Comparison of Three Approaches for Scenario Classification for the Automotive Field

Nicola Bernini, Massimo Bertozzi, Luca Devincenzi, and Luca Mazzei

Dip. Ing Informazione, Parco Area delle Scienze 181A, 43124 Parma, Italy
bertozzi@vislab.it
www.vislab.it

Abstract. To extend the functionalities of Advanced Driver Assistance Systems (ADAS) and have a more accurate control on the parameters of sensors mounted on an intelligent vehicle, a tool that can classify the scenarios which the vehicle moves in, is needed.

This article presents a comparison of three classification techniques (PCA, ANN and SVM) to obtain a fast and robust scene classifier based only on images. The systems presented in this paper have been trained on three different categories of traffic scenarios: urban, highway, and rural, on a total of more than 23 hours of driving in different countries.

Keywords: scenario classification, intelligent vehicles, automotive.

1 Introduction

During recent years the number of integrated sensors on an intelligent vehicle is progressively increased. In this context ADAS have became significantly relevant allowing more comfortable and safe driving.

The reliability of a sensor is strongly dependent on the scenario in which the vehicle is moving. As an example, if the vehicle is driving in the center of a town among tall buildings or trees, or even inside a tunnel, the positioning system based on GNSS (Global Navigation Satellite System) may have difficulty in giving accurate information. The laserscanner is considered reliable in tight areas or at night, but it is generally negatively affected by the rain, that reflects the light beam, and by dust.

The knowledge of the context in which the vehicle is moving is of primary importance to adjust the sensor configuration in order to improve ADAS reliability and expanding the scope of such systems, freeing them from the restriction to operate only in a restricted number of environments.

Moreover, the knowledge of the scene is valuable to improve high-level applications, allowing fine tuning of planning, such as increasing controls reactivity on the safety distance (on highways, moving at high speed, it should be higher than in city, moving at lower speed) or decide whether or not to rely on GNSS to detect position.

This article presents a comparison of three machine learning techniques for scenario classification in the automotive field. These Systems must to be able

A. Petrosino (Ed.): ICIAP 2013, Part I, LNCS 8156, pp. 582–591, 2013.

(a) (b) (c)

Fig. 1. Examples for different scene classes: (a) City, (b) Highway, and (c) Country Rural Road

to recognize the correct scenario among three possible classes; figure 1 shows examples images of the possible choices: city, highway, and country rural roads.

This paper is organized as follows. The section 2 shows a brief report on the state of art about images segmentation and classification. The section 3 analyzes the whole system structure and shows the developed classifiers in depth. The section 4 presents the results obtained in terms of correct detection and time consumption for the three classification methods, with a brief comparison. Finally, section 5 ends the paper with some final remarks.

2 Related Works

Concerning scenes classification on images, many works are related to image retrieval techniques on image databases. In sever works indoor scenes classification is considered as [5]. In [4] an approach with deformable part model, DPM, is presented. On the other hand, outdoor scenarios classification is performed by low level features and selected categories of scenes, indoor/outdoor, landscape/city, etc. as in [6–8]. Also object bank approach, a high level features method, is used to classify outdoor scenarios, as in [2]. In [3] classification of the scenes is applied to the entire image, with a global scene footprint based on information obtained by the extraction of FFT (Fast Fourier Transform) spectrum from the image.

In [1], he suggested to divide the image into equal parts and, on each of these, to apply a frequency transformation to obtain the features according to the hypothesis that every scenario corresponds to a specific Fourier footprint. These features are reduced using a technique called HPCC (Hierarchical Principal Component Classification), based on PCA, and then LDA and ANNs are used with good results: more than 80% accuracy.

The system detailed in this paper uses a similar approach for the processing, because we believe it's an effective strategy to exploit properly the informative content of the image that's mainly located in subparts of it (the sky and the street in front of the car are generally not very useful to detect the context

the vehicle is moving in), and compares additional classifiers for a performance confrontation, under a real-time constraint..

The novelty of the approach is related to the automotive scenario application and real time performance. Generally, scenario classification techniques are performed on really different ones, while, concerning the automotive environment, the fact that the differences amongst the selected contexts are really limited makes the task tougher.

3 Algorithm Description

The system, which is based exclusively on the analysis of a single frame, is composed of a first common layer that is in charge of pre-process frames with the purpose of obtaining features in the frequency domain. These features are then processed using PCA and analyzed with a Gaussian probability distribution function, artificial neural networks (ANN), and Support Vector Machine (SVM) with Gaussian RBF (Radial Basis Function) Kernel to test which of these three systems is the most reliable, fast and robust forthe classification.

