PROCEEDINGS OF SPIE

SPIE—The International Society for Optical Engineering

Internet Multimedia Management Systems

John R. Smith
Chinh Le
Sethuraman Panchanathan
C.-C. Jay Kuo
Chairs/Editors

6–7 November 2000
Boston, USA

Volume 4210

PROCEEDINGS OF SPIE
SPIE—The International Society for Optical Engineering

Internet Multimedia Management Systems

John R. Smith
Chinh Le
Sethuraman Panchanathan
C.-C. Jay Kuo
Chairs/Editors

6–7 November 2000
Boston, USA

Sponsored and Published by
SPIE—The International Society for Optical Engineering

SPIE
P
PROCEEDINGS SERIES

Volume 4210

SPIE is an international technical society dedicated to advancing engineering and scientific applications of optical, photonic, imaging, electronic, and optoelectronic technologies.

Please use the following format to cite material from this book:

Author(s), "Title of paper," in *Internet Multimedia Management Systems*, John R. Smith, Chinh Le, Sethuraman Panchanathan, C.-C. Jay Kuo, Editors, Proceedings of SPIE Vol. 4210, page numbers (2000).

ISSN 0277-786X
ISBN 0-8194-3875-8

Published by
SPIE—The International Society for Optical Engineering
P.O. Box 10, Bellingham, Washington 98227-0010 USA
Telephone 1 360/676-3290 (Pacific Time) • Fax 1 360/647-1445
http://www.spie.org/

Printed in the United States of America.

Conference Committee

Conference Chairs

John R. Smith, IBM Thomas J. Watson Research Center (USA)
Chinh Le, LeWiz Communications, Inc. (USA)
Sethuraman Panchanathan, Arizona State University (USA)
C.-C. Jay Kuo, University of Southern California (USA)

Program Committee

Shih-Fu Chang, Columbia University (USA)
Yuan-Chi Chang, IBM Thomas J. Watson Research Center (USA)
Nevenka Dimitrova, Philips Research (USA)
Oscar Feinstein, Stanford University (USA)
Forouzan Golshani, Arizona State University (USA)
Giridharan Iyengar, IBM Thomas J. Watson Research Center (USA)
Jesse S. Jin, University of New South Wales (Australia)
Chung-Sheng Li, IBM Thomas J. Watson Research Center (USA)
Clement H. Leung, Victoria University of Technology (Australia)
Andrew S. Lou, University of California/San Francisco (USA)
Wei-Ying Ma, Hewlett-Packard Company (USA)
Mrinal K. Mandal, University of Alberta (Canada)
B. S. Manjunath, University of California/Santa Barbara (USA)
Wido Menhardt, Sedara Software (USA)
Jose Rodriguez, RealChip Inc. (USA)
Yong Rui, Microsoft Corporation (USA)
Cyrus Shahabi, University of Southern California (USA)
A. Murat Tekalp, University of Rochester (USA)
George R. Thoma, National Library of Medicine (USA)
Asha Vellaikal, Hughes Research Laboratories (USA) and University of Southern California (USA)
Wayne H. Wolf, Princeton University (USA)
Stephen T. Wong, Philips Medical Systems (USA)
HongJiang Zhang, Microsoft Research China

Introduction

Recent advances in multimedia technologies and the Internet are driving a tremendous interest in systems for accessing multimedia content. For example, novel content distribution scenarios such as those for peer-to-peer online digital music exchange are providing challenging new problems for multimedia research. Other diverse application domains relating to distance learning, online entertainment, digital photography, medical imaging, remote sensing, and interactive digital television are driving many additional requirements for multimedia content management. All of these recent developments are stimulating the tremendous interest in advanced research topics of multimedia searching, browsing, retrieving, indexing, and delivery for a diversity of Internet applications.

It is in this context that we welcome you to the SPIE Internet Multimedia Management Systems 2000 (SPIE IMMS 2000) conference. The SPIE IMMS 2000 conference reports on the recent progress in technologies and applications for Internet Multimedia Management Systems.

