A Genetic Algorithm Approach to Ontology-driven Semantic Image Analysis

P. Panagi*†, S. Dasiopoulou*†, G.Th. Papadopoulos*† I. Kompatsiaris† and M.G. Strintzis*†

* Information Processing Laboratory
Electrical and Computer Engineering Department
Aristotle University of Thessaloniki, Greece
e-mail: {panagi, dasiop, papad, strintzis}@iti.gr

† Informatics and Telematics Institute
1st Km Thermi-Panorama Road
Thessaloniki, GR-57001 Greece
e-mail: ikom@iti.gr

Keywords: semantic image analysis, genetic algorithm, ontology, fuzzy spatial relations.

Abstract

In this paper, a hybrid approach coupling ontologies and a genetic algorithm is presented for realizing knowledge-assisted semantic image analysis. The employed domain knowledge considers both high-level information referring to objects of the domain of interest and their spatial relations, and low-level information in terms of prototypical low-level visual descriptors. To account for the inherent in visual information ambiguity, fuzzy spatial relations have been employed and the corresponding domain ontology definitions are obtained though training. A genetic algorithm is applied to decide the most plausible annotation. Experiments with images from the beach vacation domain demonstrate the performance of the proposed approach.

1 Introduction

Given the amount of available visual content, effective and efficient access at semantic level has become the key-enabling factor to allow users to benefit from such content and meet their information needs. As a result, the challenging issue of bridging the so called *semantic gap* between the content descriptions that can be automatically extracted and the ones adhering to user perception has received strong interest, leading to the emergence of numerous methodologies for handling efficiently the tasks of image analysis, indexing and retrieval [13, 8, 14].

Initially, the main focus was on the definition of suitable descriptors that could be automatically extracted and of appropriate metrics in the descriptor space that would allow for efficient image retrieval. These approaches did not target directly the extraction of semantics from visual content, but rather attempted to imitate the way users assess visual similarity. To address the resulting limitations, research interest was shifted from data-driven methodologies to knowledge-driven ones that utilize high-level domain knowledge in terms of guiding features extraction, descriptions derivation and symbolic inference. The relevant literature considers roughly two types of approaches, depending on the knowledge acquisition and representation process: implicit, realized by machine learning methods, and explicit, realized by model-based approaches.

The main characteristic of learning-based approaches is their capability to adjust their internal structure according to input and corresponding output data pairs. Consequently, the use of machine learning techniques provides a relatively powerful method for discovering complex and hidden relationships between image data and higher-level descriptions, resulting to a variety of image applications targeting semantic analysis. Among the most commonly used machine learning techniques are Neural Networks (NNs), Hidden Markov Models (HMMs) and Support Vector Machines (SVMs) [2, 11]. On the other hand, model-based image analysis approaches make use of explicitly defined prior knowledge such as models, rules, etc., thus providing a coherent semantic domain model to support "visual" inference in the context specified [4, 7, 9]. However, an issue considering approaches that use explicit knowledge is the complexity that increases exponentially with the number of objects of interest.

In order to benefit from the advantages of both categories of knowledge-assisted approaches and overcome their individual limitations, a hybrid approach to domain-specific automatic analysis is proposed in this paper. The employed domain knowledge considers both high- and low-level information, and in accordance with the recent Semantic Web advances, the ontology paradigm has been followed for representation. High-level knowledge refers to the domain objects of interest and their spatial relations, whereas low-level knowledge consists of low-level visual information required for the actual analysis process. To account for the inherent ambiguity in visual information, fuzzy spatial relations, obtained through training, have been included in the domain knowledge. Initially, a set of graded hypothesis is produced based on visual similarity. These hypotheses along with the extracted segments' spatial relations are passed in the sequel to a genetic algorithm that based on the provided domain knowledge decides the optimal image interpretation.

Section 2 presents the overall system architecture and details the individual components. Section 3 presents experimental results and evaluation in the domain of beach vacation images, and Section 4 concludes the paper.

