BEHAVIOR OF DIFFERENT IMAGE CLASSIFIERS WITHIN A BROAD DOMAIN

B. Clemente
Sicubo Ltd., Avda. V. de la Montana, 18, 10002 Cáceres, Spain

M. L. Durán, A. Caro, P. G. Rodríguez
Escuela Politécnica, 10071 Cáceres, Spain; Universidad de Extremadura, Cáceres, Spain

Keywords: Machine learning, Image classification, Human perception.

Abstract: Image classification is one of the most important research tasks in the Content-Based Image Retrieval area. The term image categorization refers to the labeling of the images under one of a number of predefined categories. Although this task is usually not too difficult for humans, it has proved to be extremely complex for machines (or computer programs). The major issues concern variable and sometimes uncontrolled imaging conditions. This paper focuses on observation of behavior for different classifiers within a collection of general purpose images (photos). We carry out a contrastive study between the groups obtained from these mathematical classifiers and a prior classification developed by humans.

1 INTRODUCTION

Research on image retrieval has steadily gained high recognition over the past few years as a result of the great increase in digital image productivity. Research in Content-Based Image Retrieval (CBIR) is today a very active discipline, concentrating on in depth issues, such as learning or management access to information content in images. One of these issues is content-based categorization of images. Some CBIR applications aim to the retrieval of an arbitrary image that is representative of a specific class. In general, for CBIR systems, classifiers should be viewed as its own subfield of machine learning. The construction of systems capable of learning from experience (or from examples) has for a long time been the object of both philosophical and technical debates. This aspect has received great appraisal, while some researchers have demonstrated that machines can display a significant level of learning ability.

The term image categorization refers to the labeling of images under one of a number of predefined categories. The input/output pairings reflect a functional relationship that maps inputs to outputs. When an underlying function from inputs to outputs exits it is referred to as the *decision function*. This is chosen from a set of candidate functions which map from the input space to the output domain. The algorithm which takes the training data as input and selects a *decision function* is referred to as the *learning algorithm*, and, in this particular case the process is called *supervised learning* (Cristianini and Shawe-Taylor, 2000).

Although classification is not a very difficult task for humans, it proves to be an extremely difficult problem for machines (or computer programs). The main difficulties include variable and sometimes uncontrolled imaging conditions, complex and hard-to-describe image objects, objects occluding other objects, and the gap between the arrays of numbers representing physical images and the conceptual information perceived by humans.

Classification techniques are usually applied in the area of CBIR systems. Image categorization contributes to performing more effective searches. In the repertoire of images under consideration there is a gradual distinction between narrow and broad domains. A broad domain has an unlimited and unpredictable variability in its appearance even for the same semantic meaning (Smeulders et al., 2000). The good performance of classifiers has been demonstrated when the image domain is specific, i.e, it is a narrow domain which has a limited and predictable variability in relevant aspects for the specific purpose.

Clemente B., Durán M., Caro A. and Rodríguez P. (2009).
BEHAVIOR OF DIFFERENT IMAGE CLASSIFIERS WITHIN A BROAD DOMAIN.
In *Proceedings of the International Conference on Knowledge Discovery and Information Retrieval*, pages 278-283
DOI: 10.5220/0002297002780283

An example is the case of medical environments (El-Naqa et al., 2004), or text categorization (Dumais et al., 1998). There are other various studies that classify images only by means of other types of image features, such as color (Saber et al., 1996) or texture features (Fernández et al., 2003). Some approaches combine different features such as color and shapes, e.g., proposals by (Forczmanski and Frejlichowski, 2008) and (Mehtre et al., 1998). However, it is less common to find studies that focus on combining color, texture and shape features for classification purposes.

When classification methods are applied to general-purpose image collections the results are not positive, even less so if we hope that the performance of the classifier may match with the classification developed by non-expert humans. We find some examples in (Vailaya et al., 1999) and (Li and Wang, 2005).

