

## Research Article

# Combining Global and Local Information for Knowledge-Assisted Image Analysis and Classification

G. Th. Papadopoulos,<sup>1,2</sup> V. Mezaris,<sup>2</sup> I. Kompatsiaris,<sup>2</sup> and M. G. Strintzis<sup>1,2</sup>

<sup>1</sup>Department of Electrical and Computer Engineering, Aristotle University of Thessaloniki, Thessaloniki 54006, Greece

<sup>2</sup>Centre for Research and Technology Hellas (CERTH), Informatics and Telematics Institute, Thessaloniki 57001, Greece

Received 8 September 2006; Revised 23 February 2007; Accepted 2 April 2007

Recommended by Ebroul Izquierdo

A learning approach to knowledge-assisted image analysis and classification is proposed that combines global and local information with explicitly defined knowledge in the form of an ontology. The ontology specifies the domain of interest, its subdomains, the concepts related to each subdomain as well as contextual information. Support vector machines (SVMs) are employed in order to provide image classification to the ontology subdomains based on global image descriptions. In parallel, a segmentation algorithm is applied to segment the image into regions and SVMs are again employed, this time for performing an initial mapping between region low-level visual features and the concepts in the ontology. Then, a *decision function*, that receives as input the computed region-concept associations together with contextual information in the form of concept frequency of appearance, realizes image classification based on local information. A fusion mechanism subsequently combines the intermediate classification results, provided by the local- and global-level information processing, to decide on the final image classification. Once the image subdomain is selected, final region-concept association is performed using again SVMs and a genetic algorithm (GA) for optimizing the mapping between the image regions and the selected subdomain concepts taking into account contextual information in the form of spatial relations. Application of the proposed approach to images of the selected domain results in their classification (i.e., their assignment to one of the defined subdomains) and the generation of a fine granularity semantic representation of them (i.e., a segmentation map with semantic concepts attached to each segment). Experiments with images from the personal collection domain, as well as comparative evaluation with other approaches of the literature, demonstrate the performance of the proposed approach.

Copyright © 2007 G. Th. Papadopoulos et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

## 1. INTRODUCTION

Recent advances in both hardware and software technologies have resulted in an enormous increase of the number of images that are available in multimedia databases or over the internet. As a consequence, the need for techniques and tools supporting their effective and efficient manipulation has emerged. To this end, several approaches have been proposed in the literature regarding the tasks of indexing, searching, classification, and retrieval of images [1, 2].

The very first attempts to address these issues concentrated on visual similarity assessment via the definition of appropriate quantitative image descriptions, which could be automatically extracted, and suitable metrics in the resulting feature space [1]. Whilst low-level descriptors and metrics are fundamental building blocks of any image manipulation technique, they evidently fail to fully capture by themselves the semantics of the visual medium. Achieving the lat-

ter is a prerequisite for reaching the desired level of efficiency in image manipulation tasks. To this end, research efforts have concentrated on the semantic analysis and classification of images, often combining the aforementioned techniques with *a priori* domain specific knowledge, so as to result in a high-level representation of them [2]. Domain specific knowledge, when utilized, guides low-level feature extraction, higher-level descriptor derivation, and symbolic inference.

Image classification is an important component of semantic image manipulation attempts. Several approaches have been proposed in the relevant literature regarding the task of the categorization of images in a number of predefined classes. In [3], SVMs are utilized for discriminating between indoor/outdoor images, while a graph decomposition technique and probabilistic neural networks (PNN) are adopted for the task of supervised image classification in [4]. In [5], multicategory image classification is realized

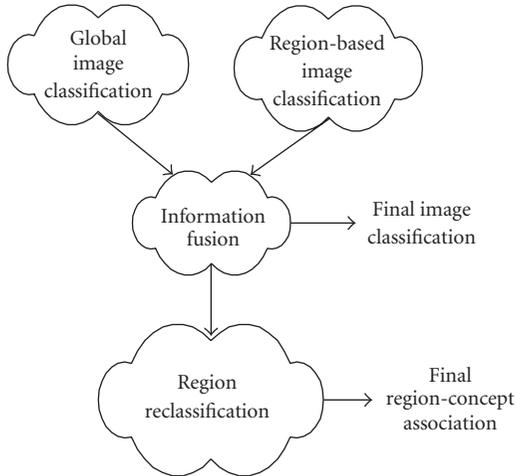


FIGURE 1: General system architecture.

based on an employed parametric mixture model (PMM), which is adopted from the corresponding multicategory text-classification task, and the exploitation of the image color histogram. In [6], classification of images is performed on the basis of maximum cross correlation estimations and retrieval of images from an existing database against a given query image.

The aforementioned methods are based on global visual descriptions that are automatically extracted for every image. However, image manipulation based solely on global descriptors does not always lead to the best results [7]. Coming one step closer to treating images the way humans do, image analysis tasks (including classification) shifted to treating images at a finer level of granularity, that is, at the region or local level, taking advantage of image segmentation techniques. More specifically, in [8], an image classification method is proposed, which uses a set of computed multiple-level association rules based on the detected image objects. In [9], it is demonstrated through several applications how segmentation and object-based methods improve on pixel-based image analysis/classification methods, while in [10], a region-based binary tree representation incorporating with adaptive processing of data structures is proposed to address the problem of image classification.

Incorporating knowledge into classification techniques emerges as a promising approach for improving classification efficiency. Such an approach provides a coherent semantic domain model to support “visual” inference in the specified context [11, 12]. In [13], a framework for learning intermediate-level visual descriptors of objects organized in an ontology is presented to support the detection of them. In [14], a priori knowledge representation models are used as a knowledge base that assists semantic-based classification and clustering. Moreover, in [15], semantic entities, in the context of the MPEG-7 standard, are used for knowledge assisted multimedia analysis and object detection, thus allowing for semantic level indexing.

In this paper, a learning approach to knowledge-assisted image analysis and classification is proposed that combines global and local information with explicitly defined knowledge in the form of an ontology. The ontology specifies the domain of interest, its subdomains, the concepts related to each subdomain as well as contextual information. SVMs are employed in order to provide image classification to the ontology subdomains based on global image descriptions. In parallel, a segmentation algorithm is applied to segment the image into regions and SVMs are again employed, this time for performing an initial mapping between region low-level visual features and the concepts in the ontology. Then, a *decision function*, that receives as input the computed region to concepts associations together with contextual information in the form of frequency of appearance of each concept, realizes image classification based on local information. A fusion mechanism combines the intermediate classification results, provided by the local- and global-level information processing, and decides on the final classification. Once the image subdomain is selected, final region-concept association is performed using again SVMs and a genetic algorithm (GA) for optimizing the mapping between the image regions and the selected subdomain concepts taking into account contextual information in the form of spatial relations. The values of the parameters used in the final image classification and final region-concept association processes are computed according to a parameter optimization procedure. The general architecture of the proposed system for semantic image analysis and classification is illustrated in Figure 1. Application of the proposed approach to images of the selected domain results in their classification (i.e., their assignment to one of the defined subdomains) and the generation of a fine granularity semantic representation of them (i.e., a segmentation map with semantic concepts attached to each segment). Experiments with images from the personal collection domain, as well as comparative evaluation with other approaches of the literature, demonstrate the performance of the proposed approach.

As will be seen by the experimental evaluation of the proposed approach, the elegant combination of global and local information as well as contextual and ontology information leads to improved image classification performance, as compared to classification based solely on either global or local information. Furthermore, this image to subdomain association is used to further improve the accuracy of region to concept association, as compared to region-concept association performed without using knowledge about the former.

The paper is organized as follows: Section 2 presents the overall system architecture. Sections 3 and 4 describe the low-level information extraction and the employed high-level knowledge, respectively. Section 5 details the image classification process and Section 6 presents the region-concept association procedure. Section 7 describes the methodology followed for the optimization of the proposed system parameters. Experimental results and comparisons are presented in Section 8 and conclusions are drawn in Section 9.

## 2. SYSTEM OVERVIEW

The first step in the development of the proposed knowledge-assisted image analysis and classification architecture is the definition of an appropriate knowledge infrastructure. This is defined in the form of an ontology suitable for describing the semantics of the selected domain. The proposed ontology comprises of a set of subdomains, to which images of the domain can be classified, and a set of concepts, each associated with at least one of the aforementioned subdomains. The latter represent objects of interest that may be depicted in the images. In addition to the above, the proposed ontology also defines contextual information in the form of the frequency of appearance of each concept in the images of each subdomain, as well as in the form of spatial relations between the defined concepts. The defined ontology is discussed in Section 4 and the subdomains and concepts it includes are shown in Figure 4.

At the signal level, low-level global image descriptors are extracted for every image and form an *image feature vector*. This is utilized for performing image classification to one of the defined subdomains based on global-level descriptions. More specifically, the computed vector is supplied as input to a set of SVMs, each trained to detect images that belong to a certain subdomain. Every SVM returns a numerical value which denotes the degree of confidence to which the corresponding image is assigned to the subdomain associated with the particular SVM; the maximum of the degrees of confidence over all subdomains indicates the image classification using global-level information.

