# A Comparative Study of Classification Techniques for Knowledge-Assisted Image Analysis*

G. Th. Papadopoulos[1,2], K. Chandramouli[3], V. Mezaris[2], I. Kompatsiaris[2], E. Izquierdo[3] and M.G. Strintzis[1,2]

[1]Information Processing Laboratory
Electrical & Computer Eng. Dep.
Aristotle University of Thessaloniki
Thessaloniki, GR-54124, Greece

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

[3]Multimedia and Vision Research Group
Queen Mary, University of London
Mile End Road, E1 4NS
London, UK

## Abstract

*In this paper, four individual approaches to region classification for knowledge-assisted semantic image analysis are presented and comparatively evaluated. All of the examined approaches realize knowledge-assisted analysis via implicit knowledge acquisition, i.e. are based on machine learning techniques such as Support Vector Machines (SVMs), Self Organizing Maps (SOMs), Genetic Algorithm (GA) and Particle Swarm Optimization (PSO). Under all examined approaches, each image is initially segmented and suitable low-level descriptors are extracted for every resulting segment. Then, each of the aforementioned classifiers is applied to associate every region with a predefined high-level semantic concept. An appropriate evaluation framework has been employed for the comparative evaluation of the above algorithms under varying experimental conditions.*

## 1 Introduction

Given the continuously increasing amount of image content generated everyday and the richness of the available means for sharing and distributing it, the need for efficient and advanced methodologies regarding image manipulation emerges as a challenging and imperative issue. To this end, intense research efforts have concentrated in the development of sophisticated and user-friendly systems for skilful management of images. Most emerging approaches adopt the fundamental principle of shifting image manipulation techniques towards the process of the visual content at a semantic level, thus attempting to bridge the so called *semantic gap* [11]. Among the approaches of this category, techniques that exploit *a priori* knowledge have received particular interest and have presented promising results.

Depending on the adopted knowledge acquisition and utilization process, two types of approaches can be identified in the relevant literature: explicit, realized by model-based approaches, and implicit, realized by machine learning methods. The characteristic advantage of the latter is that they have proven to be a robust methodology for discovering complex relationships and interdependencies between numerical image data and perceptually higher-level semantic concepts. Moreover, they achieve to elegantly handle problems of high dimensionality. Among the most commonly adopted machine learning techniques are Neural Networks (NNs), SOMs[7], GAs[8] and SVMs[13].

In this paper, four individual approaches to knowledge-assisted semantic image analysis are presented. All of the approaches are based on machine learning techniques via implicit knowledge acquisition, namely a SVM-, a GA-, a SOM- and a PSO-based classifier are considered. Initially, the examined image is segmented and suitable low-level descriptors are extracted for every resulting segment. Then, the estimated descriptors are provided as input to each of the aforementioned classifiers in order to associate every region with a predefined high-level semantic concept. The latter is used for denoting a real-world object that can be present in the examined image. Moreover, an appropriate evaluation framework has been developed for investigating the behavior and the corresponding performance of each algorithm under varying experimental conditions.

The remainder of the paper is organized as follows: Sections 2 and 3 describe the visual information processing and the developed classification methods, respectively. Section 4 details the formulated evaluation framework. Experimental results are presented in Section 5 and conclusions are

*The work presented in this paper was supported by the European Commission under contracts FP6-027026 K-Space, FP6-001765 aceMedia and FP6-027685 MESH.

978-0-7695-3130-4/08 $25.00 © 2008 IEEE
DOI 10.1109/WIAMIS.2008.36

drawn in Section 6.

## 2 Low-Level Visual Information Processing

In order to perform the region-concept association procedure, the examined image has to be segmented into regions and suitable low-level descriptions have to be extracted for every resulting segment. In the current implementation, an extension of the Recursive Shortest Spanning Tree (RSST) algorithm has been used for segmenting the image [1]. Output of this segmentation algorithm is a segmentation mask, where the created spatial regions $s_n$, $n = 1, ...N$, are likely to represent meaningful semantic objects. For every generated image segment, the following MPEG-7 descriptors are extracted and form a *region feature vector*: *Scalable Color*, *Homogeneous Texture*, *Region Shape*, *Edge Histogram*, *Color Structure* and *Color Layout*. The above descriptors are utilized by the classification algorithms, i.e. they constitute a common data set, for performing the region-concept assignment.

