A REGION-BASED APPROACH TO CONCEPTUAL IMAGE CLASSIFICATION *

Symeon Papadopoulos[1], Vasileios Mezaris[1,2], Ioannis Kompatsiaris[2], Michael G. Strintzis[1,2]

[1] Information Processing Laboratory, Electrical and Computer Engineering Dept., Aristotle University of Thessaloniki, Thessaloniki 54006, Greece

[2] Informatics and Telematics Institute/Centre for Research and Technology Hellas, 1st Km Thermi-Panorama Rd, Thessaloniki 57001, Greece

Abstract

Classifying images into a set of semantic categories that are meaningful to humans has proved to be a challenging and attractive problem in the field of content-based retrieval. Addressing this problem is typically based on the initial extraction of low-level features for the images and the subsequent application of a pattern recognition technique, to divide the feature space in a number of subspaces corresponding to the semantic categories. An extension to this framework is presented in this paper, aiming at the improvement of the efficiency of image classification systems. This is based on the introduction of an unsupervised still image segmentation algorithm to the process and its combination with MPEG-7 low-level descriptors and a Bayes classifier. Experimental results using different pairs of classes and corresponding data sets demonstrate the efficiency of the proposed approach.

1 Introduction

Due to multimedia data becoming abundant in personal computers, local networks and the internet, the need for efficient image indexing and retrieval, especially at a semantic level, has emerged as an important challenge. Both the necessity for effective search algorithms and the prospect of implementing more autonomous AI systems have led to the development of algorithms and approaches that attempt to address this problem.

The first image retrieval systems that were presented in this field were based on the Query-by-Example (QbE) paradigm, according to which the user provides to the retrieval system a piece of multimedia data that roughly represents what the users wishes to retrieve, which serves as the example for performing similarity search. Systems based upon this scheme include QBIC [4], Photobook [12] and Visualseek [13]. Newer approaches to image retrieval additionally considered the conceptual characterization of multimedia entities, using a set of predefined classes to which images were classified. The most usual approach to conceptual characterization is based on binary classification, as in [16] and [14], which is typically integrated in a hierarchical scheme, [15]. Other approaches to visual information retrieval are based on more complex schemes, e.g. MediaNet [1], that incorporate lexical characterization, instance-based representations as well as feature description of the multimedia entities and make use of complicated logical structures, e.g. ontologies [8], in order to infer knowledge from data.

In this work, an extension to the general image classification framework is presented, aiming at the improvement of the efficiency of image classification systems by the use of an unsupervised still image segmentation algorithm as part of the classification process. Feature extraction, classifier training and classification are thus performed at the region level. MPEG-7 standardized low-level descriptors are used as region features for the classification. A general overview of the proposed classification system architecture is presented in section 2. Brief description of the segmentation algorithm and the low-level features is provided in section 3, while in section 4 the theoretical basis for the classification process is illustrated. Comparison of the efficiency of the proposed classification system with one that does not make use of a segmentation algorithm, i.e. treats an image as a whole, is reported in the experimental results section (section 5), and finally, conclusions are drawn in section 6.

2 System Architecture

The architecture that was adopted in terms of the implementation and testing of the proposed classification system is based on the model proposed in [5, 6]. In the proposed approach, this is extended with the insertion of an additional step in the process of classification. More specifically, instead of applying the classification algorithm to the images, we first apply an automatic image segmentation algorithm to them and then classify the produced regions. These regions are homogeneous in color and texture, so they typically correspond to meaningful entities of the image, i.e. objects or parts of them. Subsequently, the class membership of the images is determined based on the classification of their constituent regions.

Initially, a set of classes $C = \{\omega_1, \omega_2, ..., \omega_N\}$ is defined, so

*This work was supported by the EU projects SCHEMA "Network of Excellence in Content-Based Semantic Scene Analysis and Information Retrieval" (IST-2001-32795) and aceMedia "Integrating knowledge, semantics and content for user centred intelligent media services" (FP6-001765).

that the classification problem becomes specific. A set E of suitable images is assembled, in order to be used as a training and test set.

The automatic image segmentation algorithm of [10, 9] is then applied on that set of images, thus producing a set of regions P. A number of standardized MPEG-7 low-level descriptors are then extracted, in order to serve as classification features for each region $p \in P$. The features that are used by the employed high-level classifier are the Dominant Color, Edge Histogram and Contour Shape descriptors; subsets of these features are used for classification.

