A Hierarchical Fuzzy Integration of Local and Global Feature-based Classifiers to Recognize Objects in Autonomous Vehicles

Luciano Oliveira, Paulo Peixoto and Urbano Nunes

Abstract— **Sensing is a paramount task for autonomous vehicles. In this field, machine vision is usually applied owing to its high dense information per sensing area. This characteristic contributes for the development of classification algorithms that extract information of objects in images. Although individual classifiers have been increasingly enhanced lately, effective integration of classification characteristics from different classifiers is still in a recent stage of investigation. To contribute for early step of classification fusion, we propose a method to combine the decisions of a local and a global feature-based classifiers. The main goal is to use a Label Set Reduction strategy based on a Hierarchical Fuzzy Integration system to make a decision about the results of the classification labels. The proposed architecture has been applied to recognize pedestrians and cars in outdoor environments, but it is easily expansible to include other types of classifiers and object classes. Promising results, in the final classifier, have been achieved.**

I. INTRODUCTION

Machine vision applied to vehicles is an underlying field of research, responsible to provide the expected autonomy by means of sensing the surroundings more densely and identifying concerning objects from image. In this realm, several methods have been proposed to detect objects in image with a satisfactory performance [1]-[4].

Throughout the years, many works have been carried out in order to extract relevant information using individual classifiers. Since this approach seems to have constraints because the particular nature of applications, several ways have been suggested toward to combine the results of different classifiers in order to increase the overall performance of the systems [5]-[7].

In [8], a survey about general methods to combine classifiers is introduced. The authors suggest two main kinds of combination methods: those that are performed on the structures of the classifiers and those that are performed on the results of the classifiers. The first ones try to make featurewise classification while the second ones analyze the final decision made by the component classifiers. Here, we propose a Label Set Reduction (LSR) method, based on the second class of fusion, which tries to reduce the set of considered labels, achieved by the two classifiers, ensuring that the correct label is still in the reduced set.

Concerning to the type of feature, classifiers fall into two main categories: those based on global feature extraction [1], [5] and those based on local feature extraction [9], [10], [11]. The first group assumes that objects are intrinsically related to the background and tries to generalize this situation, keeping a context and learning the relationship between them. Latter one assumes that objects have their particular characteristics aside from the background. Each of them has their specificities, strengths and shortcomings, and, in this work, it is intended to cope with the problem of combining the superiorities of each method to build a more robust final classification applying it in the early step of fusion.

In [12], Monteiro *et al* have introduced a complete framework to detect pedestrians and vehicles. In order to enhance robustness of our vision-based system, here we propose a novel architecture to integrate the decisions of a global and a local feature-based classifiers: a Haar-like feature / Adaboost [1] and a Histogram of Oriented Gradient (HOG) / linear SVM [2]. To effectively combine the output of these two classifiers, a Hierarchical Fuzzy Integration (HFI) system is proposed. The goal of this system is to decide whether the output labels should be maintained or not according to the neighborhood intersection and the posterior probability of the two class labels (each class label is represented by a bounding box).

The underlying goal of our work is to obtain a trade-off between false alarm decrease and true positive detection rate increase, using a LSR strategy (combination of the outputs of classifiers) to detect pedestrians and vehicles in cybercar scenario [13]. The proposed architecture is easily expansible to include other types of classifiers and object classes. Detection Error Tradeoff (DET) curves were built and promising results have been achieved, showing that our system has outperformed component classifiers.

Contents. In §II, some related works are briefly resumed. §III presents some points of comparison between global and local feature-based classifiers. In §IV, we explain our method and implementation details. §V presents our validation method and final results, using DET curves. Finally, §VI draws some conclusions and future works.

II. RELATED WORKS

A combination of multiple classifiers using fuzzy templates in order to aggregate the output labels according to their posterior probability is proposed in [5]. It performs with only global feature-based classifiers.