This paper aims to observe the behavior of different kinds of classifiers within a collection of general-purpose images (photos). We thus describe a contrastive study between the groups made from these mathematical classifiers and a prior classification performed by humans.

To apply mathematical classifiers, it is necessary that each image be represented by a feature vector, i.e., each image is a point in a multidimensional space, called the feature space. In narrow environments with a defined purpose, feature extraction method are restricted to those that highlight what is relevant and necessary for the application. Because this paper focuses on natural images in a broad domain, we obtain texture, color and shapes features. The reason is that these require human knowledge in their perception.

2 MATERIALS AND METHODS

Our collection consists of more than 2000 images retrieved from Internet, all withdrawn from different sites. With the aiding guidance of untrained users, these images are grouped according to perceptual criteria, that is, 10 groups are made, because this is what seemed logical from the standpoint of the people involved. In the initial distribution, the number of images within each class was not homogeneous. Yet, in order to conduct a more rigorous testing, each class is maintained with a total of 200 images. It is important that the distribution of the samples be uniform across all groups. The grouping is entirely based on the perceptual criteria related to how content is valued by humans. Thus, we have 2000 images classified into 10 classes, namely trees, people, cars, flowers, buildings, shapes, textures, animals, sunsets, and circles. Some samples of each group are shown in Fig. 1, one respectively per row.

2.1 Feature Extraction

Each image is represented by a feature vector of 110 features. This feature vector is divided into three groups: the first 60 are labeled as color features, the following 41 as texture features, and the last nine as shape features.

2.1.1 Color Features

The color features used in this work are based on the HLS model (Hue, Saturation, Luminosity), since human perception is quite similar to this model. We are using color discretization (MacDonald and Luo, 2002) in 12 colors in Hue and in addition three other colors, white, grey and black in the luminosity axis (15 colors), indicating the ratio of pixels for each one. On the other hand, local color features are used in order to achieve information about the spatial distribution (Cinque et al., 1999): in particular the barycenter of every 15 discrete colors with its coordinates in the image (x,y), with 30 other features. Finally, the standard deviation information from barycenter is also computed, and therefore there are 15 additional features. Summarizing, the total number of color features are 60.

2.1.2 Texture Features

These have been obtained by applying two well known methods. The first one works on a global processing of images, it is based on the Gray Level Coocurrence Matrix proposed by Haralick (Haralick and Shapiro, 1993). This matrix is computed by counting the number of times that each pair of gray levels occurs at a given distance and for all directions. Features obtained from this matrix are: energy, inertia, contrast, inverse difference moment, and number non-uniformity. The second method is focused to detect only linear texture primitives. It is based on features obtained from the Run Length Matrix proposed by (Galloway, 1975), where a textural primitive is a set of consecutive pixels in the image having the same gray level value. Four matrices, one for each direction, are made, computed by counting all runs into the image. Every item in these matrices indicates the number of runs with the same length and gray level. There are four matrices obtained from angles quantized to 45^o intervals. One for horizontal runs (0^o), one per vertical runs (90^o) and the other two for the two diagonals (45^o and 135^o). The features obtained from these matrices are long run emphasis (LRE), short run emphasis (SRE), gray level non-uniformity (GLNU),

