Knowledge-Assisted Semantic Video Object Detection

Stamatia Dasiopoulou, Vasileios Mezaris, *Member, IEEE*, Ioannis Kompatsiaris, *Member, IEEE*, Vasileios-Kyriakos Papastathis, and Michael G. Strintzis, *Fellow, IEEE*

***Abstract*—An approach to knowledge-assisted semantic video object detection based on a multimedia ontology infrastructure is presented. Semantic concepts in the context of the examined domain are defined in an ontology, enriched with qualitative attributes (e.g., color homogeneity), low-level features (e.g., color model components distribution), object spatial relations, and multimedia processing methods (e.g., color clustering). Semantic Web technologies are used for knowledge representation in the RDF(S) metadata standard. Rules in F-logic are defined to describe how tools for multimedia analysis should be applied, depending on concept attributes and low-level features, for the detection of video objects corresponding to the semantic concepts defined in the ontology. This supports flexible and managed execution of various application and domain independent multimedia analysis tasks. Furthermore, this semantic analysis approach can be used in semantic annotation and transcoding systems, which take into consideration the users environment including preferences, devices used, available network bandwidth and content identity. The proposed approach was tested for the detection of semantic objects on video data of three different domains.**

***Index Terms*—Knowledge-assisted analysis, multimedia ontologies, video analysis.**

I. Introduction

THE RECENT progress in hardware and telecommunication technologies has resulted to a rapid increase of the available amount of multimedia information. Multimedia content is used in a wide range of applications in areas such as content production and distribution, telemedicine, digital libraries, distance learning, tourism, distributed CAD/CAM, GIS, and, of course, on the World Wide Web. The usefulness of all these applications is largely determined by the accessibility of the content and as such, multimedia data sets present a great challenge in terms of storing, transmitting, querying, indexing, and retrieving. To face such challenges it is not sufficient to just develop faster hardware or to design more sophisticated algorithms. Rather, a deeper understanding of the information at the semantic level is required [1]. This is of particular importance in many emerging applications such as semantic transcoding [2], where it is assumed that the user does not want to access all data, but only data semantically useful. This requires the semantic identification of the objects and events appearing in the content so as to be in a position to match them with the user preferences. In this way, the part of the content which is of interest to the user is identified, isolated, and transmitted.

Although new multimedia standards, such as MPEG-4 and MPEG-7 [3], provide the needed functionalities in order to manipulate and transmit objects and metadata, their extraction, most importantly at a semantic level, is out of the scope of the standards and is left to the content developer. In the last two decades, significant results have been reported regarding the successful implementation of several prototypes [4]. However, the lack of precise models and formats for object and system representation and the high complexity of multimedia processing algorithms make the development of fully automatic semantic multimedia analysis and management systems a challenging task [1].

This is due to the difficulty, often referred to as the *semantic gap*, of mapping semantic concepts into a set of image and/or spatiotemporal features that can be automatically extracted from video data without human intervention [5]. The use of domain knowledge is probably the only way by which higher level semantics can be incorporated into techniques that capture the semantics through automatic analysis. The various approaches exploiting domain-specific knowledge for multimedia analysis consider two types of knowledge: implicit, as realized by stochastic methods that exploit automatic learning capabilities, and explicit, where knowledge encoded in some formal representation is used to drive the extraction of high-level semantics. The work in [6], where the problem of bridging the gap between low-level representation and high-level semantics is formulated as a probabilistic pattern recognition problem, and [7], where fuzzy ontological relations and context aware fuzzy hierarchical clustering are employed to interpret multimedia content for the purpose of automatic thematic categorization of multimedia documents, fall within the first category. Previous work in the literature that exploits explicit domain knowledge includes among others the approaches presented in [8], [9], where *a priori* knowledge representation models are used as a knowledge base that assists semantic-based classification and clustering, and [10], in which semantic entities, in the context of the MPEG-7 standard, are used for knowledge-assisted video analysis and object detection, thus, allowing for semantic level indexing.

Manuscript received May 28, 2004; revised December 19, 2004. This paper was recommended by Associate Editor P. van Beek.

S. Dasiopoulou, V. Mezaris, and M. G. Strintzis are with the Informatics and Telematics Institute, Centre for Research and Technology Hellas, 57001 Thermi—Thessaloniki, Greece and also with the Information Processing Laboratory, Electrical and Computer Engineering Department. University of Thessaloniki, 54124 Thessaloniki, Greece (e-mail: dasiop@iti.gr; bmezaris@iti.gr; strintzi@eng.auth.gr).

I. Kompatsiaris and V.-K. Papastathis are with the Informatics and Telematics Institute, Centre for Research and Technology Hellas, 57001 Thermi—Thessaloniki, Greece (e-mail: ikom@iti.gr; vkpapa@iti.gr).

Digital Object Identifier 10.1109/TCSVT.2005.854238

Most of the approaches presented include domain and application dependent multimedia processing tools and algorithms for performing the required analysis tasks [11], [12]. However, the use of specific approaches that are heavily dependent on the targeted domain application reduces the chances of exploiting or extending such knowledge-assisted systems to different domains. Furthermore, while some of them manage to formally represent the knowledge about the domain, i.e., the objects and the associated low-level features, they do not include algorithms in this knowledge representation, thus, reducing their interoperability and reusability. Work on building a unifying model of all these aspects includes [8], where domain knowledge is used to realize application and media independent content-based retrieval for multimedia databases, and [13], where Semantic Web technologies are used in system integration to describe how tools for analysis and visualization can be applied to different data-type sources.

In this paper, an approach to knowledge-assisted semantic video object detection based on a multimedia ontology infrastructure is presented. More specifically, semantic and low-level attributes of the objects to be detected in combination with appropriately defined rules, determine the set of algorithms and parameters required for the detection of the objects. Semantic concepts in the context of the examined domain are defined in an ontology, enriched with qualitative attributes (e.g., color homogeneity), low-level features along with numerical data generated via training (e.g., color models, also defined in the ontology), object spatial relations and multimedia processing methods (e.g., color clustering). Semantic Web technologies are used for knowledge representation in the resource description framework schema (RDFS) language [14]. F-logic [15] rules are defined to describe how tools for multimedia analysis should be applied according to different object attributes and low-level features, aiming at the detection of video objects corresponding to the semantic concepts defined in the ontology. Object detection considers the exploitation of object characteristic features in order to apply the most appropriate detection steps for the analysis process in the form of algorithms and numerical data generated off-line by training (e.g., color models). Furthermore, the incorporation of object spatial descriptions in the domain ontology reduces the search space, allowing for improved semantics extraction from multimedia content. Thus, the proposed framework provides the needed functionality in semantic transcoding [16] systems to match semantic objects and events with the user's interests and depending on the network and device capabilities, to apply different types of transcoding [17]. For example, it may be desirable to code only those parts of the video containing interesting objects or to code such interesting objects with higher quality.

The remainder of the paper is organized as follows. In Section II, a detailed description of the proposed knowledge-assisted analysis system for semantic video object detection is given. Sections III and V describe the developed ontology framework and its application to the Formula One domain respectively, while in Section IV the low-level processing algorithms employed for the analysis are described. Experimental results are presented in Section VI. Finally, conclusions are drawn in Section VII.

II. Knowledge-Assisted Multimedia Content Analysis

Intense past research in the domains of knowledge representation and reasoning with knowledge has, over the last decade, gained new interest in the context of the Semantic Web. New initiatives such as RDFS [14] and OWL (Web Ontology Language) [18] have been defined by the World Wide Web consortium (W3C) in order to render meaning to information on the web and allow for better methods of search and retrieval. As a next step, inference rules and logic are to be used by intelligent applications to derive new information from the already existing one. Among the possible knowledge representation formalisms, ontologies [19], providing a formal framework for a shared and common understanding of a domain that can be communicated between people and application systems, ontologies have become key enabling technology for the Semantic Web.

Content-based analysis of multimedia requires methods which will automatically segment video sequences and key frames into image areas corresponding to salient objects (e.g., cars, road, people, field, etc), track these objects in time, and provide a flexible framework for object recognition, indexing, retrieval and for further analysis of their relative motion and interactions. This problem of semantic video object detection can be viewed as relating symbolic terms to visual information by utilizing syntactic and semantic structure in a manner related to approaches in speech and language processing [20]. More specifically, low-level multimedia features (e.g., MPEG-7 descriptors) are assigned to semantic concepts and visual processing algorithms are assigned to object attributes thus forming an *a priori* knowledge base. Processing may then be performed by relating high-level symbolic representations to extracted features in the signal domain. Basing such a representation on an ontology, one can capture both concrete and abstract relationships between salient visual properties. Additionally, the ontological representation of the domain-specific knowledge allows for the creation of interoperable machine-processable video content semantic annotations.

Ontology modeling and ontology-based metadata creation has addressed mainly textual resources for the past decades, while in multimedia, ontologies have been mostly used in the form of thesauri-aided approaches for manual multimedia content annotation [21], [22]. However, acknowledging the importance of coupling domain-specific and multimedia low-level description vocabularies for analysis purposes has recently set focus on using ontologies to drive the extraction of semantic descriptions instead of only providing a formal structure for annotations. In [23], an object ontology coupled with a relevance feedback mechanism, is introduced to facilitate the mapping of low-level to high-level features and allow the definition of relationships between pieces of multimedia information for retrieval purposes. In [24] and [25], Semantic Web technologies have been used to attach formal semantics to MPEG-7 metadata, thus making them accessible, re-usable and interoperable with other domains.