A set $P_{tr} \subset P$ of regions which are manually classified (i.e. a semantic tag is attached to each one of them) is formed, in order to serve as a training set for the classifier. Following training, the remaining regions, belonging to set $P - P_{tr}$, can be classified by use of the classification algorithm, which is presented in the sequel. Finally, a rule is used to classify the images based on the classification of the regions that constitute them.

3 Image Segmentation and Low-level Features

3.1 Segmentation Algorithm

The segmentation algorithm that was adopted is an unsupervised color image segmentation algorithm proposed in [10, 9]. It is based on a modified K-Means-with-Connectivity-Constraint (KMCC) algorithm that groups pixels into regions, based on their weighted sum of Euclidean distances from the region centers in the combined intensity-texture-position feature space. Consequently, using this algorithm each image is segmented to connected regions homogeneous with respect to intensity and texture characteristics.

In general, the segmentation process can be summarized in the following steps:

At first, the number of regions that should be formed as well as their spatial centers are estimated, so that a set of initial values are supplied to the modified KMCC algorithm. This initialization, based on clustering at a lower resolution, aims at the faster convergence of the KMCC algorithm. Then, preprocessing by means of the conditional application of a moving average filter that alters pixel intensities, controlled by the corresponding pixel texture features, is performed, followed by the assignmet of pixels to regions using the modified KMCC algorithm.

Extensive experimental evaluation of the aforementioned segmentation algorithm showed that the generated regions typically correspond to meaningful semantic objects depicted in the image, or parts of them. This observation led to the hypothesis that classifying regions instead of images and using these classification results to infer image cluster membership could result in a gain in performance.

3.2 MPEG-7 low-level descriptors

The well known MPEG-7 standard contains specifications for multimedia data and content description, including detailed specifications for the description of digital audio, image and video, as well as speech, graphics, and their combinations [3]. The descriptors that were chosen for use by the employed high-level classifier are the Dominant Color [7], Edge Histogram [7] and Contour Shape descriptors [2]; subsets of these descriptors are used for classification. These were chosen due to their conciseness and discriminative power.

More specifically, the MPEG-7 descriptors are extracted in XML form using the MPEG-7 XM [11]. Then, a customized parser isolates the necessary values and concatenates them in a single vector. The exacted values used, are:

- Dominant Color: The color indices for the two most dominant colors.
- Edge Histogram: After acquiring the 80 values, which actually correspond to 16 sub-images (5 dimensions each), the corresponding values are added so that only 5 dimensions result. Thus, the high dimensionality of the descriptor is reduced to a compact description of each region's edge directionality.
- Contour Shape: Global curvature, prototype curvature, highest peak and the number of peaks of the region (6 dimensions in total).

4 Classification Algorithm

The classification process is viewed as a typical pattern recognition problem, consisting of a training and a testing process. Given a set of measurements, i.e. the low-level features that are extracted from an arbitrary image region and constitute the input feature vector x, the classification goal is to assign the pattern x to one of N predefined classes, ω_i, $i = 1, ..., N$. A decision rule divides the measurement space in N regions Ω_i , $i = 1, ..., N$. If an observed vector lies in region Ω_i, it is assumed to belong to class ω_i.

4.1 Training

The training process can be summarized in two major steps:

First, the conditional probability of the vector x given that it belongs to class ω_i is assumed to be a multivariate Gaussian distribution.

$$p(x|\omega_i) = N(x|m, \Sigma) = \frac{1}{2\pi^{\frac{p}{2}} \cdot |\Sigma|^{\frac{1}{2}}} \cdot e^{-\frac{1}{2}\cdot(x-m)^T\cdot\Sigma^{-1}\cdot(x-m)} \tag{1}$$

where $x = [x_1\ x_2 \ldots x_p]^T$ is the p-dimensional feature vector. Then, the mean vector m and the covariance matrix Σ are estimated for each class ω_i from the training set.

$$\hat{m} = \frac{1}{n} \cdot \sum_{i=1}^{n} x_i \tag{2}$$

$$\hat{\Sigma} = \frac{1}{n} \cdot \sum_{i=1}^{n} (x_i - \hat{m})) \cdot (x_i - \hat{m})^T \tag{3}$$

4.2 Classification

The calculation of the a posteriori probability $p(x|\omega_i)$ of an unknown vector is performed using the mean vector and the covariance matrix that were computed for class ω_i, during the training process.