2 Ontology-driven genetic algorithm-based image analysis

In accordance with the principles of knowledge-driven approaches and acknowledging the individual limitations of both explicit and implicit knowledge representation, a hybrid approach to ontology-driven image analysis using a genetic algorithm (GA) is proposed in this paper. The choice of the genetic algorithm approach is based on the premise that image interpretation can be perceived as an optimization problem of searching the most plausible assignment between a set of concepts and the respective image segments. Furthermore, as the mapping between segments and concepts is not a well-defined one, i.e., concepts cannot in general be associated with unique nor complete visual definitions, determining a mapping reduces in finding the mapping that best satisfies the set of provided concept descriptions. Thereby, since genetic algorithms [10] have been widely applied in many fields involving optimization problems and have proved to outperform other traditional methods, following such an approach provides certain rationales. Additionally, the authors' previous experience with genetic-based visual content semantic analysis has showed promising results [15].

The overall architecture of the proposed framework is illustrated in Fig. 1. First segmentation is applied, and subsequently low-level descriptors and fuzzy spatial relations are extracted for the generated image segments. Once the low-level descriptors are available, an initial set of hypotheses is generated for each image segment based on the distance between each segment extracted descriptors and the prototypical descriptors of the domain concepts included in the domain knowledge base. The resulting hypotheses with the associated degrees of confidence are then passed to the genetic algorithm along with the segments extracted spatial relations. Utilizing the provided domain spatial-related knowledge, the genetic algorithm decides the optimal semantic interpretation of the examined image, i.e. the semantic concept assigned to each image segment. In the following, the details of the individual components are presented.

2.1 Knowledge infrastructure

Among the possible knowledge representations, ontologies [6] present a number of advantages. They provide a formal framework for supporting explicit, machine-processable, semantics definition and enable the derivation of new knowledge through automated inference. Thus, ontologies appeal to expressing multimedia content semantics in a formal machine-processable representation that will allow automatic semantic analysis and further processing of the extracted semantic annotations. Following these considerations, in the proposed framework, the ontology infrastructure introduced in [3] has been used as the means for representing the knowledge components needed. This infrastructure consists of a visual descriptor ontology that contains the representations of the employed low-level visual descriptors, a multimedia structure ontology, and the DOLCE core ontology to harmonize them with the corresponding domain ontology.

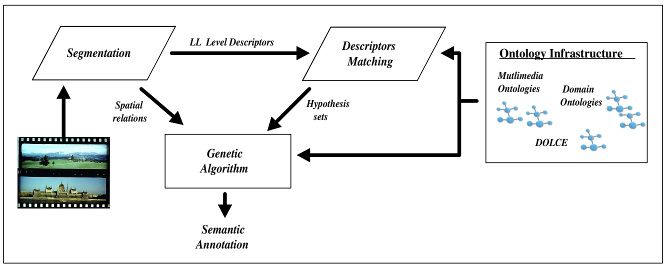

Figure 1: Overall architecture.

2.2 Hypothesis generation

As mentioned above, an initial set of graded hypotheses is generated for each image segment utilizing the provided domain knowledge prototypical visual definitions. To accomplish this, firstly segmentation needs to be applied and secondly, the required low-level descriptors have to be extracted. In the current implementation, an extension of the Recursive Shortest Spanning Tree (RSST) algorithm has been used for segmenting the image [1], while descriptors extraction is based on the guidelines given by the MPEG-7 eXperimentation Model (XM) [17]. In order to produce the hypotheses sets, appropriate measures need to be defined for qualitatively assessing visual similarity between the examined image segments and the defined domain concepts. As MPEG-7 does not provide a standardized method for combining different descriptors distances or for estimating a single distance based on more than one descriptor, a simple weighted sum approach was followed. Thereby, for each segment a similarity degree is produced against each of the defined domain concepts. The pairs of domain concept and corresponding degree of confidence that result for each segment comprise its hypothesis set.

2.3 Fuzzy spatial relations extraction

Exploiting domain-specific spatial knowledge in image analysis tasks is a common practice as objects tend to occur within a particular spatial context, and additionally, spatial knowledge can assist in discriminating objects exhibiting

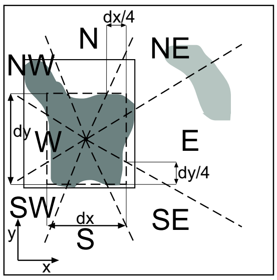

Figure 2: Reduced MBR spatial relations definition.

