

Research Article

Eigen-Gradients for Traffic Sign Recognition

**Sheila Esmeralda Gonzalez-Reyna, Juan Gabriel Avina-Cervantes,
Sergio Eduardo Ledesma-Orozco, and Ivan Cruz-Aceves**

*Division de Ingenierias Campus Irapuato-Salamanca, Universidad de Guanajuato,
Carretera Salamanca-Valle de Santiago km 3.5 + 1.8 km, Comunidad de Palo Blanco, 36885 Salamanca, GTO, Mexico*

Correspondence should be addressed to Sheila Esmeralda Gonzalez-Reyna; se.gonzalezreyna@ugto.mx

Received 19 July 2013; Revised 1 November 2013; Accepted 8 November 2013

Academic Editor: Daniel Zaldivar

Copyright © 2013 Sheila Esmeralda Gonzalez-Reyna et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Traffic sign detection and recognition systems include a variety of applications like autonomous driving, road sign inventory, and driver support systems. Machine learning algorithms provide useful tools for traffic sign identification tasks. However, classification algorithms depend on the preprocessing stage to obtain high accuracy rates. This paper proposes a road sign characterization method based on oriented gradient maps and the Karhunen-Loeve transform in order to improve classification performance. Dimensionality reduction may be important for portable applications on resource constrained devices like FPGAs; therefore, our approach focuses on achieving a good classification accuracy by using a reduced amount of attributes compared to some state-of-the-art methods. The proposed method was tested using German Traffic Sign Recognition Benchmark, reaching a dimensionality reduction of 99.3% and a classification accuracy of 95.9% with a Multi-Layer Perceptron.

1. Introduction

Road signs provide visual information to drivers by warning them from possible danger on the road, setting rules for pedestrian protection, informing people about their environment, and so forth.

Traffic sign recognition (TSR) is a challenging task due to illumination changes, scene complexity, and weather conditions. Moreover, high motion velocity may cause blurred images when taken from a moving car. In addition, road signs may be disoriented, damaged, or partially occluded. All these factors (shown in Figure 1) may significantly reduce the TSR systems performance.

In recent years, many approaches for road sign detection and recognition have been developed [1]. There is yet a variety of issues to be solved: high rate recognition accuracy, false positive avoidance, color invariance, partial occlusion, and real-time processing, among some common examples.

In this work, a method for road signs identification is proposed. This approach is based on oriented gradient maps and the Karhunen-Loeve transform as preprocessing steps. Oriented gradient maps provide a color and lighting

invariant, strong source of information about traffic signs general appearance, while the Karhunen-Loeve transform makes dimensionality reduction possible without significant loss of information; furthermore, it allows a good traffic sign characterization using a minimal number of attributes, which is an important factor for applications on resource constrained devices like FPGAs. Artificial neural networks perform the classification stage, due to their noise robustness and generalization abilities.

The rest of the document is organized as follows. In the next section, relevant work in the field of traffic sign recognition is briefly mentioned. The techniques related to the proposed approach are described in Section 3. In Section 4, the results are presented and discussed. Finally, in Section 5, conclusions are stated.

2. Relevant Work

Traffic signs are pictograms that provide warnings, rules, and general information to the driver. They are designed with specific colors and geometric shapes in order to be easily detected and recognized by humans. Moreover, intelligent

