MIR Benchmarking: Lessons Learned from the Multimedia Community

Josh Reiss
Department of Electronic Engineering
Queen Mary, University of London
Mile End Road,
London E1 4NS UK
+44-207-882-5528
josh.reiss@elec.qmul.ac.uk

Mark Sandler
Department of Electronic Engineering Queen Mary,
University of London
Mile End Road,
London E1 4NS UK
+44-207-882-7680
mark.sandler@elec.qmul.ac.uk

ABSTRACT

Music Information Retrieval may be perceived as part of the larger Multimedia Information Retrieval research area. However, many researchers in Music Information Retrieval are unaware that the problems they deal with have analogous problems in image and video retrieval. Many issues concerning the creation of testbed digital libraries and effective benchmarking of information retrieval systems are common to all multimedia retrieval systems. We examine the approaches used in the image and video communities and show how they are applicable to testbed creation and information retrieval system evaluation when the media is music.

1. INTRODUCTION

In recent years, the Music Information Retrieval(MIR) community has been concerned with issues concerning the creation of Music Digital Libraries and the benchmarking and evaluation of MIR systems. For the most part, they have been working alone on these issues. Notable exceptions to this come from the Library and Information Science community, such as the liasing with TREC researchers on ideas for benchmarking, and inclusion of MDLs into larger digital libraries, e.g., incorporation of MelDex into the New Zealand Digital Library[1]. Still, the MIR community have encountered few researchers who deal with related challenges of a depth and complexity similar to their own.

This isolation is unnecessary. Both the Video Information Retrieval (VIR) and Image Information Retrieval (IIR) communities have struggled with the difficulties in creating large digital libraries, designed appropriately for IR evaluation and free from copyright issues. VIR is now a track at TREC, and so they have a large amount of experience in evaluating their retrieval systems. The IIR community have created their own methods for benchmarking and evaluation, borrowing some ideas from TREC and creating novel approaches for content-based retrieval of multimedia.

In addition, many within the Multimedia Information Retrieval (MMIR) community have identified problems that may exist in MIR evaluation of which MIR researchers are not yet aware. In IIR, there was a need to automate and streamline the evaluation process. Furthermore, they found that the lack of a common access method, analogous to SQL for relational databases, was a hindrance to providing consistency in evaluation. Thus they devised their own solutions to these problems.

VIR researchers have encountered the more abstract problems associated with the vagueness of relevance definitions for multimedia. Similarity between documents when they each have a time dependent component is a complicated issue. The choice of metadata, segmentation and the subjective evaluation of precision and recall are all issues of concern to both MIR and VIR researchers. Furthermore, video researchers have explored different approaches from those typically used with musical data.

In this paper, we study the approaches of the MMIR community and see where these approaches are applicable to MDL creation and MIR evaluation and benchmarking. From this, we propose a set of recommendations and guidelines for the MIR community. These guidelines have the benefit of requiring only small modifications from the guidelines that have been tested and streamlined for video and image problems. Finally, we suggest how the research in MIR may be augmented and used within a full Multimedia Information Retrieval System that incorporates, images, video and audio.

This document is divided into three sections. The first contains a discussion of the experiences and suggestions concerning creation of digital libraries for multimedia. The second section concerns methods and implementations of benchmarking and evaluation of

MMIR systems. Finally, in the third section, we discuss how to incorporate musical queries into a full multimedia system.

2. COMMENTS ON MDL CREATION

2.1 The Open Video Project

One of the more interesting projects in video that could be mirrored by the MIR/MDL community is the Open Video Project[2]. Anticipating a future with widespread access to large digital libraries of video, a great deal of research has focused on methods of browsing and retrieving digital video, developing algorithms for creating surrogates for video content, and creating interfaces that display result sets from multimedia queries. Research in these areas has required that each investigator acquire and digitize video for their studies since the multimedia information retrieval community does not yet have a standard collection of video to be used for research purposes. The primary goal of the Open Video Project[5 is to create and maintain a shared digital video repository and test collection to meet these research needs.

The Open Video Project aims to collect and make available video content for the information retrieval, digital library, and digital video research communities]. Researchers can use the video to study a wide range of problems, such as testing algorithms for feature extraction and the creation of metadata; or creating and evaluating interfaces that display result sets from multimedia queries. The idea is to collect video that is in the public domain, or provided by owners who grant permission to use their intellectual property for research purposes, and make that video available in a variety of standard formats, including streaming, along with a set of accompanying metadata. Because researchers attempting to solve similar problems will have access to the same video content, the repository is also intended to be used as a test collection that will enable systems to be compared, similar to the way the TREC conferences are used for text retrieval.