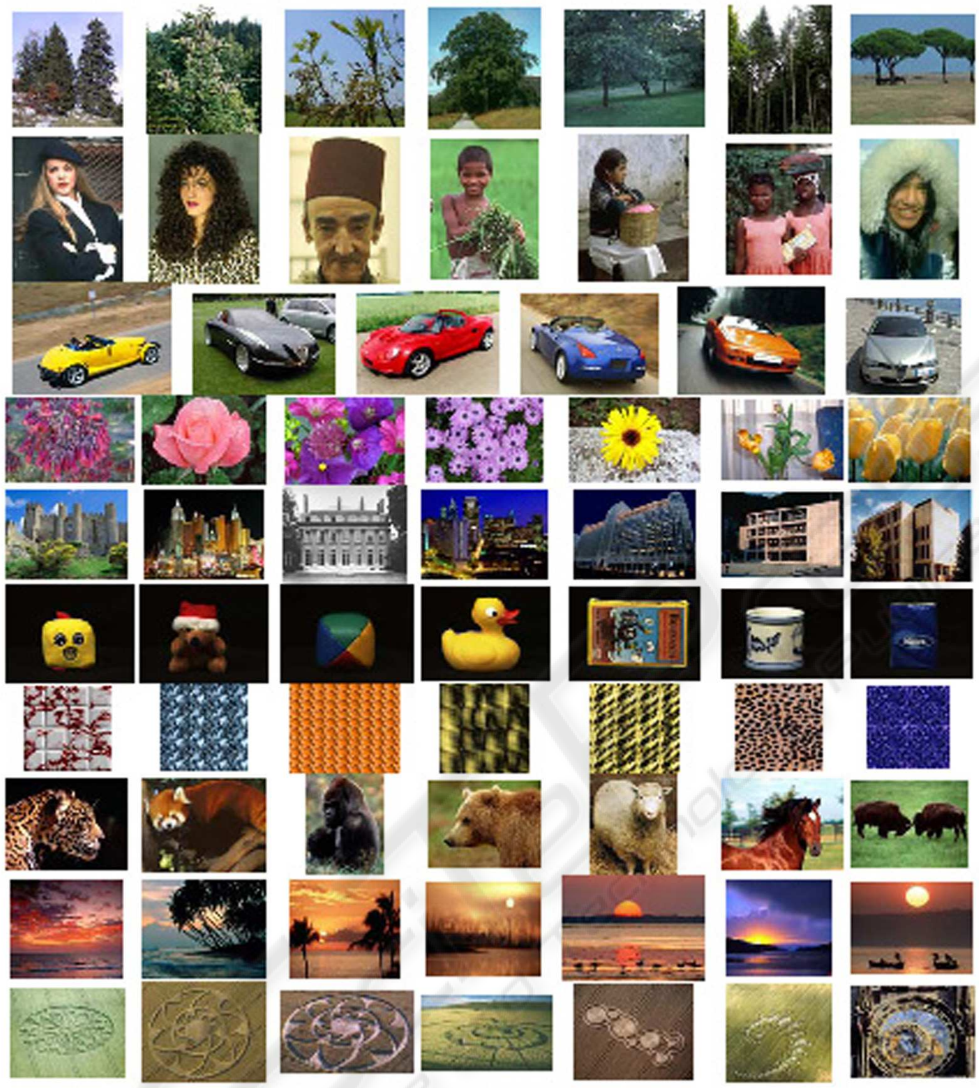

Figure 1: Some samples of the image collection.

run length non-uniformity (RLNU), run percentage (RPC), short runs in low gray emphasis (SRLGE), short runs in high gray emphasis (SRHGE), long runs in low grey emphasis (LRLGE), and long runs in high grey emphasis (LRHGE) (Chu, 1990). These nine features have been obtained four times, one per direction.

2.1.3 Shape Features

The images are processed by using Active Contours (Caro et al., 2007a) as segmentation a method, and, then some shape features are obtained from these contours. Shape features are based on Hu's moments (first and second moments), centroid (center of gravity), angle of minimum inertia, area, perimeter, ratio of area and perimeter (RAP), and major and minor axes of fitted ellipses. The methods to obtain these features are referred in (Caro et al., 2007b).

2.2 Classification

When each image is represented in the features space by its feature vector, the next step is the application of classification methods. Three applied methods belong to the supervised learning classifiers, the first one (Support Vector Machine) is one of the latest purposes on classifiers, the other two are the most traditional and frequently applied classifiers.