In parallel to this process, a segmentation algorithm is applied to the image in order to divide it into regions, which are likely to represent meaningful semantic objects. Then, for every resulting segment, low-level descriptions and spatial relations are estimated, the latter according to the relations supported by the ontology. The estimated low-level descriptions for each region are employed for generating initial hypotheses regarding the region's association to an ontology concept. This is realized by evaluating the respective low-level *region feature vector* and using a second set of SVMs, where each SVM is trained to identify instances of a single concept defined in the ontology. SVMs were selected for the aforementioned tasks due to their reported generalization ability and their efficiency in solving high-dimensionality pattern recognition problems [16, 17]. Subsequently, a *decision function*, that receives as input the computed region to concept association hypothesis sets together with the ontology-provided contextual information in the form of frequency of concept appearance, realizes image classification based on local-level information. The domain ontology drives this process by controlling which concepts are associated with a specific subdomain.

The computed hypothesis sets for the image-subdomain association based on both global- and local-level information are subsequently introduced to a fusion mechanism, which combines the supplied intermediate global- and local-based classification information and decides on the final image classification. Fusion is introduced since, depending on

the nature of the examined subdomain, global-level descriptions may represent more efficiently the semantics of the image or local-level information may be advantageous. Thus, the fusion mechanism is used for adjusting the weight of the global features against the local ones for every individual subdomain to reach a final image classification decision.

After the image subdomain is selected, generation of refined region-concept association hypotheses is performed. The procedure is similar to the one described at the previous stage, the difference being that at this stage only the SVMs that correspond to concepts of the estimated subdomain are employed and thus subdomain-specific hypothesis sets are computed. The refined hypothesis sets for every image region along with the spatial relations computed for each region, are subsequently employed for estimating a globally optimal region-concept assignment by introducing them to a genetic algorithm. The GA is employed in order to decide upon the most plausible image interpretation and compute the final region semantic annotation. The choice of a GA for these tasks is based on its extensive use in a wide variety of global optimization problems [18], where they have been shown to outperform other traditional methods, and is further endorsed by the authors' previous experience [19, 20], which showed promising results. The values of the proposed system parameters used in the aforementioned final image classification and final region-concept association processes are computed according to a parameter optimization procedure. The detailed architecture of the proposed system for semantic image analysis and classification is illustrated in Figure 2.

Regarding the tasks of SVMs training, computation of the required contextual information, parameter optimization and evaluation of the proposed system performance, a number of image sets needs to be formed. More specifically, a collection of images,  $\mathcal{B}$ , belonging to the domain of interest was assembled. Each image in this collection was manually annotated (i.e., assigned to a subdomain and, after segmentation is applied, each of the resulting image regions associated with a concept in the ontology). The collection was initially divided into two sets:  $\mathcal{B}_{tr}$ , which is made of approximately 30% of the images of  $\mathcal{B}$ , and  $\mathcal{B}_{te}$ , which comprises the remaining 70%.  $\mathcal{B}_{tr}$  is used for training the SVMs framework and computing the required contextual information. On the other hand,  $\mathcal{B}_{te}$  is used for evaluating the proposed system performance. For the case of the parameter optimization procedure,  $\mathcal{B}_{tr}$  is equally divided into two subsets, namely  $\mathcal{B}_{tr}^1$  and  $\mathcal{B}_{tr}^2$ .  $\mathcal{B}_{tr}^1$  is again used for training the SVMs framework and computing the required contextual information, while  $\mathcal{B}_{tr}^2$  serves in estimating the optimal values of the aforementioned parameters. The usage and the notation of all image sets utilized in this work are illustrated in Table 1. The main symbols used in the remainder of the manuscript are outlined in Table 2.

## 3. LOW-LEVEL VISUAL INFORMATION PROCESSING

### 3.1. Global features extraction

The image classification procedure based on global-level features, as will be described in detail in the sequel, requires that

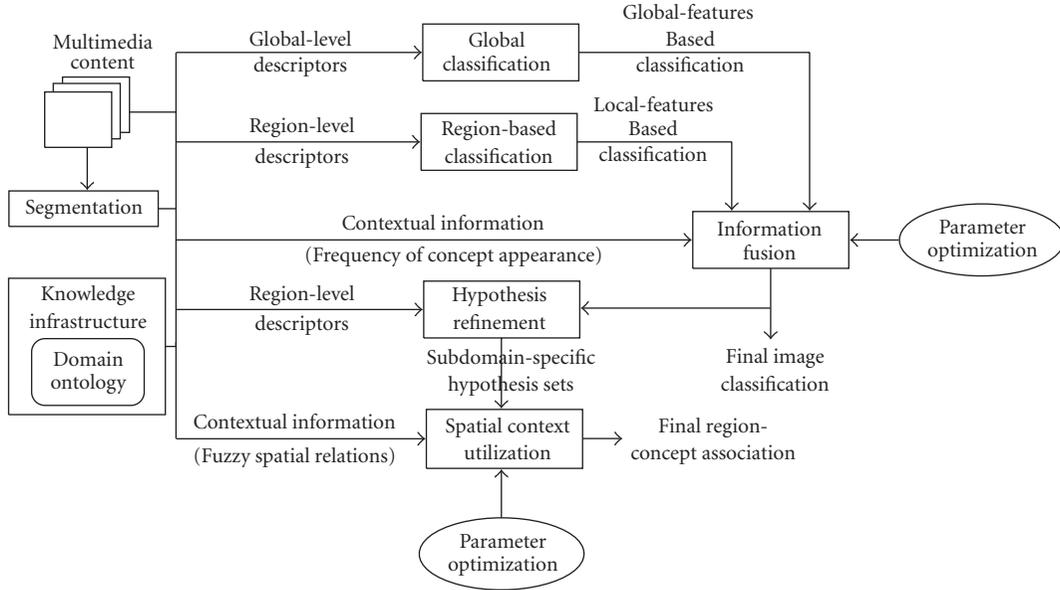


FIGURE 2: Detailed system architecture.

TABLE 1: Table of training and test sets.

$\mathcal{B}$	Entire image set used for training and evaluation.
$\mathcal{B}_{tr}$	Subset of $\mathcal{B}$ , used for training the SVMs and computing contextual information. Subdivided to $\mathcal{B}_{tr}^2$ and $\mathcal{B}_v^2$ .
$\mathcal{B}_{te}$	Subset of $\mathcal{B}$ , used for evaluation.
$\mathcal{B}_{tr}^2$	Subset of $\mathcal{B}_{tr}$ , used for training the SVMs and computing contextual information during the parameter optimization procedure.
$\mathcal{B}_v^2$	Subset of $\mathcal{B}_{tr}$ , used for estimating the parameter values during parameter optimization.

appropriate low-level descriptions are extracted at the image level for every examined image and form an *image feature vector*. The image feature vector employed in this work comprises of three different descriptors of the MPEG-7 standard, namely the *Scalable Color*, *Homogeneous Texture*, and *Edge Histogram* descriptors. Their extraction is performed according to the guidelines provided by the MPEG-7 experimentation model (XM) [21]. Following their extraction, the image feature vector is produced by stacking all extracted MPEG-7 descriptors in a single vector. This vector constitutes the input to the SVMs structure which realizes the global image classification, as described in Section 5.1.

### 3.2. Segmentation and local features extraction

In order to implement the initial hypothesis generation 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

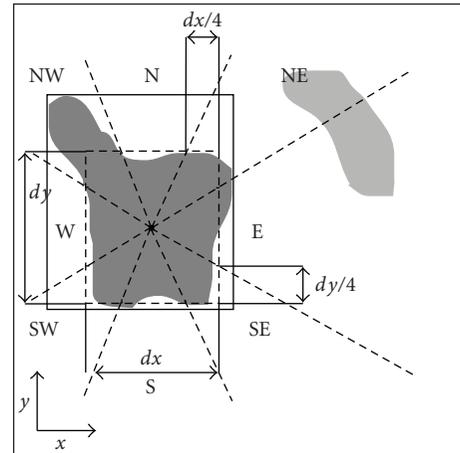


FIGURE 3: Fuzzy directional relations definition.

extension of the recursive shortest spanning tree (RSST) algorithm has been used for segmenting the image [22]. Output of this segmentation algorithm is a segmentation mask  $S$ ,  $S = \{s_i, i = 1, \dots, N\}$ , where  $s_i, i = 1, \dots, N$ , are the created spatial regions.

For every generated image segment, the following MPEG-7 descriptors are extracted, according to the guidelines provided by the MPEG-7 experimentation model (XM) [21]: *Scalable Color*, *Homogeneous Texture*, *Region Shape*, and *Edge Histogram*. The above descriptors are then combined to form a single *region feature vector*. This vector constitutes the input to the SVMs structure which computes the initial hypothesis sets for every region, as described in Section 5.2.

TABLE 2: Legend of main symbols.

