IMAGE CLASSIFICATION WITH USER DEFINED ONTOLOGY

Remi Vieux, Jean-Philippe Domenger, Jenny Benois-Pineau and Achille Braquelaire

University of Bordeaux, LaBRI UMR CNRS 5800
351 cours de la Libration, F-33405, Talence Cedex, France
phone: + (33)540006900, fax: + (33)540006669, email: {vieux,domenger,benois-p,achille}@labri.fr
web: www.labri.fr

ABSTRACT

In this paper we are interested in classification of objects in images according to user defined scenarios. We show how the user-defined ontology with a specialisation by a concrete scenario / object of interest allows for an adapted choice of methods and their tuning through the whole framework: selection of the area of interest, descriptors choice, classification of objects.

Particular attention here is payed to the classification. We use SVM classifiers for their good capacity of generalisation. We show that in an adapted descriptor space, the choice of a "light" linear kernel together with boosting of classifiers is interesting compared to more complex and computationally expensive RBF kernels.

The results on real-life images are promising. The paper results from the research we conduct in the framework of X-Media EU-funded Integrated Project.

1. INTRODUCTION

In the field of image analysis and processing, the semantic gap between the low level features computed at the signal level, and the human interpretation and understanding of an image has always been a major issue. Several works have been led to try to narrow this gap. In [7], object extraction from static images is realised using initial low level segmentation and grouping of regions accordingly to the domain ontology. Thus, compound objects such as computers can be extracted by grouping keyboard, screen, etc. segmented in the image as homogeneous regions. In [9], image annotation is realised by mapping of regions to clusters with textural levels. In [1], a high level semantic segmentation of input images is obtained due to the grouping of regions belonging to the same concept which is trained as a cluster in a adapted feature space. The objective of the present work is to move from low level image segmentation approach to extraction of meaningful objects guided by user defined ontology. On the contrary of previously cited works, our approach can be qualified as *top down*, in the sense that a generic ontology will guide the whole segmentation and classification process. The actual research is fulfilled in the framework of the European Project X-Media. The project goal is to enable the extraction and sharing of knowledge across text, images and raw data. Every medium can extract knowledge, and the fusion of each information source can enable the inference of more knowledge, and also drive the knowledge extraction for each medium. As part of the image team for the X-Media Consortium, our aim is to develop a methodology and appropriate analysis and reasoning tools to ensure the semantically driven image analysis which will bridge the gap between low level knowledge (such as shape, colour information, etc.) and high level semantic concepts. In this article, we will first introduce the motivation and needs from our industrial partner. We will then present a methodology for knowledge driven image analysis, and introduce an example with the classification of car part images. The fourth section presents the different approaches to the classification problem that we used, while the last part details the results.

2. USER SCENARIO

When designing a car, whether constructing a completely new model or a new version of an existing one, particular attention has to be payed to the conception of the car interior. As part of the *Ergonomics Department* in the FIAT company, the *Human Engineering* service is a service which is dedicated to this kind of task. Human engineering main goal is the achieving of factors like usability, comfort, ability to guaranty the user wellness while driving or sitting in a car, as well as to ensure that the user intuitively understand and correctly uses the car and its components. Human engineering is ruled by a number of standard fixing dimensions, visibility, accessibility. Above all, it must give to the user a comfortable sensation that often cannot be expressed by a number but only by a judgement. Therefore, it is important, when developing a new car, to be aware of what is happening inside the company as well as to have a look at what other competitors are proposing, e.g. new shapes, new colours, news solutions, etc. At the moment, the process of looking for solutions is very time consuming and is an entirely manual task. It involves looking at huge databases of car pictures, both internal and external, and collecting pictures from the Internet. The main difficulty of this task is that the databases are, in the best case, structured by car maker and car model name. The user has to look at all the images to pick up the ones that can be relevant for his specific task. Hence, there is room for a software solution that could release the human burden in the task of collecting relevant images. From the image analysis point of view, if we are able to recognise specific car components in an image, the classification of these images in the database could be much more pertinent. In fact, the granularity of the classification depends on the user ontology. In this scenario, the classification could be operated from the very high level (e.g. differentiate pictures of car interiors/exteriors) to the more specific (find pictures of steering wheel, or give all pictures representing cars with square air ducts, etc.).

