Preface

Today most documented information is in digital from. Digital information, in turn, is rapidly moving from textual information to multimedia information that includes images, audio and video content. Yet searching and retrieving required information is a challenging and arduous task, because it is difficult to access just the required parts of information stored in a database. In the case of text documents, the table of contents serves as an index to different sections of the document. However, creating a similar index that points to different parts of multimedia content is not an easy task. Manual indexing of audiovisual content can be subjective, as there are several ways to describe the multimedia information depending on the user, the purpose of use, and the task that needs to be performed. The problem gets even murkier, as the purpose for retrieval is often completely different from the purpose for which the content was created, annotated and stored in a database.

Work in the area of multimedia information retrieval started with techniques that could automatically index the content based on some inherent features that could be extracted from one medium at a time. For example, features that can be extracted from still images are colour, texture and shape of objects represented in the image. In the case of a video, static features such as colour, texture and shape are no longer adequate to index visual content that has been created using powerful film editing techniques that can shape viewers experiences. For audios, the types of features that can be extracted are pitch, tonality, harmonicity, and so forth, which are quite distinct from visual features.

Feature extraction and classification techniques draw from a number of disciplines such as artificial intelligence, vision and pattern recognition, and signal processing. While automatic feature extraction does offer some objective measures to index the content of an image, it is insufficient for the retrieval task, as information retrieval is based on the rich semantic notions that humans can conjecture in their minds while retrieving audiovisual information. The other alternative is to index multimedia information using textual descriptions. But this has the problem of subjectivity, as it is hard to have a "generic" way to first describe and then retrieve semantic information that is universally acceptable. This is inevitable as users interpret semantics associated with the multimedia content in so many different ways, depending on the context and use of the information. This leads to the problem of managing multiple semantics associated

with the same material. Nevertheless, the need to retrieve multimedia information grows inexorably, carrying with it the need to have tools that can facilitate search and retrieval of multimedia content at a semantic or a conceptual level to meet the varying needs of different users.

There are numerous conferences that are still addressing this problem. Managing multimedia semantics is a complex task and continues to be an active research area that is of interest to different disciplines. Individual papers on multimedia semantics can be found in many journals and conference proceedings. Meersman, Tari and Stevens (1999), present a compilation of works that were presented at the IFIP Data Semantics Working Conference held in New Zealand. The working group focused on issues that dealt with semantics of the information represented, stored and manipulated by multimedia systems. The topics covered in this book include: data modeling and query languages for multimedia; methodological aspects of multimedia database design, information retrieval, knowledge discovery and mining, and multimedia user interfaces. The book covers six main thematic areas. These are: Video Data Modeling and Use; Image Databases; Applications of multimedia systems; Multimedia Modeling; Multimedia Information retrieval; Semantics and Metadata. This book offers a good glimpse of the issues that need to be addressed from an information systems design perspective. Here semantics is addressed from the point of view of querying and retrieving multimedia information from databases.

In order to retrieve multimedia information more effectively, we need to go deeper into the content and exploit results from the vision community, where the focus has been in understanding inherent digital signal characteristics that could offer insights into semantics situated within the visual content. This aspect is addressed in Bimbo (1999), where the focus is mainly on visual feature extraction techniques used for content-based retrieval of images. The topics discussed are image retrieval by colour similarity, image retrieval by texture similarity, image retrieval by shape similarity, image retrieval by spatial relationships, and finally one chapter on content-based video retrieval. The focus here is on low-level feature-based content retrieval. Although several algorithms have been developed for detecting low-level features, the multimedia community has realised that content-based retrieval (CBR) research has to go beyond low-level feature extraction techniques. We need the ability to retrieve content at more abstract levels — the levels at which humans view multimedia information. The vision research then moved on from low-level feature extraction in still images to segment extraction in videos. Semantics becomes an important issue when identifying what constitutes a *meaningful* segment. This shifts the focus from image and video analysis (of single features) to synthesis of multiple features and relationships to extract more complex information from videos. This idea is further developed in Dorai and Venkatesh (2002), where the theme is to derive high-level semantic constructs from automatic analysis of media. That book uses media production and principles of film theory as the bases to extract higher-level semantics in order to index video content. The main chapters include applied media aesthetics, space-time mappings, film tempo, modeling colour dynamics, scene determination using auditive segmentation, and determining effective events.

In spite of the realisation within the research community that multimedia research needs to be enhanced with semantics, research output has been discipline-based. Therefore, there is no single source that presents all the issues associated with modeling,

representing and managing multimedia semantics in order to facilitate information retrieval at a semantic level desired by the user. And, more importantly, research has progressed by handling one medium at a time. At the user level, we do know that multimedia information is not just a collection of monomedia types. Although each media type has its own inherent properties, multimedia information has a coherence that can only be perceived if we take a holistic approach to managing multimedia semantics. It is our hope that this book fills this gap by addressing the whole spectrum of problems that need to be addressed in order to manage multimedia semantics, from an application perspective, that adds value to the user community.