Symbol	Description
$s_i, S = \{s_i, i = 1, \dots, N\}$	Image regions after segmentation, set of regions for an image
$c_j, C = \{c_j, j = 1, \dots, J\}$	Concept defined in the ontology, the set of all concepts
$D_l, l = 1, \dots, L$	Subdomains defined in the ontology
$r_k, R = \{r_k, k = 1, \dots, K\}$	Spatial relation, set of all spatial relations defined in the ontology
$H^D = \{h_l^D, l = 1, \dots, L\}$	Hypothesis set for global image classification
$H_i^C = \{h_{ij}^C, j = 1, \dots, J\}$	Hypothesis set for region-concept association, for region $s_i$
$g(D_l)$	Result of local-based image classification for subdomain $D_l$
$G(D_l)$	Result of final image classification for subdomain $D_l$
$\text{freq}(c_j, D_l)$	Frequency of appearance of concept $c_j$ with respect to subdomain $D_l$
$g_{ij}$	Assignment of concept $c_j$ to region $s_i$
$I_M(g_{ij})$	Degree of confidence, based on visual similarity, for $g_{ij}$ assignment
$Q$	Genetic algorithm's chromosome
$f(Q)$	Genetic algorithm's fitness function
$\text{area}(s_i)$	Area of region $s_i$
$\nu$	Region compactness value
$I_r(s_i, s_j)$	Degree to which relation $r_k$ is satisfied for the $(s_i, s_j)$ pair of regions
$I_S(g_{ij}, g_{pq})$	Degree to which the spatial constraint between the $g_{ij}, g_{pq}$ concept to region mappings is satisfied

### 3.3. Fuzzy spatial relations extraction

Exploiting domain-specific spatial knowledge in image analysis constitutes an elegant way for removing ambiguities in region-concept associations. More specifically, it is generally observed that objects tend to be present in a scene within a particular spatial context and thus spatial information can substantially assist in discriminating between concepts exhibiting similar visual characteristics. Among the most commonly adopted spatial relations, directional ones have received particular interest. They are used to denote the order of objects in space. In the present analysis framework, eight fuzzy directional relations are supported, namely *North* (N), *East* (E), *South* (S), *West* (W), *South-East* (SE), *South-West* (SW), *North-East* (NE), and *North-West* (NW). These relations are utilized for computing part of the contextual information stored in the ontology, as described in detail in Section 4, and further used for the final region-concept association of Section 6.

Fuzzy directional relations extraction in the proposed analysis approach builds on the principles of projection- and angle-based methodologies [23, 24] and consists of the following steps. First, a *reduced box* is computed from the *ground* region's (the region used as reference and is painted in dark grey in Figure 3) minimum bounding rectangle (MBR), so as to include the region in a more representative way. The computation of this *reduced box* is performed in terms of the MBR compactness value  $\nu$ , which is defined as the fraction of the region's area to the area of the respective MBR: if the initially computed  $\nu$  is below a threshold  $T$ , the ground region's MBR is reduced repeatedly until the desired threshold is satisfied. Then, eight cone-shaped regions are formed on top of this reduced box, as illustrated in Figure 3, each corresponding to one of the defined directional relations. The

percentage of the *figure* region (whose relative position is to be estimated and is painted in light grey in Figure 3) points that are included in each of the cone-shaped regions determines the degree to which the corresponding directional relation is satisfied. After extensive experimentations, the value of threshold  $T$  was set equal to 0.85.

## 4. KNOWLEDGE INFRASTRUCTURE

Among the possible domain knowledge representations, ontologies [25] present a number of advantages, the most important being that they provide a formal framework for supporting explicit, machine-processable semantics definition and they enable the derivation of new knowledge through automated inference. Thus, ontologies are suitable for expressing multimedia content semantics so that automatic semantic analysis and further processing of the extracted semantic descriptions are allowed [12]. Following these considerations, an ontology was developed for representing the knowledge components that need to be explicitly defined under the proposed approach. More specifically, the images of concern belong to the personal collection domain. Consequently, in the developed ontology, a number of subdomains, related to the broader domain of interest, are defined (such as *Buildings*, *Rockyside*, etc.), denoted by  $D_l, l = 1, \dots, L$ . For every subdomain, the particular semantic concepts of interest are also defined in the domain ontology (e.g., in the seaside subdomain the defined concepts include *Sea*, *Sand*, *Person*, etc.), denoted by  $c_j, C = \{c_j, j = 1, \dots, J\}$  being the set of all concepts defined in the ontology. Contextual information in the form of spatial relations between the concepts, as well as contextual information in the form of frequency of appearance of each concept in every subdomain, are also included. The subdomains and concepts of the ontology employed in

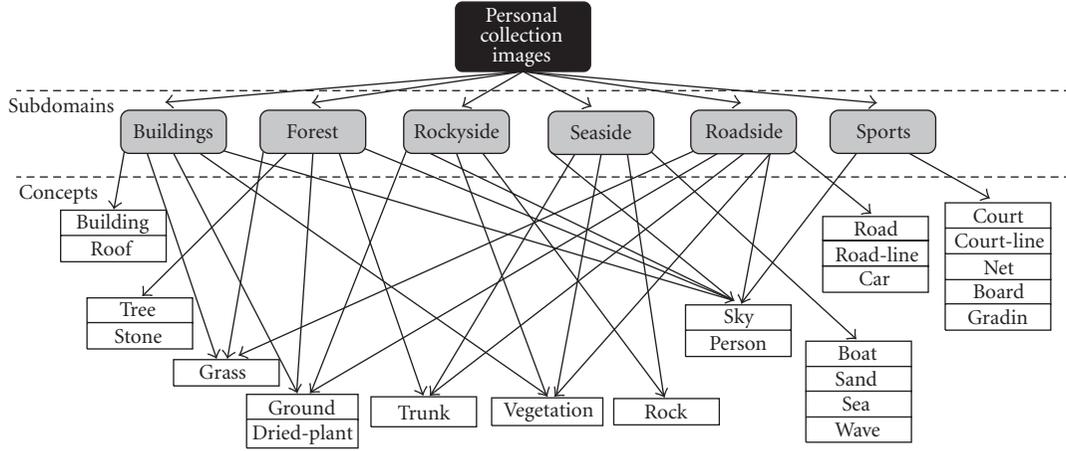


FIGURE 4: Subdomains and concepts of the ontology developed for the personal collection domain.

this work are presented in Figure 4, where can be seen that the developed ontology includes 6 subdomains and 24 individual concepts. It must be noted that the employed ontology can easily be extended so as to include additional concepts and subdomains, as well as any additional information that could be used for the analysis.

The values of the spatial relations (spatial-related contextual information) between the concepts for every particular subdomain, as opposed to the concepts themselves that are manually defined, are estimated according to the following ontology population procedure.

Let  $R$ ,

$$R = \{r_k, k = 1, \dots, K\} = \{N, NW, NE, S, SW, SE, W, E\}, \quad (1)$$

denote the set of the supported spatial relations. Then, the degree to which region  $s_i$  satisfies relation  $r_k$  with respect to region  $s_j$  can be denoted as  $I_{r_k}(s_i, s_j)$ . The values of function  $I_{r_k}$ , for a specific couple of regions, are estimated according to the procedure of Section 3.3 and belong to  $[0, 1]$ . To populate the ontology, this function needs to be evaluated over a set of segmented images with ground truth classification and annotations, that serves as a training set. For that purpose, the subset  $\mathcal{B}_{tr}$  is employed as discussed in Section 2. Then, using this training set the ontology population procedure is performed by estimating the mean values,  $I_{r_k, \text{mean}}$ , of  $I_{r_k}$  for every  $k$  over all pairs of regions assigned to concepts  $(c_i, c_j)$ ,  $i \neq j$ , and storing them in the ontology. These constitute the constraints input to the optimization problem which is solved by the genetic algorithm, as will be described in Section 6.

Regarding the contextual information in the form of frequency of appearance, the reported frequency of each concept  $c_j$  with respect to the subdomain  $D_l$ ,  $\text{freq}(c_j, D_l)$ , is defined as the fraction of the number of appearances of concept  $c_j$  in images of the training set that belong to subdomain  $D_l$  to the total number of the images of the afore-mentioned training set that belong to subdomain  $D_l$ .

## 5. IMAGE CLASSIFICATION AND INITIAL REGION-CONCEPT ASSOCIATION

### 5.1. Image classification using global features

In order to perform the classification of the examined images to one of the subdomains defined in the ontology using global image descriptions, a compound *image feature vector* is initially formed, as described in Section 3.1. Then, an SVMs structure is utilized to compute the class to which every image belongs. This comprises  $L$  SVMs, one for every defined subdomain  $D_l$ , each trained under the “one-against-all” approach. For the purpose of training the SVMs, the subdomain membership of the images belonging to the training set  $\mathcal{B}_{tr}$ , assembled in Section 2, is employed. The image feature vector discussed in Section 3.1 constitutes the input to each SVM, which at the evaluation stage returns for every image of unknown subdomain membership a numerical value in the range  $[0, 1]$ . This value denotes the degree of confidence to which the corresponding image is assigned to the subdomain associated with the particular SVM. The metric adopted is defined as follows: for every input feature vector the distance  $z_l$  from the corresponding SVM’s separating hyperplane is initially calculated. This distance is positive in case of correct classification and negative otherwise. Then, a sigmoid function [26] is employed to compute the respective degree of confidence,  $h_l^D$ , as follows:

$$h_l^D = \frac{1}{1 + e^{-t \cdot z_l}}, \quad (2)$$

where the slope parameter  $t$  is experimentally set. For each image, the maximum of the  $L$  calculated degrees of membership indicates its classification based on global-level features, whereas all degrees of confidence,  $h_l^D$ , constitute its subdomain hypotheses set  $H^D$ , where  $H^D = \{h_l^D, l = 1, \dots, L\}$ . The SVM structure employed for image classification based on global features, as well as for the region-concept association tasks described in the following sections, was realized using the SVM software libraries of [27].