Figure 1 shows typical images of cities, highways and rural roads environments which represent the more common scenarios for vehicles to move in, they constitute the three classes used for the classification in this study.

The overall system is divided in three main parts, as shown in figure 2. The first part (preprocessing) is the image pre-processing that reduces the effect of changes in illumination and eliminates small frequency variations, then the image is resized so it can be divided equally into 16 subparts, which are then separately computed in the sampling phase.

The cutting scheme problem has been heuristically addressed, identifying image ROIs suitable to solve the context detection problem. These ROIs are

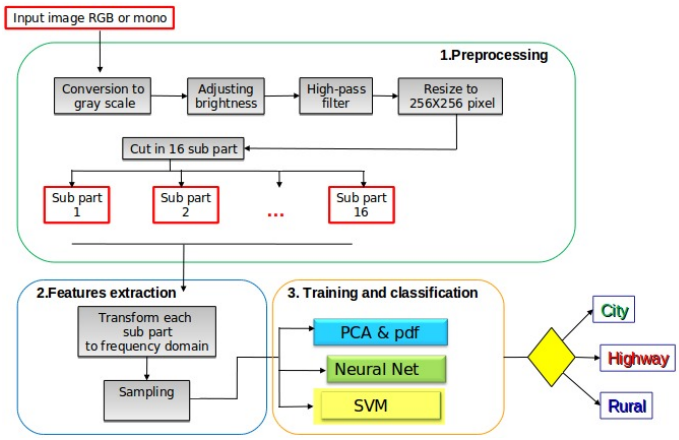

Fig. 2. Algorithm overview flow diagram

square-shaped because FFT will be calculated over them and their dimension depends on the specific image region we want to analyze (sky, front street, sides).

The subdivision phase is a key step in order to be able to exploit the images high similarity in the traffic scenario. Such images have many common parts (presence of road, lines, sky, etc) so dividing them into subparts and treating each of them as a standalone image makes the search for differences a simplified task. This step is similar to the one used in [1] and has been adapted to this specific scenario.

In the second phase (features extraction) an extraction of features to obtain relevant information from the images is performed. The system transforms each subpart of the initial image into frequency domain, using a DFT, so 16 spectra are obtained. Each spectrum is sampled using median filters with different resolutions. The sampled data is then scaled to fit in a predetermined range.

In the third step (training and classification), three methods of classification are applied on the computed samples. In the first approach data is reduced with PCA and a Gaussian probability distribution function for each class is computed. In the second approach an ANN analysis is performed on each subpart of the image, then the each partial result is passed to another ANN in charge of achieving the final classification. In the third approach a SVM is used to map the input space into a new space with higher dimensionality with respect to the previous ones, in order to be able to better discriminate the different classes.

In the following paragraph each algorithm stage is detailed.

3.1 Preprocessing

The preprocessing stage can take as input both gray level or color images; RGB images are converted to grayscale. After that, to expand the dynamics of the

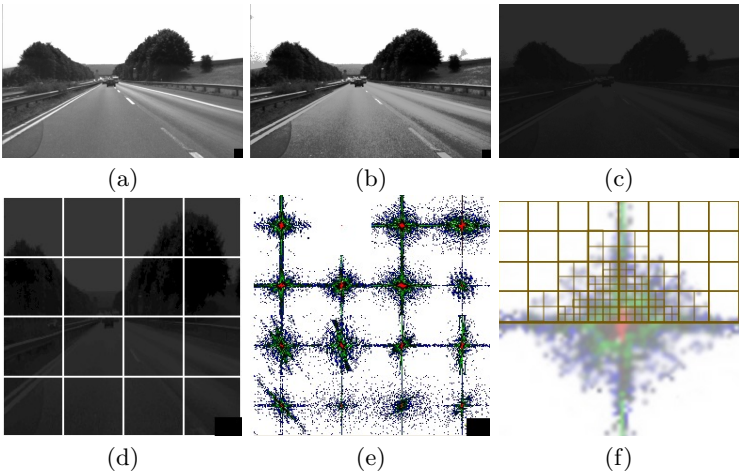

Fig. 3. Image elaboration: (a) input image, (b) histogram normalization, (c) image after suppression of low frequency, (d) rescaled image, (e) subdivision on 16 blocks, (e) DFT results, and (f) 128 masks feature extraction areas

gray levels and to reduce the influence of different illumination conditions, the histogram is normalized. In the following step, a high pass filter is applied, whose frequency response is given by the following equation,

$$H(k,l) = \begin{cases} 1 & se\ k = 0 \cap l = 0 \\ 1 - (1 - 0.9e^{-\frac{k^2+l^2}{156.25}}) & else \end{cases}$$

in order to mitigate the small differences in terms of contrast and frequency that would otherwise produce edges effects in the DFT. The image is then resized to 256×256 pixels in order to have a standard size and it is divided into 16 equal subparts of 64×64 pixels. The division is made to inspect information stored in the amplitude values in the frequency domain in different image regions. This would not be possible if the entire image is directly transformed to the frequency domain. Images 3.b-d show an example of the different preprocessing steps.