All of the tremendous opportunities for multimedia technologies and the Internet have been greatly fueled by the ever-increasing pervasiveness of affordable computing, the unprecedented growth in storage capacities, and the ubiquity of networking resources. Also feeding the exploding interest is the realization that multimedia content-based retrieval will profoundly impact many application domains. The emerging MPEG-7 standard is one good indication of this growing interest in interoperable multimedia searching and filtering. Furthermore, the design of the MPEG-7 standard in particular allows great opportunity for future innovation in developing efficient and effective technologies for multimedia content feature extraction, searching, filtering, and indexing. As a result, MPEG-7 will prove to be a significant driver of new multimedia research problems for many years.

Overall, the SPIE IMMS 2000 program was formed from recent papers that best reflect the important advances and trends in Internet Multimedia Management Systems. The proceedings of this conference includes nine major themes: Semantics and Knowledge Representation; Multimedia Content Description Standards: MPEG-7; Universal Multimedia Access and Transcoding; Video Content Analysis; Video Segmentation and Visualization; Content-Based Indexing; Image Analysis and Retrieval; Feature Extraction and Description; and Video Compression and Delivery.

The papers contribute greatly to the advancement of multimedia research by exploring a wide diversity of problems. Furthermore, the SPIE IMMS 2000 conference makes unique and significant contributions in the following: (1) bringing to the forefront recent research examining the enormous potential of intelligent content-based retrieval systems, (2) establishing new problem domains such as universal multimedia access, and (3) focusing on a wide range of multimedia research problems related to the goal of enabling interoperability of multimedia systems using the MPEG-7 standard.

In closing, we would like to express our deep appreciation to the SPIE IMMS 2000 technical program committee for organizing the outstanding and exciting technical program.

Finally, we thank the many contributing authors to the SPIE IMMS-2000 conference and we look forward to active participation of all conference attendees this year and in the years to come.

John R. Smith
Chinh Le
Sethuraman Panchanathan
C.-C. Jay Kuo

Contents

POSTER SESSION

MediaNet: A Multimedia Information Network for Knowledge Representation

Ana B. Benitez [*ab], John R. Smith [a], Shih-Fu Chang [b]

[a] IBM T. J. Watson Research Center, New York, NY 10532
[b] Dept. of Electrical Engineering, Columbia University, New York, NY 10027

ABSTRACT

In this paper, we present MediaNet, which is a knowledge representation framework that uses multimedia content for representing semantic and perceptual information. The main components of MediaNet include conceptual entities, which correspond to real world objects, and relationships among concepts. MediaNet allows the concepts and relationships to be defined or exemplified by multimedia content such as images, video, audio, graphics, and text. MediaNet models the traditional relationship types such as generalization and aggregation but adds additional functionality by modeling perceptual relationships based on feature similarity. For example, MediaNet allows a concept such as "car" to be defined as a type of a "transportation vehicle", but which is further defined and illustrated through example images, videos and sounds of cars. In constructing the MediaNet framework, we have built on the basic principles of semiotics and semantic networks in addition to utilizing the audio-visual content description framework being developed as part of the MPEG-7 multimedia content description standard.

By integrating both conceptual and perceptual representations of knowledge, MediaNet has potential to impact a broad range of applications that deal with multimedia content at the semantic and perceptual levels. In particular, we have found that MediaNet can improve the performance of multimedia retrieval applications by using query expansion, refinement and translation across multiple content modalities. In this paper, we report on experiments that use MediaNet in searching for images. We construct the MediaNet knowledge base using both WordNet and an image network built from multiple example images and extracted color and texture descriptors. Initial experimental results demonstrate improved retrieval effectiveness using MediaNet in a content-based retrieval system.