More recent works are presented in [6] and [7]. The first one introduces two methods to combine local and global features: a "stacking" ensemble technique, which performs in the output labels of the classifiers, combining their output vectors and classifying these vectors in a meta-classifier; and a Hierarchical Classification System (HCS), which first uses

Luciano Oliveira, Paulo Peixoto and Urbano Nunes are with Instituto de Sistemas e Robótica, Dep. de Engenharia Electrotecnica e de Computadores, Universidade de Coimbra, Portugal {lreboucas,peixoto,urbano}@isr.uc.pt.

global feature-based classifiers and, then, tries to validate this output with a local feature classification. In [7], a similar approach to HCS is discussed. The very big difference between them is the fact that the latter one computes the posterior probability to infer whether to apply local feature computation or not.

These three works discussed in this section have some similarities with our work, differing in two points:

- the work that explores an adaptive way of integrating classifiers just use global feature classification and, thus, fail in treating occlusion and taking advantage of other strengths encountered in local feature-based classifiers;
- those that use the combination of global and local features, tend to perform it in a sequential way, first applying a global feature classification, and hence discard some beneficial characteristics of local feature-based classifier in the first stage of classification.

III. GLOBAL VERSUS LOCAL FEATURE CHARACTERIZATION

A. Global feature characteristics

The global features are commonly used to generalize the idea of an object, i. e., one is not interested of finding "the" object, but "an" object that may belong to a category. In this characterization, a slightly change in illumination, cluttered background or occlusion of the object may cause a decrease in the performance of the classification. However, global feature-based classifiers usually may capture a context of the image, making the object localization task to be easier.

Regardless of the aforementioned characteristics of global features, it is important to calculate the posterior probability in order to integrate the outputs of the classifiers in the fuzzy stage so that the empirical evidence of the classification may be found. For Haar-like feature / Adaboost, the posterior probability for every feature to make part of an object Y is estimated by:

$$p(Y = 1|x_i) = \frac{e^{2F(x_i)}}{1+e^{2F(x_i)}} \tag{1}$$

where the final classification function $F(x_i)$ is given by:

$$F(x_i) = \sum_{k=1}^{N} \alpha_k f_k(x_i) \tag{2}$$

where x_i is the ith element of the input vector, α_k is a weight function of each weak classifier $f_k(x_i)$ and N is the number of weak classifiers.

B. Local feature characteristics

Whilst the global features try to capture the context of the object with respect to background, local features perform right on the spot, being it an object or background. This approach tries to achieve patches in the image that distinguish uniquely the object and usually have some invariance to affine transformations and illumination and, hence, it is robust for occlusion and cluttered environment.

Though these features capture a specific (local) characteristic of an object, there are ways to overcome this situation

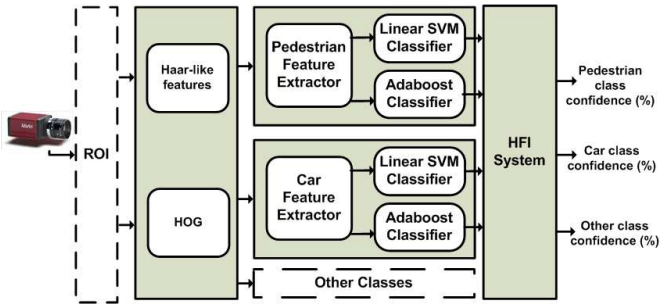


Fig. 1. Overall architecture of the proposed classification system. The Region of Interest (ROI) comes from a laserscanner, as in [12].

for a categorization task, building a visual dictionary in order to recognize objects [15], [16].

As mentioned in §III-A, here one should also estimate the posterior probability. When testing, every local feature x_m of the testing image X is associated with its nearest neighbor x_n of the training set since the Euclidean distance $argmin\{d(x_m, x_n)\}$ is less than an established threshold D. The posterior probability of x_m to make part of a class k is given by:

$$p(k|x_m) = \begin{cases} 1, & \text{if } argmin\{d(x_m, x_n)\} \leq D \\ 0, & \text{otherwise} \end{cases} \tag{3}$$

The posterior probabilities of all local features x_m of X are combined using the sum rule:

$$p(k|X) = \frac{1}{M} \sum_{m=1}^{M} p(k|x_m) \tag{4}$$

Resuming, the posterior probability of a local feature classifier is obtained by finding the minimum Euclidean distance between the local features in the image region being classified (x_m) and those contained in the visual dictionary (x_n).