### 5.2. Image classification using local features and initial region-concept association

As already described in Section 2, the SVMs structure used in the previous section for global image classification is also utilized to compute an initial region-concept association for every image segment. Similarly to the global case, at this finer level of granularity an individual SVM is introduced for every concept  $c_j$  of the employed ontology, in order to detect the corresponding association. Each SVM is again trained under the “one-against-all” approach. For that purpose, the training set  $\mathcal{B}_{tr}$ , assembled in Section 2, is again employed and the *region feature vector*, as defined in Section 3.2, constitutes the input to each SVM. For the purpose of initial region-concept association, every SVM again returns a numerical value in the range  $[0, 1]$ , which in this case denotes the degree of confidence to which the corresponding region is assigned to the concept associated with the particular SVM. The metric adopted for expressing the aforementioned degree of confidence is similar to the one adopted for the global image classification case, defined in the previous section. Specifically, let  $h_{ij}^C = I_M(g_{ij})$  denote the degree to which the visual descriptors extracted for region  $s_i$  match the ones of concept  $c_j$ , where  $g_{ij}$  represents the particular assignment of  $c_j$  to  $s_i$ . Then,  $I_M(g_{ij})$  is defined as

$$I_M(g_{ij}) = \frac{1}{1 + e^{-t \cdot z_{ij}}}, \quad (3)$$

where  $z_{ij}$  is the distance from the corresponding SVM’s separating hyperplane for the input feature vector used for evaluating the  $g_{ij}$  assignment. The pairs of all supported concepts and their respective degree of confidence  $h_{ij}^C$  computed for segment  $s_i$  comprise the region’s concept hypothesis set  $H_i^C$ , where  $H_i^C = \{h_{ij}^C, j = 1, \dots, J\}$ .

The estimated concept hypotheses sets,  $H_i^C$ , generated for every image region  $s_i$ , can provide valuable cues for performing image classification based on local-level information. To this end, a decision function for estimating the subdomain membership of the examined image on the basis of the concept hypotheses sets of its constituent regions and the ontology provided contextual information in the form of frequency of concept appearance (i.e., effecting image classification based on local-level information) is defined as follows:

$$g(D_l) = \sum_{s_i, \text{ where } c_j \in D_l} I_M(g_{ij}) \cdot E(s_i, c_j, a_l, D_l) \quad (4)$$

$$E(s_i, c_j, a_l, D_l) = a_l \cdot \text{freq}(c_j, D_l) + (1 - a_l) \cdot \text{area}(s_i),$$

where  $\text{freq}(c_j, D_l)$  is the concept frequency of appearance defined in Section 4 and  $\text{area}(s_i)$  is the percentage of the total image area captured by region  $s_i$ . Parameters  $a_l$ , where  $a_l \in [0, 1]$ , are introduced for adjusting the importance of the aforementioned frequencies against the regions’ areas for every supported subdomain. Their values are estimated according to the parameter optimization procedure described in Section 7.1. As can be seen in (4), the constructed domain ontology drives the estimation of the respective subdomain membership of the image by controlling which concepts are

associated with a specific subdomain and thus can contribute to the summation of (4). The latter is essentially a weighted summation of region-concept association degrees of confidence, the weights being controlled by both contextual information (concept frequency of appearance) as well as region visual importance, here approximated by the relative region area.

### 5.3. Information fusion for image classification

After image classification has been performed using solely global and solely local information, respectively, a fusion mechanism is employed for deciding upon the final image classification. Fusion is introduced since, depending on the nature of the examined subdomain, global-level descriptions may represent more efficiently the semantics of the image or local-level information may be advantageous. Thus, adjusting the weights of both image classification results leads to more accurate final classification decisions. More specifically, the computed hypothesis sets for the image-subdomain association based on both global- ( $h_l^D$ ) and local- ( $g(D_l)$ ) level information are introduced to a mechanism which has the form of a weighted summation, based on the following equation:

$$G(D_l) = \mu_l \cdot g(D_l) + (1 - \mu_l) \cdot h_l^D, \quad (5)$$

where  $\mu_l$ ,  $l = 1, \dots, L$  and  $\mu_l \in [0, 1]$ , are subdomain-specific normalization parameters, which adjust the magnitude of the global features against the local ones upon the final outcome and their values are estimated according to the procedure described in Section 7.1. The subdomain with the highest  $G(D_l)$  value constitutes the final image classification decision.

## 6. FINAL REGION-CONCEPT ASSOCIATION

### 6.1. Hypotheses refinement and fuzzy spatial constraints verification factor

After the final image classification decision is made, a refined region-concept association procedure is performed. This procedure is similar to the one described in Section 5.2, the difference being that only the SVMs that correspond to concepts associated with the estimated subdomain are employed at this stage and thus subdomain-specific concept hypothesis sets are computed for every image segment. Subsequently, a genetic algorithm is introduced to decide on the optimal image interpretation, as outlined in Section 2. The GA is employed to solve a global optimization problem, while exploiting the available subdomain-specific spatial knowledge, thus overcoming the inherent visual information ambiguity. Spatial knowledge is obtained for every subdomain as described in Section 4 and the resulting learnt fuzzy spatial relations serve as constraints denoting the “allowed” subdomain concepts spatial topology.

Let  $I_S(g_{ij}, g_{pq})$  be defined as a function that returns the degree to which the spatial constraint between the  $g_{ij}$ ,  $g_{pq}$  concept to region mappings is satisfied.  $I_S(g_{ij}, g_{pq})$  is set to

receive values in the interval  $[0, 1]$ , where “1” denotes an allowable relation and “0” denotes an unacceptable one, based on the learnt spatial constraints. To calculate this value the following procedure is used: let  $I_{r_k}(s_i, s_p)$  denote the degrees to which each spatial relation is verified for a certain pair of regions  $s_i, s_p$  of the examined image (as defined in Section 4) and  $c_j, c_q$  denote the subdomain defined concepts assigned to them, respectively. A normalized Euclidean distance  $d(g_{ij}, g_{pq})$  is calculated, with respect to the corresponding spatial constraint, as introduced in Section 4, based on the following equation:

$$d(g_{ij}, g_{pq}) = \frac{\sqrt{\sum_{k=1}^8 (I_{r_k \text{ mean}}(c_j, c_q) - I_{r_k}(s_i, s_p))^2}}{\sqrt{8}}, \quad (6)$$

which receives values in the interval  $[0, 1]$ . The function  $I_S(g_{ij}, g_{pq})$  is then defined as

$$I_S(g_{ij}, g_{pq}) = 1 - d(g_{ij}, g_{pq}) \quad (7)$$

and takes values in the interval  $[0, 1]$  as well.

## 6.2. Implementation of genetic algorithm

As already described, the employed genetic algorithm uses as input the refined hypotheses sets (i.e., the subdomain-specific hypothesis sets), which are generated by the same SVMs structure as the initial hypotheses sets, the fuzzy spatial relations extracted between the examined image regions, and the spatial-related subdomain-specific contextual information as produced by the particular training process. Under the proposed approach, each chromosome represents a possible solution. Consequently, the number of the genes comprising each chromosome equals the number  $N$  of the regions  $s_i$  produced by the segmentation algorithm and each gene assigns a defined subdomain concept to an image segment.

A population of 200 randomly generated chromosomes is employed. An appropriate *fitness function* is introduced to provide a quantitative measure of each solution fitness for the estimated subdomain, that is, to determine the degree to which each interpretation is plausible:

$$f(Q) = \lambda_l \cdot FS_{\text{norm}} + (1 - \lambda_l) \cdot SC_{\text{norm}}, \quad (8)$$

where  $Q$  denotes a particular chromosome,  $FS_{\text{norm}}$  refers to the degree of low-level descriptors matching, and  $SC_{\text{norm}}$  stands for the degree of consistency with respect to the provided spatial subdomain-specific knowledge. The set of variables  $\lambda_l, l = 1, \dots, L$ , and  $\lambda_l \in [0, 1]$ , are introduced to adjust the degree to which visual feature matching and spatial relation consistency should affect the final outcome for every particular subdomain. Their values are estimated according to an optimization procedure, as described in Section 7.2.

The values of  $SC_{\text{norm}}$  and  $FS_{\text{norm}}$  are computed as follows:

$$FS_{\text{norm}} = \frac{\sum_{i=1}^N I_M(g_{ij}) - I_{\min}}{I_{\max} - I_{\min}}, \quad (9)$$

where  $I_{\min} = \sum_{i=1}^N \min_j I_M(g_{ij})$  is the sum of the minimum degrees of confidence assigned to each region hypotheses set and  $I_{\max} = \sum_{i=1}^N \max_j I_M(g_{ij})$  is the sum of the maximum degrees of confidence values, respectively,

$$SC_{\text{norm}} = \frac{\sum_{l=1}^W I_S(g_{ij}, g_{pq})}{W}, \quad (10)$$

where  $W$  denotes the number of the constraints that had to be examined.