Keywords: MediaNet, concept, multimedia, content-based retrieval, MPEG-7, WordNet, semiotics, semantic networks, intelligence

1. INTRODUCTION

Audio-visual content is typically formed from the projection of real world entities through an acquisition process involving cameras and other recording devices. In this regard, audio-visual content acquisition is comparable to the capturing of the real world by human senses. This provides a direct correspondence of human audio and visual perception with the audio-visual content [16]. On the other hand, text or words in a language can be thought of as symbols for the real world entities. The mapping from the content level to the symbolic level by computer is quite limited and far from reaching human performance. As a result, in order to deal effectively with audio-visual material, it is necessary to model real world concepts and their relationships at both the symbolic and perceptual levels.

In order to address the problem of representing real world concepts using semantics and perceptual features, we propose the MediaNet multimedia knowledge representation framework. MediaNet represents the real world using concepts and relationships that are defined and exemplified using multiple media. MediaNet can be used to facilitate the extraction of knowledge from multimedia material and improve the performance of multimedia searching and filtering applications. In MediaNet, concepts represent real world entities. Furthermore, relationships can be conceptual (e.g., Is-A-Subtype-Of) and perceptual (e.g., Is-Similar-To). The framework offers functionality similar to that of a dictionary or encyclopedia and a

[*] Correspondence: Email: ana@ee.columbia.edu; WWW: http://www.ee.columbia.edu/~ana/; Telephone: +1-212-316-9136; Fax: +1-212-932-9421

Internet Multimedia Management Systems, John R. Smith, Chinh Le, Sethuraman Panchanathan, C.-C. Jay Kuo, Editors, Proceedings of SPIE Vol. 4210 (2000) © 2000 SPIE. · 0277-786X/00/$15.00

thesaurus by defining, describing and illustrating concepts, but also by denoting the similarity of concepts at the semantic and perceptual levels.

1.1. Related Work

Previous work has focused on the development of multimedia knowledge bases for information retrieval such as the multimedia thesaurus (MMT) [19], a central component of the MAVIS 2 system [3], and the texture image thesaurus [7]. The multimedia thesaurus provides concepts, which are abstract entities of the real world objects, semantic relationships such a generalization and specialization, and media representations of the concepts, which are portions of multimedia materials and associated features vectors. MediaNet extends this notion of relationships to include perceptual relationships that can also be exemplified and defined using audio-visual content. Furthermore, MMT treats semantic objects and perceptual information quite differently. MMT defines concepts that correspond to high-level, semantically meaningful objects in the real world with names in a language ("car" and "man") [5, 20]. However, this exclude concepts that represent patterns based on perceptual information that are not named, such as the texture patterns in the texture image thesaurus [7].

MediaNet extends the current knowledge representation frameworks by including multimedia information. By integrating both conceptual and perceptual representations of knowledge, MediaNet has potential to impact a broad range of applications that deal with multimedia content at the semantic and feature levels. In particular, we have found that MediaNet can improve the performance of multimedia retrieval applications by using query expansion, refinement and translation across multiple content modalities. In this paper, we report on experiments that use MediaNet in searching for images. We construct the MediaNet knowledge base using both WordNet and an image network built from multiple example images and extracted color and texture descriptors. Initial experimental results demonstrate improved retrieval effectiveness using MediaNet in a content-based retrieval system; however, more extensive experiments are needed.

1.2. Outline

This paper is organized as follows. In section 2, we describe the main components of MediaNet and their correspondence to principles in semiotics, semantic networks in AI, and MPEG-7 description tools. Section 3 describes the implementation of an extended content-based retrieval system, which uses the MediaNet framework. In particular, it focuses on the construction and the use of the MediaNet knowledge base. Section 4 reports on the experiments that compare the performance of the extended content-based retrieval system to a typical content-based retrieval system. Finally, section 5 concludes with a summary and open issues.