IV. HIERARCHICAL FUZZY INTEGRATION

The architecture of the proposed system is depicted in Fig. 1. After finding the percentage of the confidence with respect to the neighborhood intersection of the labels and their posterior probability, for each detected object, a threshold is applied to evaluate the final result according to a predefined observed behavior of the classifiers. The architecture is expansible to include other types of classifiers and object classes.

The intuition behind this approach may be resumed by three main ideas:

- find an intersection between the class labels;
- evaluate the intersection encountered according to the posterior probability of each classifier;
- compute a final confidence based on latter ones.

Fuzzy logic has been chosen to evaluate non-linearly the intersection between the class labels, based on human skills. Moreover, it is important to realize that if just the intersection

of the class label neighborhood was evaluated, true positive classes would be quickly discarded or it would be placed an emphasis in one classifier or another. To overcome this situation, the posterior probability of each classifier has been included in the fuzzy decision.

The algorithm of the final classifier may be resumed bellow:

Algoritmo 1 The HFI system for each object

for each frame **do**
- Find labels and posterior probabilities using Haar-like features / Adaboost
- Find labels and posterior probabilities using HOG / SVM
- Calculate fuzzy output for class labels intersections
- Calculate fuzzy output for posterior probability
- Give the final fuzzy confidence

end for

In each frame, obtained from a camera on front of the car, and for each object class, a classification task, using haar-like feature / adaboost and HOG / SVM, is accomplished, using a sliding window strategy. The object labels and posterior probability are, then, achieved. Finally, each fuzzy variable is sent to the respective fuzzy system in order to find the final confidence based on human skills.

A. Haar-like feature / Adaboost

An Adaboost has been utilized to classify a set of haar-like features. The features used in our Adaboost classifier are shown in Fig. 2. The haar-like features are represented by templates. Each feature is achieved by a weighted sum of two components: gray level sum over the black rectangle and the whole rectangle area.

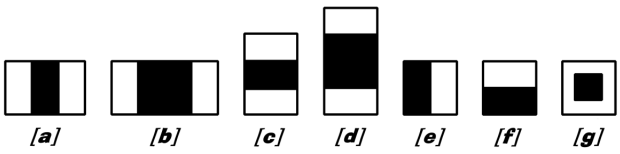


Fig. 2. Set of haar-like features used. a, b and c are the line features, e and f are the edge features and g is the center-surround feature.

For each feature, a weak classifier determines an "optimal" threshold. Each weak classifier $h_j(x)$, where x is a mxn pixel sub-window of an image, consists of a feature f_j, a threshold Θ_j and a parity p_j which indicates the direction of the inequality sign, according to:

$$h_j(x) = \begin{cases} 1 & \text{if } p_j f_j < p_j \Theta_j \\ 0 & \text{otherwise} \end{cases} \quad (5)$$

Each of $h_j(x)$ classifier reacts to a haar-like feature. The final classifier is composed by all weak classifiers as defined by (2). An object is only considered if it goes by all the stages of weak classification.

B. HOG / Linear SVM

Histogram of Oriented Gradient descriptor reached a maturity version for object matching in [14] and it was further utilized slightly modified by [2], along with a linear SVM, in a pedestrian detection system.

The main idea of the system is to build a visual dictionary, composed by local features that may describe object categories in an invariant way with respect to illumination, scale and translation, capturing the shape context of objects. The descriptors are evaluated densely by a linear SVM classifier.