The basis for classifying a pattern x to one of a set of predefined classes is the minimum error Bayes rule, which in the special case of binary classification (i.e. when the number of classes equals two) is formulated by means of the likelihood ratio l_r:

$$l_r(x) = \frac{p(x|\omega_1)}{p(x|\omega_2)} > \frac{p(\omega_2)}{p(\omega_1)} \ \rightarrow \ x \in \Omega_1 \tag{4}$$

In this work, a variation of the minimum error Bayes rule, which is called Neyman-Pearson rule is used for the assignment of a pattern x to a class.

$$\frac{p(x|\omega_1)}{p(x|\omega_2)} > \mu \rightarrow x \in \Omega_1 \tag{5}$$

The threshold μ is chosen empirically, in accordance with the probability distribution of the feature values within the region (image) training and test set. The selection of a Bayesian method was made on the basis of its simplicity and efficiency.

After all the regions of an image are classified according to the aforementioned method, the final decision on the image category is made on the basis of a simple rule: If at least one region of the image is found to belong to the "positive" class, then the whole image is considered to belong to that class. The 'positive' class is pre-defined in each problem, e.g. in the classification problem of face - non-face images, the "positive" class is intuitively chosen to be the face class.

5 Experimental Results

Comparative evaluation between the traditional classification (image-level) approach and the one proposed here (region-level) was carried out for three pairs of classes: face/non-face images, city/landscape images, and images containing/not containing a sky region. The images used in the experiments include images belonging to the Corel gallery and to the Macedonian Press Agency as well as images collected from the web. For the three aforementioned pairs of classes, the "positive" class was defined as the face, city and sky-image class, represented by the face, building and sky region correspondingly. In the results depicted in Figures 1 to 3, the regions found by the classifier to be members of the positive class are marked with a grid-like texture. In each case, the first four images of each row constitute correctly classified images, while the last one of each row constitutes a misclassification example.

Starting with the face/non-face classifier, a set of 616 images were used for evaluating its efficiency; these images can be roughly divided in three groups:

- images where the presence of human face is dominant (i.e. close-up face images),
- images where human faces are clearly distinguishable but not dominant,
- images in which no faces are depicted.

The results that were obtained are recorded in Table 1, while several sample results are depicted in Figure 1. These results indicate that the proposed region-based classification to face/non-face images is more accurate that the widely used global image classification, particularly in classifying as face images those containing distinguishable but not dominant face regions. However, in the region-based approach, accurate classification may be affected by the accuracy of segmentation, as shown in Figure 1.

For the city/landscape classifier, a set of 477 images were used for evaluating its efficiency; these were divided in two groups:

- images depicting urban areas (buildings, streets etc.),
- natural landscape and countryside images.

The results that were obtained are presented in Table 2, while several sample results are depicted in Figure 2. The results indicate that the global image approach is more suitable than the region-based approach; this is due to the particular classification problem being not as "region oriented" as the face/non-face one, i.e. judging the distinction between the two classes cannon be based solely on the detection of one object (building) but others may be needed as well (e.g. street). The diversity of building low-level characteristics as opposed to face low-level characteristics is an additional factor influencing the efficiency of the region-based approach in this case.

Region-level approach (using 2 Dominant Colors and Contour Shape, $\mu = 130$)	
Face images	Non-face images
$352/414$ (85%)	$146/202$ (72.3%)
Close-ups	Distant shots
$184/214$ (85.9%)	$168/200$ (84%)
Global image classification (using 2 Dominant Colors and Edge Histogram, $\mu = 5$)	
Face images	Non-face images
$290/414$ (70%)	$176/202$ (87.1%)
Close-ups	Distant shots
$179/214$ (83.6%)	$111/200$ (55.5%)

Table 1: Correct classification rates for face/non-face image classification

Region-level approach (using 2 Dominant Colors and Edge Histogram, $\mu = 5$)	
City images	Landscape images
$233/267$ (87.3%)	$171/210$ (81.4%)
Global image classification (using 2 Dominant Colors and Edge Histogram, $\mu = 1$)	
City images	Landscape images
$235/267$ (88%)	$199/210$ (94.8%)

Table 2: Correct classification rates for city/landscape image classification

Finally, for the classification of images to those containing/not containing a sky region, a set of 398 images were used for evaluation, divided in the two aforementioned categories. Table 3 contains the comparative evaluation results and Figure 3 depicts sample results. Due to the "region oriented" nature of this classification problem, the region-based classification scheme is shown to outperform the global image approach, as was the case with the face/non-face classifier.