2. MEDIANET

MediaNet is a semantic network for representing real world knowledge through both symbolic and audio-visual information such as images, video, audio, graphics and text, including extracted feature descriptors. In MediaNet, real world entities are represented using concepts and relationships among concepts, which are defined and exemplified by multimedia material. For example, the concept Human can be represented by the words "human", "man", and "homo", the image of a human, and the sound recording of a human talking, and can be related to the concept Hominid by a Is-A-Subtype-Of relationship; the concept Hominid can be represented by the word "hominid" and the text definition "a primate of the family Hominidae" [+]. A graphical representation of previous example is shown in Figure 1. In the following sections, we describe the main components of MediaNet and how they map to work on multiple disciplines such as semiotics, semantic networks in AI, and MPEG-7 description tools.

1.3. Concepts

Concepts are the basic units for knowledge representation in MediaNet. Concepts refer to semantic notions of objects in the world. Objects are any elements that exist in the world such as inanimate objects, living entities, events, and properties. Examples of concepts are Human (see Figure 1) and Car that refer to any living entity human and any inanimate object car, respectively; Wedding is the concept of an event and Blue, the concept of a property. Concepts can refer to classes of objects in the world such as Car; unique and identified objects such as Ronald Reagan; and abstract objects with no physical presence in the world such as Freedom.

[+] The text definitions of words in this paper were taken from the electronic lexical system WordNet [9].

It is important to point out the differences between words in a language and concepts. In the previous examples, concepts were usually named with one word; however, we humans may have no words to designate some concepts, more than one word to designate the same concept, or no words to uniquely designate a concept. An example of the first case is the unknown texture of a specific piece of fabric. The second case corresponds to synonyms, i.e., words having the same or nearly the same meaning or sense in a language, as "human" and "man" (see Human concept in Figure 1). Finally, the third case corresponds to polysemy, i.e., a word having multiple meanings in a language, as "studio" which can be a "the workroom or atelier of an artist", "a room or set of rooms specially equipped for broadcasting radio or television programs, etc.", and "an apartment consisting of one main room, a kitchen, and a bathroom".

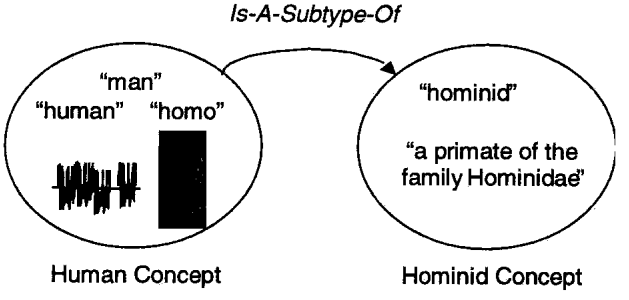

Figure 1: Concepts Human and Hominid with multiple media representations and related by an Is-A-Subtype-Of relationship.

1.4. Media Representations of Concepts

Concepts, which refer to objects in the world, can be illustrated and defined by multimedia content such as text, image, and video, among others. As an example, the concept Human is exemplified by the words "human", "man", and "homo", the image of a human, and the sound recording of a human in Figure 1.

Media representations are not necessarily whole multimedia materials such as an entire image; they can be sections of multimedia material and have associated features extracted from the media. A media representation of the concept Human can be a region of an image that contains an object human and the value of the contour shape feature for that region. Concept could also be illustrated by feature values with no multimedia material. As an example, the concept Human can be presented by the value of a contour shape feature. The value of the contour shape feature may be the average of the contour shape feature values for a set of image regions depicting a human; however, the image data is not included in the media representation only the value of the contour shape feature. Although, the image media was used to exemplify the fact that a media representation can be any section of multimedia material and/or features extracted from the multimedia material; the same applies to any other media such as text, audio, and video. Some media representations may be more or less relevant or applicable to concepts. Some example follow. Cars can be of many colors; therefore, a color feature will not be very relevant for the concept Car. Audio-based representations do not apply to the concept Blue. A relevance factor could be assigned to each representation of a concept.