C. Proposed fuzzy system

The fuzzy stages of the proposed hierarchical system owns a fuzzifier, an inference and a defuzzifier module. The inference module searches on a set of rules, examining which rules shall be triggered to achieve the final confidence, whose crisp value is found by a defuzzification method.

A Mamdani fuzzy system has been utilized [17], whose T-norms and S-norms are, respectively, *Min* and Max functions. After an evaluation, a mean of maxima defuzzification method was chosen and the crisp output u^* is obtained by:

$$u^* = \sum_{i=1}^{N} \frac{u_i}{N} \quad (6)$$

where u_i is the ith element in the universe of the discourse where the membership function of the output $\mu_{out}(u)$ is at the maximum value, and N is the total number of such elements. In each stage, (6) is applied in order to provide a crisp value for the next stage.

The choice of a Mamdani structure was made by the fact that it is more nearby human skills and, consequently, easier to be implemented and evaluated according to observation.

The goal of the proposed fuzzy system is not only to evaluate the intersection of the class label neighborhoods, represented by the bounding boxes of each detected object, but also to weigh the use of each classifier in order to discard or to maintain the class label, using one fuzzy system for intersection, one for posterior probabilities and a final fuzzy stage to classify latter ones.

Following, more details are given about the proposed fuzzy system.

1) Fuzzy variables: In Fig. 3, fuzzy variables are shown as the way they are extracted from the image or component classifiers. In the first tier, four variables have been used in parallel: perimeter ratio of the class labels (minor divided by major), (Euclidean) distance ratio between the nearest class labels divided by the major width of the bounding boxes of the classes, Adaboost posterior probability and SVM posterior probability.

2) Fuzzy architecture: The HFI system is composed by three fuzzy systems that are resumed on Table I. The universes of discourse of all variables are in the interval [0, 100], representing rate or confidence values.

All fuzzy modules are composed by nine rules with knowledge surfaces depicted in Fig. 4. These surfaces show how each rule is triggered in the system: in Fig. 4(a), rules

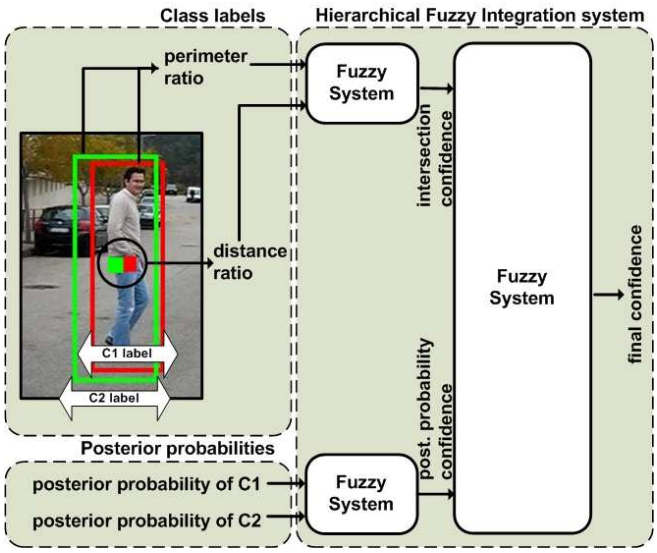


Fig. 3. Variables used in the fuzzy system. C1 is the Haar-like feature / Adaboost classifier and C2 is the HOG / Linear SVM classifier

TABLE I

CHARACTERISTICS OF THE HIERARCHICAL FUZZY SYSTEM

Fuzzy system	Variables	Fuzzy sets
Class label	Distance ratio	{Near, Medium, Far}
	Perimeter ratio	{Little, Medium, Big}
	Intersection confidence	{Low, Medium, High}
Posterior prob.	Adaboost post. prob.	{Low, Medium, High}
	HOG / SVM post. prob.	{Low, Medium, High}
	Post. prob. confidence	{Low, Medium, High}
Final	Intersection confidence	{Low, Medium, High}
	Post. prob. confidence	{Low, Medium, High}
	Final confidence	{Low, Medium, High}

in the middle are not triggered since it is not important either discarding or maintaining the class labels; in Fig. 4(b), all the rules may be triggered because it is necessary to evaluate all the range of posterior probabilities, and in the final stage, Fig. 4(c), only the last rules are considered due to the threshold used to evaluate the final confidence of the system.