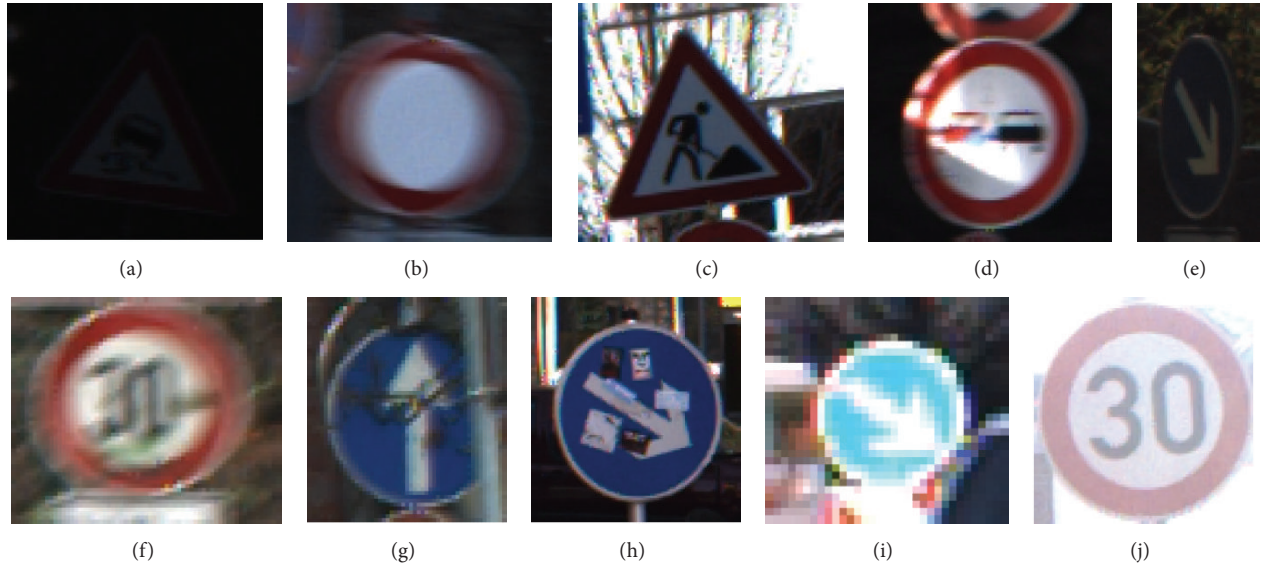


FIGURE 1: Road signs examples and the difficulties in the recognition tasks: (a) lighting, (b) blurring, (c) rotation, (d) shadows, (e) visual perspective, (f) blurring and graffiti, (g) graffiti, (h) partial occlusion, (i) nonuniform colors, and (j) fog.

systems take advantage of these particular features, in order to discriminate road signs from the background information.

The field of TSR has been mainly divided into detection and recognition stages. Detection refers to an automatic localization process while recognition intends to identify the pictograms. Even though these stages can be used independently, they are usually treated as interrelated and complementary tasks.

In traffic sign localization, RGB color is often exploited to discriminate regions of interest (RoI) [5–7]. However, since real traffic sign images are seriously affected by illumination changes, among other factors, many approaches have been proposed to use different color spaces, such as HSI [8, 9], HSV [10], and CIECAM [11].

After using RoI discrimination, many authors use shape information to correctly detect the signs. Dimension of RoIs, width/height, and edge/area ratios are used in [7] to discriminate incorrect candidates and categorize shapes. The authors of [6, 9] used size rejection and aspect ratio as geometric filtering. Blob signature is proposed in [8]. Other approaches make use of morphological filters to reduce the search space [5].

Once the detection process is complete, the next step is the road sign categorization. Two important elements in this stage will be considered: the features that describe the image and the classification tool. The discriminatory attributes in an image are decisive, because they make the classifier task more reliable, and therefore the recognition rate will be improved. In this field, a variety of features have come to attention, K-d trees, Random Forests and Support Vector Machines (SVM) classification of histograms of oriented gradients (HOG) descriptors were analyzed in [5], pixels of interest as features for SVM were used in [9]. Foveal Systems for Traffic Signs (FOSTS) models imitate mechanisms of biological visual

systems, and in [11], classification of traffic signs is performed by the means of these features. Radial Basis Function Neural Classifier was used in [12]. A deeper state of the art in TSR systems is described in [1, 2].

Some relevant work based on the use of gradient orientation as image descriptor has been developed. In [13], gesture recognition is done by exploiting Motion Gradient Orientation (MGO) images to form motion features and Principal Component Analysis (PCA) for dimensionality reduction, and Sparse Bayesian classifiers perform the identification task. Liao and Chung in [14] proposed a novel Local Binary Pattern (LBP) named Local Gradient Orientation Binary Pattern (LGOBP) which captures the neighborhood gradient orientation information for face recognition applications.

3. The Proposed Approach

3.1. Oriented Gradient Maps. Gradients have been widely used for shape recognition [5, 15, 16]. Since gradients are vectors, one of its representations is composition of magnitude and direction. In this study, gradient orientation is the main attribute; therefore, orientation takes only the active range $[0, \pi)$.