The aim of the X-Media project is to enable knowledge sharing and re-usability across heterogeneous media such as text, images and raw data. To ensure the interoperability between the different medium, a common perspective has to be defined: this is achieved by the use of an ontology. The ontology enables the user to operate at the knowledge level, e.g. performing tasks such as semantic search over text, images or data. As part of the image processing and knowledge acquisition team for the X-Media project, we have to design solutions enabling the system to populate a user defined ontology with image documents.

3. SCHEMA OF THE METHOD

To populate the ontology with instances of documents, we are proposing a method of knowledge driven image analysis as depicted in figure 1.

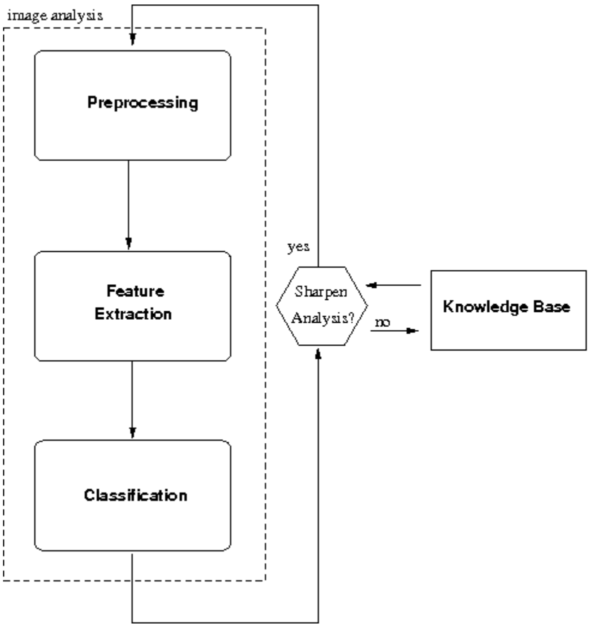

Figure 1: Methodology for knowledge driven image analysis

As a standard approach, image analysis is decomposed in three steps. First, the image is preprocessed (e.g. segmented). Then a set of features are computed on the global image or regions. Finally, classification is achieved based on these features. The specificity of the proposed method is that the image analysis algorithms to be applied are dedicated algorithms, linked to the concepts defined in the user ontology. The whole analysis can be seen as a feedback loop, where the process starts with a predefined knowledge and stops when the desired granularity of the analysis is reached. For example, if one is interested in retrieving pictures of steering wheels but have no further knowledge of input pictures (except that they are pictures of cars), the first step of the analysis would be to differentiate images of car interiors and exteriors. This could be achieved by a classification based on global features such as MPEG-7 Colour Layout Descriptor. Then the analysis of picture of car interiors could then be refined using an algorithm dedicated to steering wheel detection. In the context of cross media analysis, we could have received the first piece of information from another source, for example the caption of the image. As an example of a concrete instantiation of this methodology, we will now describe the task of air duct detection.

As we can see on figure 6, air ducts in a car can take a lot of different shapes and sizes, but are usually composed of a grid. To analyse the images of car interiors, we proposed a method of region of interest extraction based on a combination of Hough transform for line detection and clustering, the use of the MPEG-7 Edge Histogram Descriptor as a relevant feature and classification with Support Vector Machines. Since classification is the core procedure of the methodology, it will be described in more details in the next section.

Figure 2 shows the output of line detection with Hough transform, after a Canny edge detection. In this image, the air duct region is composed of a high density of horizontal lines. To extract a region of interest, we filter the result of line detection keeping only the lines with the most encountered direction in the global image. Then, we perform a clusterisation of the remaining lines using k-means algorithm on the Euclidean distance between the centre of each line. Finally, the bounding box of each cluster in the image will define a region of interest. An example of ROI extraction is given in figure 3, where each coloured point represents the centre of a line given by Hough.