similar visual characteristics. In the presented analysis framework, eight directional relations are supported, namely *above, right, below, left, above-right, above-left, below-right* and *below-left*, for adjacent regions. To account for the complexity in assessing which spatial relations best describe the examined domain concepts relations, fuzzy relations where preferred. Building on the principles of projection- and angle based methodologies [16, 12], a combined approach is employed that consists of the following steps. First, a reduced box from the ground object MBR is computed so as to include the object in a more representative way, and then cone-based regions on top of this reduced box corresponding to the eight examined relations are defined. *Figure* object denotes the one whose relative position is to be estimated, while *ground* object the reference one. The computation of the reduced MBR is performed in terms of the MBR compactness value c. If the currently computed c is greater than the given threshold, eight cone-shaped regions are defined corresponding to the eight supported relations 2. The percentage of the figure object points that are included in each of the cone-shaped regions determines the degree to which the corresponding directional relation is satisfied. If however, the computed c is below the threshold set, the MBR is reduced repeatedly until the desired threshold is satisfied. After experimentation, the value of the threshold was set to 0.85.

2.4 Genetic algorithm

As described above, first, an initial set of hypotheses is generated based solely on visual similarity (subsection 2.2) and fuzzy spatial relations are extracted for each pair of adjacent segments (subsection 2.3). This information is then passed to a genetic algorithm that exploiting the available domain spatial knowledge determines the optimal image interpretation. To account for the inherent in visual information ambiguity and the difficulty in coming up with spatial definitions that cover every possible domain concept instantiation, appropriate training was conducted in order to acquire membership values for each of the spatial relations comprising the domain knowledge. As explained in the following, these fuzzy spatial relations serve as constraints with respect to the "allowed" domain concepts spatial arrangement.

In the subsequent description, the following definitions have been used:

$$s_i\ \ ,\ i = 0,\ 1, \ldots\ N-1, \tag{1}$$

is used to denote the $i-th$ image segment of the N segments produced by the initial segmentation.

$$O = \{o_1,\ o_2,\ o_3,\ \ldots\}, \tag{2}$$

denotes the set of concepts defined in the each time employed domain ontology.

The function

$$I_M(g_i)\ \equiv\ I_M(s_i,\ o_j), \tag{3}$$

provides the degree to which the visual descriptors extracted for segment s_i match the ones of object o_j, where g_i represents the particular assignment of o_j to s_i. Thus, $I_M(g_i)$ gives the degree of confidence associated with each hypothesis and takes values in the interval $[0, 1]$.

The set of spatial relations used is denoted by R_k, and in the current implementation reduces to

$$R_k = \{N,\ NW,\ NE,\ S,\ SW,\ SE,\ W,\ E\}, \tag{4}$$

The function:

$$I_{R_k}\ (s_i,\ s_j)\quad,\ k \in \aleph\ and\ k \in [1, 8], \tag{5}$$

returns the degree to which s_i satisfies the relation R_k with respect to s_j; again the resulted degree belongs to $[0, 1]$.

The function:

$$I_S\ (g_i,\ g_j), \tag{6}$$

returns the degree to which the spatial constraints between the g_i, g_j object to segments mappings are satisfied, i.e. the degree to which the relations extracted between segments s_i and s_j comply with the spatial knowledge of the objects o_i and o_j respectively that have been assigned to these segments.

Training and fuzzy spatial constraints acquisition.

To compute the fuzzy spatial constraints between each pair of the defined domain objects, a set of images has been assembled to serve as training set, and respective ground truth annotations were constructed. For every pair of adjacent segments assigned to objects (o_i, o_j), $i \neq j$, the degrees to which each relation R_k is verified are summed over the set of training images. Thus, the mean $I_{R_k mean}$ and the standard deviation $\sigma^2_{R_k}$ (Eq. 7), are obtained for each of the relations.

$$\sigma_k^2 = \frac{\sum_{i=1}^{N}(I_{R_k i} - I_{R_k mean})^2}{N} \quad , \tag{7}$$

where N denotes the number that R_k is satisfied for objects o_i and o_j.

Fuzzy spatial relations membership.

