Discriminant Splitting of Regions in Traffic Sign Recognition

Sergio Lafuente-Arroyo[1], Roberto J. López-Sastre[1], Saturnino Maldonado-Bascón[1], and Rafael Martínez-Tomás[2]

[1] GRAM, Department of Signal Theory and Communications, UAH, Alcalá de Henares, Spain
[2] Department of Artificial Intelligence, UNED, Madrid, Spain

Abstract. Mining discriminative spatial patterns in image data is a subject of interest in traffic sign recognition. In this paper, we use an approach for detecting spatial regions that are highly discriminative. The main idea is to search the normalized size blobs for discriminative regions by adaptively partitioning the space into progressively smaller sub-regions. Thus, each cluster of signs is characterized by an unique region pattern which consists of homogeneous and discriminant 2-D regions. The mean intensities of these regions are used as features. To evaluate the discriminative power of the attributes corresponding to detected regions, we performed classification experiments using a classifier based on Support Vector Machines. The proposed method has been tested in a real traffic sign database. Results demonstrate that the method can achieve a considerable reduction of features with respect to extraction from raw images while maintaining accurate.

Keywords: feature extraction, adaptive partitioning, traffic sign classification.

1 Introduction

Within the context of smart cities, there are several applications which need a robust system for the automated detection of traffics signs. A recognition system is a module integrated in a complete traffic sign detection system (TSDS) oriented to applications related to intelligent vehicles or maintenance of highways. When a TSDS is considered in full generality, the sign detection and classification stages are usually distinguished. In a traffic scene image, a road-sign detector identifies a set of candidate regions that have been segmented previously Each region of interest (ROI) is then passed on to a classification module and either assigned to one of the known road-sign classes or rejected as a nonsign. In this paper, we focus on the design of the classification module.

In order to design a road sign classifier, candidate regions must be appropriately represented for the given classification technique. After a basic preprocessing such as scaling of the regions to equal size, or masking out the general sign

J.M. Ferrández Vicente et al. (Eds.): IWINAC 2013, Part II, LNCS 7931, pp. 395–403, 2013.

background a more specific data representation has to be constructed. So far, different data representations have been used for road sign classification. Each candidate region is represented by a vector of numerical characteristics (features). Examples of descriptors used in road sign classification are color histograms [2], wavelets [8], appearance-based features [10], or directly the subsampled pixel intensities [1][5][6].

Based on a database of labeled examples, a road-sign-recognition system is trained, minimizing the error expected on examples unseen in training. Apart from high accuracy in classification of different sign types, a recognition system should also avoid erroneous identification of nonsigns, i.e., limit the number of false alarms. Furthermore, it should be suited for real-time deployment.

In order to reduce the computational load in the recognition stage, we propose a novel approach based on the search of a descriptor which splits the sign in blocks and codifies them in different ways. The most relevant areas are encoded with a high number of features whereas other parts are described with less information. Motivated by the use of partitioning splitting in three dimensional (3D) image data [7], we propose a novel descriptor for traffic sign recognition that splits the image into significative regions. The main idea is the creation of a set of regions that are significative in the sense that these discriminant and homogeneous regions will provide adequate information in order to distinguish a certain set of images from another one. The region splitting is based on a variant of the classical image splitting technique. The features that this method uses are the mean intensities of the regions.

This paper is organized as follows. In the next section we present the database used to measure the quality of the proposed method. In section 3 we introduce the discriminant splitting approach to extract the descriptor features while in section 4 the method is evaluated. Finally, section 5 draws conclusions and prospectives.

2 Traffic Sign Dataset

Unlike other fields of pattern recognition, where researchers work with well-known datasets, there are not standard sets for training and testing for traffic sign detection and recognition system (TSDS). A few publicly available traffic sign data sets exist:

1. German TSR Benchmark (GTSRB) [11]
2. KUL Belgium Traffic Signs Data set (KUL Data set) [12]
3. Swedish Traffic Signs Data set (STS Data set) [3]

Most of these databases have emerged within the last two years and are not yet widely used. Some important limitations is that GTSRB and the KUL Data sets are oriented toward classification, rather than detection, since each image contains exactly one sign without background. However, complete scenes are necessary for detection.

