# IMKA: A Multimedia Organization System Combining Perceptual and Semantic Knowledge

Ana B. Benitez
Dept. of Electrical Engineering,
Columbia University
500 W. 120th Street
New York, NY 10027, US
+1-212-854-7473
ana@ee.columbia.edu

Shih-Fu Chang
Dept. of Electrical Engineering,
Columbia University
500 W. 120th Street
New York, NY 10027, US
+1-212-854-6894
sfchang@ee.columbia.edu

John R. Smith
Pervasive Media Management Group,
IBM T. J. Watson Research Center
30 Saw Mill River Road
Hawthorne, NY 10532, US
+1-914-784-7320
jrsmith@watson.ibm.com

## ABSTRACT
In this demo, we present the *IMKA* system, which implements the innovative approach of integrating perceptual information such as low-level features and images, and symbolic information such as words to represent the knowledge associated with a large multimedia collection for multimedia organization and retrieval. The IMKA system utilizes the unique MediaNet framework, which greatly extends existing knowledge representation tools in the text domain (e.g., semantic networks and WordNet) and the multimedia domain (e.g., Multimedia Thesaurus) by combining perceptual and semantic concepts in the same network and by supporting perceptual and semantic relationships among concepts exemplified by different media. It also brings the level of multimedia retrieval closer to users' needs by translating low-level feature queries to high-level semantic queries and vice versa. We will demonstrate the process of constructing the MediaNet knowledge base and new ways of searching multimedia in the IMKA system by presenting the current implementation of the IMKA system that uses image collections from online sources.

## Keywords
Knowledge representation, MediaNet, MPEG-7, knowledge construction, multimedia organization and retrieval.

## 1. MOTIVATION
Recent research on the analysis of audio-visual data has enabled multimedia information systems that support content-based retrieval, automatic but constrained classification of objects and scenes, and enhanced searching using relevance feedback from users. However, existing systems still lack adequate capabilities of representing diverse concepts associated with multimedia at the perceptual level (e.g., color and motion) as well as the semantic level (e.g., real-world and abstract concepts). Given a large multimedia collection and their associated metadata (e.g., annotations and audio-visual features), effective organization of the concepts associated with the content is an open problem.


*Conference '00*, Month 1-2, 2000, City, State.


Powerful tools exist in the text domain for representing and organizing comprehensive real-world knowledge; examples are semantic networks and WordNet [3]. WordNet is an electronic lexical system that organizes English words into sets of synonyms linked with relations such as similar to, specialization of, and part of, among others. Such knowledge tools are very useful in information retrieval. For example, queries can be restricted or expanded to resolve the ambiguity in the user's query and improve the final search results. One of our objectives is to apply and extend such knowledge representation frameworks to multimedia and related retrieval and visualization applications.

## 2. THE IMKA SYSTEM
The IMKA (Intelligent Multimedia Knowledge Application) system comprises the MediaNet multimedia knowledge representation framework, and techniques for constructing multimedia knowledge and for retrieving multimedia based on prior multimedia knowledge [1][2]. Our approach is based on integrating both symbolic and perceptual representations of knowledge.

Although the analysis and retrieval techniques of the IMKA system are generic to any kind of media, their current implementation specializes on partly annotated collections of images. The system uses the photograph collection of the Digital Library Project at the University of California, Berkeley, which is composed of about 50.000 images of plants, animals, people and landscapes. Most images include textual descriptions and other annotations regarding where and when the photograph was taken.

### 2.1 MediaNet
MediaNet is a unified knowledge representation framework that uses multimedia information for representing semantic and perceptual information about the world. The main components of MediaNet include concepts, which correspond to world entities, and relations among concepts. Concepts can represent either semantically meaningful objects (e.g., car) or perceptual patterns in the world (e.g., texture pattern). MediaNet models the traditional semantic relation types such as generalization and aggregation but adds additional functionality by modeling perceptual relations based on feature descriptor similarity and constraints. In MediaNet, both concepts and relationships are defined and/or exemplified by multimedia information such as images, video, audio, graphics, text, and audio-visual feature descriptors. MediaNet differs from related work such as the Multimedia Thesaurus [4] in combining perceptual and semantic

concepts in the same network and in supporting perceptual and semantic relationships exemplified by different media.

Weights and probabilities can be assigned to the concepts, relationships, and media representations in MediaNet to capture positive (i.e., positive weights) and negative (i.e., negative weights) examples of concepts, dynamic knowledge, and user feedback, in other words, the process of producing semantics from perceptual patterns. MediaNet can also include conceptual contexts defined as surrounding circumstances (e.g., location or time). An example of a MediaNet knowledge base is shown in Figure 1. MPEG-7 description tools including the semantic and the model description tools can be used to encode MediaNet knowledge bases [1].