In this paper, domain knowledge is combined with object low-level features and spatial descriptions realizing an ontology-aided video analysis framework. To accomplish this, F-logic rules are used to relate the extraction of the semantic

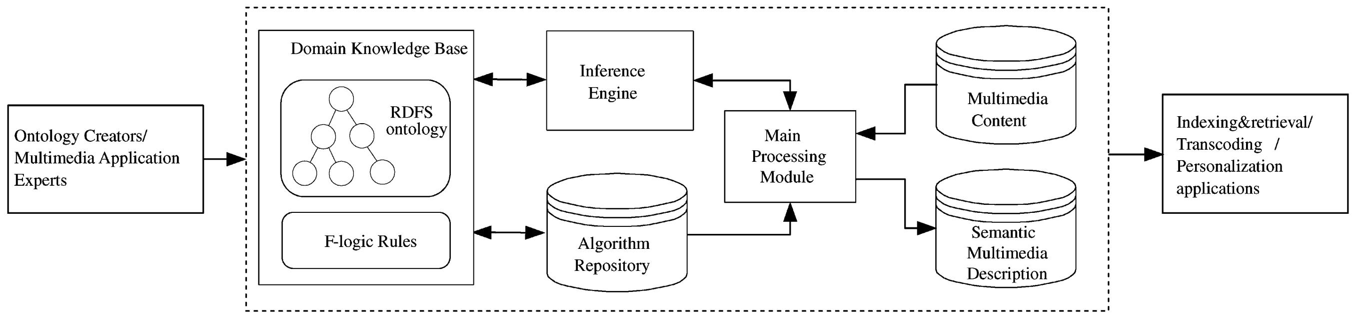


Fig. 1. Overall system architecture.

concepts, the execution order of the necessary multimedia processing algorithms and the low-level features associated with each semantic concept, thus integrating knowledge and intelligence in the analysis process. F-logic is a language that enables both ontology representation and reasoning about concepts, relations and instances [15], [26]. The general system architecture, depicted in Fig. 1, consists of a knowledge base (including the ontology schema, the respective domain instantiation data and the rules), an inference engine, the algorithm repository containing the necessary multimedia analysis tools and the system main processing module, which performs the analysis task, using the appropriate sets of tools and multimedia features, for the semantic multimedia description extraction. Ontology building is left to ontology and/or multimedia engineering experts, while the results can be used in semantic video indexing and retrieval, semantic annotation, transcoding and personalization.

Following this approach, the multimedia analysis process depends largely on the knowledge base of the system and as a result the method can be easily applied to different domains provided that the knowledge base is enriched with the respective domain information. Enriching the knowledge base with spatiotemporal object relations and event definitions would provide the means for additionally supporting domain-specific semantics extraction at event level on top of object level semantics (e.g., a car getting off the road, a player scoring a goal).

By building this unifying model of all aspects of the semantic multimedia analysis task, all its parts are treated as ontological concepts. Consequently, different multimedia processing algorithms can be tested by just defining corresponding instances in the ontology and providing appropriate interfaces. In addition, new semantic concepts can be defined and be automatically extracted by describing their attributes and corresponding low-level features in the ontology and populating the knowledge base with appropriate instances. Common ontology tools can be used to edit and handle the proposed ontology infrastructure.

III. Multimedia Analysis Ontology Development and Rules Construction

In order to implement the knowledge-assisted approach described in the previous section, an analysis ontology and a domain ontology are constructed. The *multimedia analysis ontology* is used to support the detection of domain specific objects, while the *domain specific ontology* is used to represent knowledge about the examined domain. The domain-independent, primitive classes comprising the analysis ontology serve as attachment points allowing the integration of the two ontologies.

Object detection depends on their characteristic features which are used to select the most appropriate algorithms for the analysis process. Consequently, the development of the proposed *analysis ontology* deals with the following concepts (RDFS classes) and their corresponding properties, as illustrated in Fig. 2. Solid lines have been used to represent the classes and properties currently employed by the proposed system, while dotted lines have been used to represent potential extensions, demonstrating the flexibility and adaptation capabilities offered by the system. Arrows not explicitly named correspond to the subclass relation.

- Class **Object**: the superclass of all video objects to be detected through the analysis process, i.e., the semantic objects defined in each domain ontology become subclasses of **Object** after integration of the two ontologies takes place. Each object instance comprises a model (prototype) for the corresponding semantic object. The *hasFeature* property is used to link an object instance to its visual description in terms of low-level features and spatial behavior. Appropriate properties have been defined to represent the implemented directional and topological object spatial relations, i.e., *above-of*, *below-of*, *left-of*, *right-of*, and *inside*, *adjacent-to*, respectively. In Fig. 2, in order to maintain readability the *hasSpatialRelation* property has been used to illustrate where the used spatial relations are placed in the ontology schema definition.
- Class **Feature**: the superclass of visual low-level features associated with each object. It is subclassed to **Connectivity**, **Homogeneity**, and **Size**. The **Connectivity** and **Homogeneity** classes are further subclassed to **Full Connectivity**, **Partial Connectivity**, **Non-Connectivity**, and **Motion Homogeneity**, **Color Homogeneity** classes respectively.
- Class **Feature Parameter**, which denotes the actual qualitative descriptions of each corresponding feature. It is subclassed according to the defined features, i.e., to **Connectivity Feature Parameter**, **Homogeneity Feature Parameter**, and **Size Feature Parameter**.
- Class **Limit**: it is subclassed to **Minimum** and **Maximum** and through the **hasLimit** property allows to pose range restrictions to the various parameters.

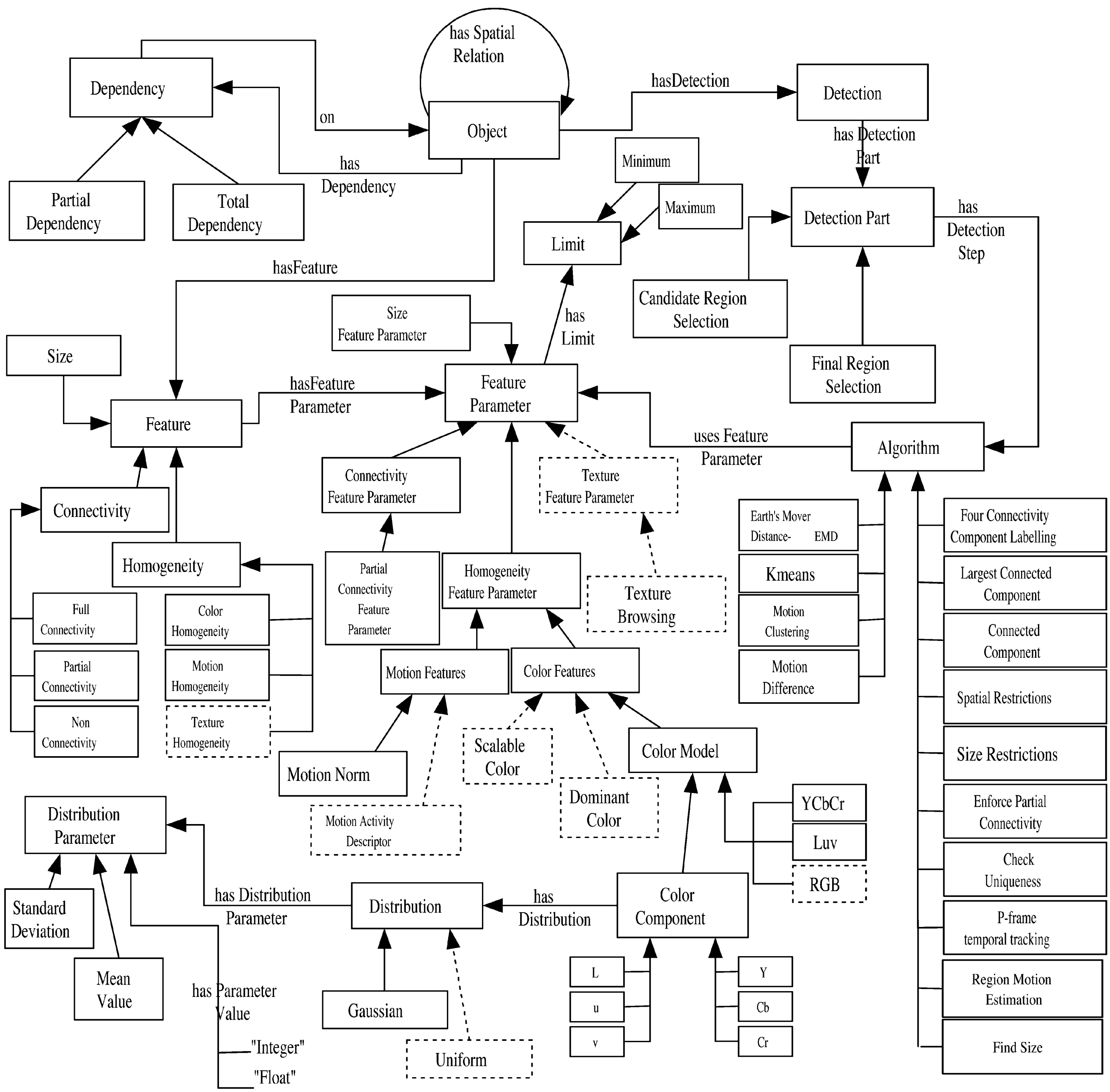


Fig. 2. Multimedia analysis ontology.