As the objective of a complete TSDS is to analyze the accuracy detection of traffic signs as well as the identification of their pictogram, Recognition and

Multi-sensorial analysis group (GRAM) at the Universidad of Alcalá has collected a complete database of Spanish traffic signs. All the samples have been extracted from sequences acquired by different video-cameras under variable lighting conditions. In general, the sizes of individual local regions used within the trainable similarity measure may vary. However, we have fixed the region size as the external meta-parameter. Input color images of variable size were first converted to a gray-level and re-scaled to 32×32 pixel raster by a nearest neighbor interpolation.

Some examples of normalized patterns of our database are shown in Figure 1 for the case of triangular signs. Note that samples include noisy blobs. For each color-shape combination the data set comprises a different number of classes with unbalanced class frequencies due to some signs are more common in traffic scenes. In Table 1 we summarized the number of classes and the number of patterns of all classes with the same color and shape. It is important to note that many signs can be segmented by two colors. For example, prohibitory signs can be extracted by two criteria corresponding to the red outer rim and the inner white area.

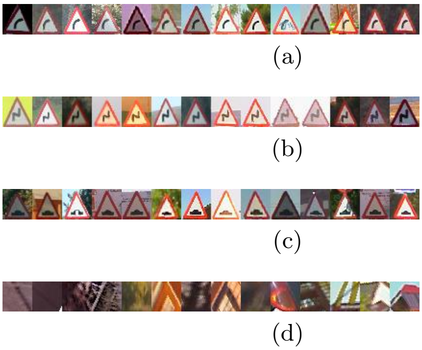

Fig. 1. Sample images of the GRAM dataset. (a), (b) and (c) Samples of red triangular signs. (d) Noisy blobs.

3 Extraction of Descriptors

We assume that there exist N subsets corresponding to the different combinations of color and shape regarding to Table 1. Thus, the dataset D is divided into subsets, $D = \cup_{i=1}^{N} U_i$, and in addition, each color-shape combination includes N_c classes.

Natural image information is not distributed uniformly over each pattern image that includes a sign: there are parts that are most relevant regarding to their contents, while other parts are far less relevant. In fact, the message of traffic signs is given by the pictogram, which covers different parts of the blob

Table 1. Number of classes for different color-shape combinations in the GRAM road-sign dataset

Combination	# Classes	# Samples
Red-Circular	62	18432
Red-Triangular	45	12498
Blue-Circular	54	2770
Blue-Rectangular	99	3600
White-Circular	114	8274
White-Triangular	44	9019
White-Rectangular	114	266
Yellow-Circular	47	1536
Yellow-Triangular	26	570

according to the class. Otherwise, background pixels of the sign does not provide relevant information for recognition task. Background of signs include big areas with pixels of the same color and texture that can be encoded by a representative value. We search homogeneous regions that are discriminant between two sets of images in order to build a map of features.

Let a set U_a containing l samples (images). If each image is of dimensions $h \times w$, these l images can be considered as a stack of slices (volume) with dimensions $l \times h \times w$, as illustrated in Figure 2. Thus for our purpose, a certain region B can be considered as being a parallelepiped volume comprising of the parts of every image in the set that fall within the region. If an image I is divided into R regions we can constitute a map of regions with a strategy similar to coarse-to-fine methods. The subsequent fine-level analysis can thus focus the attention just on the interesting parts of the blob and obtain the descriptor by considering the map. The criterium exploits the homogeneity and any volume homogeneity check method can be used. We have chosen the one based on the intensity range $|I_{max} - I_{min}|$, where I_{max}, I_{min} are the maximum and minimum intensity values of a region. If the range is smaller than a certain threshold, i.e.: $|I_{max} - I_{min}| \leq T_s$, then the region is regarded to be homogeneous, where the threshold T_s denotes the Otsu threshold [9] calculated for the current region. The homogeneity of a region is judged based on the pixels intensity values of the parts of all the training images that fall within the regions boundaries, i.e. on all pixels of the corresponding volume.

In order to determine the discriminant and homogeneous regions for each set, the classical splitting approach is applied to the l images of this set. For each set the stack of images is recursively split into four quadrants or regions (see Figure 2), until 2D homogeneous regions are encountered. The splitting is performed by bisecting the rectangular regions (in the entire image stack) in the vertical and horizontal directions. To deal with the problem of determining the right splitting resolution of the partitioning approach, it dynamically and adaptively partitions the space. For this reason we call the proposed approach Dynamic Recursive Partitioning (DRP).