3.2 Feature Extraction

The second stage of the system algorithm involves a FFT (Fast Fourier Transform) on each 16 subparts of the image. It is composed of two steps. Initially, FFT is computed on each subpart according to the following formula

$$F_k(i,j) = |F_k(i,j)|e^{\phi(i,j)},\ \text{with } k \in \{1...16\} \text{ subpart}$$

Concerning the purpose of this work, only FFT magnitude is used while the phase information is discarded.

Finally, in the second step of this stage, a sampling of the FFT is performed. Since FFT provides a huge amount of information, a sampling step becomes necessary to enable the following steps to reach real-time performances.

Each spectrum is here filtered by a mask that provides as output 128 values. In the case of road environments, the higher concentration of information is around the origin; while there is little variation at high frequencies. The mask is sized to have low resolution at high frequencies and high resolution at low frequencies, as shown in figure 3.f, Sampling is performed using 2×2 masks in the middle of the subimage, 4×4 mask in the intermediate zones, and a mask of 8×8 pixels for the more external part.

In the same way as a median filter, values of the pixels are used to obtain 128 samples by the 64×64 subpart. The final output of this step of features extraction is 16 vectors of 128 samples. In order to perform a better classification, data used by classifier are scaled in a fixed chosen range $[0, 1]$.

3.3 Training and Classification

Three classifiers are implemented to measure their robustness, performance and execution time.

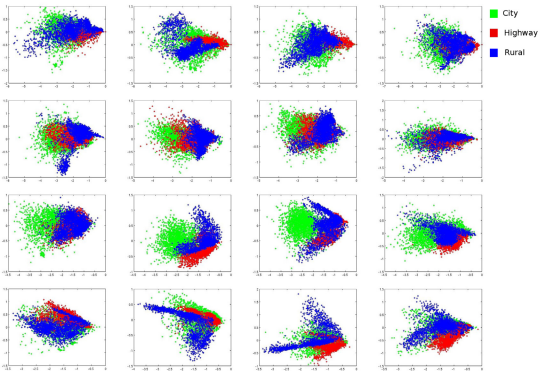

Fig. 4. 2D representation of extracted features

Classification with PCA. This first classifier uses a PCA reduction to reorganize the samples in a small features space, in a second step a vote for each subpart of the image is performed in order to take a classification decision.

The algorithm obtains the 16 matrices for each subpart of the image, as input. Each matrix has a number of row equal to the number of the training set frames, and a number of column equal to the number of the samples.

A reduction in the workspace is performed with a PCA: using the firsts 2 eigenvectors obtained from the SVD (Singular Value Decomposition) to compose a transformation matrix that maps each of the 16 vectors from 128 samples into a 2-dimensional space. Applying this reduction to all of the training set vectors for each subpart, a point cloud is obtained for each class. Figure 4 shows the 2D representation of extracted features: image shows the 16 blocks point clouds. The green points represent the city class, red points the highway class and the blue points to the rural class.

In order to describe statistically these clouds of points, they are characterized by a normal probability density function $\mathcal{N}$ calculating the mean and variance for each class.

During the classification process the same PCA reduction is applied to the image subparts, then the 3 probability density functions are computed to get 3 votes.

To perform the final classification of the entire image, a naïve Bayesian classifier is used to compute hypothesis among classes according to the following formula.

$$\arg\max_{k} \sum_{i} \mathcal{N}(\mathbf{x_i}, \mu_{k,i}, \sigma_{k,i})$$

with $k \in \{r, c, h\}$ classes, and $i \in [1...16]$ subparts of the current image.

Classification with Artificial Neural Networks. The architecture overview of the artificial neural networks is shown schematically in figure 5, and features a

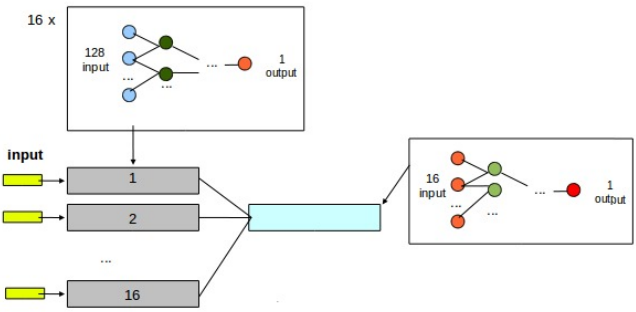

Fig. 5. Artificial neural networks structure

two levels scheme. The first level is based on 16 ANNs, one for each subpart, with 128 inputs and 3 outputs. Each ANN takes decisions using a binary encoding for the 3 class labels. Another ANN, with 48 (16×3) inputs, gathers up the results of the previous 16 networks and takes the final decision on the entire image.