- The **Color Model** and **Color Component** classes are used for the representation of the color information. The Y, Cb, Cr components of MPEG color space are transformed to the L, u, v components used for color representation in the applied low-level processing algorithms. Consequently, all above mentioned color components are subclasses of **Color Component**.
- Class **Distribution** and **Distribution Parameter** represent information about the type of distribution (e.g., Gaussian, uniform etc) and the necessary parameters for its description respectively. The specific parameters appear as subclasses of **Distribution Parameter**, introducing in the current implementation the **Mean Value** and **Standard Deviation** classes.
- Class **Motion Features** is used to represent low-level information regarding the object motion. Its subclass, **Motion Norm**, is the the applied motion descriptor.
- Class **Algorithm**: the superclass of the available processing algorithms $(A_1, A_2, \ldots, A_n)$ to be used during the analysis procedure. This class is linked to the **FeatureParameter** class through the *usesFeatureParameter* property in order to determine the argument list for each algorithm.
- Class **Detection**: used to model the detection process, which in our framework consists of two stages. The first in-

volves finding a set of regions which are potential matches for the object to be detected, while the second one leads to the selection of only one region that best matches the criteria predefined for this object (e.g., size specifications). Thus, a **DetectionPart** class is introduced and subclassed to **CandidateRegionSelection** and **FinalRegionSelection** classes. To define the set of algorithms comprising each detection part the **DetectionPart** and **Algorithm** classes are linked through the property *hasDetectionStep*.

- Class **Dependency**: this concept addresses the possibility that the detection of one object may depend on the detection of another, due to possible spatial (or temporal) relations between the two objects. For example in the Formula One domain, the detection of the car could be assisted and improved if the more dominant and characteristic region of road is detected first. In order to differentiate between the case where the detection of object O_1 requires the detection of the candidate regions of object O_2 and the case where the detection of object O_1 requires the entire final region of object O_2, the **Dependency** class is subclassed to the **PartialDependency** and **TotalDependency** classes.

  As mentioned earlier, the sequence of algorithms to be applied for the detection of each object is directly dependent on its available characteristic attributes, i.e., visual low-level features and spatial behavior. This association is determined through a set of properly defined rules represented in F-logic. There are three kind of rules currently used in the presented approach including: rules to define the mapping between algorithms and features, which implicitly determine the execution order of an object detection steps (algorithms), rules to determine each algorithm input parameters values and rules to deal with object dependencies as explained above. The general form of each of the above defined rule categories is the following.

- "IF an object O has features $F_1 \bigcap F_2 \bigcap \dots F_n$ as part of its qualitative description THEN algorithm A_1 is a step for the detection of O."
- "IF an object O has feature F AND O has algorithm A as detection step AND A uses feature F THEN A has as input the parameter values of F."
- "IF an object O_1 has dependency on object O_2 AND object O_2 has as **CandidateRegionSelection** part the set of algorithms $S_{CRS} = \{A_1, A_2, \dots, A_m\}$ AND as **FinalRegionSelection** part the set $S_{FRS} = \{A'_1, A'_2, \dots, A'_n\}$ respectively THEN IF the dependency is partial execute the set of algorithms included in S_{CRS} before proceeding with the detection of O_1 ELSE execute both S_{CRS} and S_{FRS} before proceeding with O_1 detection."

The rules as defined up to this point do not define explicitly the execution order of the sequence of the algorithms comprising a particular object detection. However, during the analysis process, specific priority values are given to the corresponding algorithms, which affect the actual order of execution. This is accomplished by defining an additional parameter for the *hasDetectionStep* property in the defined rules. Thus, when the detection of an object O requires the execution of algorithms A_1 and A_2, during rules evaluation appropriate priorities are assigned to determine which of the two should be executed first.

In Section V, the realization of the aforementioned methodology is exemplified for the Formula One domain. The low-level processing algorithms employed for the different analysis tasks are first described in the following section to allow for better understanding.

IV. Compressed-Domain Video Processing

A. Compressed-Domain Information Extraction

To enable the efficient processing of large volumes of visual data, the proposed knowledge-based approach is applied to MPEG-2 compressed streams. The information used by the proposed algorithms is extracted from MPEG sequences during the decoding process. Specifically, the extracted color information is restricted to the discrete cosine (DC) coefficients of the macroblocks of I-frames, corresponding to the Y, Cb, and Cr components of the MPEG color space. These are transformed to the CIE Luv color space and are employed for color clustering, as will be discussed in the sequel. Additionally, motion vectors are extracted for the P-frames and are used for generating motion information for the I-frames via interpolation. P-frame motion vectors are also necessary for the temporal tracking in P-frames, of the objects detected in the I-frames [27]. Due to the limited information that is required by the proposed approach, only partial decompression of the video stream is necessary.

B. I-Frame Processing

The procedure for detecting the desired objects in I-frames, for which as previously discussed both color and motion information can be extracted is based on applying a number of low-level processing algorithms. The most important such algorithms of the analysis ontology used in the proposed knowledge assisted analysis framework are described in the sequel.

- *K-means clustering algorithm*. Color clustering is performed by identifying up to eight dominant colors in the frame, as done by the MPEG-7 Dominant Color descriptor [28], and using them to initialize a simple K-means algorithm, as in [29]. For a frame t, this results to the generation of preprocessing mask R_t^{NC}, which contains a number of nonconnected color-homogeneous regions. This mask can be used for model-based selection of semantic objects for which the homogeneity attribute is described in the ontology by the **Color Homogeneity** class and additionally the connectivity attribute is described by the **Non-Connectivity** class.
- *Four connectivity component labeling algorithm*. For objects whose connectivity attribute is described in the ontology by the **Full Connectivity** class, the four connectivity component labeling algorithm is applied to the R_t^{NC} mask to generate a preprocessing mask R_t^{CC} featuring connected color-homogeneous components. In the case of partly connected objects, described by the **Partial Connectivity** class, the four connectivity component labeling algorithm is also applied and subsequently the information in masks R_t^{NC} and R_t^{CC} is combined to generate suitable preprocessing masks in accordance with the descriptions in the ontology, e.g., "color-homogeneous object b may be represented by more than one connected component,

but each should account for at least $\beta\%$ of the total area of the color cluster". This specific restriction is represented in the ontology using the *hasLimit* property of the **FeatureParameter** class and an appropriate instance of the **Min** class.

- *Earth Mover's Distance (EMD).* Color-model-based selection of the region corresponding to a semantic object described by the **Color Homogeneity** class is performed using a suitable preprocessing mask and the EMD [30]. The EMD computes the distance between two distributions, which are represented by signatures, and is defined as the minimum amount of work needed to change one signature into the other. The notion of "work" is based on the user-defined ground distance, which is the distance between two features; in this work, the Euclidean distance is employed to this end. The signatures involved in the computation of the EMD are defined as

$$S = \{s_j = (\mathbf{m}_j, w_j)\}$$

where $\mathbf{m}_j$ represents a d-dimensional point (e.g., the three mean color values corresponding to a histogram bin) and w_j is the weight associated with this point (e.g., the nonzero value of the corresponding histogram bin; empty bins can be omitted). For each examined region of the appropriate preprocessing mask (e.g., R_t^{NC}, R_t^{CC}), its histogram is calculated and is treated as its signature. Regarding the color-model signature, a set of a few points in the three-dimensional color space and the corresponding nonzero values of the continuous model (e.g., Gaussian $\{(\mu_k, \sigma_k)\}$) are easily extracted, given the continuous model. The region for which the model-cluster EMD is minimum is selected as representative of the semantic object and is marked with a distinct label in the final mask R_t^F.
- *Motion-based clustering algorithm.* For objects whose homogeneity attribute is described in the ontology by the **Motion Homogeneity** class, motion-based grouping of color homogeneous regions is performed rather than motion-based clustering. Thus, color clustering is initially performed using the K-means algorithm. The four connectivity component labeling algorithm is subsequently applied if their connectivity is described by the **Full Connectivity** class. Then, motion based grouping of the components of the corresponding preprocessing mask is performed, similarly to [31].
- *Region motion estimation algorithm.* Region motion estimation is performed using the macroblock motion vectors extracted from the compressed stream and global motion compensation. The latter is necessary for the calculation of the object motion when the camera itself is moving. To this end, the global motion is estimated from the macroblock motion vectors using an iterative rejection procedure based on the bilinear motion model [32].

C. P-Frame Processing

In order to detect semantic objects in P-frames in the absence of color information, temporal macroblock tracking can be performed using the motion information associated with them in the compressed stream and the final semantic labeling mask extracted for the preceding frame. In this work, the temporal tracking is based upon the work presented in [33], where objects are marked manually at first, by selecting their constituent macroblocks, and subsequently tracked in the compressed domain using the respective macroblock motion vectors.

More specifically, let $\tau(\cdot)$ be the tracking operator realizing the tracking process of [33], which for an input macroblock at time $t-1$ outputs the corresponding macroblock (or macroblocks) at time t. The correspondence is established by estimating the overlapping of the examined macroblock with its spatially adjacent ones, which are determined using the displacement indicated by its motion vector. Consequently, in order to perform temporal tracking of the detected semantic objects, the $\tau(\cdot)$ operator is applied to all macroblocks of the semantic labeling mask that were identified as belonging to the corresponding semantic objects.