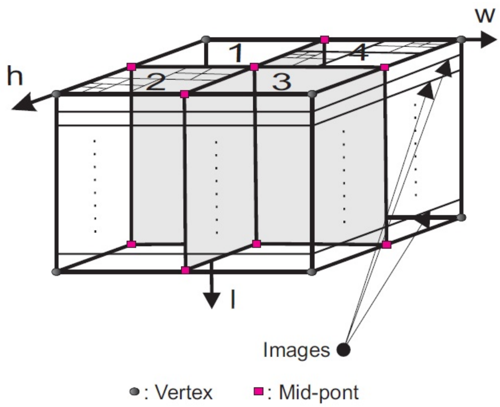

Fig. 2. Representation of a stack of images

An example of application of the method is shown in Figure 3. Note the splitting maps are obtained when the threshold T_s takes values from to 0.1 to 0.7 in steps of 0.1.

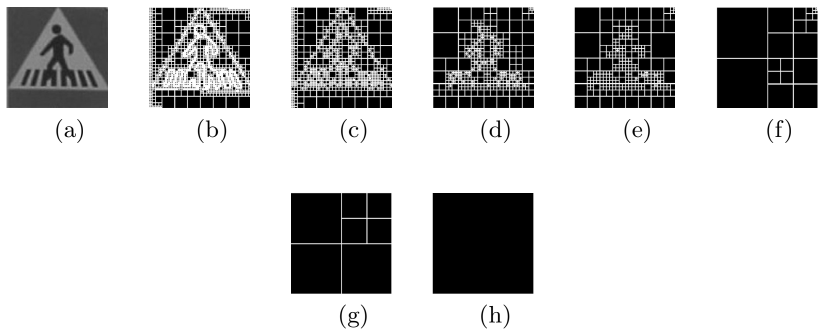

Fig. 3. Example of a traffic sign and its maps of splitting for different thresholds (T_s). (a) Original sign. (b) $T_s = 0.1$. (c) $T_s = 0.2$. (d) $T_s = 0.3$. (e) $T_s = 0.4$. (f) $T_s = 0.5$. (g) $T_s = 0.6$. (h) $T_s = 0.7$.

4 Results

4.1 Experimental Setup

We compared the performance of DRP with the descriptor used in [5], where we scanned directly the raw pixel values in gray-level images. We will refer to this last method as Raw Pixel Descriptor(RPD). Note that in the limit situation in which each block of the DRP method corresponds with a pixel of the image, DRP and RPD are the same. Both descriptors were used as inputs to an architecture based on Support Vector Machines (SVMs) with gaussian kernel. The problem under study is a multi-class one and the extension of the SVM method was done using a one-against-all strategy. From the database, half of the samples was chosen randomly for training and the other half for testing. The training was performed using half of samples from each class of the dataset. The remaining samples were used for testing. Each training-testing trial was repeated five times and we averaged the percentage of the correct predictions to obtain the reported accuracy. Processing time was averaged too.

In the RPD method, background sign is masked out and only those pixels that cover the inner area of the shape are evaluated. The number of features of 32×32 normalized blobs is 1024, 545 and 711 components, respectively, for rectangular, triangular and circular signs.

4.2 Architectures

The DRP descriptor needs the creation of a map of features in the training process. According to how we selected the sets from which we obtained the maps, two strategies were used:

– A map of features is created for each class. Figure 4 shows examples of recognition patterns, where we can see how the maps fit to the pictogram distributions. It is necessary to point out that even when a map is generated for each sample in the training process, we fusion the maps from all training samples that belong to the same class in order to obtain an unique map. This class pattern is created by considering regions more repeated in the maps of the training samples. The descriptor extraction is different for each class and this is the reason why computational cost associated to the extraction is critical in the test phase. Figure 5 illustrates the architecture of this strategy.
– A map of features is generated for each color-shape combination. The stack of images comprises the training samples corresponding to all classes that share color and geometric shape. In this sense an unique map is created for all classes and the descriptor extraction is common for all classes in the test stage. Figure 6 shows the maps associated to three specific sets: blue rectangular, red triangular and red circular. Note that since outer pixels are noisy in many samples, regions of the periphery are strongly split sometimes as it is illustrated in Figures 6(a) and 6(c) even when these areas are quite uniform. Figure 7 illustrates the architecture, which is chosen in this work due to the common process of descriptor extraction for all the classes.