The function $I_S(g_i,\ g_j)$ returns the degree to which the spatial constraint between the objects involved in the g_i and g_j mappings is satisfied. $I_S(g_i,\ g_j)$ is set to receive values in the interval $[-1, 1]$, where '1' denotes an allowable relation and '-1' denotes an unacceptable one based on the learnt spatial constraints. To calculate this value for a specific pair of objects o_i and o_j the following procedure is used. For every computed triplet $[R_k\ I_{R_k mean}\ s_k^2]$ of the corresponding spatial constraint where $I_{R_k mean} \neq 0$, a triangular fuzzy membership function is formed to compute the corresponding degree.

Let d_k denote the value resulted from applying the membership function with respect to relation R_k. Once, the corresponding d_k values have been computed for each relations, i.e. for $k = 0, 1, ..8$, they are combined to form the degree $I_S(g_i,\ g_j)$ to which the corresponding spatial constraint is satisfied. The calculation is based on based on Eq. (8), where R_m are the relations for which $I_{R_m mean} \neq 0$, and R_n are the extracted relations for which $I_{R_n mean} = 0$.

$$I_S\ (g_i,\ g_j) = \frac{\sum_m \frac{I_{R_m} * d_m}{IR}}{\sum_m d_m} - \sum_n I_{R_n}\ ,\quad i \neq j \tag{8}$$

where IR stands for

$$IR = \sum_m I_{R_m} \tag{9}$$

Implementation of genetic algorithm.

As described above, the developed algorithm uses as input the initial set of hypotheses, the spatial relations extracted between between the examined image adjacent segments, and spatial domain knowledge as produced by the above described 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 segments s_i produced by the segmentation and its chromosome assigns a domain concept to an image segment.

A population of 200 chromosomes is used, and it is initialized with respect to the input set of hypotheses. An appropriate *fitness function* is introduced to provide a quantitative measure of each solution fitness, i.e. to determine the degree to which each interpretation is plausible:

$$f(C)\ =\ \lambda \times FS_{norm}\ +\ (1-\lambda) \times SC_{norm}\ , \tag{10}$$

where C denotes a particular Chromosome, FS_{norm} refers to the degree of low-level descriptors similarity, and SC_{norm} stands for the degree of consistency with respect to the provided spatial domain knowledge. The variable λ is introduced to adjust the degree to which visual similarity and spatial consistency should affect the final outcome. After thorough experimentation, λ was set to 0.35, which points out the importance of spatial context.

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

$$FS_{norm}\ =\ \frac{\sum_{i=0}^{N-1} I_M(g_i) - I_{min}}{I_{max} - I_{min}}\ , \tag{11}$$

where I_{min} is the sum of the minimum degrees of confidence assigned of each region hypotheses set and I_{max} the sum of the maximum degrees of confidence values respectively.

$$SC_{norm}\ =\ \frac{SC+1}{2}\ \ and\ \ SC\ =\ \frac{\sum_{r=1}^{M} I_{s_r}(g_i, g_j)}{M}\ , \tag{12}$$

where M denotes the number of relations in 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.

- 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 [5], with replacement, is used.
- Crossover: two selected chromosomes serve as parents for the computation of two new offsprings. Uniform crossover with probability of 0.7 is used.
- 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 keeping unchanged the degree of confidence.

To ensure that chromosomes with high fitness will contribute in the next generation, the overlapping populations approach was adopted. More specifically, assuming a population of m chromosomes, m_s chromosomes are selected following the employed selection method, and through the application of the crossover and mutation operators m_s new chromosome are produced. Upon the resulted $m + m_s$ chromosomes, the selection operator is applied once again in order to select the m chromosomes that will comprise the new generation. After experimentation, it proved that selecting $m_s = 0.4m$ resulted in higher performance and faster convergence. The above iterative procedure continues until the diversity of the current generation is equal or less than 0.001.

3 Experimental results

In this section, we present experimental results from testing the proposed approach in the domain of beach vacation images. First, a domain ontology had to be developed to represent the domain concepts of interest and their spatial relations. The beach vacation domain was selected for experimentation, and under the current implementation four concepts, namely *Sky*, *Sea*, *Sand* and *Person*. A variety of images from the beach vacation domain was selected then in order to assemble a training set of 50 images for the acquisition of low-level visual descriptors prototypes and membership values for the spatial relations. Each image of the training set was manually annotated according to the domain ontology. Subsequently, segmentation was performed as described above, and the Dominant Color and the Region Shape descriptors, i.e., the two currently supported descriptors, of the annotated segments were extracted. Approximately, 10 prototype descriptor instances resulted for each of the defined domain concepts after the elimination of the redundant ones, i.e., of prototypes almost identical to each other that do not offer any additional discriminative power. Additionally, for each pair of adjacent segments the degree to which each spatial relation is satisfied was estimated and thus, following the procedure described in Section 2.4 for each possible combination of the defined domain concepts, the domain ontology spatial relations were enhanced with fuzzy degrees.