## 3 Classification Methods

### 3.1 Support Vector Machines

SVMs have been widely used in semantic image analysis tasks due to their reported generalization ability [6]. Under the proposed approach, SVMs are employed for performing the association of the computed image regions to one of the defined high-level semantic concepts based on the estimated region feature vector. An individual SVM is introduced for every defined concept $C_l$, $l = 1, ...L$, to detect the corresponding instances, and is trained under the '*one-against-all'* approach. Each SVM at the evaluation stage returns for every segment a numerical value in the range $[0, 1]$ denoting the degree of confidence, $h^C_{nl}$, to which the corresponding region is assigned to the concept associated with the particular SVM. The degree of confidence is calculated according to the following equation:

$$h^C_{nl} = \frac{1}{1 + e^{-p \cdot z_{nl}}} \quad , \tag{1}$$

where $z_{nl}$ is the distance of the input feature vector from the corresponding SVM's separating hyperplane and $p$ is a slope parameter set experimentally. For each region, $argmax(h^C_{nl})$ indicates its concept assignment. A detailed description of this procedure can be found in [9][4].

### 3.2 Genetic Algorithm

GAs have been extensively used in a wide variety of optimization problems [8], where they have been shown to outperform other traditional methods. In the present analysis framework, a GA is employed on top of the SVM-based approach of Section 3.1 for refining the classification results by treating semantic image analysis as a global optimization problem. More specifically, the GA receives as input the estimated degrees of confidence, $h^C_{nl}$, for all possible region-concept pairs (Section 3.1), spatial relations among the image segments and spatial-related contextual information. The latter is in the form of fuzzy directional relations and is obtained according to a simple learning process [9]. The GA is provided with an appropriate fitness function for denoting the plausibility of every possible image semantic interpretation, which is represented with a particular chromosome, and has the form:

$$f(V) \;=\; \lambda \times FS_{norm} \;+\; (1-\lambda) \times SC_{norm} \;, \tag{2}$$

where $V$ denotes a particular chromosome, $FS_{norm}$ refers to the degree of visual features similarity and $SC_{norm}$ stands for the degree of spatial relations consistency. Output of this procedure is a final region-concept association, which corresponds to the solution with the highest fitness value. Parameter $\lambda$ is introduced to adjust the weight of $FS_{norm}$ and $SC_{norm}$ on the final outcome and its value is estimated according to a separate optimization procedure. A detailed description of the GA implementation can be found in [9].

### 3.3 Self Organizing Maps

Neural network based clustering and classification has been dominated by SOMs [7] and adaptive resonance theory (ART) [12]. The objective of SOM is to represent high-dimensional input patterns with prototype vectors that can be visualized in a usually two-dimensional lattice structure. Input patterns are fully connected to all neurons via adaptable weights and during the training process, neighboring input patterns are projected into the lattice, corresponding to adjacent neurons. Similar to the SVM classifier presented in Section 3.1, an individual SOM network is employed to detect instances of the defined high-level semantic concepts. Each SOM is trained under the one against all approach. In the basic training algorithm, the prototype vectors are trained according to the following equation:

$$m_d(t+1) \;=\; m_d(t) + g_{cd}(t)[x - m_d(t)] \;, \tag{3}$$

where $m_d$ is the weight of the neurons in the SOM network, $g_{cd}(t)$ is the neighborhood function and $d$ is the dimension of the input feature vector. Outcome of the trained set of SOM classifiers is a hypothesis set for region-concept association similar to the one of Section 3.1.

### 3.4 Particle Swarm Optimization

To further improve the performance of the SOM classifier, the weight of the neurons $m_d$ in Eq. 3 is optimized with Particle Swarm Optimization. The PSO algorithm is one of the evolutionary computation techniques [3]. It was originally inspired by the social behavior of a flock of birds [10]. In the PSO algorithm, the birds in a flock are considered to be "flying" through a problem space searching for a solution. The solution obtained by the particles is evaluated by a fitness function that provides a quantitative value of the solution's utility. The PSO consists of at each time step changing the velocity (accelerating) of each particle toward its personal best ($pbest$) and global best ($gbest$). The velocity and position of the particles are governed by Eq. 4 and 5.

$$v_{id} = v_{id} + c_1(pbest_i - x_{id}) + c_2(gbest_d - x_{id}) \quad (4)$$

$$x_{id}(t) = x_{id}(t-1) + v_{id}(t-1) \;, \quad (5)$$

where $v_{id}$ and $x_{id}$ represent the velocity and position of individual particles $i$ in each dimension $d$, respectively. More specifically, PSO receives the weight ($m_d$) from Eq. 3 of the winner node neuron from SOM network as input and the optimized value of $m_d$ is reassigned to the winner node of SOM network. A detailed description of the PSO implementation can be found in [2].

