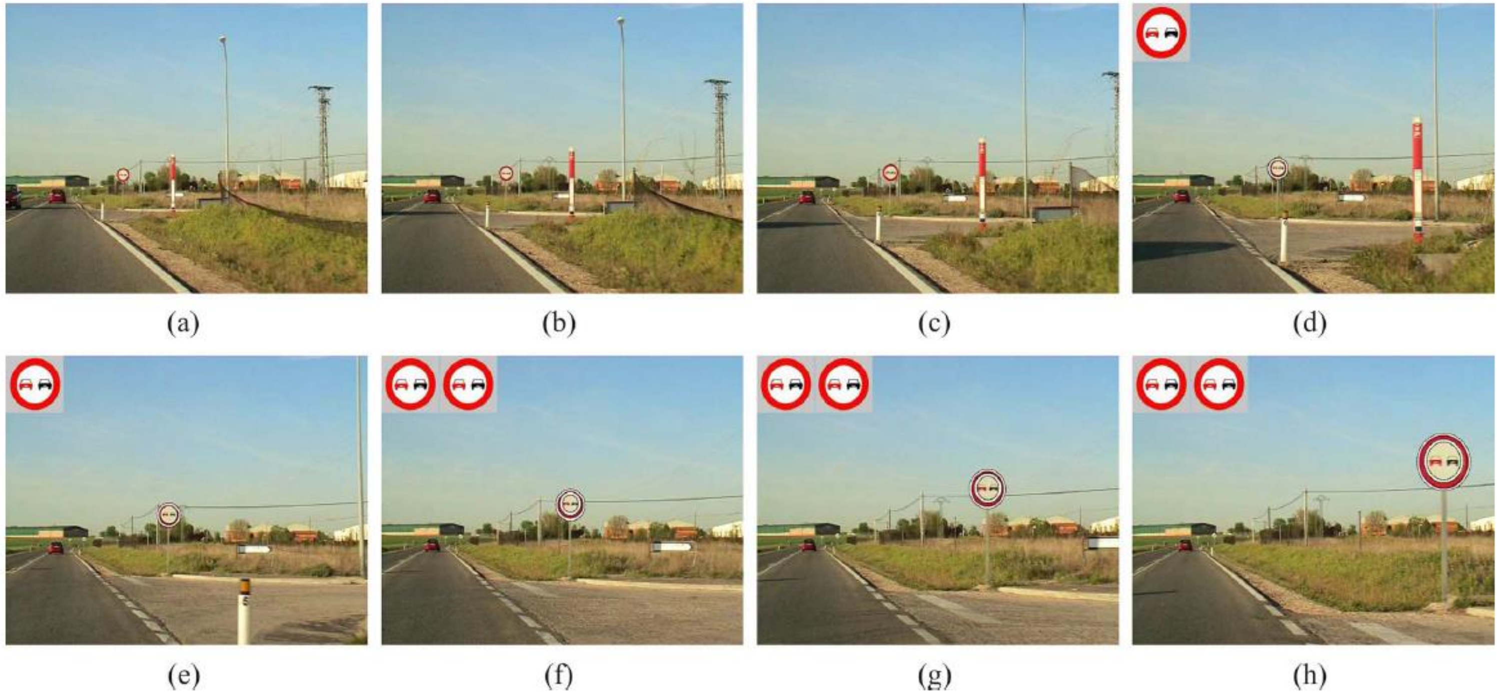


Fig. 21. Experimental results with eight frames for a circular road sign. The size of each image is 720 × 576 pixels, and the time between two successive images is 0.2 s.

RESULTS

- System also works when the signs are not placed perpendicular to the movement of the vehicle (3-D rotations)

Fig. 23. Experimental results with 3-D rotation.

- The results of system are similar during both day and night.

Fig. 24. Experimental results at night.

RESULTS

- System also works when arrays of two or more traffic signs exist in the image.