Having available the needed domain knowledge, semantic annotation of images can be performed following the proposed approach. Based on the extracted prototype instances, initial hypotheses are generated for the examined image segments as described in Section 2.2, which are then passed in the genetic algorithm along with the fuzzy spatial constraints in order to determine the final interpretation. In Fig. 3 indicative results are given showing the input image, the annotation resulted from the initial hypotheses set taking for each image segment the hypothesis with the highest degree of confidence and the final interpretation after the genetic algorithm application. As illustrated, the proposed system achieves satisfactory results, justifying the use of a genetic algorithm to reach an optimal image interpretation given degrees of confidence for visual similarity and spatial consistency against the domain definitions. In Table 3, quantitative performance measures are given in terms of precision and recall over a test set of 200 beach vacation images. It must be noted that for the numerical evaluation, any concept present in the examined test set images that was not included in the domain ontology concept definitions, such as umbrellas, sailing boats, etc., was not taken into account. As this affects directly the evaluation results, it is within the authors' future intentions to introduce a threshold in the acquired degrees of confidence to represent the "unknown" concept in order to account for such cases and obtain more accurate indications of the attained performance.

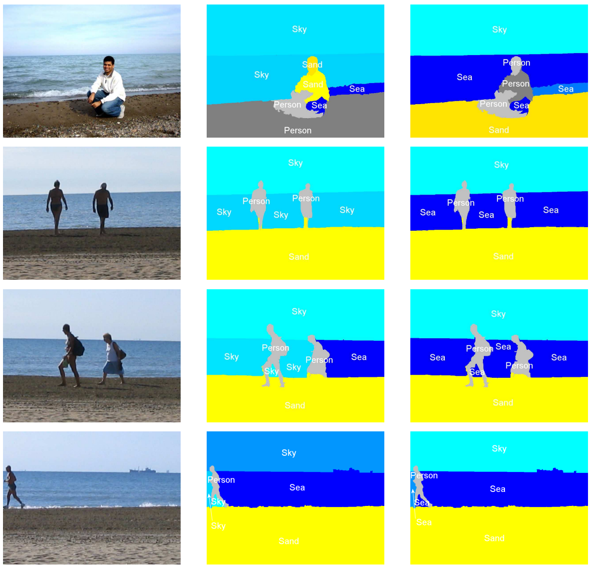

Figure 3: Exemplar results for the beach domain.

	Initial Hypothesis		Final interpretation	
Object	precision	recall	precision	recall
Sky	56.0%	77.7%	80.0%	98.8%
Sea	83.6%	60.0%	85.7%	70.6%
Sand	43.0%	82.2%	56.9%	91.1%
Person	78.9%	56.9%	91.7%	67.9%

Table 1: Numerical evaluation for the beach vacation domain.

4 Conclusions

In this paper, an approach to semantic image analysis that couples ontologies with a genetic algorithm is presented. The employed knowledge considers both high- and low-level information, and in accordance with the recent Semantic Web advances, the ontology paradigm has been followed for representation. High-level knowledge includes the domain objects of interest and their spatial relations, whereas low-level knowledge consists of low-level visual descriptors required for the analysis process. To account for the inherent ambiguity in visual information, fuzzy spatial relations are employed and the corresponding domain relations definitions are obtained though training. Based on the low-level domain ontology definitions, initial hypotheses sets are generated including the domain concepts possibly depicted to each of the image segments and their respective degrees of confidence. These hypotheses along with the segments' spatial relations form the input to a genetic algorithm that decides the optimal image interpretation based on the provided spatial domain knowledge.

Future work includes the exploitation of spatial constraints involving more than two objects, so that the employed spatial knowledge provides contextual knowledge not only at local level, i.e., adjacent regions, but instead, at a more global level taking into account objects co-occurrence. By encoding such knowledge into the domain knowledge and adapting appropriately the defined fitness function, the extraction of the image semantics would become more reliable as the presence of certain objects in a given topology would provide hints for reaching a correct decision in cases of higher ambiguity.