The basic idea behind choosing oriented gradients for image representation consists of finding a local feature for appearance and shape which can correctly characterize the object, in spite of the lighting and weather conditions of the original image.

The computation of oriented gradient maps was done according to Dalal and Triggs [15] and Hinterstoisser et al. [16] methodology. Since RGB and CIE-Lab color spaces are reported to produce similar results [15], RGB color space is used for this research. Similarly, gradients calculation is

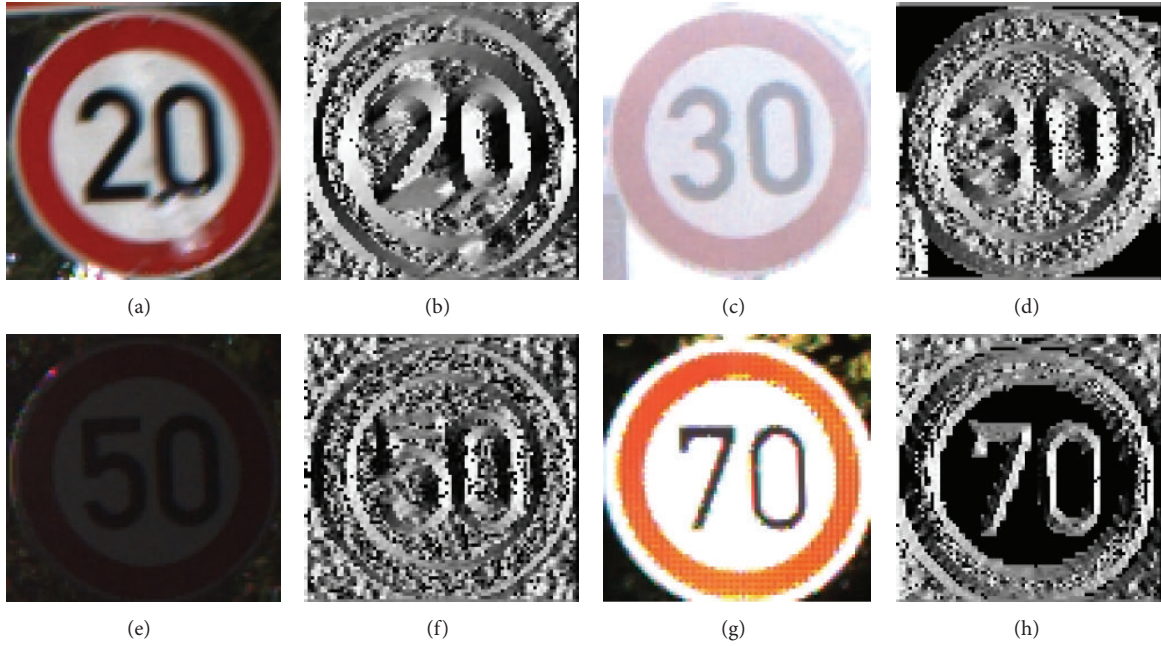


FIGURE 2: Maps of oriented gradients are invariant to illumination.

achieved by the simplest convolution masks, as can be seen in

$$\begin{aligned}\frac{\partial J}{\partial x} &= J * [-1, 0, 1], \\ \frac{\partial J}{\partial y} &= J * [-1, 0, 1]^T,\end{aligned}\quad (1)$$

where $J \in \{R, G, B\}$. Gradients norm L_1 r and orientation θ can be calculated as shown in

$$\begin{aligned}r &= \left| \frac{\partial J}{\partial x} \right| + \left| \frac{\partial J}{\partial y} \right|, \\ \theta &= \arctan \left(\frac{\partial J / \partial y}{\partial J / \partial x} \right), \quad \theta \in [0, \pi).\end{aligned}\quad (2)$$

For color images, gradients are calculated for each color channel, and then the one with the largest Manhattan norm on the pixel's gradient vector remains, forming a single gradient orientation map as follows:

$$\Theta = \theta_C, \quad C \in \max \{r_R, r_G, r_B\}.\quad (3)$$

Figure 2 shows orientation maps for different images, it can be seen the features robustness to varying color and illumination. Gradient orientation maps are the system's main source of information, but since the number of attributes is still large, a dimensionality reduction scheme will be performed.