4.3 Results

We compared the performance of the proposed method (DRP) using a map common to all the classes with that based on scanning raw pixels (RPD). The mean intensity values of regions were used as inputs to the classifier (SVMs). Table 2 summarizes the results obtained in the test stage. The number of support vectors (SV) obtained in the training phase is denoted in the third column.

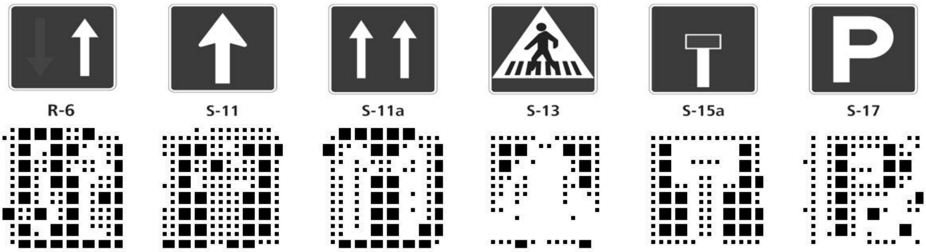

Fig. 4. DRP maps for blue rectangular signs with T_s=0.5

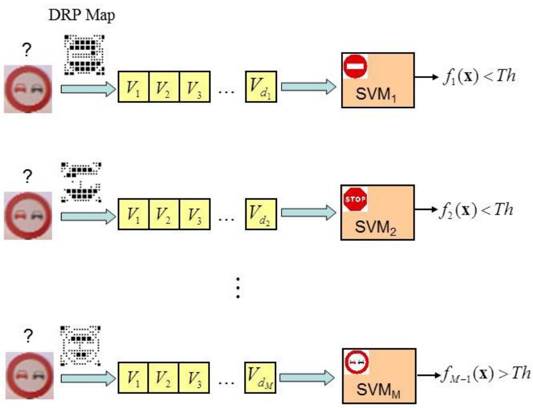

Fig. 5. Architecture DRP for the testing stage when a map of features is generated for each class

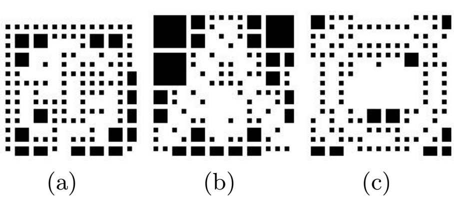

Fig. 6. Global maps for color-shape combinations. (a) Blue rectangular map. (b) Red triangular map. (c) Red circular map.

Table 2. Results with with blue rectangular road signs

Combination	Descriptor	# Features	# SV	$P_d(\%)$	T_{ext}	T_{svm}	T_{total}
Blue Rectangular	RPD	1024	2241	96.01	28.88	54.16	83.04
	DRP	433	1841	95.8	12.76	26.46	39.22
Red Triangular	RPD	545	7133	96.3	108.5	384.3	492.8
	DRP	352	6788	95.52	76.3	328.43	404.73
Red Circular	RPD	711	9440	94.9691	525.87	849.5	1375.37
	DRP	511	9419	94.839	433.29	739.56	1172.85

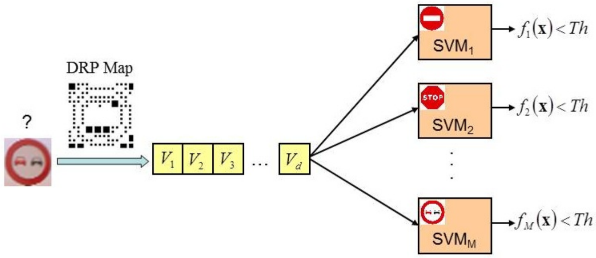

Fig. 7. Architecture DRP for the testing stage when a map of features is generated for each color-shape combination

The computational cost, including computation and storage, increases as the number of support vector does. The classification accuracy across the whole test dataset is included as the percentage of samples classified correctly by SVMs. Finally, the overall execution speed is quantified through the time (in seconds) that algorithm needs to extract the descriptors of training sets (T_{ext}) and classify them with SVMs (T_{svm}), respectively. The total time of classification is given in the last column of the table.