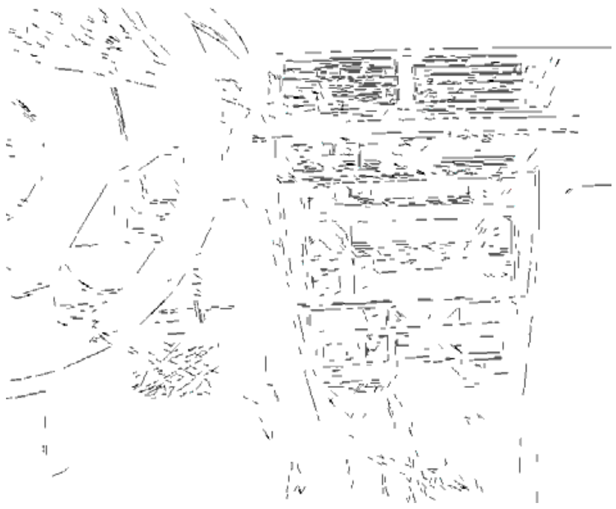

Figure 2: Segment line detection using Hough transform.

The second step of the methodology is to extract features in the regions that are characteristic enough for the classifier to perform well. As we already said, air ducts are composed of a grid, hence have a characteristic spatial edge distribution.

The MPEG-7 standard [12] defines three descriptors for texture: Homogeneous Texture Descriptor (HTD), Texture Browsing Descriptor (TBD) and Edge Histogram Descriptor (EHD)[15]. The Edge Histogram Descriptor is used to represent the local spatial distribution of edges according to five different orientations (vertical, horizontal, 45 degree edges, 135 degree edges and non directional edges) in 16 sub-images (see figure 4), hence composing a $16*5=80$ bins histogram. This descriptor has been used efficiently for various image similarity matching problems [17]. Moreover, this descriptor has the advantage of being scale independent. Figure 5 presents the mean edge histogram computed over a set of manually selected air ducts (top), versus the mean

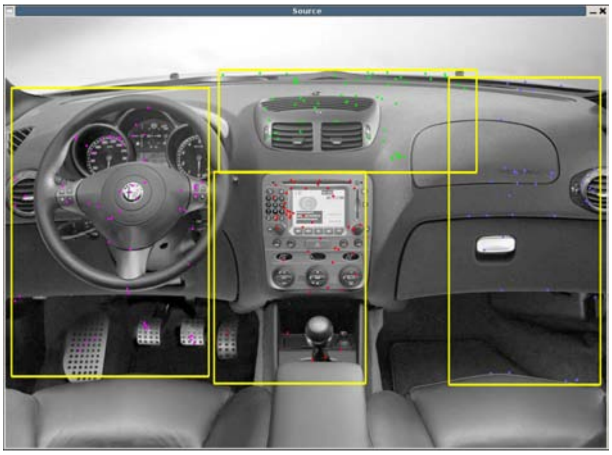

Figure 3: ROI detection with Hough transform and k-means clustering

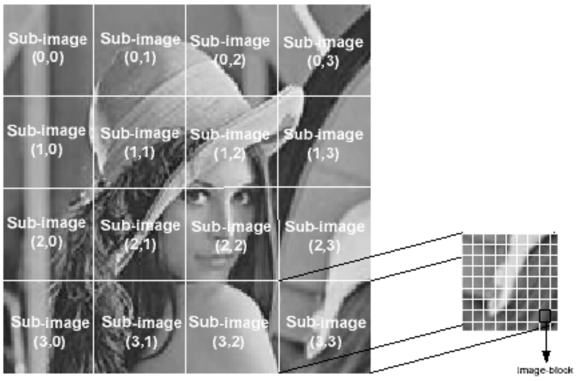

Figure 4: Definition of sub images and image blocks (from [17])

edge histogram of randomly chosen regions in the same picture not containing air ducts. It is clear from this diagram that the two categories have different edge distribution.

In the next section we will study the use of different approaches for air duct classification based on Edge Histogram.

4. CLASSIFICATION

In this research we work in a supervised learning paradigm. Furthermore, we aim to build an optimal classifier by a boosting approach.