In order to improve this system performances, a proper tuning is need in order to be able to find the optimum setup for parameters like the number of hidden layers, neurons, activation functions, average error, and training time. The proposed approach involves a backpropagation training scheme with an incremental algorithm. In the training test phase, the best results in terms of MSE (Mean Squared Error) (less than 0.005) are obtained with 3-layer networks and a sigmoid hidden layer activation function. The output layer has a stepwise linear approximation to sigmoid activation function. The neurons number of the 16 ANNs on the first level are 97, 66, and 35; while the ANN on the second level has 49, 34, and 19 neurons.

Classification with SVM. The structure of the SVM classifier consists of 16 parallel SVMs that process data from the 16 vectors of subimage samples, producing 16 output, and a second layer formed by another SVM that classifies the entire image using the first layer output: a structure very similar to the ANN-based solution one. The SVM classifier uses a gaussian radial basis kernel function (RBF) to map the input space into a high dimensional feature space, looking for a separating hyperplane. The following formula shows the gaussian radial base function: $k(x_i, x_j) = e^{(-\gamma \|x_i - x_j\|^2)}$ for $\gamma > 0$ (SVM classifier uses $\gamma = 1/2\sigma^2$).

RBF kernel is the best choice because linear kernel and sigmoid kernel can be seen as special RBF cases and they have the same performance with specific parameters. Finally the kernel RBF has the lowest number of computing complexity, moreover cross-validation and grid search can be used to determine the best values for its two parameters: C (used to weight the penalty related to misclassification in soft margin SVM) and γ (Gaussian RBF Kernel free parameter).

4 Results

The training set has been built recording 23 hours of driving and manually assigning parts of sequences to the right category: city, highway and rural. Images of the training set has been sampled with a frequency of 1 FPS so they were not too similar. Images have been taken in Italy, Germany, Russia, Austria and U.S. to have a very broad and general set. Samples for the city class consist of up to 2698 frames including images that show narrow streets, no more than two lanes wide, tall buildings, as well as roundabouts and intersections. The training set for highway class includes 2046 frames where the road is wide, with two or more lanes, few or none tall trees on the sides and there are no subways or tunnels which can occlude the sky. The samples for the rural scenarios refer to typical rural landscape, with fields and vegetation on the sides of the road that is at least two lanes wide, with few or no intersections, roundabouts, houses and buildings, for a total of 2205 frames. The test set therefore sums up to 19231 images which were also introduced noisy elements in, such as plants and highway overpass, squares and boulevards in towns.

Table 1 shows the correct detection on test set and figure 6 shows correct detection for the three classes. On the left corner there are three box that encode the vote of PCA, ANNs and SVMs. A vote for city is shown as green, while votes for highway or rural scenarios are rendered in red or blue respectively.

All the three classifiers correctly discriminate urban environment, where the SVM percentage reaches 100%. The best method for the highway classification is the ANN. Concerning the rural context only the SVMs, that work in a larger feature space proved to be able to identify an environment that is someway really similar to the other two ones. The PCA usually mismatches rural environment as urban scenario, while the ANN as highways. SVMs, however, managed to reach a high percentage of correctness of about 95% in all scenarios.

The confusion matrices for all classifiers are detailed in table 1. The ANN confusion matrix clearly shows that ANN were not able to distinguish rural images in the test set, always mismatching them as highways.

Concerning the execution time, a large part of runtime is required by the image pre-processing stage, because of the filters, the DFT, and the sampling. The total

Table 1. Correct Detections and confusion matrix for the 3 classification methods

Method	Correct Detections		
	City	Highway	Rural
PCA	95.6%	96.9%	7.0 %
ANN	91.6%	100%	6.0 %
SVM	100%	84.4%	95.0%

Method	**Scenario**	City	Highway	Rural
PCA	City	95.6%	4.0%	0.4%
	Highway	2.4%	96.9%	0.7%
	Rural	64.0%	29.0%	7.0%
ANN	City	91.6%	8.1%	0.3%
	Highway	0.2%	93.3%	0.5%
	Rural	0.3%	93.7%	6.0%
SVM	City	100.0%	0.0%	0.0%
	Highway	15.6%	84.4%	0.0%
	Rural	5.0%	0.0%	95.0%