V. EXPERIMENTAL RESULTS

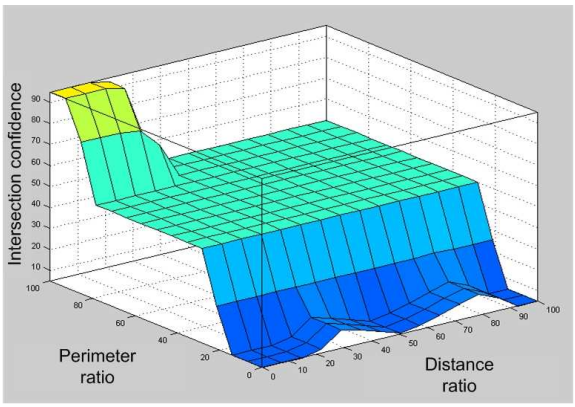
A. Methodology

Preliminary experiments were conducted in order to test the system, and two training data sets have been used: for pedestrians, the INRIA dataset, proposed by [2], and, for cars, the CALTECH dataset [19]. Some samples are shown in Fig. 5.

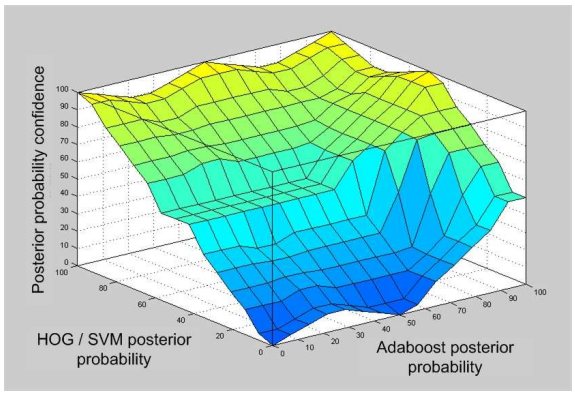
The INRIA dataset is composed by 2416 pedestrian images and 12180 background images, for the training phase; 997 pedestrian images in a variety of realistic scenes, for the test phase. The CALTECH dataset is composed by 3698 car images and 13690 background images, for the training phase and 337 car images, in the test phase.

B. Validation

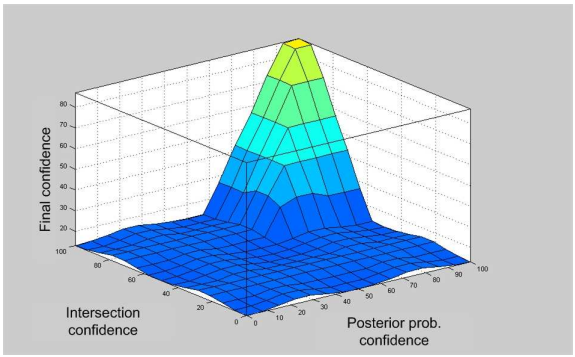
After gathering the datasets involved to validate the system, we have run both component classifiers against the



(a) Knowledge surface of the fuzzy system that evaluates the class label intersection neighborhood.



(b) Knowledge surface of the fuzzy system that evaluates the posterior probability of the classifiers.



(c) Knowledge surface of the final stage.