4.1 Boosting

In machine learning theory, the *ensemble learning* paradigm consists in combining several classifiers to build a highly accurate classifier. Among the ensemble techniques, boosting as been shown to be an efficient method [2]. Freund and Schapire have defined a boosting algorithm called AdaBoost [8] for *Adaptive Boosting*. The aim of AdaBoost is to train iteratively an ensemble of component classifiers (also called weak classifiers) by re-sampling the training samples over a distribution. First, each sample is assigned the same weight. A *weak* classifier is trained on these data. Then weight of the training samples which are correctly classified by this classifier is decreased, while the weight of the training samples which are incorrectly classified is increased, and the process is iterated. Hence, the algorithm focuses on the most difficult cases by assigning an important weight to those examples which are hard to classify. In AdaBoost, the weak classifier can be any learning algorithm more precise than average. The output of classification is given by a weighted sum of each component classifier answers.

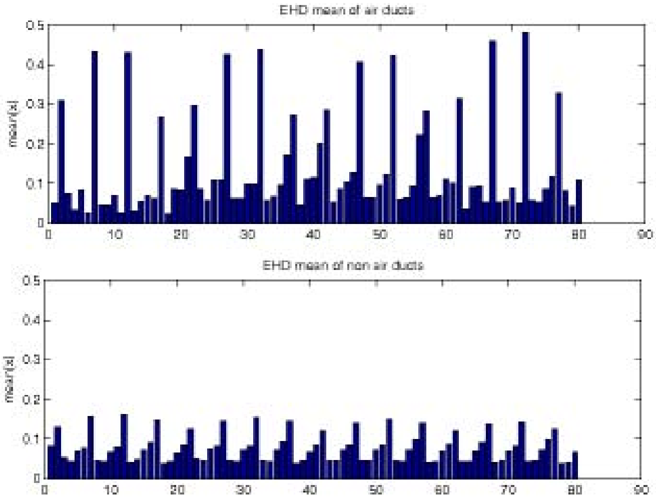

Figure 5: Mean of the EHD of negative and positive examples

Since the early work of Cortes and Vapnik on handwritten character recognition [5], Support Vector Machines have been widely used in several machine learning problems. We use SVM classifiers as component classifier trained in the AdaBoost fashion.

4.2 Kernel selection

When using SVM classifiers, kernel selection is an important process which can lead to significant difference in the classification performances. Though many research works are being led on kernel elaboration such as [13], there are four basic SVM kernels in use: linear, polynomial, RBF and sigmoid (see [4]).

Hsu *et al.* [11] state that using the RBF kernel is a reasonable first choice when you have no particular knowledge of the problem. The RBF kernel non linearly maps the samples into a higher dimensional space so that it can, unlike the linear kernel, handle the case when the training samples are not linearly separable. It has been shown that the sigmoid kernel behaves like RBF for certain parameters [14]. Figure 5 shows the mean of the EHD computed on our training set. It can be seen in the top graph that there are regular picks every $(5i+1)^{th}, i = 0,1,\ldots,15$ bin. This actually corresponds to the number of horizontal edges in the sub-images. The regions not containing air ducts are usually smoother, and there are no such picks in the histogram of the negative examples. We can make the assumption that the classes will be easily separated by an SVM with linear kernel. Whatismore, it is better for computational cost to use a lighter kernel when training a classifier in the AdaBoost fashion. In [3], we pro-

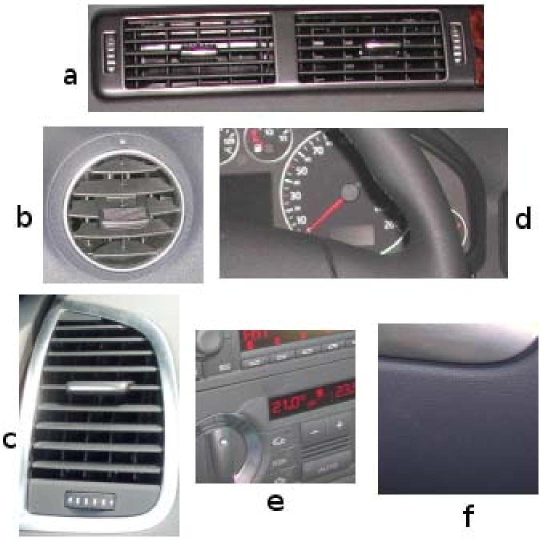

Figure 6: Some positive (a, b, c) and negative (d, e, f) examples

posed an optimisation of computational cost in the case of polynomial kernel, while this reduction for linear kernel was proposed in [6]. We show in the results that the choice of the simplest linear kernel is justified.