V. Domain Knowledge Infrastructure

Following the proposed methodology, the *multimedia analysis ontology* described in Section III and the low-level processing algorithms described in Section IV can be applied to a specific video stream in order to automatically detect objects appearing in the stream. For this purpose, a *domain ontology* is needed to provide the vocabulary and background knowledge of the examined domain, i.e., the semantically significant object concepts and their properties. In the context of video understanding this maps to the important objects, their qualitative and quantitative attributes and their spatial relations. Populating the *domain ontology* results in obtaining a set of instances that comprise the prototypes (models) for the semantic objects to be detected. To perform semantic object detection, the domain ontology is merged with the *analysis ontology* on the basis of the semantically equivalent concepts of **Object**, **Feature**, and **Feature Parameter** defined in both ontologies as will be explained in the following.

As mentioned earlier, the proposed approach will be demonstrated on the Formula One domain. Alternative domains such as soccer video (with objects like field, player, ball) and beach vacations (with objects like sea, sky, sand) can be analyzed using the same multimedia analysis ontology and the appropriate domain ontology, as presented in the experimental results section. The set of visual low-level and spatial descriptions associated with each semantic object drive the applied analysis tasks, i.e., the differences in the domain object definitions indicate the different processing methods that should be applied for their identification. Consequently, the attributes to be included in each object domain ontology definition are selected on the basis of their suitability as distinctive characteristics for the analysis to follow.

A. Formula One Domain Ontology

Following the aforementioned, the object concepts definitions, which subclass the **Object** class defined in the *multimedia analysis ontology*, for the Formula One domain are the following.

- **Car**: a motion homogeneous, fully connected region *inside* the **Road** object whose motion norm must be above a minimum value and whose size can not exceed a predefined maximum value.
- **Road**: a color homogeneous, fully connected region, whose size has to exceed a predefined minimum value and additionally to be the largest such region in the video.
- **Grass**: a color homogeneous, partly connected region with the requirement that each of its components has a minimum predefined size and is spatially related to the **Road** and **Sand** objects through the *adjacent-to* relation.
- **Sand**: a color homogeneous, partly connected region, with the requirement that each of its components has a size exceeding a predefined minimum and additionally is *adjacent-to* the **Road** and **Grass** objects.

A snapshot of OntoEdit [34] containing the result of merging the multimedia analysis ontology described in Section III and the described Formula One domain ontology, is illustrated in Fig. 3. As can be seen, the developed domain ontology focuses mainly on the representation of object visual attributes (e.g., connectivity, homogeneity) and spatial relations and in the current version does not include spatiotemporal modeling and event definitions. The left part visualizes the concept hierarchy, while on the right part instantiations of different subparts of the resulting ontology are shown. More specifically, for each of the defined domain objects the features, spatial relations and dependencies can be seen, as well as how the actual values corresponding to the color prototypes of object **Grass** are associated with it. Support is provided for multiple visual low-level and spatial definitions, since real world objects tend to have different instantiations. Consequently, providing more than one object models, e.g., color models, proves advantageous in terms of allowing for more complete object description as further discussed in the experimental results section.

B. Rules for the Formula One Domain

The mapping of the generic content rules to domain specific ones is quite straightforward and derives directly from the video processing methodology detailed in Section IV. As follows from Section IV, color clustering is the first step for the detection of any of the three objects. Consequently, a rule of the first category without any condition is used to add the *k-means* algorithm to the **CandidateRegionSelection** part of all domain objects detection, and more specifically as the first one to be executed.

Similarly, since the EMD algorithm is used for evaluating the degree of matching between two color models, the detection for the defined as color homogeneous domain objects includes the EMD algorithms in their **CandidateRegionSelection** part.

"IF [(Object O hasFeature ColorHomogeneity) AND (O hasDetection D)

AND (D hasDetectionPart P) AND (P is instance of CandidateSelectionPart)]

THEN [P hasDetectionStep EMD]."

As another example let us consider the construction of a rule to determine the parameters used as input for the EMD algorithm when applied for the detection of a specific color homogeneous object O_1. As was mentioned before, the EMD computes the distance between two distributions represented by signatures.

"IF [(Object O hasDetection D) AND (D hasDetectionPart P)

AND (P hasDetectionStep EMD) AND (O hasFeatureParameter C)

AND (C is instance of ColorModel) AND (C hasColorComponent T)

AND (T hasDistribution R) AND (R hasDistributionParameter V)]

THEN [EMD usesFeatureParameter V]."

Having an ontology representation of the *a priori* knowledge regarding the object qualitative (i.e., visual and spatial descriptions) and quantitative descriptions as well as properly defined rules to deal with the algorithmic issues of the actual processing, the way to communicate this knowledge to the video analysis system has to be established. To accomplish this, appropriate queries have to be formulated and communicated to the knowledge base in order to retrieve the required knowledge. Queries are also written in F-logic. The queries are of three types, in correspondence to the three rule categories: queries to retrieve the algorithms required for the detection of a particular object, queries for obtaining the values required as input for an algorithm, and queries considering the potential object dependencies.

In the following, an example query asking the algorithms required for the detection of object **Car** and its output are presented, assuming that the object **Road** has already been detected.

"FORALL Object O,Part P, Order N, Algorithm A such as (O is instance of Car)

AND (O hasDetection D) AND (D hasDetectionPart P) AND (P hasDetectionStep$_N$ A)

RETURN O, A, P, N."

The answer consists of a list of quartets. The variable *Part* is used to identify whether an algorithm belongs to the **CandidateRegionSelection** or the **FinalRegionSelection** detection part, using the values 1.0 and 2.0, respectively, while the variable *Order* refers to execution order. After obtaining the list of algorithms required for the detection of object **Car**, the values comprising each algorithm input have also to be retrieved from the knowledge base, by appropriate queries formulation.

Object = TemplateCar Algorithm = Kmeans_ins
Part = 1.0 Order = 1.0
Object = TemplateCar Algorithm = FourConnectivityComponentLabeling_ins
Part= 1.0 Order = 2.0
Object = TemplateCar Algorithm = RegionMotionEstimation_ins
Part= 1.0 Order = 4.0
Object = TemplateCar Algorithm = MotionDifference_ins
Part= 2.0 Order = 5.0
Object = TemplateCar Algorithm = FindSize_ins

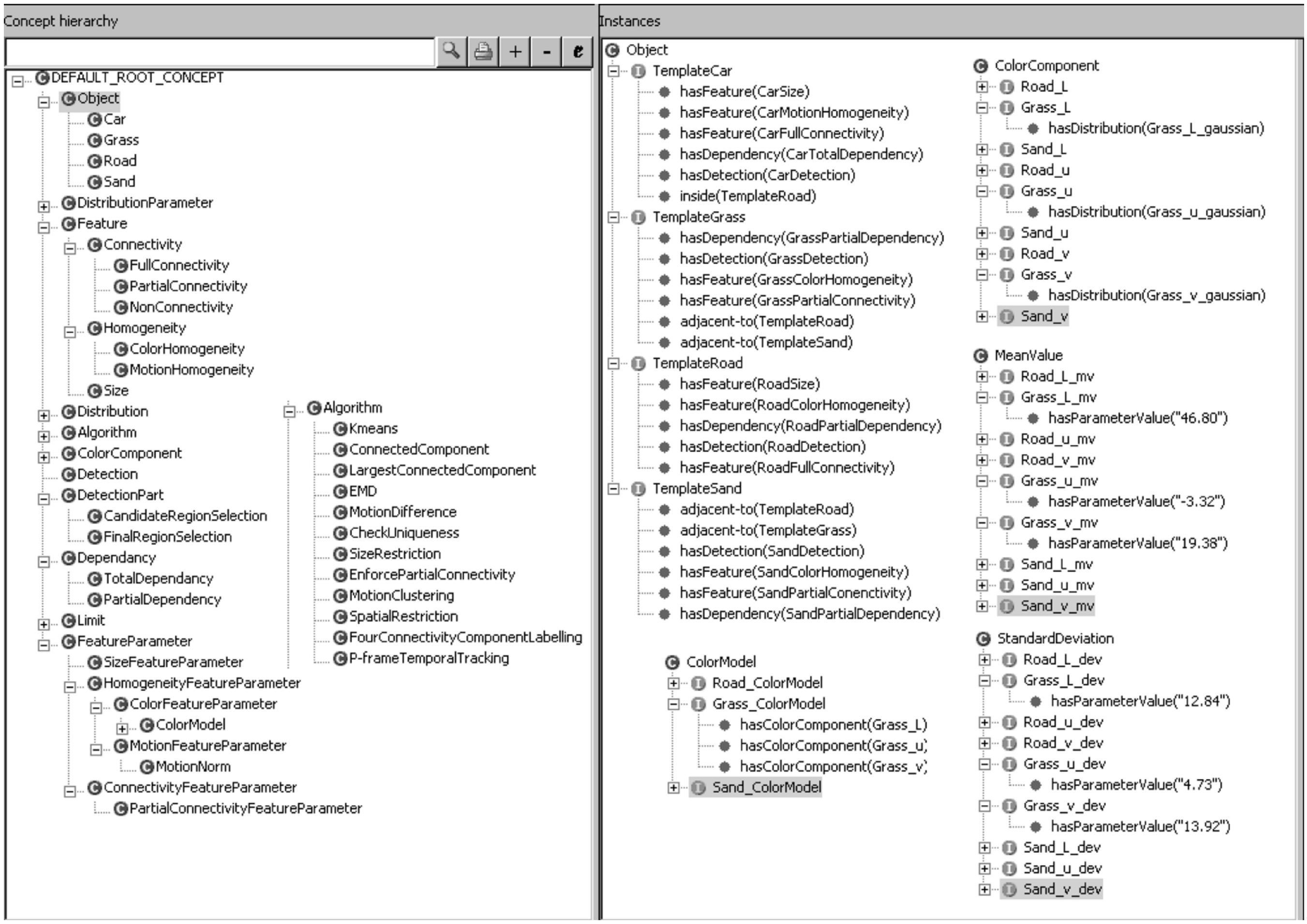


Fig. 3. OntoEdit snapshot illustrating the result of merging the multimedia analysis ontology and the Formula One domain ontology.