(a) (b) (c)

Fig. 6. Results on classification: correct detection of (a) city, (b) highway, (c) and rural class.

execution time of the pre-elaboration stage is 15.2 ms. For the classification stage, PCA and ANN processing algorithms proved to be both quick requiring only 2 ms and 1.3 ms respectively. SVM was the classifier that required longer time with 18.5 ms. Even in the slowest case, the complete algorithm was still below 38 ms and seems therefore suited for the use in the automotive field. The tests were performed on a Intel Core 2 P7450@2.133 GHz Duo, with 2 GBytes of RAM at 1600 MHz.

5 Conclusions

This paper presents the implementation of a classifier for traffic scenarios through the comparison of different machine learning techniques, PCA, ANN, and SVM.

The classification tasks are performed only on images, of any kind, that are pre-processed to extract spectral information. Results obtained demonstrate that it is necessary to work in a larger space in order to be able to discriminate among classes with elements so similar, as shown by the SVM classifier. In all the cases, except for ANN with rural scenario, high correct detection rate was achieved and results were accompanied by excellent execution times well below the human average reaction time.

In city or highway scenarios, either PCA and ANN classifier can be used, due to their high rates of accuracy. Vice-versa, if the purpose is to discriminate amongst all of the three classes, only SVM classifier is reliable.

From the data collected it emerges that SVM is generally the best method and reaches the best classification performance while respecting the time constraint related to the Automotive Environment, that's a mandatory requirement.

SVM outperforming ANN is a fact that has already been observed in other kind of application [9, 10]. The reason is probably related to the Non-Linear SVM ability to discriminate in a non linear way, due to use of the kernel trick that transforms the problem of a non-linear separating curve identification, in a problem of a separating hyperplane identification in high dimensionality. In the present case, a Gaussian RBF Kernel has been used, so the classification has occurred in an Infinite Dimensional Hilber Space.

The proposed system is well suited for a variety of different applications in the automotive fields, as examples: the support to other ADAS for the automatic setting of driving parameters or the environment analysis on completely unmanned intelligent vehicles.

Acknowledgments. The work described in this paper has been developed in the framework of the OFAV Project funded by the European Research Council (ERC) within an Advanced Investigators Grant.

References

1. Kastner, R., Schneider, F., Michalke, T., Fritsch, J., Goerick, C.: Image-based classification of driving scenes by hierarchical principal component classification (hpcc). In: 2009 IEEE Intelligent Vehicles Symposium, pp. 341–346 (June 2009)
2. Li, L.-J., Xing, E.P., Su, H., Fei-Fei, L.: Object bank: A high-level image representation for scene classification & semantic feature sparsification. In: Neural Information Processing Systems (NIPS), Vancouver, Canada, pp. 1378–1386 (December 2010)
3. Oliva, A., Torralba, A., Dugue, A.G., Herault, J.: Global semantic classification of scenes using power spectrum templates (1999)
4. Pandey, M., Lazebnik, S.: Scene recognition and weakly supervised object localization with deformable part-based models. In: 2011 IEEE International Conference on Computer Vision (ICCV), pp. 1307–1314 (November 2011)
5. Quattoni, A., Torralba, A.: Recognizing indoor scenes. In: IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2009, pp. 413–420 (June 2009)
6. Serrano, N., Savakis, A., Luo, A.: A computationally efficient approach to indoor/outdoor scene classification. In: Proceedings of the16th International Conference on Pattern Recognition, vol. 4, pp. 146–149 (2002)
7. Szummer, M., Picard, R.: Indoor-outdoor image classification. In: IEEE International Workshop on Content-Based Access of Image and Video Database, pp. 42–51 (January 1998)
8. Vailaya, A., Jain, A., Zhang, H.J.: On image classification: city vs. landscape. In: IEEE Workshop on Content-Based Access of Image and Video Libraries, pp. 3–8 (June 1998)
9. Dal Moro, F., Abate, A., Lanckriet, G.R.G., Arandjelovic, G., Gasparella, P., Bassi, P., Mancinim, M., Pagano, F.: A novel approach for accurate prediction of spontaneous passage of ureteral stones: Support vector machines. Kidney International 69, 157–160 (2006)
10. Byvatov, E., Fechner, U., Sadowski, J., Schneider, G.: Comparison of support vector machine and artificial neural network systems for drug/nondrug classification. Journal of Chemical Information and Computer Sciences 43(6), 1882–1889 (2003)