OUR APPROACH TO ADDRESS THIS CHALLENGE

The objective of the book — managing multimedia semantics — is to assemble in one comprehensive volume the research problems, theoretical frameworks, tools and technologies that contribute towards managing multimedia semantics. The complexity of managing multimedia semantics has given rise to many frameworks, models, standards and solutions. The book aims to highlight both current techniques and future trends in managing multimedia semantics.

We systematically define the problem of multimedia semantics and present approaches that help to model, represent and manage multimedia content, so that information systems deliver the promise of providing access to the rich content held in the vaults of multimedia archives. We include topics from different disciplines that contribute to this field and synthesise the efforts towards addressing this complex problem. It is our hope that the technologies described in the book could lead to the development of new tools to facilitate search and retrieval of multimedia content at a semantic or a conceptual level to meet the varying needs of the user community.

ORGANISATION OF THIS BOOK

The book takes a close look at each piece of the puzzle that is required to address the multimedia semantic problem. The book contains 16 chapters organised under five sections. Each section addresses a major theme or topic that is relevant for managing multimedia semantics. Within a section, each chapter addresses a unique research or technology issue that is essential to deliver tools and technologies to manage the multimedia semantics problem.

Section 1: Semantic Indexing and Retrieval of Images

Chapters 1, 2 and 3 deal with semantic indexing, classification and retrieval techniques related to images.

Chapter 1 describes a feature-based indexing technique that uses low-level feature vectors to index and retrieve images from a database. The interesting aspect of the architecture here is that the feature vector carries some semantic properties of the image along with low-level visual properties. This is moving one step towards semantic indexing of images using low-level feature vectors that carry image semantics.

Chapter 2 addresses the semantic gap that exists between a user's query and low-level visual features that can be extracted from an image. This chapter presents a state-of-the-art review of pattern classifiers in content-based image retrieval systems, and then extends these ideas from pattern recognition to object recognition. The chapter presents three new indexing schemes that exploit pattern classifiers for semantic indexing.

Chapter 3 takes the next step in the object recognition problem, and proposes a self-supervised learning algorithm called KDEM - Kernel Discriminant-EM to speed up semantic classification and recognition problems. The algorithms are tested for image classification, hand posture recognition and fingertip tracking.

We then move on from image indexing to context-based interpretation and indexing of videos.

Section 2: Audio and Video Semantics: Models and Standards

Chapter 4 describes the characterisation of video data using the temporal behaviour of features, using context provided by the application domain in the situation of a shot. A framework based on Dynamic Bayesian Network is presented to position the video segment within an application and provide an interpretation within that context. The framework learns the temporal structure through the fusion of all features, and removes the cumbersome task of manually designing a rule-based system for providing the high-level interpretation.

Chapter 5 moves on to audio and presents a comprehensive survey of content-based music summarisation and classification. This chapter describes techniques used in audio feature extraction, music representation, and summarisation for both audio and music videos. The chapter further identifies emerging areas in genre classification, determining song structure, rhythm extraction, and semantic region extraction in music signals.

Chapter 6 takes a holistic approach to video semantics, presenting a multidimensional model for describing and representing video semantics at several levels of abstraction from the perceptual to more abstract levels. The video metamodel VIMET supports incremental description of semantics, and presents a framework that is generic and not definitive, while still supporting the development of application-specific semantics that exploit feature-based retrieval techniques. Although the chapter addresses video semantics, it provides a nice framework that encompasses several aspects of multimedia semantics.

Chapter 7 presents Continuous Media Web — an approach that enables the searching of time-continuous media such as audio and video using extensions to standard Web-based browsing tools and technology. In particular, the chapter presents the Annodex file format that enables the creation of webs of audio and video documents using the continuous media markup language (CMML). Annodex extends the idea of surfing the web of text documents to an integrated approach of searching, surfing and managing the web of text and media resources.

Chapter 8 examines the new role of the new MPEG-7 standard in facilitating the management of multimedia semantics. This chapter presents an overview of the MPEG-7 Content description Interface and examines the Descriptions Schemes (DS) and Descriptors (Ds) to address multimedia semantics at several levels of granularity and

abstraction. The chapter presents a discussion on application development using MPEG-7 descriptions. Finally the chapter discusses some strengths and weaknesses of the standard in addressing multimedia semantics.

Section 3: User-Centric Approach to Manage Semantics

Chapters 9, 10, 11 and 12 move away from a media-centric approach and take a user-centric perspective while creating and interacting with multimedia content.