(a) (b) (c) (d)

Fig. 25. Experimental results with array road signs.

RESULTS

- System also works in occlusion scenarios.

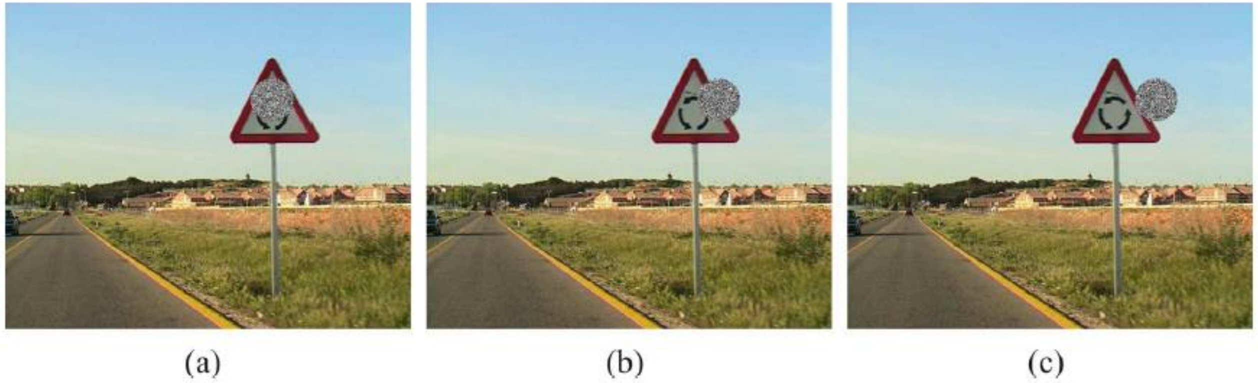


Fig. 26. Masks of occlusion. Three sizes are used to test the robustness of the system against occlusions with different areas covered. (a)–(c) Masks whose diameters are one half, one third, and one fourth of the major dimension of the bounding box, respectively.

- The recognition success probabilities are 93.24%, 67.85%, and 44.90% for the small, medium-sized, and large masks, respectively.

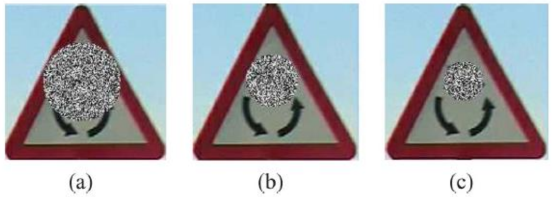


Fig. 27. Displacements followed by a mask for the first orientation.

OUTLINE

- Introduction
- Proposed Method
- Results
- Conclusions

CONCLUSIONS

- The paper describes a complete method to detect and recognize Spanish traffic signs from a video sequence.
- The system can be useful for the maintenance of traffic signs and visual driver assistance system.
- The system consists of three main stages including **segmentation**, **shape classification** and **content recognition**.
  - The shape classification is based on linear SVMs and a content recognition is developed with Gaussian kernels.

THE STRENGTHS AND WEAKNESSES

- **Strengths:**
  - The system detects different geometric shapes, i.e., circular and octagonal, and triangular and rectangular
  - It works correctly in difficult situations, i.e., when array signs appear on the scene or when images are taken under night
  - It is invariant to rotations, changes of scale, and different positions.
  - It can also detect signs that are partially occluded.

- **Weaknesses**:
  - The system is not real-time (high computational complexity)
    - A mean processing time of 1.77 sec per frame on a 2.2-GHz Pentium 4-M
  - Some road signs such as rectangular route-guidance signs for navigation are not recognized.
  - In some scenes, the sign is not detected due to the similarity of the sign and the background.
  - The recognition traffic signs depends on the color segmentation process which can be affected by motion blur, geometric distortions, and complex background ant it is influenced greatly by light and weather.
  - The system operates on single images. Exploiting image sequences improves robustness and may provide additional information, such as the sign location

THANK YOU!