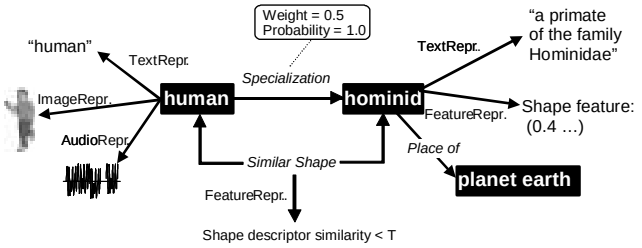


**Figure 1: Example of a MediaNet knowledge base.**

## 2.2 The Construction Process

The process of constructing multimedia knowledge from a partially annotated collection of images, consists on extracting relevant concepts and relationships among the concepts of both perceptual and semantic kinds. At the perceptual level, the images are segmented into homogeneous regions; the images and regions are then clustered hierarchically based on their visual features and the visual features of neighboring regions. Neighboring regions are considered because objects perceived in the surroundings of other objects help the recognition of objects. A perceptual concept is created for each of the resulting clusters; the hierarchical relationships among clusters are instances of perceptual relations.

The analysis at the semantic level generates semantic concepts and relationships among these concepts by processing the available textual annotations using common natural language processing techniques (tokenizer, part-of-speech tagger, parser, and stop-word remover) and disambiguating the sense of the remaining words using an original technique that uses the annotations of images that have similar low-level features and the lexical database WordNet. Concepts in the MediaNet knowledge base are created for the most relevant remaining senses. The semantic relationships provided by WordNet are used in the MediaNet knowledge base to relate those concepts.

The results of the perceptual and semantic analyses of the annotated image collection are then integrated and represented using a MediaNet knowledge base. Relationships among perceptual and semantic concepts can be found based on co-occurrences and classifiers built from the available knowledge such as Bayesian networks. Although the image collection has some textual annotations, the IMKA system combines both the perceptual and semantic knowledge of the annotated images in order to enable the discovery the concepts present in the images that were not described in the textual annotations.

## 2.3 Applications in Multimedia Retrieval

The IMKA system allows users to retrieve images from one or more search engines using multimedia knowledge. The search engines may use different descriptors (e.g., color and texture), have different searching capabilities (e.g., visual query by example and text query by keyword), and use different image collections. Initial experiments with one search engine have demonstrated improved retrieval effectiveness for the IMKA system [1][2].

The IMKA system uses the MediaNet knowledge base to preprocess incoming text or visual queries from users. First, the IMKA system classifies incoming queries into relevant semantic and perceptual concepts based on the media representations of the concepts. The initial set of relevant concepts is extended with, or reduced of, semantically and perceptually similar or dissimilar concepts, respectively. During this process, weights can be assigned to concepts, relationships and media representations to reflect user feedback, among others.

The IMKA system is designed to intelligently select and interface with multiple search engines by ranking their performance for concepts of past queries. The IMKA system issues a visual or text query to each selected search engine according to its searching capabilities, for the initial user query and for each relevant concept obtained in the preprocessing phase. The media representations of the concepts are used to generate visual and/or text queries for a concept. Finally, the results of all the queries are merged into a unique list by considering the concept(s) that generated those results and by finding the most relevant concepts of the results. The system evaluates the quality of the results returned by each search engine based on the user's feedback and updates the performance database for future queries.

# 3. CONCLUSIONS

In this demo, we present the IMKA system, which is empowered by an innovative multimedia knowledge representation framework and new tools for multimedia analysis and retrieval. The IMKA approach is based on integrating both perceptual and symbolic representations of knowledge. Therefore, it has the potential to impact a broad range of applications that deal with multimedia data at the symbolic and perceptual levels such as query, navigation, summarization and synthesis of multimedia.

# 4. REFERENCES

[1] Benitez, A. B., and Smith, J. R. New Frontiers for Intelligent Content-Based Retrieval, in Proceedings of the SPIE 2001 Conf. on Storage and Retrieval for Media Databases (San Jose CA, January 2001).

[2] Benitez, A. B., Smith, J. R., and Chang, S.-F. MediaNet: A Multimedia Information Network for Knowledge Representation. Paper in Proceedings of the SPIE 2000 Conf. on Internet Multimedia Management Systems (Boston MA, November 2000).

[3] Miller, G. A. WordNet: A Lexical Database for English. Communication of the ACM 38, 11, 39-41.

[4] R. Tansley. The Multimedia Thesaurus: Adding A Semantic Layer to Multimedia Information. Ph. D. Thesis, Computer Science, University of Southampton, Southampton UK, August 2000.