When the gradient of an image is computed, two equally important components can be calculated: magnitude and orientation. Gradient magnitude is affected by illumination conditions, and therefore contrast normalization is usually

needed. On the other hand, gradient orientation is illumination invariant.

In [15], Dalal and Triggs used gradient magnitude for two purposes: (1) for color images, gradient orientation is computed for each color channel and the gradient orientation with the largest norm L_2 is taken, and (2) for histogram calculation, a weighted vote is performed on the pixels orientation by using gradient magnitude as weighting factor. Our approach considers the gradient magnitude only to compute gradient orientation for color images. Gradient orientation is the main source of information and it is not further affected by the gradient magnitude. The Karhunen-Loeve transform is a tool that can be used to decide which features carry important information. As a consequence, our approach can be considered illumination invariant; furthermore, the number of features per image is lower than HOG too. Additionally, although both methods use gradient orientation, the contrast normalization, cell voting and histograms computing steps are not required; therefore, computation time is considerably lower than HOG. Our method can be improved by both the use of a hierarchical design and a classifier that considers frequency of observations per class.

3.2. Karhunen-Loeve Transform for Image Classification. The Karhunen-Loeve transform (also known as Hotelling transform, Eigenvector transform, or Principal Component Analysis) analyses a data set representing observations composed of several correlated variables. Its goal is to extract the relevant information and express it by a set of orthogonal new variables, the principal components. The new representation is supposed to be in a more tractable, lower-dimensional form, without losing too much information.

Turk and Pentland [17] proposed a popular method for face recognition based on PCA, the so-called Eigenfaces. The present study makes use of their methodology in order to identify road signs.

Let \vec{X} be an N^2 -dimensional vector, corresponding to an image with $N \times N$ pixels. Since we have M different images, the data train set can be represented as an input matrix \mathbf{A} , size $N^2 \times M$. In order to center the data, the mean vector will be computed as shown in

$$\mu = \frac{1}{M} \sum_{i=1}^M \vec{X}_i. \quad (4)$$

So when the mean vector is subtracted from the rest of the data, the resulting matrix $\bar{\mathbf{A}}$ will have zero mean in every component feature. Now, having the new centered data $\bar{\mathbf{A}}$, the covariance matrix can be defined as

$$\mathbf{C} = \frac{1}{M-1} \bar{\mathbf{A}} \bar{\mathbf{A}}^T, \quad (5)$$

where the resulting matrix \mathbf{C} will have $N^2 \times N^2$ elements. The rest of the algorithm is only feasible when the original images are small, and N^2 is less than M .

The next step consists of the calculation of the eigenvalues λ and the eigenvectors \vec{v} of the covariance matrix. Subsequently, since the purpose of this study is data reduction, the most significant vectors, which correspond to the greatest eigenvalues, will be chosen. Originally, the eigenvectors were represented by an $N^2 \times N^2$ matrix, but the k most significant vectors will remain, so \mathbf{V} will be $N^2 \times k$. The projection of any \vec{X} vector onto the new feature space is now given by

$$\vec{Y} = \mathbf{V}^T \vec{X}, \quad (6)$$

where the new vector \vec{Y} is now k -dimensional.

4. Results

All the experiments were performed on the German Traffic Sign Recognition Benchmark (GTSRB) [2]. The database contains more than 39,000 images for training and over 12,600 images for evaluation, with 43 different classes of traffic signs. Image sizes may vary between 15×15 and 128×128 pixels. The GTSRB provides images with varying illumination, weather conditions, shadows, partial occlusion, graffiti, blurring, perspective, and rotated images. The database also offers precomputed attributes per image: three different histogram of oriented gradient features (HOG 1, HOG 2, and HOG 3) and Haar-like features.

A Multi-Layer Perceptron (MLP) performed the categorization stage. MLP is a type of feedforward artificial neural network that can be used to develop classification tasks. It is robust to noise and presents satisfactory nonlinear discriminative properties.