We shall provide some more examples of media representations of concepts now. The concept Car can have the following media representations: the word "car", the text definition "an automobile", an image depicting a car together with shape features extracted from it, and the sound recording of a running car. The concept Blue can have the following representations: the English word "blue", the Spanish word "azul", the text definition "the pure color of a clear sky; the primary color between green and violet in the visible spectrum", and the value of the color histogram corresponding to blue. Text representations may be in different languages.

1.5. Relationships Among Concepts

Concepts can be related to other concepts by semantic and perceptual relationships. Semantic relationships relate concepts based on their meaning. All the semantic relationships in the lexical system WordNet [9] except for synonymy apply to

concepts; these are listed in Table 1 together with definitions and examples. Antonymy is a binary, symmetric relationship among concepts; the rest of the relationships in Table 1 are binary and transitive. There is usually one hypernym per concept so this relationship organizes the concepts into a hierarchical structure.

Table 1: Examples of semantic relationships with definitions and examples.

Relationship	Definition	Example
Antonymy	To be opposite to	White Is-Opposite-To Black
Hypernymy/ Hyponymy	To be a super type of To be a sub type of	Hominid Is-A-Supertype-Of Human Human Is-A-Subtype-Of Hominid
Meronymy/ Holonymy	To be a part, member, or substance of To have a part, member, or substance of	Ship Is-Member-Of Fleet Martini Has-Substance Gin
Entailment	To entail or To cause or involve by necessity or as a consequence	Divorce Entails Marry
Troponymy	To be a manner of	Whisper Is-Manner-of Speak

Concepts refer to objects that are perceived by senses. Therefore, concepts can also be related by perceptual relationships, which are based on impressions obtained by the use of senses. Examples of perceptual relationships are visual relationships (e.g., Sky Has-Similar-Color-To Sea) and audio relationships (e.g., Stork Sounds-Similar-To Pigeon). Audio-visual relationships are special because they are recorded in the audio-visual media and, therefore, can have media representations as concepts. These relationships can usually be generated and inferred by automatic tools and expressed as constraints and similarity on audio-visual features. They can also be exemplified by two media representations related with that relationship. Figure 2 exemplifies the relationship "Sounds-Similar-To" using two images.

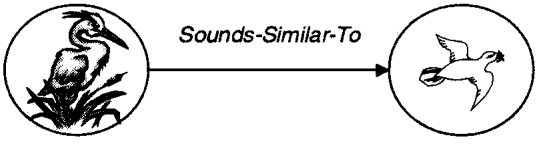

Figure 2: Example of audio relationship Sounds-Similar-To.

1.6. Semiotics View of MediaNet

Semiotics [8] is the science that analyzes signs and puts them in correspondence with particular meanings (e.g. word "car" and notion of real object car). Examples of sign systems are conversational and musical languages. We will focus on two of the basic principles of semiotics that are most useful for multimedia applications. The first principle is the decomposition of semiotics into three domains: syntax (sign; "car"), semantic (interpretant; notion of car), and pragmatics (object; real object car). The second principle, the multi-resolution, arises from modeling a unit of intelligence as three cognitive processes applied repeatedly: focusing attention, combinatorial search, and grouping (or generalization). First, attention is focused on a subset of the available data. Then, different combinations of this data are generated based on some similarity criteria. The combination or grouping providing the best results generates data at the next level. Higher-level thinking skills such as rational reasoning and intuitive thinking are a composition of these basic skills.

In MediaNet, multimedia materials are considered signs of objects in the real world and features extracted from the multimedia material are considered signs of the multimedia material (as proposed in [5]). We identify concepts in our framework with the semantics (interpretant) conferred upon these signs when users interpret them. Although the interpretation of a sign is relative to users and tasks at hand [5], MediaNet aims at recording some interpretations of multimedia and feature signs to enable multimedia applications that intelligently satisfy the user needs to browsing and search multimedia content. The information about how users would interpret a sign in one way or another for this or that task could be also included in MediaNet in the form of context information similar to the modality/disposition/epistemology context dimension proposed in [6], which specifies if a fact represent a belief or a desire, and the people who believe in it. The generalization/specialization (or hypernymy/hyponymy) relationship enables the creation of hierarchies of concepts at multiple levels implementing the semiotic principle of multi-resolution.