Fig. 4. Knowledge surfaces of each fuzzy stage.

test set to evaluate their performance (some results are depicted, in Fig. 7). This was made by plotting DET curves of each classifier and final fusion, according to Fig. 6, where the lower is the curve, the better the performance of the classifier is. Although the proposed fusion clearly outperform its component classifiers, some points must be clarified about the criteria of validation methodology:

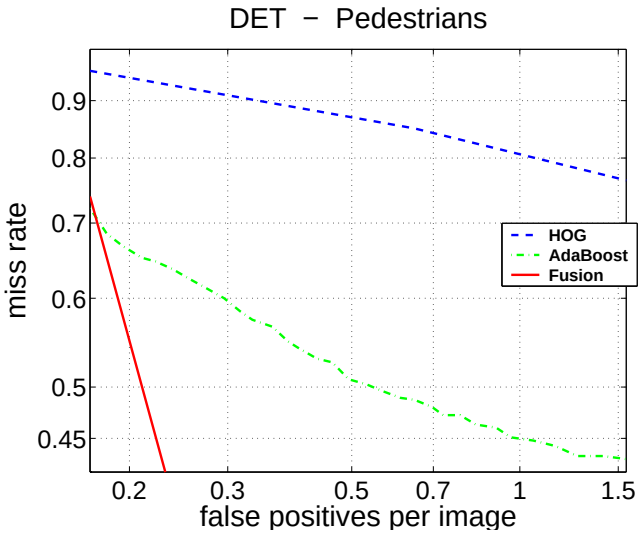
- **Ground truth** - we use the annotations provided by the two datasets. In order to assign a true detection, the PASCAL challenge criterion has been used [18], according to:

Fig. 5. Samples from INRIA and CALTECH datasets used for the training and testing. The upper row shows training examples, the other rows show testing examples.

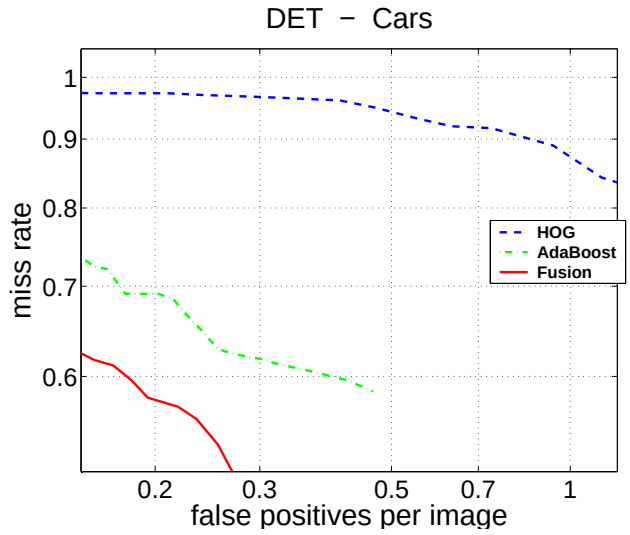
$$OA = \frac{A_{gt} \cap A_c}{A_{gt} \cup A_c} \geq 0.5 \qquad (7)$$

where OA stands for Overlapped Area, A_{gt} is the area of the bounding box assigned for the annotation (ground truth) and the A_c is the area of the bounding box assigned for the classifier. The OA must be greater or equal to 0.5 (50% of overlapped area), for a given class c labels correctly an object in the image.

- **HOG/SVM Curve** - The curves encountered here differ from [2]. The reason for this is the criterion that we used for counting the false positive rate. It is defined by the ratio between all false positives achieved (in all images) and the total number of images (not per windows as defined by [2]). Other point that should be highlighted is that several labels, which was visually realized as a true detection, were discarded because of a small deviation bellow the threshold adopted (50%), when (7) has been applied.



(a) DET curve of pedestrian classification.



(b) DET curve of car classification.

Fig. 6. Evaluation of the component classifiers and fusion.

VI. CONCLUSIONS AND FUTURE WORK

A. Conclusions

A Hierarchical Fuzzy Integration system to combine the outputs of local and global feature-based classifiers has been presented in this paper. The proposed system tries to combine the strengths of two well-established classifiers to incorporate an invariance concerning to illumination, translation and scale, and a global context in autonomous vehicle tasks of categorization.