2.2.1 Support Vector Machines

Support Vector Machines (SVM) are learning structures based on the statistical learning theory to solve classification, regression, and probability estimate problems. SVMs working on a space of linear functions hypotheses with high dimensionality attributes space. The learning algorithm has to solve a quadratic programming problem to return the hypothesis that separates, with a maximum margin, the positive examples set of the negative examples. The margin is defined as the distance from the hyperplane to positive and negative examples closest to it. SVMs induce a linear hyperplane in the input space, therefore they belong to the method family called linear learning machines. Among all possible separation hyperplanes, the SVM choose the maximum margin (Cristianini and Shawe-Taylor, 2000).

The LIBSVM software has been used to test the performance of SVM (Chang and Lin, 2001). This provides an efficient multi-class classification using the ”one-against-one” approach in which $k(k-1)/2$ classifiers are constructed and each one trains data from two different classes. In classification, LIBSVM uses a voting strategy.

2.2.2 Multilayer Perceptron

The advantage of using neural networks in pattern recognition is based on the fact that regions of nonlinear decision can be separated depending on the number of neurons and layers. Therefore, the artificial neural networks are used to solve classification problems with high complexity (Cristianini and Shawe-Taylor, 2000).

Within the neural networks, the ones most commonly used are the networks with multiple layers that work forward. This type of neural network is composed of a layer of input neurons, a set of one or more hidden layers and a output layer. The input signal starts from the input layer and spreads forward, going through the hidden layer until it reaches the output layer.

Multilayer Perceptron (MLP) is based on the Back Propagation algorithm. This is a generalization of the rule of least squares, which is also based on error correction. The Back Propagation algorithm provides an efficient method to train such networks. Importance is in ability to adapt the weights of intermediate neurons to learn the relations between the input set and its corresponding output, and that relations can be applied to new patterns. The network must find an internal representation that allows to generate the desired outputs for the training stage, and later during the test phase it must be capable of generating outputs for which entries were not shown during learning, but that resemble one which was shown.

2.2.3 Bayesian Classification

This method is based on statistics that use the calculus of probabilities from the Bayes Theorem. Given a set of training examples and a priori knowledge about the probability of each hypothesis, Bayesian learning can be seen as the process to find the most probable hypothesis. The way of applying the Bayes theorem for classification consists of calculating the most probable posteriori hypothesis (Domingos and Pazzani, 1996). This method presents some difficulties such as the need to have previous knowledge and the high computational cost. Moreover, this method has a restriction as strong as the independence of the attributes. To prevent this restriction, a preprocess stage has been applied to the features. This preprocess has consisted of the principal components analysis and the result is a vector with 46 features. 32 color features, 8 texture features and 6 shape features, all of them supporting 90% of the variability.

Software provided by WEKA has been used to apply these two last types of classifiers (http://www.cs.waikato.ac.nz/ml/weka/).

3 RESULTS

Considering the methods described in the previous section, a comparative study to determine the performance of the different classification methods is considered an important issue. To achieve this, the application of such classification methods is tested on the digital image collection.

In all the achieved experiments, a training set of 800 images (40% of the samples) is selected, and the remaining 1200 images (60% of the image collection) are used as test set. All the classifiers are trained by only using color, texture or shape features. A fourth possibility is then considered, taking into account all the color, texture and shape features at once. Fig.2 summarizes the results for all the experiments.

Color-based feature vectors are used in the first experiments of the ones achieved. Fig.2.a shows the obtained results. SVM obtains the best rate, which reaches 62.4%. Multilayer Perceptron achieves an acceptable rate of 58.7%, while the Bayes classifier obtains the worst result (50.2%), by 12% below the results of SVM.