Part= 2.0 Order = 7.0
Object = TemplateCar Algorithm = SizeRestriction_ins
Part= 2.0 Order = 8.0
Object = TemplateCar Algorithm = MotionBasedClustering_ins
Part= 2.0 Order = 9.0
Object = TemplateCar Algorithm = SpatialRestrictions_ins
Part= 2.0 Order = 15.0

As illustrated, the query resulted in the sequence of algorithms that have to be executed in order to detect objects belonging to the semantic class **Car**. The nonsequential values of the variable *Order* reflect the fact that not all algorithms available in the repository are needed for a particular object detection. Instead, as previously mentioned, the required algorithms are determined according to the definitions included in the domain ontology and the corresponding rules. Furthermore, although there is a dependency between the **Car** and **Road** objects, no algorithms are returned to account for the detection of the **Road** object since it is assumed to have already been detected; thus, the algorithm that checks consistency with the spatial descriptions as defined in the domain ontology can be applied directly.

The ontological representation of the domain knowledge allows the formulation of other queries as well, such as "which of the objects are connected and color-homogeneous" or "what type of distribution characterizes the color components of the road" etc., as would for any ontology repository. However, this

TABLE I
OBJECT COLOR MODELS USED FOR THE FORMULA ONE DOMAIN

Object	μ_L	μ_u	μ_v	σ_L	σ_u	σ_v
Road	53.51	0.43	0.77	6.61	3.06	4.64
	42.94	-1.61	-1.61	5.84	1.54	4.01
	47.08	-0.97	1.61	6.06	3.20	3.79
	59.94	1.90	1.79	5.86	2.19	4.28
	70.58	-5.89	-17.32	7.15	1.63	4.35
	64.72	0.16	2.97	6.59	1.07	3.86
	56.30	-4.05	-2.52	15.87	3.05	5.48
	50.06	-2.92	-9.01	12.28	2.29	8.79
	59.53	3.79	1.09	9.58	21.93	9.19
	57.14	-18.75	-32.35	14.84	8.60	13.59
Grass	46.80	-3.32	19.38	12.84	4.73	13.92
Sand	76.64	2.96	8.25	10.78	2.85	6.56

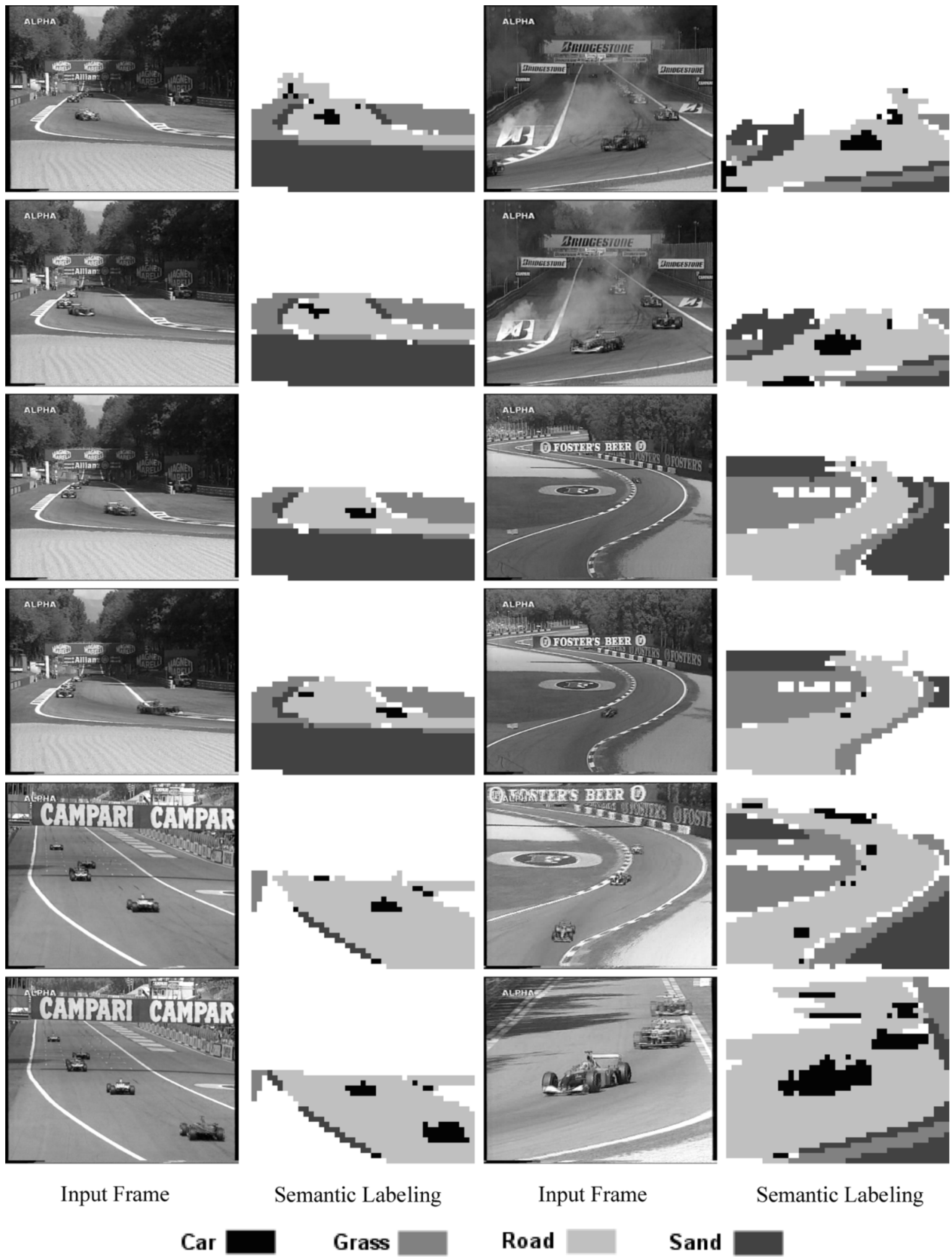


Fig. 4. Formula One video results. Macroblocks identified as belonging to no one of the domain semantic objects are shown in white.

kind of information is not practically useful in terms of the presented analysis. It must be noted that query formulation, as well as ontology development is left to knowledge and/or multimedia engineers, as also mentioned in Section II. Consequently, the system complexity is transparent to end users who exploit the system output to perform search and retrieval tasks in the semantically annotated content.

VI. Experimental Results

The proposed approach was tested in the Formula One, soccer and beach vacations domains. For these domains, appropriate domain ontologies were defined. In all cases, the exploitation of the knowledge contained in the system ontology and the associated rules resulted to the application of the appropriate analysis algorithms using suitable parameter values, for the detection of the domain specific objects. The OntoEdit ontology engineering environment [34] was used for ontology creation. Inference and query services are realized using the OntoBroker inference engine [35]. Since OntoBroker uses F-Logic as its internal representation language and OntoEdit supports the representation of the developed ontologies in RDFS, DAML, or F-Logic, the latter was chosen as the output language. It must be noted that

TABLE II
NUMERICAL EVALUATION FOR THE FORMULA ONE DOMAIN

	Proposed approach			Region-based extension of [36]		
Object	correct detections	false detections	missed	correct detections	false detections	missed
Road	97%	2%	1%	81%	5%	14%
Grass	90%	7%	3%	70%	19%	11%
Sand	88%	8%	4%	63%	23%	14%
Car	74%	22%	7%	43%	24%	33%

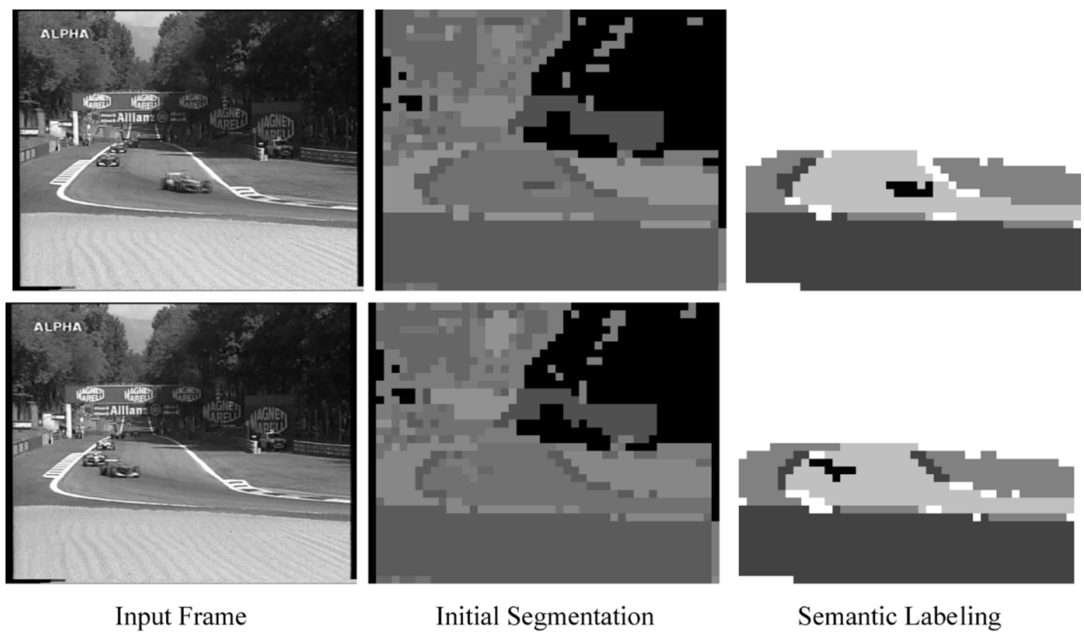


Fig. 5. Indicative initial segmentation results without the use of domain knowledge and corresponding final results of the proposed framework.

this is only a matter of notation for representation uniformity. The RDFS ontologies could have been used directly as OntoBroker input as well.