Chapter 9 presents a user-centric algorithm for visualisation and layout for content-based image retrieval from a large photo library. The framework facilitates an intuitive visualisation that adapts to the user's time-varying notions of content, context and preferences in navigation and style. The interface is designed as a touch-sensitive, circular table-top display, which is being used in the Personal Digital Historian project that enables interactive exploratory story telling.

Chapter 10 deals with a holistic approach to multimedia authoring and advances the idea of creating multimedia authoring tools for the amateur media creator. The chapter proposes that in order to understand media semantics, the media author needs to address a number of issues. These involve a deep understanding of the media creating process; knowledge of the deeper structures of content; and the surface manifestations in the media within an application domain. The chapter explores software and human interactions in the context of implementing a multimedia authoring tool in a target domain and presents a future outlook on multimedia authoring.

Chapter 11 presents MM4U, a software framework to support the dynamic composition and authoring of personalised multimedia content. It focuses on how to assemble and deliver multimedia content personalised to reflect the user's context, specific background, interest and knowledge, as well as the physical infrastructure conditions. Further, the application of MM4U framework is illustrated through the implementation of two applications: a personalised city guide delivered on a mobile device, and a personalised sports ticker application that combines multimedia events (audio, video and text-based metadata) to compose a coherent multimedia application delivered on the preferred device.

Chapter 12 considers the role of the mature relevance feedback technology, which is normally used for text retrieval, and examines its applicability for multimedia retrieval. The chapter surveys a number of techniques used to implement relevance feedback while including the human in the loop during information retrieval. An analysis of these techniques is used to develop the requirements of a relevance feedback technique that can be applied for semantic multimedia retrieval. The requirements analysis is used to develop a user-centric framework for relevance feedback in the context of multimedia information retrieval.

Section 4: Managing Distributed Multimedia

Chapters 13 and 14 explore multimedia content retrieval and presentation in a distributed environment.

Chapter 13 addresses the problem that occurs due to the separation of content from its description and functionality while exchanging or sharing content in a collaborative multimedia application environment. The chapter proposes a content modeling

formalism based on enhanced multimedia metaobjects (Emmo) that can be exchanged in their entirety covering the media aspect, the semantic aspect and the functional aspect of the multimedia content. The chapter further outlines a distributed infrastructure and describe two applications that use *Emmo* for managing multimedia objects in a collaborative application environment.

Chapter 14 shows how even a limited description of multimedia object can add semantic value in the retrieval and presentation of multimedia. The chapter describes a framework Delaunay^{View} that supports distributed and heterogeneous multimedia sources based on a semantically driven approach for the selection and presentation of multimedia content. The system architecture is composed of presentation, integration and data layers, and its implementation is illustrated with a case study.

Section 5: Emergent Semantics

The next two chapters explore an emerging research area — emergent semantics — where multimedia semantics emerges and evolves dynamically responding to unanticipated situations, context and user interaction.

Chapter 15 presents an overview of emergent semantics. Emergence is the phenomenon of complex structures arising from interactions between simple units. Emergent semantics is symbiosis of several research areas and explores experiential computing as a way for users to interact with the system at a semantic level without having to build a mental model of the environment.

Chapter 16 provides a practical foundation to this emerging research area. It explores the computation of emergent semantics from integrative structures that blend media into creative compositions in the context of other media and user interaction with the media as they deal with the semantics embedded within the media. The chapter presents a media blending framework that empowers the media producer to create complex new media assets by leveraging control over emergent semantics derived from media blends. The blending framework for discovering emerging semantics uses ontologies that provide a shared description of the framework, operators to manage the computation models and an integration mechanism to enable the user to discover emergent structures in the media.

CONCLUDING REMARKS

In spite of large research output in the area of multimedia content analysis and management, current state-of-the-art technology offers very little by way of managing semantics that is applicable for a range of applications and users. Semantics has to be inherent in the technology rather than an external factor introduced as an afterthought. Situated and contextual factors need to be taken into account in order to integrate semantics into the technology. This leads to the notion of emergent semantics which is user-centered, rather than technology driven methods to extract latent semantics. Automatic methods for semantic extraction tend to pre-suppose that semantics is static, which is counterintuitive to the natural way semantics evolves. Other interactive technologies and developments in the area of semantic web also address this problem. In future, we hope to see the convergence of different technologies and research disciplines in addressing the multimedia semantic problem from a user-centric perspective.

REFERENCES

Bimbo, A.D. (1999). In M. Kaufmann (Ed.), *Visual information retrieval*. San Francisco.

Dorai, C., & Venkatesh, C. (2002). *Computational media aesthetics*. Boston: Kluwer Academic Publishers.

Meersman, R., Scott, Z., & Stevens, M. (1999, January 4-8). Database semantics - Semantic issues in multimedia systems, IFIP TC2/WG2.6. *Eighth Working Conference on Database Semantics (DS-8)*, Rotorua, New Zealand.