The results recorded in tables 1 and 3 for the classification problems of face/non-face and sky/no-sky images supply substantial evidence that the proposed method performs equally or even better than the traditional image-level classification method. Also, the fact that the "object" of classification is an image region and not the whole image provides possibilities of adding more complex logic in order to infer the class of the image from the classes of the regions that constitute it. Additionally, the solution supplied to the classification problems studied here is economical in the sense that only concise low-level features are used, achieving small dimensionality, limited storage needs and quick completion of the classification process.

Nevertheless, together with the proposed method two drawbacks are brought. First, the classification performance is greatly influenced by the success of the segmentation algorithm. As shown in the result figures, there are some cases where the segmentation has failed to form a region accurately corresponding to the object of interest depicted in the image, e.g. one of the face regions in Figure 1. In this case, an erroneous classification result is to be expected.

Furthermore, there are certain classification problems that are not "region oriented". For instance, the classification between city and landscape images is a hard task to perform by means of a region-level approach, as can be seen in table 2. This becomes obvious from the comparison of the success rates of the typical global-image-based and the proposed region-based approach. Apparently, the proposed classification system is expected to feature improved performance in object detection problems as well as in cases where one of the two classes of a classification scenario is satisfyingly represented by a single object.

6 Conclusions and Future Work

In this work, the introduction of the proposed region-based approach to classification applications was shown to contribute towards the improvement of efficiency and robustness of classification, especially when semantically meaningful entities are adequately represented by image regions. However, there are three major issues which call upon extensive research. First of all, segmentation and classification algorithms should take advantage of each other in the form of a closed-loop procedure. The output produced by the classification subsystem should be exploited in the form of feedback to the segmentation algorithm, so that the produced regions better

Region-level approach (using Most Dominant Color and Edge Histogram, $\mu = 1$)	
Images containing sky	Images not containing sky
$90/103$ (87.4%)	$273/295$ (92.6%)
Global image classification (using Most Dominant Color and Edge Histogram, $\mu = 0.1$)	
Images containing sky	Images not containing sky
$75/103$ (72.8%)	$251/295$ (85.1%)

Table 3: Correct classification rates for images containing/not containing sky

approximate semantically meaningful objects. Moreover, low and intermediate level features should be refined so that the gap between low-level data and high-level meaning can be bridged. Last, the application of more complex logic in order to extract knowledge for the whole image based on an estimation for the regions that constitute it is necessary for improving the efficiency of the proposed approach.

7 Acknowledgement

The authors would like to thank the Macedonian Press Agency S.A. for contributing some of the images used in this work.

References

[1] A.B. Benitez, J.R. Smith, and S.-F. Chang. MediaNet: A Multimedia Information Network for Knowledge Representation. *Proceedings of SPIE*, 4210, October 2000.

[2] M. Bober. MPEG-7 Visual Shape Descriptors. *IEEE Transactions on Circuits and Systems for Video Technology*, 11(6), 2001.

[3] S.F. Chang, T. Sikora, and A. Puri. Overview of the MPEG-7 Standard. *IEEE Transactions on Circuits and Systems for Video Technology*, 11(6), 2001.

[4] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yanker. Query by image and video content: the QBIC system. *Computer*, 28(9):23–32, Sept. 1995.

[5] J.W. Han, L. Guo, and Y.S. Bao. A Novel Image Retrieval Model. In *Proc. 6th International Conference on Signal Processing*, volume 2, pages 953–956, August 2002.

[6] J. Luo and A. Savakis. Indoor vs. outdoor classification of consumer photographs using low-level and semantic features. *Proceedings of International Conference on Image Processing*, 2:745–748, 2001.

[7] B.S. Manjunath, J.-R. Ohm, V.V. Vasudevan, and A. Yamada. Color And Texture Descriptors. *IEEE Transactions on Circuits and Systems for Video Technology*, 11(6), 2001.

[8] V. Mezaris, I. Kompatsiaris, and M. G. Strintzis. An Ontology Approach to Object-based Image Retrieval. In *Proc. IEEE International Conference on Image Processing (ICIP)*, volume 2, pages 511–514, Barcelona, Spain, September 2003.