As described in the following, a variety of MPEG-2 videos of 720×576 pixels were used for testing and evaluation of the knowledge assisted semantic video object detection system. The presented experimental results are for both I-frames and P-frames, where for the latter the tracking algorithm of [33] as described in Section IV has been applied.

For the Formula One domain, our approach was tested on a one-hour video. As was discussed in Section V, four objects are defined for this domain. For those objects whose homogeneity attribute is described in the ontology by the **Color Homogeneity** class, the employed color models are presented in Table I. These models were extracted from a training set of approximately 5 min of manually annotated Formula One video consisting of various selected parts including different shots where the objects of interest appear. Since we assume the model to be a Gaussian distribution for each one of the three components of the color space, the color models were calculated from the annotated regions of the training set accordingly. Regarding the road object, significant color variations can be observed, thus, better detection performance may be attained using multiple models. For this purpose, ten color models were used in this work for the road object, each calculated from different temporal segments of the employed training set. In this case, the distance used for evaluating the similarity of a region with the road model defined in the ontology is defined as the minimum of the EMD distances between the region color signature and the road color models. The employed color models are presented in Table I. Results for the Formula One domain are presented both in terms of sample segmentation masks with the extracted semantic labels showing the different objects detected in the corresponding frames (Fig. 4) as well as numerical evaluation of the results over a twenty-minute segment of the test set (Table II).

In the masks of Fig. 4, macroblocks identified as belonging to no one of the four defined objects are shown in white. The initial

TABLE III
OBJECT COLOR MODELS USED FOR THE SOCCER DOMAIN

Object	μ_L	μ_u	μ_v	σ_L	σ_u	σ_v
Field	58.00	-14.39	54.59	4.79	2.70	4.13
Spectators	36.08	-2.14	0.81	9.71	7.15	14.11

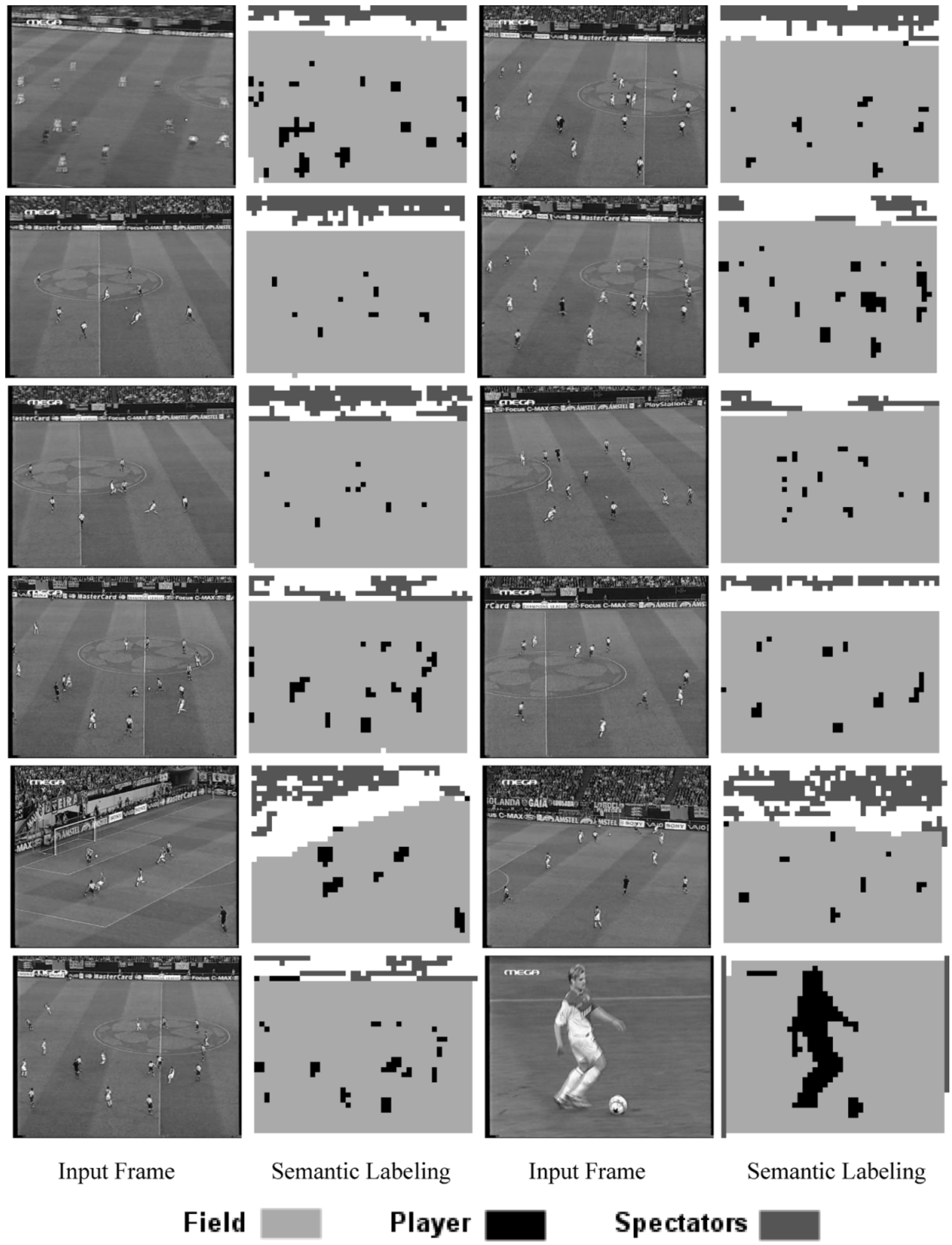


Fig. 6. Soccer video results. Macroblocks identified as belonging to no one of the domain semantic objects are shown in white.

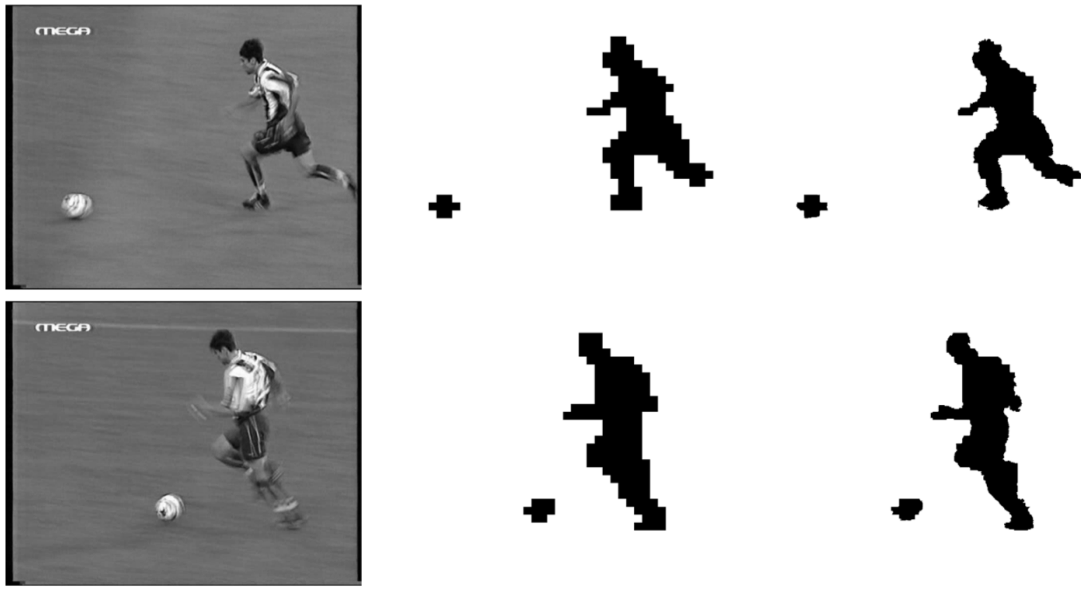
segmentation results before application of the knowledge-assisted processing are illustrated in Fig. 5 to demonstrated the added value of the proposed knowledge-assisted analysis. Comparison of the latter with the results of a region-based extension of a learning approach, based on a Bayes classifier [36], are also shown in Table II.

For the soccer domain, the following semantic objects were introduced to the domain specific ontology.

- **Player**: A motion homogeneous, fully connected region *inside* object **Field** whose motion norm must be above a minimum value and whose size can not exceed a predefined maximum value.

TABLE IV
NUMERICAL EVALUATION FOR THE SOCCER DOMAIN

Object	correct detections	false detections	missed
Field	100%	0%	0%
Player	82%	5%	13%
Spectators	70%	2%	28%



Input Frame Block-level Labelling Pixel-level Labelling

Fig. 7. Pixel-domain boundary refinement of moving objects.