5. RESULTS AND COMMENTS

5.1 Building the training and testing base

We have led our experimentation on images of car interiors given by FIAT. The images are photos taken with a digital camera by a human operator, usually representing the dashboard, or close shots of air ducts. Hence, we have a collection of pictures with no restrictions about size or resolution, and no particular point of view from where the picture is taken.

We have built two different training and testing bases. For the first base, we have manually annotated regions in the image containing air ducts as positive examples, and randomly chosen in the same image regions not containing air ducts as negative examples. An example is given in figure 6. We have selected a ratio of about three negative examples for one positive. The second base was built by annotating the regions given by the ROI extraction algorithm depicted in the previous section. In each image, k-means was set up to find four different clusters, hence giving four ROI per image. On each region, we computed the EHD and store the results as input for the SVM training and testing. Table 1 sums up the composition of the training and testing bases.

We experimented AdaBoost algorithm with linear SVM, as well as single RBFSVM classification. Boosting was done over the whole set of examples, with a uniform weight initialisation, and weight update after each iteration. We used Chang and Lin's implementation of Support Vector Machine [4].

5.2 Results

Th results obtained for each base are summed up in tables 3 and 2. For RBF classification, the optimal parameters for the Gaussian width γ and the penalty cost C were determined by performing a grid search cross validation as in [11].

	# pos. examples	# neg. examples
manual training base	200	592
manual testing base	137	449
automatic training base	349	435
automatic testing base	267	323

Table 1: Composition of training and testing bases

On the automatically extracted ROI, best classification is obtained by the RBFSVM with 0.53 recall and 0.76 precision. Single linear SVM is performing quite poorly with 0.41 recall and 0.70 precision. The boosted linear SVM, while increasing the performance compared to single linear SVM, still not compete with the RBF kernel. On the manually extracted ROI, classification performance of the RBFSVM classifier is 0.796 for recall, and 0.964 for precision. Without boosting, the classification results obtained by the linear SVM classifier (LSVM) are almost similar to the one obtained with RBFSVM, with 0.788 for recall, and 0.96 for precision (that is one true positive example less than RBFSVM). We performed boosting with linear kernel over 10, 30 and 50 iterations.

Classifier	Recall	Precision
RBFSVM	0.53	0.76
LSVM	0.41	0.70
Boosted LSVM (10 iterations)	0.46	0.66
Boosted LSVM (20 iterations)	0.47	0.62
Boosted LSVM (30 iterations)	0.48	0.63

Table 2: Classification on the automatically extracted ROIs

5.3 Comments

The weak classification results obtained for the automatically extracted ROI are due to the fact that the regions extracted are often large regions, containing an air duct but also large part of the dashboard or other car components. Hence, the air duct plays a little contribution to the overall EHD of the region,

Classifier	Recall	Precision
RBFSVM	0.796	0.964
LSVM	0.788	0.964
Boosted LSVM (10 iterations)	0.832	0.934
Boosted LSVM (30 iterations)	0.818	0.896
Boosted LSVM (50 iterations)	0.81	0.902

Table 3: Classification results on the manually extracted ROIs

and its particular edge distribution is hidden. Since most of the regions obtained did contain air ducts but were still too large for accurate classification, several solutions have to be investigated. A simple one would be to extract more regions in the image with k-means. This could work for image representing the whole dashboard but could lead to errors for closer shots of car components. Another solution could be to segment the obtained ROI in order to separate the different components inside.

The result obtained on the manually extracted ROI are more promising. First we note that as we expected, the classes of air duct/non air duct regions are almost linearly separable, since the linear kernel without boosting achieves a classification with almost the same performances as the SVM using the RBF kernel.