1.7. MediaNet as a Semantic Network

The primary objectives of the field of artificial intelligence, or AI, are to understand intelligent entities and to construct intelligent systems. Important contributions of AI have been the development of computer-based knowledge representation models such as semantic networks. Semantic networks [14] use nodes to represent objects, concepts, or situations; and arcs to represent relationships between nodes (e.g. the state "Human is a subtype of Homonid" could be represented by the chain: Human Node - Is-A-Subtype-Of Arc – Homonid Node).

The mapping of the components of MediaNet to semantic networks is as follows: concepts and media representations are represented by nodes; arcs link concepts and media representations and are labeled based on the type of representation (e.g., an image representation is linked to a concept with an Image-Representation edge); semantic relationships can be represented by arcs among the concept nodes; and, finally, perceptual relationships that have media representations should be represented by arcs instead of nodes. The example shown in Figure 1 is represented as a semantic network in Figure 3.

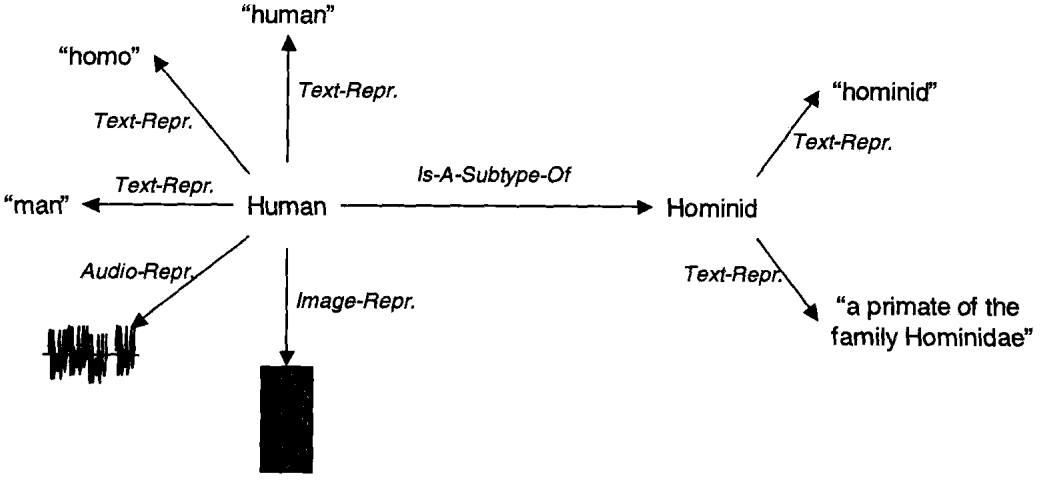

Figure 3: Semantic network corresponding to the MediaNet example shown in Figure 1.

1.8. MPEG-7 Encoding of MediaNet

The MPEG-7 standard [12] aims at standardizing tools for describing the content of multimedia material in order to facilitate a large number of multimedia searching and filtering applications. In this section, we describe and give examples of how MediaNet bases could be encoded using MPEG-7 description tools, which would greatly benefit the exchange and re-use of multimedia knowledge and intelligence among applications.

Relevant MPEG-7 tools for describing MediaNet bases are the semantic and conceptual descriptions schemes, which describe semantic and concepts of the real world as they relate to audio-visual content, and the collection description schemes, which describe collections related to audio-visual content [10,11]. As of the writing of this paper, the collection description schemes are in a more mature state than the semantic and conceptual description schemes so we have picked them for this example. However, it is important to note that these description tools are still under development within the MPEG-7 standard and their specification may change in the near future.

Collection structure description schemes allow describing a multimedia collection as multiple collection clusters (or groupings) of content description elements from the multimedia collection and relationships among these collection clusters. Content description elements can be segments, objects, events, images, and videos, among others. Collection clusters can include text annotations, statistical information (e.g. feature centroids of clusters), and relationships to other collection clusters or content description elements.