- **Field**: A color homogeneous, fully connected region, whose size has to exceed a predefined minimum value and additionally to be the largest such region in the video.
- **Spectators**: A color homogeneous, partly connected region, *adjacent-to* object **Field**, with the requirement that each of its components has a minimum predefined size.

The proposed semantic analysis and annotation framework was tested on a half-hour video. A training set of approximately 5 min was employed and a single color model was extracted for each of the color homogeneous objects (Table III) in a fashion similar to the one illustrated for the Formula One domain. Segmentation masks with semantic labeling for this domain are shown in Fig. 6. Numerical evaluation of the results over a fifteen-minute segment of the test set for this domain is given in Table IV.

Since the proposed approach was applied to MPEG-2 compressed streams, the produced masks are rather coarse as illustrated by the presented results. However, the approach can be easily applied to uncompressed streams as well, if the targeted application poses such requirements. In order to provide pixel-level accuracy, appropriate algorithm instances need to be defined, either for a pixel-level segmentation to replace the compressed domain one currently used or for a pixel-level refining algorithm to be applied as a post-processing step. Indicative results of the application of the pixel-level mask refinement method of [27] to the blocky masks generated by the proposed compressed domain object detection framework are illustrated in Fig. 7.

Segmentation masks and corresponding labels are presented for the beach vacations domain in Fig. 8, using the following domain definitions:

- **Sea**: A color homogeneous, fully connected region which is *below-of* object **Sky**.
- **Sky**: A color homogeneous, fully connected region which is *above-of* objects **Sea** and **Sand**.
- **Sand**: A color homogeneous, partly connected region, which is *adjacent-to* object **Sea**.

The beach vacation domain provides, additionally, a concrete example of the benefits gained from enriching domain knowledge with spatial information. As can be seen the semantic objects **Sea** and **Sky** are correctly recognized, although having similar low-level features, due to their different spatial behavior which is reflected in the domain ontology definitions.

It must be emphasized that only the domain ontology has to be redefined in order to apply the analysis and annotation tasks to a different domain, whereas the multimedia analysis ontology and the rules remain unaltered.

VII. Conclusion

In this paper, we have presented an ontology-based approach for knowledge-assisted domain-specific video analysis. Such knowledge involves qualitative object attributes, quantitative low-level features generated by training, object spatial relations as well as multimedia processing methods. Rules in F-logic are defined to describe how tools for multimedia analysis should be applied, depending on object attributes and low-level features, for the detection of video objects corresponding to the semantic concepts defined in the ontology. A coherent architecture is achieved by using an ontology to describe both the analysis process and the domain of the examined videos. The analysis procedure, as defined by the developed ontology and rules, was applied to Formula One, soccer, and beach vacations videos and was seen to produce satisfactory results. The same methodology could be easily applied to different domains by using appropriate domain ontologies. The followed multimedia analysis approach provides a framework for ontology-based annotation of multimedia content enabling semantic transcoding and key Semantic Web functionalities such as querying, retrieval and reasoning.

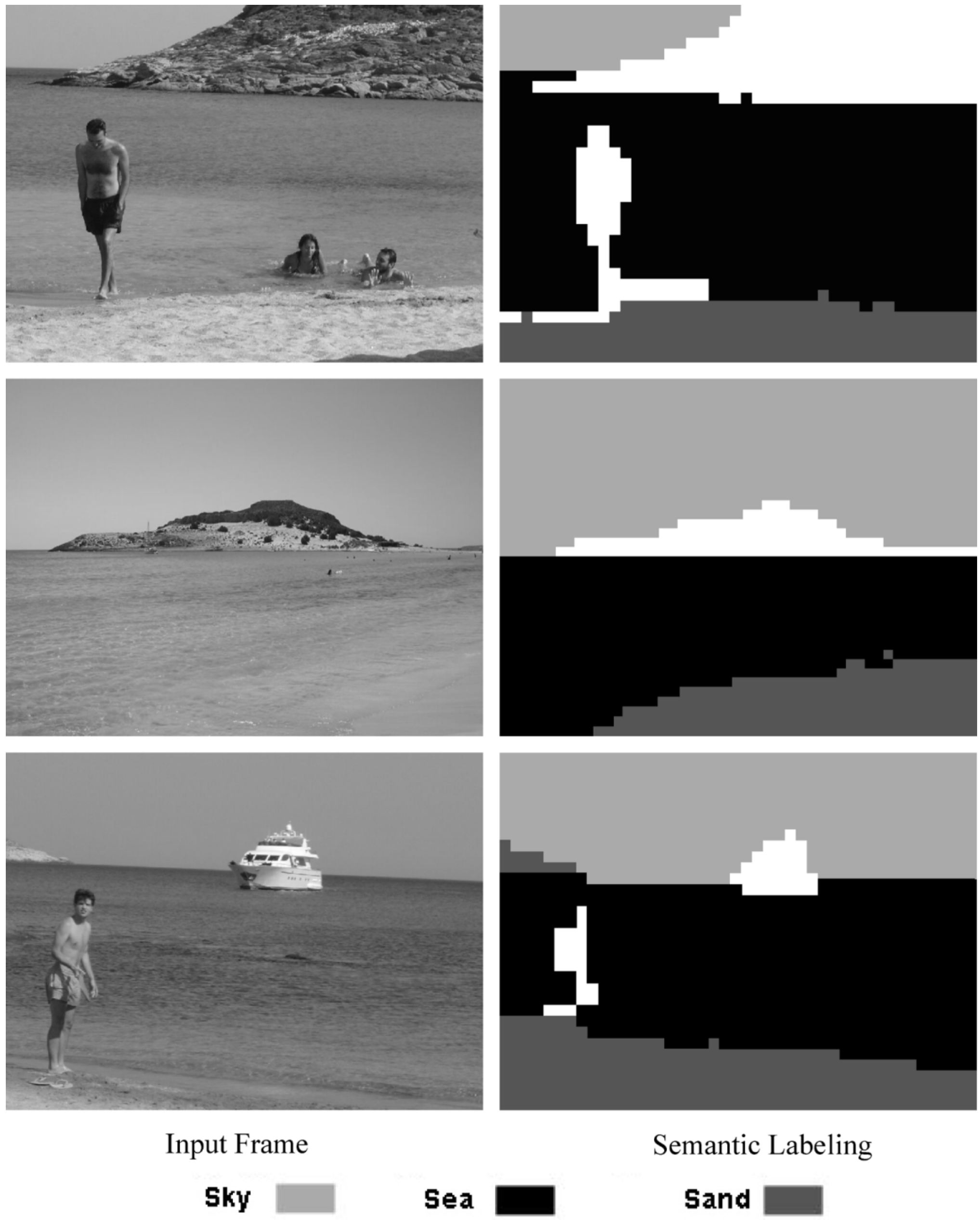


Fig. 8. Beach vacations video results. Macroblocks identified as belonging to no one of the domain semantic objects are shown in white.

Future work includes the enhancement of the domain ontology with more complex object representations, including additional low-level descriptors and spatiotemporal relations that will allow the definition of semantically important events in the domain of discourse. Further exploration of low-level multimedia features (e.g., use of the MPEG-7 standardized descriptors) is expected to lead to more accurate and, thus, efficient representations of semantic content. The above mentioned enhancements will allow for more meaningful reasoning, thus, improving the efficiency of multimedia content understanding. Another possibility under consideration is the use of a more expressive language, e.g., OWL, in order to capture a more realistic model of the specific domain semantics.

REFERENCES

[1] S.-F. Chang, "The holy grail of content-based media analysis," *IEEE Multimedia*, vol. 9, no. 2, pp. 6–10, Apr./Jun. 2002.
[2] M. Bertini, A. D. Bimbo, A. Prati, and R. Cucchiara, "Semantic annotation and transcoding for sport videos," in *Proc. Int. Workshop Image Analysis for Multimedia Interactive Services (WIAMIS 2004)*, Lisboa, Portugal, Apr. 2004, CD-ROM.
[3] S.-F. Chang, T. Sikora, and A. Puri, "Overview of the MPEG-7 standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 6, pp. 688–695, Jun. 2001.
[4] P. Salembier and F. Marques, "Region-based representations of image and video: Segmentation tools for multimedia services," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, no. 8, pp. 1147–1169, Dec. 1999.
[5] W. Al-Khatib, Y. F. Day, A. Ghafoor, and P. B. Berra, "Semantic modeling and knowledge representation in multimedia databases," *IEEE Trans. Knowl. Data Eng.*, vol. 11, no. 1, pp. 64–80, Jan./Feb. 1999.