Table 3 shows that after 10 iterations of boosting, the results are significantly improved, in terms of recall, and hardly decreased in terms of precision compared to the classification without boosting. The 10 iterations boosted classifier is the most balanced of all in terms of recall and precision. After 30 and 50 iterations we observed the phenomenon of overfitting: the classifiers performances decrease both in terms of recall and precision. Those classifiers are tuned to the classification of the examples of the training base, but do not generalise well when confronted to new examples. Given that the linear classifier performs well without boosting on this problem, it is not surprising that only a few boosting iterations are necessary.

6. CONCLUSION

Based on the study of a concrete scenario proposed by our industrial partner FIAT, we defined a methodology for knowledge driven image analysis in which the concept of feedback loop enables the uses of methods and algorithms suitable for a specific problem. We instantiated an example of this methodology with the detection of air ducts in images of car interiors.

We have developed a strategy for ROI extraction with promising but still to be improved results. We have shown that the use of the EHD as a feature can be relevant for classification of air ducts. We have have studied and trained several classifiers for this problem, and we have been able to increase the single classifier performances by applying the AdaBoost ensemble learning method combined with linear SVM classifier.

7. ACKNOWLEDGEMENT

X-Media is partly sponsored by the European Commission under contract no. FP6-26978.

REFERENCES

[1] Thanos Athanasiadis, Phivos Mylonas, and Yannis Avrithis. Context-based region labelling approach for semantic image segmentation. In *First International Conference on Semantics and Digital Media Technologies*, 2006.

[2] Eric Bauer and Ron Kohavi. An empirical comparison of voting classification algorithms: bagging, boosting, and variants. *Machine Learning*, 36(1):105–139, 1999.

[3] Lionel Carminati and Jenny Benois-Pineau. Knowledge based supervised learning methods in a classical problem of video object tracking. In *IEEE ICIP*, 2006.

[4] Chih-Chung Chang and Chih-Jen Lin. *LIBSVM: a library for support vector machines*, 2001. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

[5] C. Cortes and V.Vapnik. Support-vector networks. *Machine Learning*, 20:273–297, 1995.

[6] T. Downs, K. Gates, and A. Masters. Exact simplification of support vector solutions. *Journal of Machine Learning Research*, 2(1):293–297, 2001.

[7] Christian Ferran, Xavier Giro, Ferran Marques, and Josep Ramon Casas. BPT Enhancement Based on Syntactic and Semantic Criteria. In *First International Conference on Semantics and Digital Media Technologies*, 2006.

[8] Yoav Freund and Robert E. Schapire. A decision-theoretic generalization of on-line learning and an application to boosting. *Journal of Computr and System Sciences*, 55(1), 1997.

[9] Hichem Frigui and Joshua Caudill. Region based image annotation. In *IEEE ICIP*, 2006.

[10] T.R. Gruber. Toward principles for the design of ontologies used for knowledge sharing. Technical Report KSL 93-04, Knowledge Systems Laboratory, Stanford University, 1992.

[11] Chih-Wei Hsu, Chih-Chung Chang, and Chih-Jen Lin. *A practical guide to support vector classification.* `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

[12] ISO/IEC 15938-3:2001. *Multimedia Content Description Interface - Part 3: Visual*, 2001.

[13] Hsuan-Tien Lin and Ling Li. Infinite ensemble learning with support vector machine. *Lecture Notes in Computer Sciences*, 3720:242–254, 2005.

[14] H.T. Lin and C.J Lin. A study on sigmoids kernels for svm and the training of non-PSD kernels by SMO-type methods. Technical report, National Taiwan University, 2003.

[15] B.S. Manjunath, Philippe Salembier, and Thomas Sikora. *Introduction to MPEG-7*. John Wiley and Sons, Inc., 2002.

[16] Keerthi S.S. and C.-J Lin. Asymptotic behaviors of support vector machines with gaussian kernel. *Neural Computation*, 15(7):1667–1689, 2003.

[17] Chee Sun Won, Dong Kwon Park, and Soo-Jun Park. Efficient use of MPEG-7 Edge Histogram Descriptor. *ETRI Journal*, 24(1), 2002.