[9] V. Mezaris, I. Kompatsiaris, and M. G. Strintzis. Still Image Segmentation Tools for Object-based Multimedia Applications. *International Journal of Pattern Recognition and Artificial Intelligence*, 18(4):701–725, June 2004.

[10] V. Mezaris, I. Kompatsiaris, and M.G. Strintzis. A framework for the efficient segmentation of large-format color images. In *Proc. IEEE International Conference on Image Processing (ICIP)*, volume 1, pages 761–764, Rochester, NY, September 2002.

[11] MPEG-7 XM software. http://www.lis.ei.tum.de/research /bv/topics/mmdb/e_mpeg7.html.

[12] A. Pentland, R. Picard, and S. Sclaroff. Photobook: Content-based manipulation of image databases. *Int. J. Computer Vision*, 18(3):233–254, 1996.

[13] J.R. Smith and S.-F. Chang. Visualseek: A fully automated content-based image query system. In *ACM Multimedia*, pages 87–98, 1996.

[14] M. Szummer and R.W. Picard. Indoor-Outdoor Image Classification. In *IEEE International Workshop on Content-based Access of Image and Video Databases, in conjunction with ICCV*, pages 42–51, January 1998.

[15] A. Vailaya, M.A. Figueiredo, A.K. Jain, and H.J. Zhang. Image Classification for Content-Based Indexing. *IEEE Transactions on Image Processing*, 10(1):117–130, January 2001.

[16] A. Vailaya, A.K. Jain, and H.J. Zhang. Image Classification: City Images vs. Landscapes. *Pattern Recognition*, pages 1921–1935, January 1998.