After the population initialization, new generations are iteratively produced until the optimal solution is reached. Each generation results from the current one through the application of the following operators:

- (i) selection: a pair of chromosomes from the current generation are selected to serve as parents for the next generation. In the proposed framework, the tournament selection operator [28] with replacement is used;
- (ii) crossover: two selected chromosomes serve as parents for the computation of two new offsprings. Uniform crossover with probability of 0.7 is used;
- (iii) mutation: every gene of the processed offspring chromosome is likely to be mutated with probability of 0.008. If mutation occurs for a particular gene, then its corresponding value is modified, while updating the respective degree of confidence to the one of the new concept that is associated to it.

To ensure that chromosomes with high fitness will contribute to the next generation, the overlapping populations approach was adopted. More specifically, assuming a population of  $m$  chromosomes,  $m_s$  chromosomes are selected according to the employed selection method, and by application of the crossover and mutation operators,  $m_s$  new chromosomes are produced. Upon the resulting  $m + m_s$  chromosomes, the selection operator is applied once again in order to select the  $m$  chromosomes that will comprise the new generation. After experimentation, it was shown that choosing  $m_s = 0.4 m$  resulted in higher performance and faster convergence. The above iterative procedure continues until the diversity of the current generation is equal to/less than 0.001 or the number of generations exceeds 50. The above GA-based final region-concept association procedure was realized using the GA software libraries of [29].

## 7. PARAMETER OPTIMIZATION

In Sections 5.2 and 5.3, parameters  $a_l$  (4) and  $\mu_l$  (5) are introduced for adjusting the importance of the frequency of appearance against the region's area and the global versus local information on the final image classification decision for every particular ontology defined subdomain, respectively. Additionally, in Section 6.2 parameters  $\lambda_l$  (8) are introduced for adjusting the degree to which visual feature matching and spatial relation consistency should affect the final region-concept association outcome for every individual subdomain. In this section, we describe the methodology followed to estimate the values for the afore-mentioned parameters. This methodology is based on the use of a GA,

previously introduced for final region-concept association (Section 6.2). For the purpose of parameter value optimization, the chromosomes and the respective fitness function are defined accordingly.

Subject to the problem of concern is the computation of the values of

- (i) parameters  $a_l$  and  $\mu_l$  that lead to the highest correct image classification rate,
- (ii) parameters  $\lambda_l$  that lead to the highest correct concept association rate.

For that purpose, *Classification Accuracy*,  $CiA$ , is used as a quantitative performance measure and is defined as the fraction of the number of the correctly classified images to the total number of images to be classified, for the first case. Moreover, *Concept Accuracy*,  $CoA$ , which is defined as the fraction of the number of the correctly assigned concepts to the total number of image regions to be examined, is used for the second case. Then, for each problem the GA's chromosome,  $Q$ , is suitably formed, so as to represent a corresponding possible solution, and is further provided with an appropriate fitness function,  $f(Q)$ , for estimating each solution fitness, as described in the sequel.

### 7.1. Optimization of image classification parameters

For the case of optimizing parameters  $a_l$  and  $\mu_l$ , each chromosome  $Q$  represents a possible solution, that is, a candidate set of values for the parameters. In the current implementation, the number of genes of each chromosome is set equal to  $2 \cdot l \cdot 2 = 4 \cdot l$ . The genes represent the decimal coded values of parameters  $a_l$  and  $\mu_l$  assigned to the respective chromosome, according to the following equation:

$$Q = \begin{bmatrix} q_1 & q_2 & \cdots & q_{4 \cdot l} \\ \mu_1^1 & \mu_1^2 & \cdots & \mu_l^1 & \mu_l^2 & a_1^1 & a_1^2 & \cdots & a_l^1 & a_l^2 \end{bmatrix}, \quad (11)$$

where  $q_i \in \{0, 1, \dots, 9\}$  represents the value of gene  $i$  and  $\mu_i^t$ ,  $a_l^t$  represent the  $t$ th decimal digits of parameters  $\mu_l$ ,  $a_l$ , respectively. Furthermore, the genetic algorithm is provided with an appropriate *fitness function*, which is used for evaluating the suitability of each solution. In this case, the fitness function is defined as equal to the  $CiA$  metric already defined, where  $CiA$  is calculated over all images that comprise the validation set  $\mathcal{B}_v^2$ , after applying the fusion mechanism (Section 5.3) using for parameters  $a_l$  and  $\mu_l$  the values denoted by the genes of chromosome  $Q$ .

Regarding the GA's implementation details, an initial population of 100 randomly generated chromosomes is employed. New generations are successively produced based on the same evolution mechanism as described in Section 6.2. The differences are that the maximum number of generations is set equal to 30 and the probabilities of mutation and crossover are set equal to 0.4 and 0.2, respectively. The divergence in the value of the probability of the mutation operator denotes its increased importance in this particular optimization problem. The final outcome of this optimization procedure

are the optimal values of parameters  $a_l$  and  $\mu_l$ , used in (4) and (5).

### 7.2. Optimization of region-concept association parameters

For the case of optimizing parameters  $\lambda_l$ , the methodology described in this section is followed for every individual subdomain defined in the ontology. More specifically, under the proposed approach, each chromosome  $Q$  represents a possible solution, that is, a candidate  $\lambda_l$  value. The number of genes of each chromosome is set equal to 5. The genes represent the binary coded value of parameter  $\lambda_l$  assigned to the respective chromosome, according to the following equation:

$$Q = \begin{bmatrix} q_1 & q_2 & \cdots & q_5 \end{bmatrix} \quad \text{where} \quad \sum_{i=1}^5 q_i \cdot 2^{-i} = \lambda_l, \quad (12)$$

where  $q_i \in \{0, 1\}$  represents the value of gene  $i$ . The corresponding fitness function is defined as equal to the  $CoA$  metric already defined, where  $CoA$  is calculated over all images that belong to the  $D_l$  subdomain and are included in the validation set  $\mathcal{B}_v^2$ , after applying the genetic algorithm of Section 6.2 with  $\lambda_l = \sum_{i=1}^5 q_i \cdot 2^{-i}$ . Regarding the GA's implementation details, these are identical to the ones discussed in Section 7.1.

## 8. EXPERIMENTAL RESULTS

In this section, experimental results of the application of the proposed approach to images belonging to the personal collection domain, as well as comparative evaluation results with other approaches of the literature, are presented. The first step to the experimental evaluation was the development of an appropriate ontology in order to represent the selected domain, that is, the personal image collection domain, defining its subdomains, the concepts of interest associated with every subdomain and the supported contextual information. The developed ontology was described in detail in Section 4 and the subdomains and concepts of it can be seen in Figure 4.

Then, a set of 1800 randomly selected images belonging to the aforementioned domain were used to assemble the image collection  $\mathcal{B}$  and its constituent subsets used for training the different system components and for evaluation, as described in Section 2. Each image was manually annotated (i.e., manually generated image classification and, after segmentation is applied, region-concept associations) according to the ontology definitions. The content used was mainly obtained from the Flickr online photo management and sharing application [30] and includes images that depict cityscape, seaside, mountain, roadside, landscape, and sport-side locations. For content acquisition, the keyword-based search functionalities of [30] were employed. For every ontology defined subdomain, a corresponding set of suitable keywords was formed (e.g., regarding the *Rockyside* subdomain, the keywords *Rock*, *Rockyside*, *Mountain* were adopted) and used to drive the content acquisition process. Thus, the

Input image				
Global image classification	Buildings: 0.44 Rockside: <b>0.58</b> Forest: 0.56 Seaside: 0.30 Roadside: 0.51 Sports: 0.22	Buildings: <b>0.62</b> Rockside: 0.33 Forest: 0.32 Seaside: 0.21 Roadside: 0.27 Sports: 0.14	Buildings: 0.22 Rockside: 0.29 Forest: <b>0.84</b> Seaside: 0.31 Roadside: 0.27 Sports: 0.05	Buildings: 0.21 Rockside: 0.34 Forest: <b>0.54</b> Seaside: 0.12 Roadside: 0.37 Sports: 0.11
Local (i.e., region-based) image classification	Buildings: <b>0.64</b> Rockside: 0.32 Forest: 0.24 Seaside: 0.18 Roadside: 0.34 Sports: 0.21	Buildings: 0.23 Rockside: 0.29 Forest: 0.12 Seaside: 0.14 Roadside: <b>0.34</b> Sports: 0.11	Buildings: 0.32 Rockside: 0.29 Forest: 0.31 Seaside: <b>0.39</b> Roadside: 0.24 Sports: 0.18	Buildings: 0.24 Rockside: 0.28 Forest: 0.33 Seaside: 0.27 Roadside: <b>0.39</b> Sports: 0.11
Final image classification using information fusion	<b>Buildings</b>	<b>Roadside</b>	<b>Forest</b>	<b>Forest</b>

FIGURE 5: Indicative image-subdomain association results.

developed ontology concepts are compatible with concepts that are defined by a large number of users, which renders the whole evaluation framework more realistic.

Following the creation of the image sets, image set  $\mathcal{B}_{\text{tr}}$  was utilized for SVMs training. The training procedure for both the global image classification and the region-concept association cases was performed as described in Sections 5.1 and 5.2. The Gaussian radial basis function was used as a kernel function by each SVM, to allow for nonlinear discrimination of the samples. The low-level *image feature vector*, as described in detail in Section 3.1, is composed of 398 values, while the low-level *region feature vector* is composed of 433 values, calculated as described in Section 3.2. The values of both vectors are normalized in the interval  $[-1, 1]$ . On the other hand, for the acquisition of the required contextual information, the procedure described in Section 4 was followed for every subdomain.