Moreover, Fig.2.b illustrates the results obtained in the second experiment. In that case, texture-based

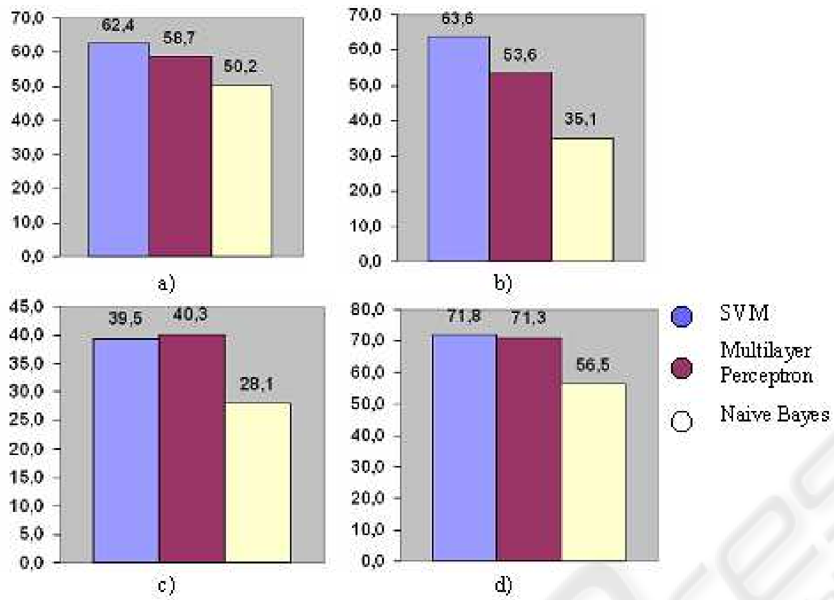

Figure 2: Results of classification with a) color features, b) with texture features, c) with shape features, and d) with all features.

feature vectors are used. Again, the best rate is obtained by SVM (63.6%), similar to that achieved by color-based feature vectors. The performance of the Multilayer Perceptron was inferior to the previous experiment (53.6%), and so was the Bayes classifier (35.1%). Particularly striking are the results obtained by this last classifier, which correspond to almost half of the SVM marks.

The third test classifies the images according to the shape features. Multilayer Perceptron reaches a percentage (40.3%) slightly higher than the one obtained by SVM (39.5%). Again, the worst result is obtained by the Naive Bayes classifier, as Fig.2.c. illustrates.

The last one of the experiments is based on all the features (color, texture and shape). The images are classified by considering all the combined features, and the final results are shown in Fig.2.d. The best results are obtained by SVM (71.8%), followed by Multilayer Perceptron (71.3%). In contrast, the Naive Bayes classifier achieves the worst marks (56.5%).

As aforementioned, both SVM and Multilayer Perceptron yield positive success rates, considering the high complexity of the images on the database used in the experiments. Average results of these two classifiers are quite similar, for vectors composed of color, texture and shape features, as well as for a combination of all the features.

4 DISCUSSION AND CONCLUSIONS

This paper has demonstrated the improvement implied in the SVM application to manage progress in such image categorization. In addition, we have contrasted this performance of the SVMs with other robust methods of classification such as the neural networks and the Naive Bayes classifiers. We should highlight the fact that, even though they are only three, these classifiers are diverse and deal with very different aspects.

We consider that the results obtained with LIBSVM are quite positive if we account for differences in color among images of the same class. Such is the case of the animal group, where we can find a photo of a white horse and at the same time an image of a black pig. Another example can be found in the images of the people group where the color is very rich and varied. In this sense, this type of occurrences takes place when we apply shapes and textures to broad domains.

SVM has demonstrated to be an algorithm with a significant level of learning ability, even within a broad domain, with photos or images with general purposes. We should emphasize that the images in each class are very different, and, about all, we are dealing with a very wide multi-category classification with ten classes.

Finally, we wish to add, for future studies in this line of work, that there must be a need for human perception in the evaluation of the system. In addition, we should attempt to integrate high-level features originating from low-level features in the research. Then another important aspect would be the phase for the relevance feedback processes, and the increase of the image database size for the testing of the system as a multilevel classification process.