Figure 1: Sample results of face/non-face image classification experiments

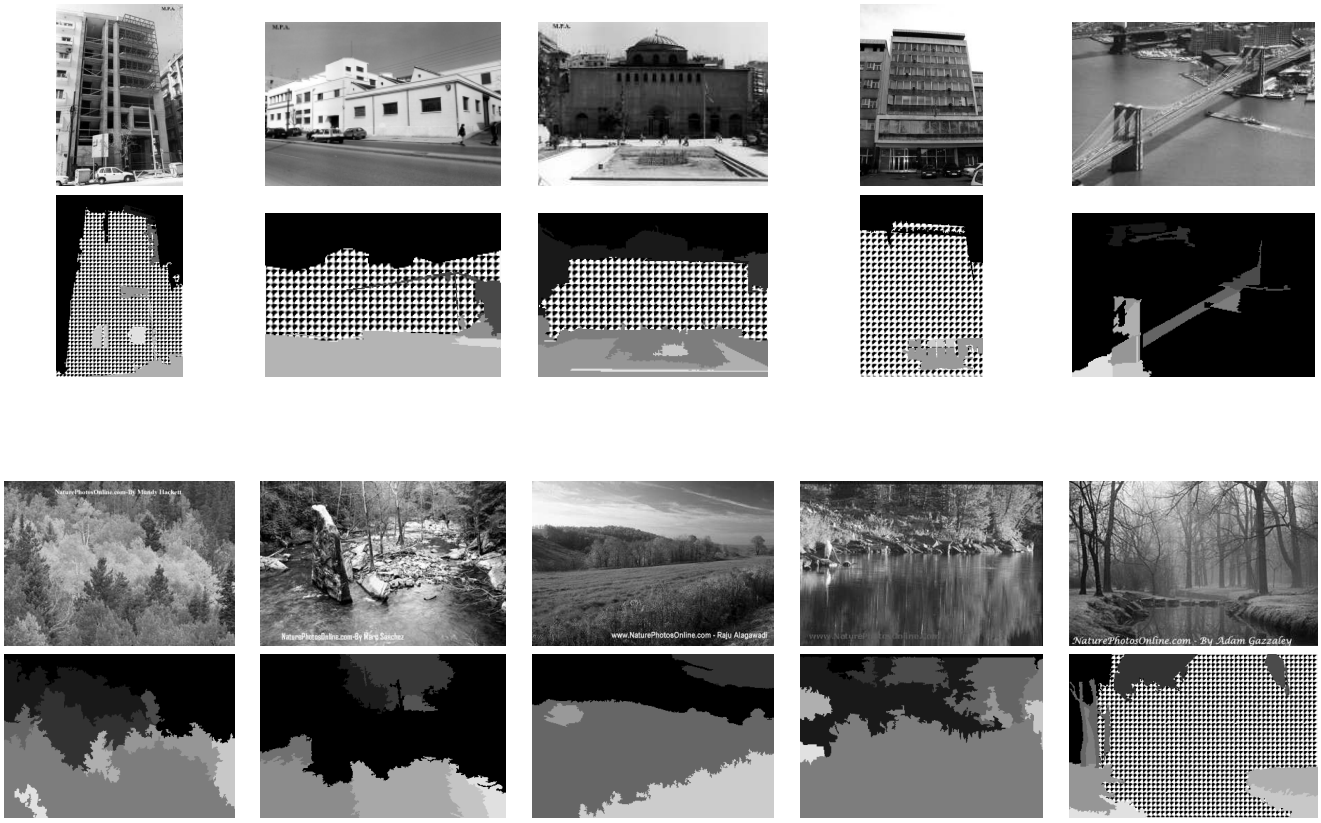

Figure 2: Sample results of city/landscape image classification experiments

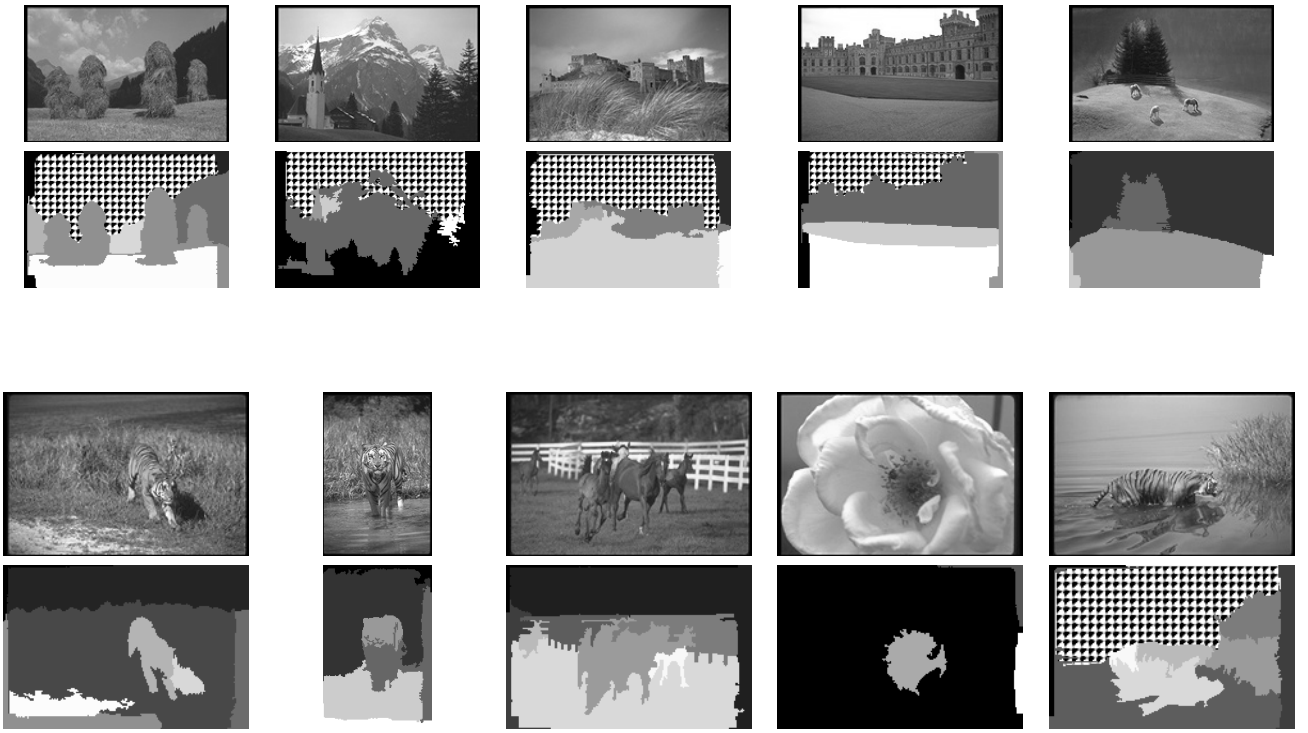

Figure 3: Sample results of classification experiments for images containing/not containing sky