## 4 Evaluation Framework

In this section, the framework that was developed and used for evaluating the performance of the aforementioned classification methods is described in detail. Aim of the introduction of this framework is the in depth investigation of the behavior and the resulting performance of the developed algorithms through extensive experimentations under varying experimental conditions.

Initially, regarding the content used for experimentation, a set of $500$ images, $\mathcal{Q}$, belonging to the general category of *vacation images* was assembled. The content was mainly obtained from the Flickr online photo management and sharing application [5] and includes images that depict cityscape, seaside, mountain and landscape locations. Then, the following set of 15 concepts, $C$, which represent meaningful real-world objects that can be present in images of the formed set, was defined: *Sand*, *Sea*, *Vegetation*, *Person*, *Sky*, *Rock*, *Tree*, *Grass*, *Ground*, *Trunk*, *Wave*, *Boat*, *Dried-Plant*, *Building* and *Pavement*. The concepts are listed in descending order with respect to their frequency of appearance in the images of $\mathcal{Q}$. Every image was manually annotated, i.e. after the segmentation algorithm described in Section 2 is applied, a single concept was associated with each resulting image segment.

The aforementioned image set $\mathcal{Q}$ was divided into two sub-sets, namely $\mathcal{Q}_{tr}$ and $\mathcal{Q}_{te}$. The first one, $\mathcal{Q}_{tr}$, was used for training the classification algorithms described in Section 3, while the second, $\mathcal{Q}_{te}$, was used for evaluating their performance. In order to effectively examine the generalization ability and the algorithms' behavior when variable amount of data is available for training purposes, the number of images forming the training set, $\mathcal{Q}_{tr}$, was set to $10\%$, $30\%$ and $50\%$ of the total images of $\mathcal{Q}$ in the various experiments, while the rest of the images were used for evaluation. The images forming in each case sub-set $\mathcal{Q}_{tr}$ were randomly chosen.

In order to investigate the classification performance of the developed methods under varying problem complexity, the supported concepts were divided into three subsets $R_r$, $r = \{1, 2, 3\}$, of variable size. Each sub-set $R_r$ is made of 5, 10 and 15 concepts, respectively. The concepts comprising each subset were selected according to the following procedure: Initially, the frequency of appearance, i.e. the percentage of the images where a particular concept is present, was calculated for every defined concept. Then, the concept set $R_1$ was formed from the $5$ concepts with the highest frequency of appearance, $R_2$ was formed from the $10$ concepts with the highest frequency, etc. For every concept set, $R_r$, the procedure followed for estimating the classification performance of every method is repeated with respect only to the concepts that belong to that set, i.e. the concepts that are not included in $R_r$ for each experiment are ignored during both training and evaluation.

## 5 Experimental Results

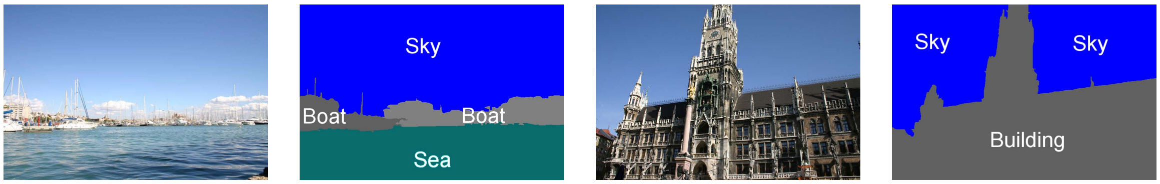
In this section, experimental results regarding the evaluation process of the developed classifiers are presented. Initially, an appropriate evaluation framework has been formed, as described in Section 4. Each image belonging to the formulated set $\mathcal{Q}$ was segmented and low-level descriptors were estimated for every image region (Section 2). Then, each of the classification algorithms described in Section 3 was applied to perform the association of every image region of the test set with one of the predefined semantic concepts.