Based on the trained SVMs structure, global image classification is performed as described in Section 5.1. Then, after the segmentation algorithm is applied and initial hypotheses are generated for every resulting image segment, the decision function is introduced that realizes image classification based on local-level as well as contextual information in the form of concept frequency of appearance, as outlined in Section 5.2. Afterwards, the fusion mechanism is employed which implements the fusion of the intermediate classification results based solely on global- and solely on local-level information and computes the final image classification (Section 5.3). In Figures 5 and 6 indicative classification results are presented, showing the input image, the image classification effected using only global (row 2) and only local (row 3) information, as indicated by the maximum of the  $h_l^D$  and of  $g(D_l)$ ,  $l = 1, \dots, L$ , respectively, and the final classification after the

evaluation of the fusion mechanism,  $G(D_l)$ . It can be seen in these figures that the final classification result, produced by the fusion mechanism, may differ from the one that is implied by the overall maximum of  $h_l^D$  and  $g(D_l)$  (e.g., second image of Figure 5).

In Table 3, quantitative performance measures of the image classification algorithms are given in terms of accuracy for each subdomain and overall. Accuracy is defined as the percentage of the images, belonging to a particular subdomain, that are correctly classified. The results presented in Table 3 show that the global classification method generally leads to better results than the local one. For the image classification based on local information, (4) is used to combine region-concept associations and contextual information in an ontology-driven manner as discussed in Section 5.2. It must be noted that the performance of both algorithms is subdomain dependent, that is, some subdomains are more suitable for classification based on global features (e.g., *Rockside* and *Forest*), whereas for other subdomains the application of a region-based image classification approach is advantageous. For example, in the *Rockside* subdomain the presented color distribution and texture characteristics are very similar among the corresponding images. Thus, image classification based on global features performs better than the local-level case. On the other hand, for subdomains like *Buildings*, where the color distribution and the texture characteristics of the depicted real-world objects may vary significantly (i.e., buildings are likely to have many different colors and shapes), the image classification based on local-level information presents increased classification rate. Furthermore, it can be verified that the proposed global and local classification information fusion approach leads to a significant performance improvement. Moreover, in Table 3 the

Input image				
Global image classification	Buildings: 0.31 Rockside: <b>0.84</b> Forest: 0.41 Seaside: 0.37 Roadside: 0.21 Sports: 0.19	Buildings: 0.17 Rockside: 0.32 Forest: <b>0.56</b> Seaside: 0.45 Roadside: 0.31 Sports: 0.22	Buildings: 0.38 Rockside: 0.18 Forest: <b>0.61</b> Seaside: 0.22 Roadside: 0.17 Sports: 0.12	Buildings: 0.12 Rockside: 0.21 Forest: 0.25 Seaside: 0.28 Roadside: 0.21 Sports: <b>0.91</b>
local (i.e., region-based) image classification	Buildings: 0.21 Rockside: 0.19 Forest: <b>0.22</b> Seaside: 0.17 Roadside: 0.12 Sports: 0.09	Buildings: 0.22 Rockside: 0.19 Forest: 0.27 Seaside: <b>0.52</b> Roadside: 0.19 Sports: 0.17	Buildings: <b>0.81</b> Rockside: 0.19 Forest: 0.29 Seaside: 0.24 Roadside: 0.12 Sports: 0.09	Buildings: 0.14 Rockside: 0.12 Forest: 0.12 Seaside: 0.21 Roadside: 0.11 Sports: <b>0.37</b>
Final image classification using information fusion	<b>Rockside</b>	<b>Seaside</b>	<b>Building</b>	<b>Sports</b>

FIGURE 6: Indicative image-subdomain association results.

TABLE 3: Subdomain detection accuracy.

Method	Accuracy						Overall
	Buildings	Rockside	Forest	Seaside	Roadside	Sports	
Global image classification	38.00%	68.63%	76.67%	<b>85.71%</b>	<b>68.42%</b>	92.00%	71.91%
Local (i.e., region-based) image classification	78.00%	50.98%	35.00%	60.71%	47.01%	94.00%	58.77%
Final image classification using information fusion	<b>84.00%</b>	<b>72.55%</b>	70.00%	<b>85.71%</b>	68.15%	95.00%	<b>78.70%</b>
SVM classifier proposed in [31]	56.00%	72.05%	73.33%	82.14%	63.15%	<b>98.50%</b>	74.07%
K-NN classifier proposed in [32]	62.00%	58.83%	<b>81.67%</b>	73.21%	42.85%	97.50%	69.13%

performance of the proposed approach is compared with the algorithms presented in [31], where an SVM-based multi-class classifier is used for image classification based on global features, and in [32], where a K-NN classifier combined with an appropriately trained feed-forward neural network realizes image categorization based on global-level descriptions. It can be easily observed that the proposed approach outperforms the aforementioned algorithms in most subdomains as well as in overall classification accuracy.

Using the final image classification decision, a concept hypothesis refinement procedure is performed (Section 6.1). The results are then passed to a genetic algorithm along with the subdomain-specific contextual knowledge in the form of spatial relations, in order to determine the globally optimal image interpretation (Section 6.2). In Figures 7 and 8 representative concept detection results are illustrated showing the original image, the annotation resulting from the initial hypotheses set, considering for each image segment  $s_i$  the hypothesis with the highest degree of confidence  $h_{ij}^C$ ,  $j = 1, \dots, J$ , and the final interpretation after the subdomain specification and the exploitation of the provided spatial related information. In Table 4, performance measures are

given for the concept detection task along the sequential steps of the proposed approach in terms of accuracy, similarly to the ones defined in Table 3. It must be noted that for the numerical evaluation of the concept detection accuracy, any concept present in the examined image test set that was not included in the ontology subdomain concepts definitions, for example, umbrella in the seaside subdomain, was not taken into account. From the results presented in Table 4, an increase in the performance introduced by the proposed approach can be observed. More specifically, the overall as well as most subdomain specific concept detection accuracies are improved after the implementation of the proposed classification algorithm, compared to the performance which corresponds to the initial region-concept association (derived by the initial hypotheses generation step). This increase in performance justifies the assumption that the reduction of the total number of concepts to be detected, after image classification is performed, leads to better concept detection results. Moreover, the use of the genetic algorithm, which is provided with the particular subdomain spatial constraints, introduces a further performance increase in most subdomains as well as overall. The latter demonstrates the effectiveness of using

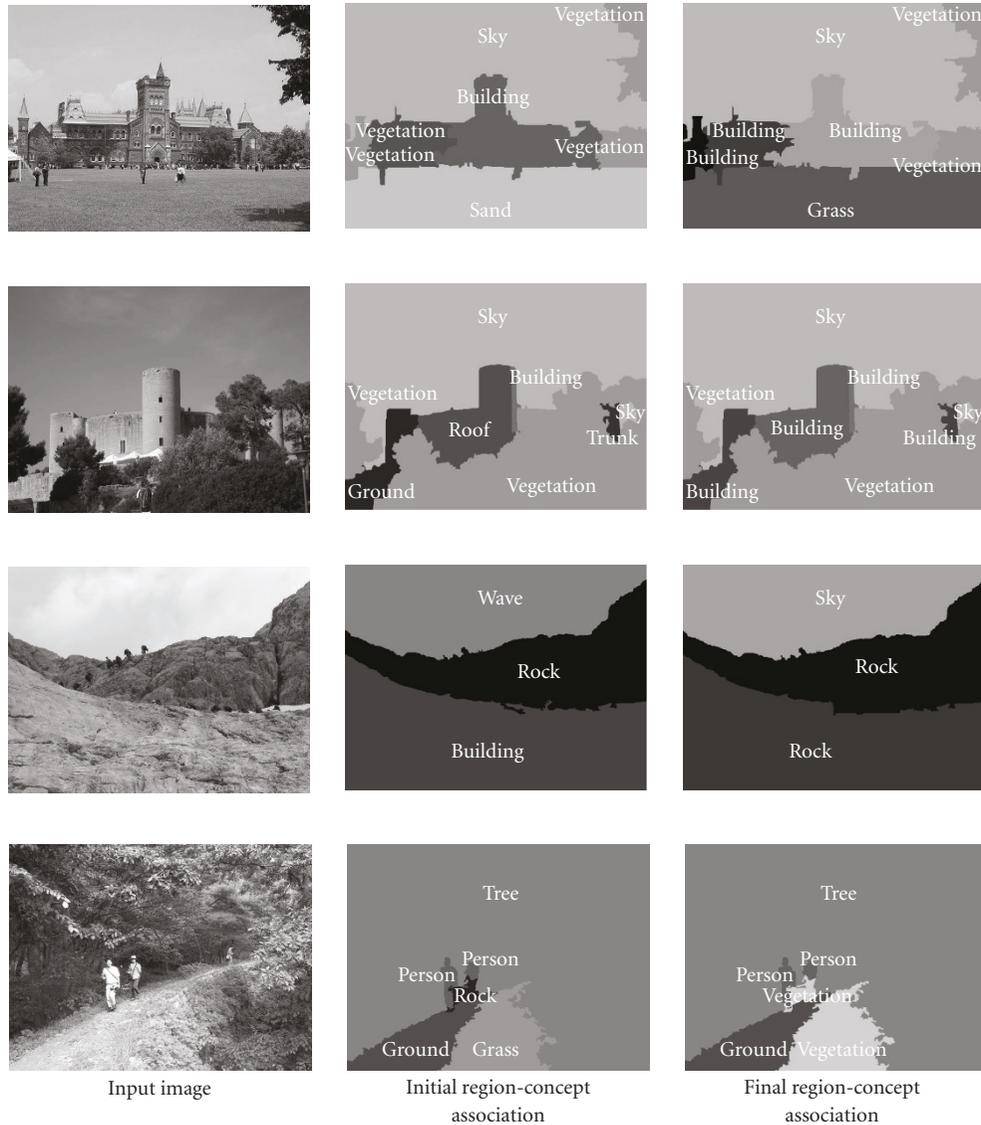


FIGURE 7: Indicative region-concept association results.

a genetic algorithm to reach an optimal image interpretation given degrees of confidence for visual similarity and spatial consistency against the domain definitions. The values of the parameters used in the final image classification and region-concept association process are computed according to the parameter optimization procedure described in Section 7.