The idea of including the posterior probability of the classifiers overcame the problem of discarding labels that do not intersect each other in certain circumstances. In this way, the proposed fusion system has shown a better performance than its components classifiers, although the final tradeoff between false positive and miss detection rate has not reached a desirable value, what lead us to better tune the classifiers and to investigate other ways of classification.

Hence, we have been implementing our own local feature classifier with the goal of making a featurewise integration, handling with the structural nature of each classifier.

B. Future Work

One point that was not included in the performance evaluation was the CPU time, which shall be incorporated

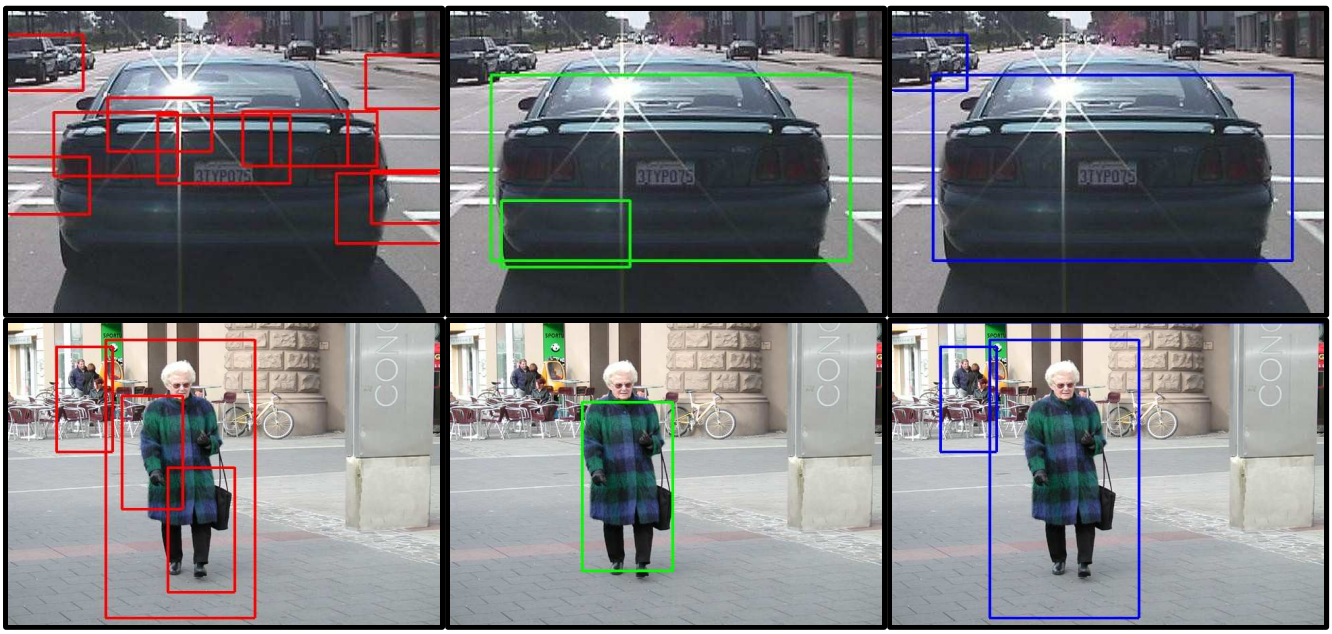


Fig. 7. Final results: left column shows the result of the HOG / SVM classifier; central column shows the result of the Adaboost classifier and the right column shows the fusion result. An example of occlusion classification may be observed in the first image of the second row, which was classified by HOG / SVM.

in a further evaluation when the system is embedded in a real-time operating system.

We are now researching other ways of combining global and local feature-based classifiers, including an extended evaluation of other classifiers and ways of combination. Since we use a laserscanner in our complete framework [12], the idea of using more classifiers may not affect drastically the performing time of the overall system, as the assigned classification area of the image comes from the laserscanner ROI.

VII. ACKNOWLEDGMENTS

This work is supported in part by Fundação para Ciência e Tecnologia (FCT) de Portugal, under Grant POSC/EEA-SRI/58279/2004, and by CyberC3 project (European Asia IT&C Programme). Luciano Oliveira is supported by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), doctorate programme of Ministry of Education of Brazil, scholarship n^o BEX 4000-5-6.