By inspecting the results we can conclude that the goal of this descriptor is the reduction of the number of features used to describe the signs, specially in the case of rectangular ones due to the geometry used in the decomposition. The reduction of processing time is approximately 52.7%, 17.8% and 14.7%, respectively, for blue rectangular, red triangular and red circular traffic signs even when the proposed algorithm was found to be able to recognize traffic signs with similar accuracy compared to RPD. Note that values are the average of 5 times of tests.

5 Conclusions

A scheme based on texture analysis is proposed in this paper in order to get a compact descriptor for the traffic sign recognition task. To show advantages of our approach we have conducted experiments on a real traffic sign dataset.

The mean intensity values of the different pattern blocks are integrated into a vector to characterize the images. Results show that the proposed technique is able to recognize the objects with very good accuracy and low computational burden. Further investigations are ongoing about the applicability of blocks with different geometry, such as radial descriptors.

Acknowledgment. This work was partially supported by projects TIN2010-20845-C03-03, IPT-2011-1366-390000 and IPT-2012-0808-370000.

References

1. de la Escalera, A., Armigol, J.M., Pastor, J.M., Rodriguez, F.J.: Visual sign information extraction and identification by deformable models for intelligent vehicles. IEEE Transactions on Intelligent Transportation Systems 5(2), 57–68 (2004)
2. Hsien, J.C., Liou, Y.S., Chen, S.Y.: Road sign detection and recognition using hidden Markov model. Asian Journal of Health and Information Sciences 1, 85–100 (2006)
3. Larsson, F., Felsberg, M.: Using fourier descriptors and spatial models for traffic sign recognition. In: Heyden, A., Kahl, F. (eds.) SCIA 2011. LNCS, vol. 6688, pp. 238–249. Springer, Heidelberg (2011)
4. Liu, Y., Ikenaga, T., Goto, S.: Geometrical, physical and text/symbol analysis based approach of traffic sign detection system. In: Proceedings of the IEEE Intelligent Vehicle Symposium, pp. 238–243 (2006)
5. Maldonado-Bascon, S., Lafuente-Arroyo, S., Gil-Jiménez, P., Gómez-Moreno, H., López-Ferreras, F.: Road-Sign Detection and Recognition Based on Support Vector Machines. IEEE Trans. on Intelligent Transportation Systems 8(2), 264–278 (2007)
6. Maldonado-Bascon, S., Acevedo-Rodríguez, J., Lafuente-Arroyo, S., Fernández-Caballero, A., López-Ferreras, F.: An optimization on pictogram identification for the road-sign recognition task using SVMs. Computer Vision and Image Understanding 114(3), 373–383 (2010)
7. Megalooikonomou, V., Kontos, D., Pokrajac, D., Lazarevic, A., Obradovic, Z.: An adaptive partitioning approach for mining discriminant regions in 3D image data. Journal of Intelligent Information Systems Archive 31(3), 217–242 (2008)
8. Nguwi, Y.Y., Cho, S.Y.: Two-tier self-organizing visual model for road sign recognition. In: Proceedings of the IEEE International Joint Conference on Neural Networks (IJCNN), pp. 794–799 (2008)
9. Otsu, N.: A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man, and Cybernetics 9, 62–66 (1979)
10. Shaposhnikov, W.D., Shaposhnikov, D.G., Lubov, N., Golovan, E.V., Shevtsova, A.: Road sign recognition by single positioning of space-variant sensor window. In: Proceedings of the 15th International Conference on Vision Interface (2002)
11. Stallkamp, J., Schlipsing, M., Salmen, J., Igel, C.: The german traffic sign recognition benchmark: a multi-class classification competition. In: Proceedings of the International Joint Conference on Neural Networks (IJCNN), pp. 1453–1460 (2011)
12. Timofte, R., Zimmermann, K., Van Gool, L.: Multi-view traffic sign detection, recognition and 3D localisation. Journal of Machine Vision and Applications, 1–15 (2011)