ACKNOWLEDGEMENTS

The authors wish to acknowledge and thank the Sicubo Ltd. for their strong belief and firm support in this work. This work is financed by the Spanish Government (National Research Plan) and the European Union (FEDER founds) by means of the grant ref. TIN2008-03063 and the Junta de Extremadura (Regional Government Board - Research Project #PDT08A021).

REFERENCES

Caro, A., Alonso, T., Rodríguez, P., Durán, M., and Ávila, M. (2007a). Testing geodesic active contours. LNCS 4478:64–71.

Caro, A., Rodríguez, P., Antequera, T., and Palacios, R. (2007b). Feasible application of shape-based classification. LNCS 4477:588–595.

Chang, C. and Lin, C. (2001). *LIBSVM: a Library for Support Vector Machines*. Department of Computer Science and Information Engineering, National Taiwan University, http://www.csie.ntu.edu.tw/ cjlin/libsvm.

Chu, A. (1990). Use of grey value distribution of run lengths for texture analysis. *Pattern Recognition Letters*, 11:415–420.

Cinque, L., Levialdi, S., Pellicano, A., and Olsen, K. (1999). Color-based image retrieval using spatial-chromatic histograms. In *IEEE. International Conference on Multimedia Computing and Systems*, volume 2, pages 969–973.

Cristianini, N. and Shawe-Taylor, J. (2000). *An Introduction to Support Vector Machines and other kernel-based learning methods*. Uiversity Press, Cambridge.

Domingos, P. and Pazzani, M. (1996). Beyond independence: conditions for the optimality of the simple bayesian classifier. In *Machine Learning: Proceedings of the Thirteenth International Conference*, pages 105–112. Morgan Kaufmann.

Dumais, S., Platt, J., Heckerman, D., and Sahami, M. (1998). Inductive learning algorithms and representations for text categorization. In *Proc. of 7th International conference on Information and Knowledge Management*.

El-Naqa, I., Yongyi, Y., Galatsanos, N., Nishikawa, R., and Wernick, M. (2004). A similarity learning approach to content-based image retrieval: application to digital mammography. In *IEEE Transactions on Medical Imaging*, volume 23, pages 1233–1244.

Fernández, M., Carrión, P., Cernadas, E., and Gálvez, J. (2003). Improved classification of pollen texture images using svm and mlp. In *3rd IASTED international conference on visualization, imaging and image processing*.

Forczmanski, P. and Frejlichowski, D. (2008). *Computer Recognition Systems 2*, volume 45 of *Advances in Soft Computing*, chapter Strategies of Shape and Color Fusions for Content Based Image Retrieval, pages 3–10. Springer Berlin / Heidelberg.

Galloway, M. (1975). Texture analysis using grey level run lengths. *Computer Graphics and Imag. Processing*, 4:172–179.

Haralick, R. and Shapiro, L. (1993). *Computer and Robot Vision*. Addison-Wesley.

Li, J. and Wang, J. (2005). Alip: the automatic linguistic indexing of pictures system. *Computer Vision and Pattern Recognition*, 2:1208–1209.

MacDonald, L. and Luo, M. (2002). *Colour Image Science: Exploiting Digital Media*. Wiley.

Mehtre, B., Kankanhalli, M., and Lee, W. (1998). Content-based image retrieval using a composite color-shape approach. *Information Processing and Management*, 34(1):109–120.

Saber, E., Tekalp, A. M., Eschbach, R., and Knox, K. (1996). Automatic image annotation using adaptive color classification. *Graph. Models Image Process.*, 58(2):115–126.

Smeulders, A. W. M., Worring, M., Santini, S., Gupta, A., and Jain, R. (2000). Content-based image retrieval at the end of the early years. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(12):1349–1380.

Vailaya, A., Figueiredo, M., Jain, A., and Hong Jiang, Z. (1999). Content-based hierarchical classification of vacation images. In *Multimedia Computing and Systems, IEEE International Conference*, pages 518–523.