REFERENCES

[1] Viola, P. and Jones, M.: Rapid Object Detection Using a Boosted Cascade of Simple Features. IEEE International Conference on Computer Vision and Pattern Recognition, (2001), 511–518.

[2] Dalal, N. and Triggs, B.: Histograms of Oriented Gradients for Human Detection. IEEE International Conference on Computer Vision and Pattern Recognition, vol. 2, (2005), 886–893.

[3] Schneiderman, H. and Kanade, T.: Object Detection Using the Statistics of Parts. International Journal of Computer Vision, vol. 56, no. 3, (2004), 151–177.

[4] Sotelo, M., Parra; I. and Naranjo, E.: Pedestrian Detection Using SVM and Multi-feature Combination. 9th IEEE Intelligent Transportation Systems Conference, (2006), 103–108.

[5] Kuncheva, L.I.; Bezdek, J.C.; Sutton, M.A.: On Combining Multiple Classifiers by Fuzzy Templates. NAFIPS Conference of EDS., (1998), 193–197.

[6] Lisin, D.A.; Mattar, M.A.; Blaschko, M.B.; Learned-Miller, E.G. and Benfield, M.C.: Combining Local and Global Image Features for Object Class Recognition. IEEE International Conference on Computer Vision and Pattern Recognition, (2005), 47–55.

[7] Murphy, K.; Torralba, A.; Eaton, D. and Freeman, W. T.: Object Detection and Localization Using Local and Global Features. Sicily Workshop on Object Recognition, Lecture Notes in Computer Science, (2005), 393–413.

[8] Ruta, D. and Gabrys, B.: An Overview of Classifier Fusion Methods. Computing and Information Systems, vol 7, no. 1, (2000), 1–10.

[9] Dorko, G. and Schmid, C.: Selection of Scale-invariant Parts for Object Class Recognition. IEEE International Conference on Computer Vision, (2003), 634–639.

[10] Fergus, R.; Perona, P.; Zisserman, A.: Object class recognition by unsupervised scale-invariant learning. IEEE International Conference on Computer Vision, (2003), 264–271.

[11] Nowak, E. and Jurie, F.; Triggs, B.: Sampling Strategies for Bag-of-Features Image Classification. European Conference on Computer Vision, (2006), 490–503.

[12] Monteiro, G.; Premebida, C.; Peixoto, P. and Nunes, U.: Tracking and Classification of Dynamic Obstacles Using Laser Range Finder and Vision. Workshop on "Safe Navigation in Open and Dynamic Environments - Autonomous Systems versus Driving Assistance Systems", held at the International Conference on Intelligent Robots and Systems (IROS 2006), (2006), 213–219.

[13] Cybercars: Cybernetic Technologies for the Car in the City [online], http://www.cybercars.org.

[14] Lowe, D.G.: Object Recognition from Local Scale-invariant Features. IEEE International Conference on Computer Vision, (1999), 1150–1157.

[15] Mutch, J. and Lowe, D.G.: Multiclass Object Recognition with Sparse, Localized Features. IEEE International Conference on Computer Vision and Pattern Recognition, (2006), 11–18.

[16] Winn, J., Criminisi, A., Minka, T.: Object Categorization by Learned Universal Visual Dictionary. IEEE International Conference on Computer Vision and Pattern Recognition, (2005), 1800 – 1807.

[17] Zadeh, L.: Fuzzy Sets. Information Sciences, (1965), 338–353.

[18] Pascal: Pattern Analysis, Statistical Modelling and Computational Learning. http://www.pascal-network.org/challenges/VOC/databases.html.

[19] L. Fei-Fei, R. Fergus and P. Perona: One-Shot Learning of Object Categories. IEEE Trans. Pattern Recognition and Machine Intelligence. In press, (2004).