The main components of MediaNet could be mapped to the MPEG-7 description tools as follows: a MediaNet knowledge base to a collection structure, each concept to a collection cluster, and each concept relationship to a cluster relationship. The

text representations of a concept could be described as text annotations of collection clusters; while other media representations could be described as cluster elements (e.g., videos and images) and/or cluster statistics (e.g. centroid for concepts). The XML [21] description that encodes the example in Figure 1 is included below. We assume the reader is familiar with the markup language XML.

```
<CollectionStructure id="MediaNet0">
        <CollectionCluster id="ConceptHuman"> <!-- Concept Human -->
                <Text Annotation> <!-- Three textual representations of concept Human -->
                        <FreeTextAnnotation> human </FreeTextAnnotation>
                        <FreeTextAnnotation> homo </FreeTextAnnotation>
                        <FreeTextAnnotation> man </FreeTextAnnotation>
                </Text Annotation>
                <ClusterElements number="2"> <!-- Two media representations, image and audio, of concept Human -->
                        <Segment xsi:type="Image"> <MediaLocator> Human.jpg </MediaLocator> </Segment>
                        <Segment xsi:type="Audio"> <MediaLocator> Human.wav </MediaLocator> </Segment>
                </ClusterElements>
        </CollectionCluster>
        <CollectionCluster id="ConceptHominid"> <!-- Concept Hominid -->
                <Text Annotation> <!-- Two text representations of concept Human -->
                        <FreeTextAnnotation> hominid </FreeTextAnnotation>
                        <FreeTextAnnotation> a primate of the family Hominidae </FreeTextAnnotation>
                </Text Annotation>
        </CollectionCluster>
        <!-- Graph describing Is-A-Subtype-Of relationship from concept Human to concept Hominid -->
        <Graph> <ClusterRelation name="Is-A-Subtype-Of" source="ConceptHuman" target="ConceptHominid"/> </Graph>
</CollectionStructure>
```

3. INTELLIGENT CONTENT-BASED RETRIEVAL SYSTEM

MediaNet extends current knowledge representation frameworks by including multimedia information. By integrating both conceptual and perceptual representations of knowledge, MediaNet has potential to impact a broad range of applications that deal with multimedia content at the semantic and feature levels. In this section, we describe the implementation of an extended content-based retrieval system that uses MediaNet to expand, refine, and translate queries across multiple content modalities. In this section, we focus on describing the construction and use of the MediaNet knowledge base in a content-based retrieval system.

Typical content-based retrieval systems index and search image and video content by using low-level visual features (for example, see [1,4,17]). These systems automatically extract low-level features from the visual data, index the extracted descriptors for fast access in a database, and query and match descriptors for the retrieval of the visual data. The extended content-based retrieval system is a typical content-based retrieval system that also includes a MediaNet knowledge base with media representation of concepts and a query processor that uses the MediaNet knowledge base to refine, extend, or translate user queries (see Figure 4).

In the extended content-based retrieval system, we use color and texture features. The color features are color histogram and color coherence; the texture features are wavelet texture and tamura texture. Queries to the CB search engine can only be visual queries as the name of an image in the database or the value for any of the low-level features in the image database. The CB search engine uses weighted Euclidean distance function to obtain distance scores between the query and each image in the database to the query and returns an ordered list of images based on the distance scores. No text annotations are stored in the database or are used in the retrieval.

4.1. Creation of the MediaNet Base

A MediaNet knowledge base could be created manually or automatically using classification, machine learning, and artificial intelligence tools among others. The MediaNet knowledge base for the extended content-based retrieval system was created semi-automatically using existing text annotations for some of the images in the database, the electronic lexical system WordNet, and human input. In this section we describe the procedure followed to construct the MediaNet base.