[6] M. R. Naphade, I. V. Kozintsev, and T. S. Huang, "A factor graph framework for semantic video indexing," *IEEE Trans. Circuits. Syst. Video Technol.*, vol. 12, no. 1, pp. 40–52, Jan. 2002.
[7] M. Wallace, G. Akrivas, P. Mylonas, Y. Avrithis, and S. Kollias, "Using context and fuzzy relations to interpret multimedia content," in *Proc. 3rd Int. Workshop Content-Based Multimedia Indexing*, Rennes, France, Sep. 2003, pp. 389–396.
[8] A. Yoshitaka, S. Kishida, M. Hirakawa, and T. Ichikawa, "Knowledge-assisted content-based retrieval for multimedia databases," *IEEE Multimedia*, vol. 1, no. 4, pp. 12–21, Apr. 1994.
[9] V. Mezaris, I. Kompatsiaris, and M. G. Strintzis, "Region-based image retrieval using an object ontology and relevance feedback," *EURASIP J. Appl. Signal Process., Special Issue on Object-Based and Semantic Image and Video Analysis*, vol. 2004, no. 6, pp. 886–901, Jun. 2004.
[10] G. Tsechpenakis, G. Akrivas, G. Andreou, G. Stamou, and S. D. Kollias, "Knowledge-assisted video analysis and object detection," in *Proc. Eur. Symp. Intelligent Technologies, Hybrid Systems and Their Implementation on Smart Adaptive Systems (Eunite02)*, Algarve, Portugal, Sep. 2002, pp. 497–504.
[11] A. Hanson and E. Riseman, "The visions image-understanding system," in *Proc, Advances in Computer Vision*, vol. I, C. Brown, Ed., 1988, pp. 1–114.
[12] M. J. T. Reinders, "Model adaptation for image coding," Ph.D. dissertation, Delft University, Delft, The Netherlands, 1995.
[13] M. Addis, M. Boniface, S. Goodall, P. Grimwood, S. Kim, P. Lewis, K. Martinez, and A. Stevenson, "SCULPTEUR: Toward a new paradigm for multimedia museum information handling," in *Proc. Semantic Web*, 2003, pp. 582–596.
[14] *RDF Schema Specification 1.0, W3C Recommendation*, D. Brickley and R. V. Guha. (2004). [Online]. Available: http://www.w3.org/TR/rdf-schema/
[15] J. Angele and G. Lausen, "Ontologies in F-logic," in *International Handbooks on Information Systems* Berlin, Germany, 2004.
[16] M. Bertini, R. Cucchiara, A. D. Bimbo, and A. Prati, "Object and event detection for semantic annotation and transcoding," in *Proc. IEEE Int. Conf. Multimedia and Expo*, Baltimore, MD, Jul. 2003, pp. 421–424.
[17] K. Nagao, Y. Shirai, and K. Squire, "Semantic annotation and transcoding: Making web content more accessible," *IEEE Multimedia*, vol. 8, no. 2, pp. 69–81, Apr./Jun. 2001.
[18] *OWL Web Ontology Language Overview, W3C Recommendation*, D. L. McGuinness and F. van Harmelen. (2004). [Online]. Available: http://www.w3.org/TR/owl-features/
[19] S. Staab and R. Studer, "Handbook on ontologies," in *Int. Handbooks on Information Systems*. Berlin, Germany: Springer-Verlag, 2004.
[20] C. Town and D. Sinclair, "A self-referential perceptual inference framework for video interpretation," in *Proc. Int. Conf. Vision Systems*, vol. 2626, 2003, pp. 54–67.
[21] A. T. Schreiber, B. Dubbeldam, J. Wielemaker, and B. J. Wielinga, "Ontology-based photo annotation," *IEEE Trans. Intell. Transp. Syst.*, vol. 16, no. 3, pp. 66–74, Jul. 2001.
[22] E. Hyvonen, A. Styrman, and S. Saarela, "Ontology-based image retrieval," in *Proc. XML Finland 2002 Conf. Toward the Semantic Web anD Web Services, HIIT Publication 2002–2003*, 2002, pp. 15–27.
[23] 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*, G. Stamou, Ed. New York: Wiley, 2004.
[24] S. L. J. Hunter and J. Drennan, "Realizing the hydrogen economy through semantic web technologies," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 1, pp. 40–47, Jan. 2004.
[25] C. Tsinaraki, P. Polydoros, and S. Christodoulakis, "Interoperability support for ontology-based video retrieval applications," in *Proc. CIVR*, 2004, pp. 582–591.
[26] M. Kifer, G. Lausen, and J. Wu, "Logical foundations of object-oriented and frame-based languages," *J. ACM*, vol. 42, no. 4, pp. 741–843, 1995.
[27] V. Mezaris, I. Kompatsiaris, N. V. Boulgouris, and M. G. Strintzis, "Real-time compressed-domain spatiotemporal segmentation and ontologies for video indexing and retrieval," *IEEE Trans. Circuits. Syst. Video Technol.*, vol. 14, no. 5, pp. 606–621, May 2004.
[28] B. S. Manjunath, J.-R. Ohm, V. V. Vasudevan, and A. Yamada, "Color and texture descriptors," *IEEE Trans. Circuits. Syst. Video Technol., Special Issue on MPEG-7*, vol. 11, no. 6, pp. 703–715, Jun. 2001.
[29] V. Mezaris, I. Kompatsiaris, and M. G. Strintzis, "A framework for the efficient segmentation of large-format color images," in *Proc. Int. Conf. Image Processing*, vol. 1, 2002, pp. 761–764.
[30] Y. Rubner, C. Tomasi, and L. J. Guibas, "A metric for distributions with applications to image databases," in *Proc. IEEE Int. Conf. Computer Vision*, Bombay, India, Jan. 1998, pp. 59–66.
[31] V. Mezaris, I. Kompatsiaris, and M. G. Strintzis, "Video object segmentation using Bayes-based temporal tracking and trajectory-based region merging," *IEEE Trans. Circuits. Syst. Video Technol.*, vol. 14, no. 6, pp. 782–795, Jun. 2004.
[32] T. Yu and Y. Zhang, "Retrieval of video clips using global motion information," *Electron. Lett.*, vol. 37, no. 14, pp. 893–895, Jul. 2001.
[33] L. Favalli, A. Mecocci, and F. Moschetti, "Object tracking for retrieval applications in MPEG-2," *IEEE Trans. Circuits. Syst. Video Technol.*, vol. 10, no. 3, pp. 427–432, Apr. 2000.
[34] Y. Sure, J. Angele, and S. Staab, *OntoEdit: Guiding Ontology Development by Methodology and Inferencing*. New Yorkf: Springer-Verlag, 2002.
[35] S. Decker, M. Erdmann, D. Fensel, and R. Studer, "Ontobroker: Ontology based access to distributed and semi-structured information," in *41.Database Semantics: Semantic Issues in Multimedia Systems, Proceedings TC2/WG 2.6 8th Working Conference on Database Semantics (DS-8)*, R. Meersman, Ed. Norwell, MA: Kluwer, 1999.
[36] J. W. Han, L. Guo, and Y. S. Bao, "A novel image retrieval model," in *Proc. 6th Int. Conf. Signal Processing*, vol. 2, Aug. 2002, pp. 953–956.

Stamatia Dasiopoulou received the Diploma degree in electronic and computer engineering from the Electronic and Computer Engineering Department, Technical University of Crete, Hania, Greece in 2003. She is currently working toward the Ph.D. degree at the Electrical and Computer Engineering Department, Aristotle University of Thessaloniki, Greece.

She is also a Graduate Research Assistant with the Informatics and Telematics Institute, Thessaloniki, Greece. Her research interests include knowledge-assisted multimedia analysis, multimedia semantic annotation and knowledge discovery, semantic web technologies, and knowledge representation and reasoning.

Vasileios Mezaris (M'98) received the Diploma degree and the 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. Prior to this, he was a Postgraduate Research Fellow with the Informatics and Telematics Institute/Centre for Research and Technology Hellas, Thessaloniki, Greece, and a Teaching Assistant with the Electrical and Computer Engineering Department of the Aristotle University of Thessaloniki. His research interests include still image segmentation, video segmentation and object tracking, multimedia standards, knowledge-assisted multimedia analysis, and knowledge extraction from multimedia, content-based, and semantic indexing and retrieval.

Dr. Mezaris is a member of the Technical Chamber of Greece.

Ioannis Kompatsiaris (S'94–M'02) received the Diploma degree in electrical engineering and the Ph.D. degree in 3-D model based image sequence coding from Aristotle University of Thessaloniki (AUTH), Thessaloniki, Greece, in 1996 and 2001, respectively.

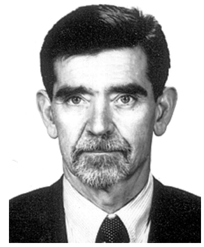
He is a Senior Researcher (Researcher D') with the Informatics and Telematics Institute, Thessaloniki. Prior to his current position, he was a Leading Researcher on two-dimensional (2-D) and three-dimensional (3-D) imaging at AUTH. His research interests include 2-D and 3-D monoscopic and multiview image sequence analysis and coding, semantic annotation ofmultimedia content, multimedia information retreival and knowledge discovery, and MPEG-4 and MPEG-7 standards. His involvement with those research areas has led to the coauthoring of two book chapters, 14 papers in refereed journals and more than 50 papers in international conferences. He has served as a regular reviewer for a number of international journals and conferences. Since 1996, he has been involved in more than 15 projects in Greece, funded by the EC, and the Greek Ministry of Research and Technology.

Dr. Kompatsiaris is a member of the IEE Visual Information Engineering Technical Advisory Panel and a member of the Technical Chamber of Greece.

Vasileios-Kyriakos Papastathis was born in Thessaloniki, Greece, in 1975. He received the Diploma degree in electrical and computer engineering from Aristotle University of Thessaloniki (AUTH), Thessaloniki, Greece, in 1999, the M.Sc. degree in advanced computing from Imperial College, University of London, London, U.K., in 2000, and the M.Sc. degree in international management from King's College, University of London, London, U.K. in 2001, respectively.

He is a Research Assistant with the Informatics and Telematics Institute, Thessaloniki, Greece. His research interests include semantic annotation of multimedia content, multimedia information retrieval and knowledge discovery, and MPEG-4 and MPEG-7 standards.

Dr. Papastathis is a member of the Technical Chamber of Greece.



Michael G. Strintzis (S'68–M'70–SM'80–F'03) 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 over 100 journal publications and over 200 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.