Acknowledgements

The work presented in this paper was partially supported by the European Commission under contracts FP6-001765 aceMedia and FP6-027026 K-Space, and COST 292 action.

References

[1] T. Adamek, N. O'Connor, N. Murphy, "Region-based Segmentation of Images Using Syntactic Visual Features," *Workshop on Image Analysis for Multimedia Interactive Services, (WIAMIS), Montreux, Switzerland,(2005).*

[2] J. Assfalg, M. Berlini, A. Del Bimbo, W. Nunziat, P. Pala, "Soccer Highlights Detection and Recognition using HMMs," *IEEE International Conference on Multimedia & Expo (ICME), pp. 825-828,(2005).*

[3] S. Bloehdorn, K. Petridis, C. Saathoff, N. Simou, V. Tzouvaras, Y. Avrithis, I. Kompatsiaris, S. Staab, M. G. Strintzis, "Semantic Annotation of Images and Videos for Multimedia Analysis," *IProc. 2nd European Semantic Web Conference, ESWC 2005, Heraklion, Greece, May 2005).*

[4] S. Dasiopoulou, V. Mezaris, V. K. Papastathis, I. Kompatsiaris, M.G. Strintzis, "Knowledge-Assisted Semantic Video Object Detection," *IEEE Transactions on Circuits and Systems for Video Technology, Special Issue on Analysis and Understanding for Video Adaptation, vol. 15, no. 10, pp. 1210-1224, (2005).*

[5] D. Goldberg, K. Deb, "A comparative analysis of selection schemes used in genetic algorithms," *In Foundations of Genetic Algorithms, G. Rawlins, 69-93, (1991).*

[6] T. R. Gruber, "A translation approach to portable ontologies," *Knowledge Acquisition, 5(2), pp.199-220, (1993).*

[7] L. Hollink, S. Little, J. Hunter, "Evaluating the Application of Semantic Inferencing Rules to Image Annotation," *3rd International Conference on Knowledge Capture (K-CAP05), Banff, Canada, (2005).*

[8] W. Al-Khatib, Y. F. Day, A. Ghafoor, P. B. Berra, "Semantic Annotation of Images and Videos for Multimedia Analysis," *2nd European Semantic Web Conference (ESWC), Herakleion, Greece, (2005).*

[9] N. Maillot, M. Thonnat, "A Weakly Supervised Approach for Semantic Image Indexing and Retrieval," *4th International Conference on Image and Video Retrieval (CIVR), Singapore, pp. 629-638, (2005).*

[10] M. Mitchell, "An introduction to genetic algorithms," *MIT Press, (1995).*

[11] M. R Naphade, I. V. Kozintsev, T. S. Huang, "A factor graph framework for semantic video indexing," *IEEE Transactions On Circuits and Systems for Video Technology, vol. 12, iss. 1, pp. 40-52, (2002).*

[12] S. Skiadopoulos, C. Giannoukos, N. Sarkas, P. Vassiliadis, T. Sellis, M. Koubarakis, "2D topological and direction relations in the world of minimum bounding circles," *IEEE Transactions on Knowledge and Data Engineering, vol. 17, iss. 12, pp. 1610-1623, (2005).*

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

[14] C. Snoek, M. Worring, "Multimodal Video Indexing: A Review of the State-of-the-art," *Multimedia Tools and Applications, Springer Science and Business Media, (2005).*

[15] N. Voisine, S. Dasiopoulou, F. Precioso, V. Mezaris, I. Kompatsiaris and M. G. Strintzis, "A Genetic Algorithm-based Approach to Knowledge-assisted Video Analysis," *Proc. IEEE International Conference on Image Processing (ICIP 2005), Genova, (2005).*

[16] Y. Wang, F. Makedon, J. Ford, L. Shen, D. Golding, "Generating Fuzzy Semantic Metadata Describing Spatial Relations from Images using the R-Histogram," *JCDL '04, June 7-11, Tucson, Arizona, USA, (2004).*

[17] "MPEG-7 Visual Experimentation Model (XM)", *Version 10.0, ISO/IEC/JTC1/SC29/WG11, Doc. N4062, Mar. (2001).*