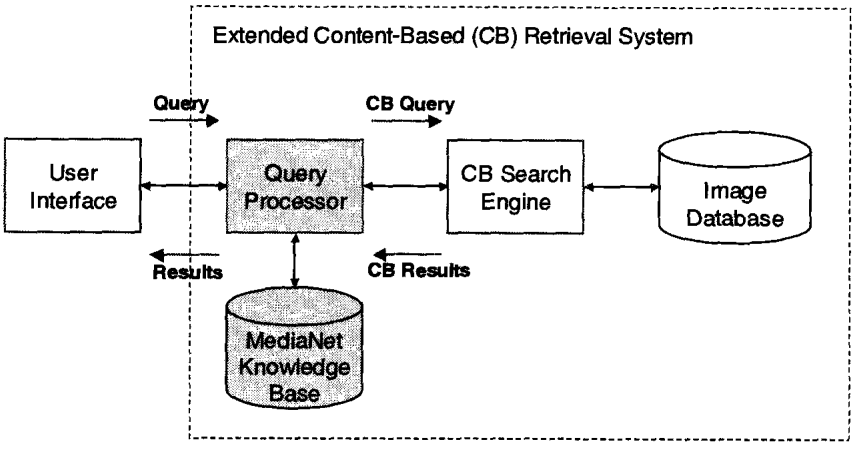

Figure 4: Components of the extended content-based retrieval engine. The new components with respect to typical content-based retrieval engines are shown in gray.

WordNet [9] is an electronic lexical system that organizes English words into set of synonyms, each representing a lexicalized concept. Semantic relationships link the synonym sets. The semantic relationships between words and words senses incorporated by WordNet includes synonymy, antonymy, hypernymy/hyponymy, meronymy/holonymy, entailment, and troponymy (these relationships are defined in Table 1 except for synonymy). WordNet contains more than 118,000 different word forms and more than 90,000 word senses. Being available in electronic form, it is one of the most convenient tools for generating concepts, text representations of concepts, and relationships among concepts at the conceptual level.

The first step for constructing the MediaNet knowledge was to create concepts and text representations of the concepts using WordNet and human assistance. First, the text annotations were serialized into words that were ordered alphabetically. Dummy words such a prepositions and articles and duplicated words were removed. Then, WordNet was used to generate the senses and the synonyms for each word. A human supervisor selected the correct sense (or senses) of each word in the context of the text annotations and the image data. As an example, for the annotation "singer in studio", the correct sense for the word "studio" was selected as "a room or set of rooms specially equipped for broadcasting radio or television programs, etc.". The supervisor also specified the syntactic category for each word (noun, verb, etc.). A concept was created for each word/sense pair, and was assigned the sense and synonyms provided by WordNet as text representations.

The next step was to generate relationships among concepts. We decided to use the top three relationships listed in Table 1. We used WordNet to automatically generate all of the antonyms, the hypernyms/hyponyms, and the meronyms/holonyms for each concept (i.e., each word-sense pair), automatically parsed the output of WordNet, obtained the relationships among the concepts, and stored them in the MediaNet knowledge base.

Finally, visual representations of concepts were generated automatically using color and texture feature extraction tools. For all the images associated with a concept, we extracted color and texture features and computed the feature centroids (centroid of group of images). The visual representation of each concept was the list of images for the concept with associated feature values, and the feature centroid.

For each application, the list of concepts and relationships in the MediaNet knowledge base should be representative of the content in the database and the goal of the application task. We used the textual annotations provided by MPEG-7, which were quite brief and general. More specific and representative text annotations could have been generated and used to construct the knowledge base. The process of generating the concepts from the words in the textual annotations could be automated by processing the text annotations using natural language techniques, or by using latent semantic analysis to match each word and surrounding words in the annotations to the different senses of the word provided by WordNet, among others.

More advanced techniques could have been used to generate more suited visual representations of the concepts. Some ideas are selecting more than one feature representation for each concept using Kohonen feature map on the feature vectors of the images associated to the concept [7], latent semantic analysis techniques applied to feature vectors as in [22], assigning weights to