The proposed methodology considers that all images have the same size. Since the GTSRB provides images with varying sizes, all of them were scaled to a fixed size using bilinear

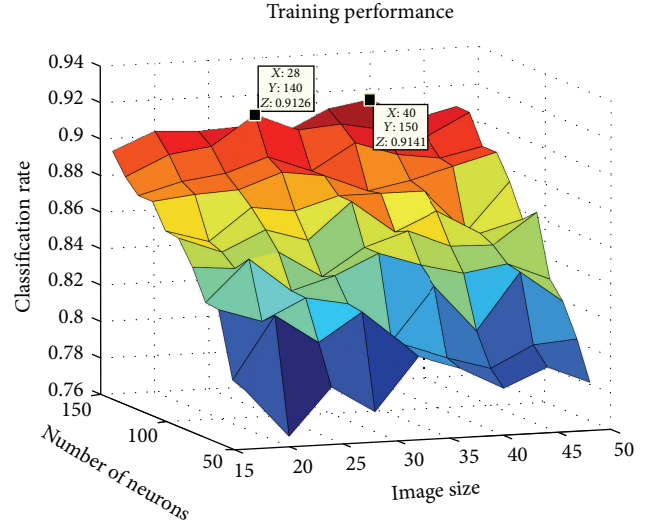


FIGURE 3: Classification performance test for varying image sizes and number of neurons in MLP hidden layer.

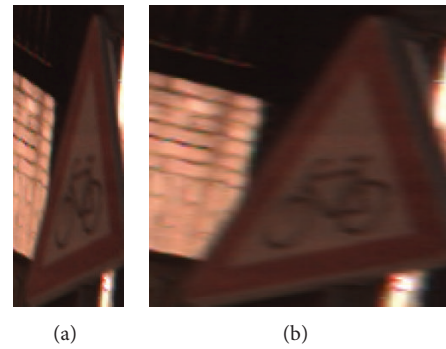


FIGURE 4: (a) Image with strong geometric variation in perspective. (b) Image correcting by bilinear interpolation resizing.

interpolation. Furthermore, this operation helps correcting some geometric transformations like perspective in images [10] (Figure 4). Although some affine transformations cannot be corrected without the use of perspective additional attributes, the classifier may achieve correct classification by considering images with these transformations in the training set.

In order to test the proposed technique, several analyses were performed. Firstly, the system was tested for different image sizes and different number of neurons in the hidden layer. Figure 3 shows that the best classification rates were reached for two different sizes, 28×28 and 40×40 pixels, and for 140 and 150 neurons in the MLP hidden layer, respectively. For data reduction purposes, the minimal size represents the best choice.

The aim of dimensionality reduction leads to the analysis of the minimum number of eigenvectors required to obtain the highest classification rate. The number of eigenvectors is decisive, because it provides the best representation of the data with the minimal amount of information and the highest discriminatory efficiency. It can be observed in

TABLE 1: General comparison with the methods reported in [2]. In the central column a brief description of the method is given: classification tool, image size, number of attributes per image, and features used for recognition.

Team	Method	Accuracy %
IDSIA [3]	CNN + MLP, image size 48×48 , 8 attributes per pixel (18,432 attributes per image), color features	99.46
INI-RTCV	Human performance (best individual)	99.22
INI-RTCV	Human performance (average)	98.84
Sermanet [4]	CNN, image size 32×32 , 3,072 attributes per image, color features	98.31
CAOR [5]	k -NN + Random Forests, image size 40×40 , 1,568 attributes per image, HOG 2 features	96.14
INI-RTCV	LDA, image size 40×40 , 1,568 attributes per image, HOG 2 features	95.68
INI-RTCV	LDA, image size 40×40 , 1,568 attributes per image, HOG 1 features	93.18
INI-RTCV	LDA, image size 40×40 , 2,916 attributes per image, HOG 3 features	92.34
Eigen-Gradients	MLP, image size 28×28 , 140 attributes per image, gradient-orientation + PCA features	95.9

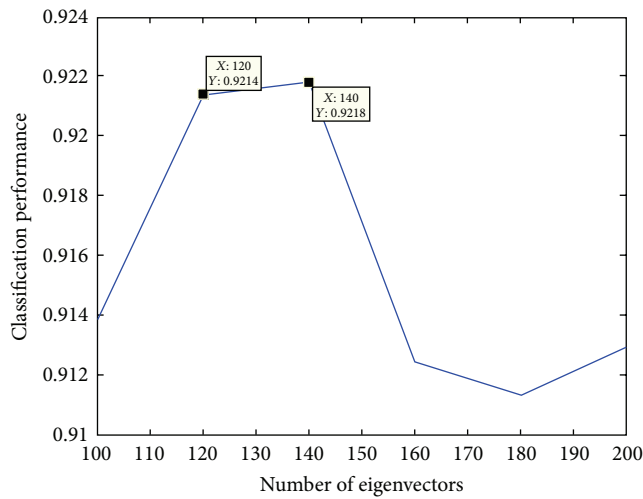


FIGURE 5: Classification performance test for a varying number of eigenvectors (dimensionality reduction).