Regarding the computational complexity of the proposed system, the times along the sequential steps of the algorithm for a  $600 \times 800$  pixels image are illustrated in Table 5. For the experimental evaluation we used a Pentium IV PC with 3 GHz CPU and 1 GB RAM. It must be noted that during the global classification step, the time needed for global descriptions extraction was considered. Similarly, for the region-based classification case, the time needed for segmentation and region-level descriptions extraction was also taken into account.

## 9. CONCLUSIONS

In this paper, an approach to knowledge-assisted image analysis and classification that combines global and local information with explicitly defined knowledge in the form of an ontology was presented. The proposed system was tested for the domain of personal collection images and produced promising results in this relatively broad domain. The effect of the different components of the proposed system in classification and analysis efficiency was clearly illustrated, documenting their usefulness in a knowledge-assisted image analysis and classification framework. As shown by the experimental evaluation of the proposed approach, the elegant combination of global and local information as well as contextual information leads to improved image classification performance, as compared to classification based solely

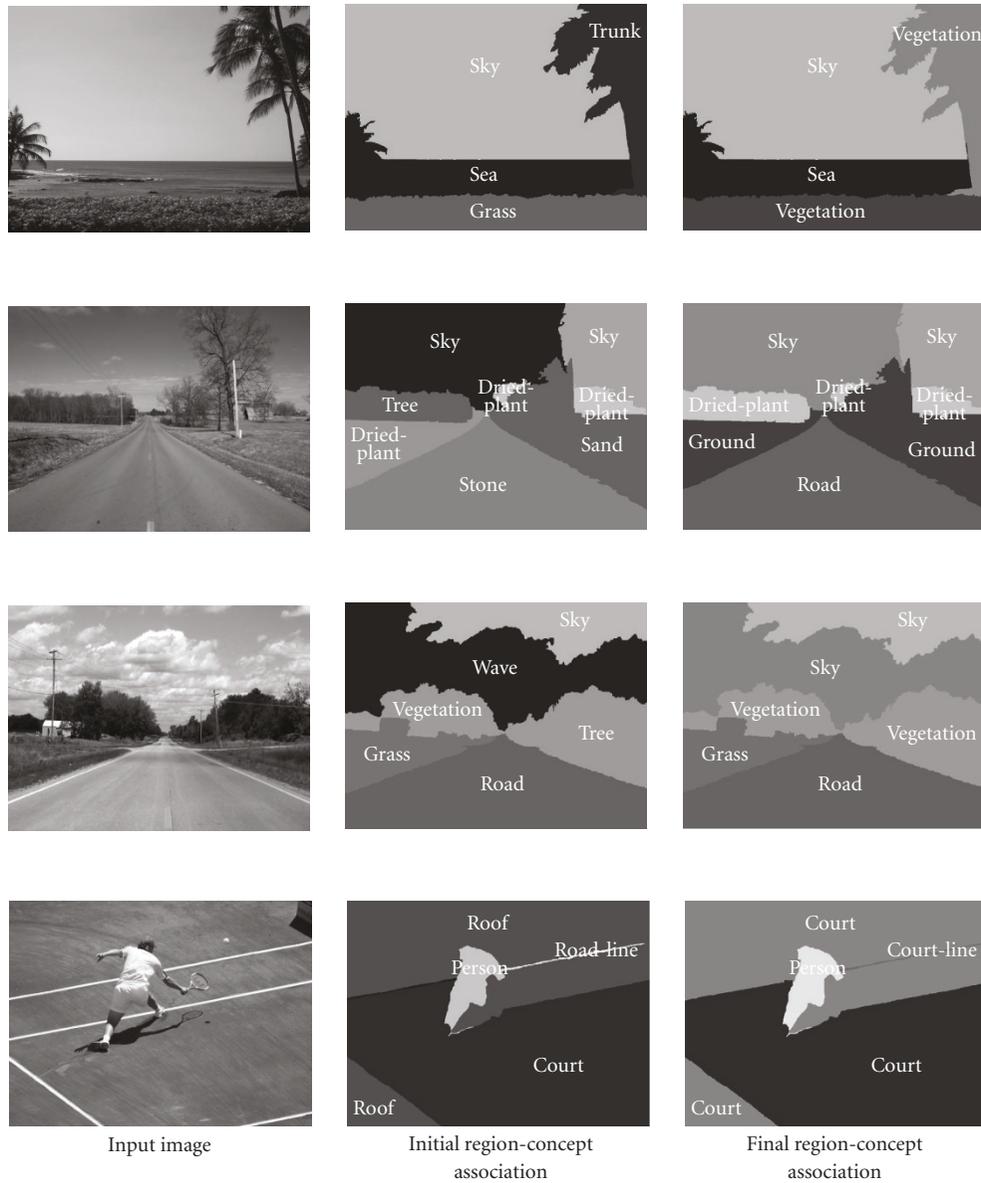


FIGURE 8: Indicative region-concept association results.

TABLE 4: Concept detection accuracy.

Algorithm stage	Accuracy						
	Buildings	Rockyside	Forest	Seaside	Roadside	Sports	Overall
Initial region-concept association	48.55%	47.45%	47.66%	63.33%	50.18%	74.55%	55.05%
Refined region-concept association	<b>50.92%</b>	49.68%	51.46%	65.19%	50.18%	<b>79.04%</b>	57.60%
Final region-concept association	50.39%	<b>50.00%</b>	<b>52.33%</b>	<b>67.77%</b>	<b>54.44%</b>	76.94%	<b>58.33%</b>

TABLE 5: Processing time for  $800 \times 600$  pixels image.

	Global classification	Region-based classification	Information fusion	Final region-concept association
Time (s)	8.77	42.89	0.001	24.46

on either global or local information. Furthermore, this image to subdomain association is used to further improve the accuracy of region to concept association, as compared to region to concept association performed without using knowledge about the former. The proposed framework is not restricted to the domain used in this work for evaluation purposes, but can easily be extended by including additional subdomains and concepts, providing that the employed knowledge representation is appropriately extended so as to account for these additional subdomains and concepts, and that the employed training set is enriched with suitable training samples.

## ACKNOWLEDGMENT

This work was supported by the European Commission under contracts FP6-001765 aceMedia, FP6-027685 MESH, and FP6-027026 K-Space, and by the GSRT under project DELTIO.