In Fig. 1, indicative region-concept association results are presented, showing the input image and the resulting annotation. In Table 1, quantitative performance measures are given in terms of accuracy for all possible combinations of percentage of images used for training and number of supported concepts. Accuracy is defined as the percentage of the image regions that were assigned to the correct semantic concept.

From the careful observation of the above results, it can be seen that the combined use of an optimization method (GA, PSO) with a more traditional classifier (SVM and

**Table 1. Concept detection accuracy**

| Training set proportion(%) | 5 supported concepts | | | | 10 supported concepts | | | | 15 supported concepts | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 10 | 52.06% | 51.28% | 33.73% | **57.77%** | 45.18% | 29.10% | 21.98% | **49.26%** | 41.87% | 37.88% | 9.87% | **43.80%** |
| 30 | **64.60%** | 63.77% | 27.21% | 62.99% | 56.49% | **56.99%** | 21.15% | 52.53% | 53.76% | **53.85%** | 20.69% | 47.85% |
| 50 | **70.67%** | 69.58% | 25.45% | 63.72% | 61.88% | **62.37%** | 19.76% | 50.25% | 59.05% | **59.91%** | 20.87% | 47.03% |
| Method | SVM | GA | SOM | PSO | SVM | GA | SOM | PSO | SVM | GA | SOM | PSO |




(a) (b) (c) (d)

**Figure 1. Indicative region-concept association using the GA (a) & (b) and PSO (c) & (d) approaches**

SOM, respectively) generally leads to increased classification accuracy as compared to using the latter classifiers alone. Furthermore, the use of an increased number of images for training the classifiers is generally beneficial, highlighting the need for the availability of large annotated media sets for appropriately training any classification method. However, even in the absence of a rich training set, meaningful classification results can be produced, as indicated by the presented experiments; the PSO classification scheme is shown to be particularly suitable in this case, while when more samples are available for training purposes the GA classifier is advantageous. Finally, the experimentation with a varying number of concepts indicated that the performance of the employed methods degrades gracefully when the number of concepts increases.

## 6. Conclusions

In this paper, four individual approaches to region classification for knowledge-assisted semantic image analysis are presented and comparatively evaluated. An appropriate evaluation framework has been developed for investigating the behavior and the corresponding performance of each algorithm under varying experimental conditions. Future work includes the examination of additional methodologies for knowledge acquisition that will facilitate image analysis tasks.

## References

[1] T. Adamek, N. O'Connor, and N. Murphy. Region-based segmentation of images using syntactic visual features. *Workshop on Image Analysis for Multimedia Interactive Services,(WIAMIS), Montreux, Switzerland*, 2005.

[2] K. Chandramouli and E. Izquierdo. Image classification using self organising feature maps and particle swarm optimisation. In *Proc. 7th Int'lWorkshop on Image Analysis for Multimedia Interactive Services (WIAMIS'06)*, pages 313–316, 2006.

[3] R. Eberhart and Y. Shi. Tracking and optimizing dynamic systems with particle swarms. *Evolutionary Computation, 2001. Proceedings of the 2001 Congress on*, 1, 2001.

[4] http://mklab.iti.gr/svmcf.

[5] http://www.flickr.com/.

[6] K. I. Kim, K. Jung, S. H. Park, and H. J. Kim. Support vector machines for texture classification. *IEEE Trans. Pattern Anal. Mach. Intell.*, 24(11):1542–1550, 2002.

[7] T. Kohonen. The self organizing map. *Proceedings of IEEE*, 78(4):1464–1480, September 1990.

[8] M. Mitchell. *An Introduction to Genetic Algorithms*. MIT Press, 1996.

[9] G. T. Papadopoulos, V. Mezaris, I. Kompatsiaris, and M. G. Strintzis. Combining Global and Local Information for Knowledge-Assisted Image Analysis and Classification. *EURASIP Journal on Advances in Signal Processing*, 2007:1–15, 2007.

[10] C. Reynolds. Flocks, herds and schools: a distributed behavioural model. In *Computer Graphics*, pages 25–34, 1987.

[11] A. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain. Content-based image retrieval at the end of the early years. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 22(12):1349–1380, 2000.

[12] R. Xu and D. Wunsch. Survey of clustering algorithms. *IEEE Trans. Neural Network*, 6(3):645–678, May 2005.

[13] L. Zhang, F. Lin, and B. Zhang. Support vector machine learning for image retrieval. *Image Processing, 2001. Proceedings. 2001 International Conference on*, 2, 2001.