Figure 5 that the best performance was obtained by both 120 and 140 eigenvectors. Data reduction is therefore in the range 99.2% ~ 99.3%.

After the calculation of optimal size and optimal number of eigenvectors, the debug process for the number of hidden neurons in the MLP was performed, with the main purpose of increasing the classification accuracy. The results of this step are depicted in Figure 6.

Finally, the highest performance is obtained by 240 neurons, image size 28×28 , and 140 eigenvectors, achieving a 95.9% of identification accuracy. A comparison test was performed for a system with the same image size, number of hidden neurons, and eigenvectors, but with raw grayscale images (no oriented gradient maps were computed), and the maximum identification accuracy was 93.39%, demonstrating the effectiveness of the proposed approach.

Once the neural network was trained, for evaluation purposes, the average image processing time is about 0.0054 s. This was tested using MATLAB v.2011 and an iMac with Intel Core i5 processor, 4 GB in RAM.

In 2012, a competition for TSR was launched using the

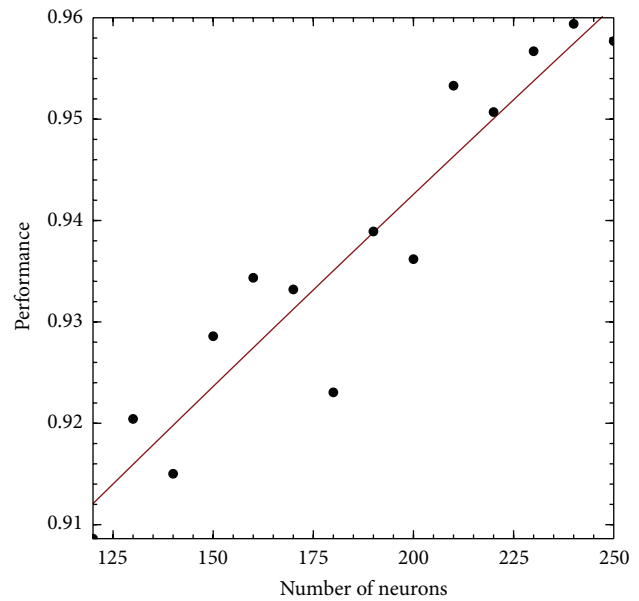


FIGURE 6: Classification performance test for a varying number of neurons in hidden layer. The tested system reached a maximum classification accuracy when using 240 hidden neurons, and remains unchanged after that.

GTSRB [2]. The authors report three baseline results: Linear Discriminant Analysis (LDA) on HOG features, k -nearest neighbors (k -NN) on HOG features, and human performance. In the competition, many teams participated, but only the top-ranking methods were reported. Table 1 shows a brief description and accuracies for those teams, the baseline results, and our approach. Convolutional neural networks (CNN) are used by two top-ranking methods. It is important to highlight that our approach uses a number of attributes at least 10 times lower than the compared methods, and the accuracy of the proposed method is comparable to that of those that use HOG attributes.

Table 1 shows four different methodologies that use HOG as a source of feature information. It can be seen that their accuracy is similar to that of the proposed method, but our method uses less amount of attributes and is working in real time (0.0054 s per image).

5. Conclusions

When attempting to recognize objects, data representation is decisive. Patterns can or cannot be well discriminated depending on the choice of the features that will characterize those objects.

In the present work, road sign images were identified. Because of all the difficulties previously mentioned, it was important to generate discriminative and robust data. Gradient orientations provide robust representation for extreme lighting condition images. Additionally, images carry a huge amount of information that can be hard to process, and as a consequence, dimensionality reduction plays an important role in this approach.