## REFERENCES

- [1] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 12, pp. 1349–1380, 2000.
- [2] S. Bloehdorn, K. Petridis, C. Saathoff, et al., "Semantic annotation of images and videos for multimedia analysis," in *Proceedings of the 2nd European Semantic Web Conference (ESWC '05)*, pp. 592–607, Herakleion, Greece, May–June 2005.
- [3] A. Barla, F. Odone, and A. Verri, "Old fashioned state-of-the-art image classification," in *Proceedings of the 12th International Conference on Image Analysis and Processing (ICIAP '03)*, pp. 566–571, Mantova, Italy, September 2003.
- [4] J. Tang, C.-Y. Zhang, and B. Luo, "A graph and PNN-based approach to image classification," in *Proceedings of International Conference on Machine Learning and Cybernetics (ICMLC '05)*, vol. 8, pp. 5122–5126, Guangzhou, China, August 2005.
- [5] H. Tanaka, H. Sakano, and S. Ohtsuka, "Retrieval method for multi-category images," in *Proceedings of the 17th International Conference on Pattern Recognition (ICPR '04)*, vol. 2, pp. 965–968, Cambridge, UK, August 2004.
- [6] I. Ahmad and M. T. Ibrahim, "Image classification and retrieval using correlation," in *Proceedings of the 3rd Canadian Conference on Computer and Robot Vision (CRV '06)*, p. 60, Quebec City, Canada, June 2006.
- [7] S. Papadopoulos, V. Mezaris, I. Kompatsiaris, and M. G. Strintzis, "A region-based approach to conceptual image classification," in *Proceedings of IEE International Conference on Visual Information Engineering (VIE '05)*, pp. 141–147, Glasgow, UK, April 2005.
- [8] V. S. Tseng, M.-H. Wang, and J.-H. Su, "A new method for image classification by using multilevel association rules," in *Proceedings of the 21st International Conference on Data Engineering Workshops (ICDE '05)*, p. 1180, Tokyo, Japan, April 2005.
- [9] T. Blaschke, "Object-based contextual image classification built on image segmentation," in *Proceedings of IEEE Workshop on Advances in Techniques for Analysis of Remotely Sensed Data (WARS'D '03)*, pp. 113–119, Washington, DC, USA, October 2003.
- [10] Z. Wang, D. Feng, and Z. Chi, "Region-based binary tree representation for image classification," in *Proceedings of International Conference on Neural Networks and Signal Processing (ICNNSP '03)*, vol. 1, pp. 232–235, Nanjing, China, December 2003.
- [11] S. Dasiopoulou, V. Mezaris, I. Kompatsiaris, V.-K. Papastathis, and M. G. Strintzis, "Knowledge-assisted semantic video object detection," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 10, pp. 1210–1224, 2005.
- [12] L. Hollink, S. Little, and J. Hunter, "Evaluating the application of semantic inferencing rules to image annotation," in *Proceedings of the 3rd International Conference on Knowledge Capture (K-CAP '05)*, pp. 91–98, Banff, Canada, October 2005.
- [13] N. Maillot, M. Thonnat, and C. Hudelot, "Ontology based object learning and recognition : application to image retrieval," in *Proceedings of the 16th IEEE International Conference on Tools with Artificial Intelligence (ICTAI '04)*, pp. 620–625, Boca Raton, Fla, USA, November 2004.
- [14] I. Kompatsiaris, V. Mezaris, and M. G. Strintzis, "Multimedia content indexing and retrieval using an object ontology," in *Multimedia Content and Semantic Web-Methods, Standards and Tools*, pp. 339–371, John Wiley & Sons, New York, NY, USA, 2004.
- [15] R. Tansley, C. Bird, W. Hall, P. Lewis, and M. Weal, "Automating the linking of content and concept," in *Proceedings of the 8th ACM International Multimedia Conference and Exhibition (MULTIMEDIA '00)*, pp. 445–447, Los Angeles, Calif, USA, October–November 2000.
- [16] K. I. Kim, K. Jung, S. H. Park, and H. J. Kim, "Support vector machines for texture classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 11, pp. 1542–1550, 2002.
- [17] O. Chapelle, P. Haffner, and V. N. Vapnik, "Support vector machines for histogram-based image classification," *IEEE Transactions on Neural Networks*, vol. 10, no. 5, pp. 1055–1064, 1999.
- [18] M. Mitchell, *An Introduction to Genetic Algorithms*, MIT Press, Cambridge, Mass, USA, 1995.
- [19] N. Voisine, S. Dasiopoulou, F. Precioso, V. Mezaris, I. Kompatsiaris, and M. G. Strintzis, "A genetic algorithm-based approach to knowledge-assisted video analysis," in *Proceedings of IEEE International Conference on Image Processing (ICIP '05)*, vol. 3, pp. 441–444, Genova, Italy, September 2005.
- [20] G. Th. Papadopoulos, P. Panagi, S. Dasiopoulou, V. Mezaris, and I. Kompatsiaris, "A learning approach to semantic image analysis," in *Proceedings of the 2nd International Mobile Multimedia Communications Conference (MobiMedia '06)*, Alghero, Italy, September 2006.
- [21] "MPEG-7 Visual Experimentation Model (XM)," Version 10.0, ISO/IEC/JTC1/SC29/WG11, Doc. N4062, March 2001.
- [22] T. Adamek, N. O'Connor, and N. Murphy, "Region-based segmentation of images using syntactic visual features," in *Proceedings of the 6th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS '05)*, Montreux, Switzerland, April 2005.
- [23] S. Skiadopoulos, C. Giannoukos, N. Sarkas, P. Vassiliadis, T. Sellis, and M. Koubarakis, "Computing and managing cardinal direction relations," *IEEE Transactions on Knowledge and Data Engineering*, vol. 17, no. 12, pp. 1610–1623, 2005.
- [24] Y. Wang, F. Makedon, J. Ford, L. Shen, and D. Goldin, "Generating fuzzy semantic metadata describing spatial relations

from images using the R-Histogram,” in *Proceedings of the 4th ACM/IEEE-CS Joint Conference on Digital Libraries (JCDL '04)*, pp. 202–211, Tucson, Ariz, USA, June 2004.

- [25] S. Staab and R. Studer, Eds., *Handbook on Ontologies*, International Handbooks on Information Systems, Springer, Berlin, Germany, 2004.
- [26] D. M. J. Tax and R. P. W. Duin, “Using two-class classifiers for multiclass classification,” in *Proceedings of the 16th International Conference on Pattern Recognition (ICPR '02)*, vol. 2, pp. 124–127, Quebec City, Canada, August 2002.
- [27] C.-C. Chang and C.-J. Lin, “LIBSVM: a library for support vector machines,” 2001, <http://www.csie.ntu.edu.tw/~cjlin/libsvm/>.
- [28] D. E. Goldberg and K. Deb, “A comparative analysis of selection schemes used in genetic algorithms,” in *Foundations of Genetic Algorithms*, G. Rawlins, Ed., pp. 69–93, Morgan Kaufmann Publishers, San Mateo, Calif, USA, 1991.
- [29] M. Wall, “GAlib: A C++ Genetic Algorithm Library,” MIT, 2005, <http://lancet.mit.edu/ga/>.
- [30] Flickr, <http://www.flickr.com/>.
- [31] J. Ren, Y. Shen, S. Ma, and L. Guo, “Applying multi-class SVMs into scene image classification,” in *Proceedings of the 17th International Conference on Innovations in Applied Artificial Intelligence*, pp. 924–934, Ottawa, Canada, May 2004.
- [32] E. Spyrou, H. Le Borgne, T. Mailis, E. Cooke, Y. Avrithis, and N. O’Connor, “Fusing MPEG-7 visual descriptors for image classification,” in *Proceedings of International Conference on Artificial Neural Networks (ICANN '05)*, pp. 847–852, Warsaw, Poland, September 2005.

**G. Th. Papadopoulos** was born in Thessaloniki, Greece in 1982. He received the Diploma degree in electrical and computer engineering from Aristotle University of Thessaloniki (AUTH), Thessaloniki, Greece in 2005. Currently he is pursuing his Ph.D. degree at the former University and he is a Postgraduate Research Fellow with the Informatics and Telematics Institute (ITI)/Centre for Research and Technology Hellas (CERTH), Thessaloniki, Greece. His research interests include still image segmentation, knowledge-assisted multimedia analysis, content-based and semantic multimedia indexing and retrieval, information extraction from multimedia, multimodal analysis, and adaptive learning techniques. He has published 2 international journals and he is the coauthor of 5 papers in international conferences. He is a member of the Technical Chamber of Greece.



**V. Mezaris** received the Diploma degree and Ph.D. degree in electrical and computer engineering from the Aristotle University of Thessaloniki, Thessaloniki, Greece, in 2001 and 2005, respectively. He is a postdoctoral research fellow with the Informatics and Telematics Institute/Centre for Research and Technology Hellas, Thessaloniki, Greece. His research interests include image and video analysis, content-based and semantic image and video retrieval, ontologies, multimedia standards, knowledge-assisted multimedia analysis, knowledge extraction from multimedia, medical image analysis. He is a Member of the IEEE and the Technical Chamber of Greece.



**I. Kompatsiaris** received the Diploma degree in electrical engineering and the Ph.D. degree in 3D model-based image sequence coding from Aristotle University of Thessaloniki (AUTH), Thessaloniki, Greece in 1996 and 2001, respectively. He is a Senior Researcher with the Informatics and Telematics Institute, Thessaloniki and currently he is leading the Multimedia Knowledge Group. His research interests include multimedia content processing, multimodal techniques, multimedia and the semantic web, multimedia ontologies, knowledge-based analysis, context aware inference for semantic multimedia analysis, personalization and retrieval. He is the coauthor of 6 book chapters, 18 papers in refereed journals, and more than 60 papers in international conferences. He has served as a regular reviewer for a number of international journals and conferences. He is a Member of IEEE and of the IEE VIE TAP.



**M. G. Strintzis** received the Diploma in electrical engineering from the National Technical University of Athens, Athens, Greece, in 1967 and the M.A. and Ph.D. degrees in electrical engineering from Princeton University, Princeton, NJ, in 1969 and 1970, respectively. He joined the Electrical Engineering Department, University of Pittsburgh, Pittsburgh, Pa, where he served as an Assistant Professor from 1970 to 1976 and an Associate Professor from 1976 to 1980. During that time, he worked in the area of stability of multidimensional systems. Since 1980, he has been a Professor of electrical and computer engineering at the Aristotle University of Thessaloniki, Thessaloniki, Greece. He has worked in the areas of multidimensional imaging and video coding. Over the past ten years, he has authored over 110 journal publications and over 280 conference presentations. In 1998, he founded the Informatics and Telematics Institute, currently part of the Centre for Research and Technology Hellas, Thessaloniki. Dr. Strintzis was awarded the Centennial Medal of the IEEE in 1984 and the Empirikeion Award for Research Excellence in Engineering in 1999. He is an IEEE Fellow since 2004.