Experimental results demonstrate that the proposed method is robust and accurate compared to raw image process. When compared to some state-of-the-art methods, the number of attributes considered is at least ten times lower, which is a desirable condition for portable applications. Classification accuracy is similar to that of the methods that use HOG attributes. However, recognition performance can be further improved. In future work, classification task could be separated in two substages: (1) shape classification and (2) road sign recognition; therefore, identification accuracy may increase.

Acknowledgments

This work has been supported by the National Council of Science and Technology of Mexico (CONACYT) under the Grant no. 329483/229696 and the Universidad de Guanajuato through the Program of Institutional Enhancement PIFI-2012.

References

- [1] A. Mogelmose, M. M. Trivedi, and T. B. Moeslund, "Vision-based traffic sign detection and analysis for intelligent driver assistance systems: perspectives and survey," *IEEE Transactions on Intelligent Transportation Systems*, vol. 13, no. 4, pp. 1484–1497, 2012.
- [2] J. Stallkamp, M. Schlipsing, J. Salmen, and C. Igel, "Man versus computer: benchmarking machine learning algorithms for traffic sign recognition," *Neural Networks*, vol. 32, pp. 323–332, 2012.
- [3] D. Cireşan, U. Meier, and J. Schmidhuber, "A committee of neural networks for traffic sign classification," in *Proceedings of the International Joint Conference on Neural Network (IJCNN '11)*, pp. 1918–1921, August 2011.
- [4] P. Sermanet and Y. Lecun, "Traffic sign recognition with multi-scale convolutional networks," in *Proceedings of the International Joint Conference on Neural Network (IJCNN '11)*, pp. 2809–2813, August 2011.
- [5] F. Zaklouta, B. Stanculescu, and O. Hamdoun, "Traffic sign classification using K-d trees and random forests," in *Proceedings of the International Joint Conference on Neural Network (IJCNN '11)*, pp. 2151–2155, August 2011.
- [6] B. Soheilian, N. Paparoditis, and B. Vallet, "Detection and 3D reconstruction of traffic signs from multiple view color images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 77, pp. 1–20, 2013.
- [7] L.-W. Tsai, J.-W. Hsieh, C.-H. Chuang, Y.-J. Tseng, K.-C. Fan, and C.-C. Lee, "Road sign detection using eigen colour," *IET Computer Vision*, vol. 2, no. 3, pp. 164–177, 2008.
- [8] S. Maldonado Bascón, J. Acevedo Rodríguez, S. Lafuente Arroyo, A. Fernández Caballero, and F. López-Ferreras, "An optimization on pictogram identification for the road-sign recognition task using SVMs," *Computer Vision and Image Understanding*, vol. 114, no. 3, pp. 373–383, 2010.
- [9] S. Maldonado-Bascón, S. Lafuente-Arroyo, P. Gil-Jiménez, H. Gómez-Moreno, and F. López-Ferreras, "Road-sign detection and recognition based on support vector machines," *IEEE Transactions on Intelligent Transportation Systems*, vol. 8, no. 2, pp. 264–278, 2007.
- [10] H. Fleyeh and E. Davami, "Eigen-based traffic sign recognition," *IET Intelligent Transport Systems*, vol. 5, no. 3, pp. 190–196, 2011.
- [11] X. W. 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," *Journal of Visual Communication and Image Representation*, vol. 17, no. 4, pp. 675–685, 2006.
- [12] H. L. King, P. S. Kah, and A. Li-Minn, "MIMO Lyapunov theory-based RBF neural classifier for traffic sign recognition," *Applied Computational Intelligence and Soft Computing*, vol. 2012, Article ID 793176, 7 pages, 2012.
- [13] S.-F. Wong and R. Cipolla, "Real-time interpretation of hand motions using a sparse bayes classifier on motion gradient orientation images," in *Proceedings of the British Machine Vision Conference*, vol. 1, pp. 379–388, 2005.
- [14] S. Liao and A. C. S. Chung, "Face recognition with salient local gradient orientation binary patterns," in *Proceedings of the IEEE International Conference on Image Processing (ICIP '09)*, pp. 3317–3320, November 2009.
- [15] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '05)*, pp. 886–893, June 2005.
- [16] S. Hinterstoisser, C. Cagniart, S. Ilic et al., "Gradient response maps for real-time detection of textureless objects," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 5, pp. 876–888, 2012.
- [17] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