This repository is hosted as one of the first channels of the Internet 2 Distributed Storage Infrastructure Initiative[3], a project that supports distributed repository hosting for research and education in the Internet 2 community.

2.1.1 Project Overview

The Open Video Project began in 1998 with the development of a basic framework and the digitization of the initial content. Additional video was contributed by various other projects. The first stage also included entering metadata for each segment into a database, and creating a Web site to enable researchers to access the available video.

The next stage of the project involves adding additional video segments to the repository, and expanding both the available formats and genre characterics (news, entertainment, and home videos) of the video. As the size of the repository is expanded in this stage of the project, the database schema is extended to incorporate more metadata fields. Further work concerns creating innovative interfaces to the video repository that enable users to more easily search, browse, preview, and evaluate the video in the collection.

2.1.2 Copyright Issues

The Open Video repository provides video clips from a variety of sources, especially various video programs obtained from U.S. government agencies such as the U.S. Records and Archives Administration and NASA. Although the government agency videos were produced with public funds and are freely available from the Archives, no copyright clearance has been obtained for audio or video elements in these productions. They encourage researchers to use the data under fair use for research purposes. Those wishing to use these video clips in any commercial enterprise must bear the burden of obtaining copyright clearances.

This inevitably will create problems since the fair use clause may differ from country to country, and sometimes includes additional restrictions for international use outside the country of origin. Furthermore, it is unclear how liable the creators of the library may be if the copyright is violated by an MDL user or anyone else. Recording companies may want a stronger guarantee than the pledge that the material will just be used for research purposes. Thus, an Open Audio Project, analogous to the Open Video Project, is not sufficient for MDL creation.

Two proposals from within the MIR community may be viable alternatives. The first[4] is similar to the Open Video Project, but would include no copyrighted or restricted material. Instead, material would be contributed by the MIR community and amateur musicians. The idea was suggested for use specifically with MIDI, but could easily be expanded to other formats.

The second proposal[5] offers the ability to use copyrighted material while still ensuring that it is not misused. Here, the testbed is kept secure, and it may be accessed such that the analysis, processing and data mining is all performed within the secure area, and no restricted material is released.

2.2 Lessons learned from the IIR community concerning multimedia database collection

The image information retrieval community suffer from many of the same problems concerning testbesds and data collection that the MIR community does. The most commonly used images in IIR come from the Corel Photo CDs; a collection of copyrighted, commercially available images. Most research groups can only purchase a small subset of the collection. Since each CD contains a set of similar images, most testbeds contain several dissimilar image groups. This leads to overoptimistic benchmarking of IIR systems.

An alternative collection is the images from MPEG-7, which has the benefit of being used in an official standard. Unfortunately, the MPEG-7 collection is also expensive and may not be used on the web.

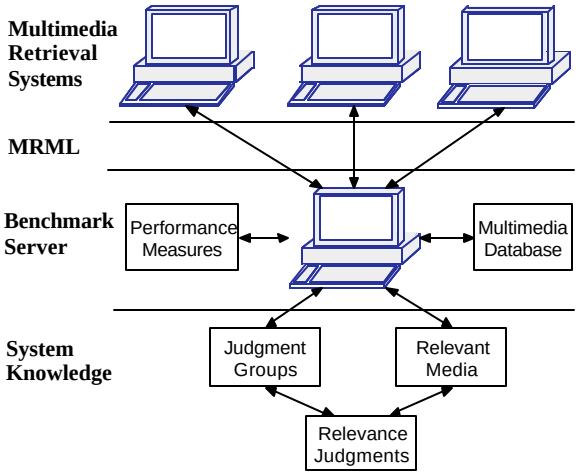

Figure 1. Flowchart depicting an automated benchmark for MMIR systems (adapted from [6]).

Thus, many in the IIR community have resorted to developing their own collections, much the same as is done in the MIR community. The Annotated Groundtruth Database[7] is one such system. This is a freely available, uncopyrighted collection of annotated photographs from different regions and about different topics. This is still a small collection (approximately 1,000 images), and it has been suggested that this collection be used as the starting point for a larger testbed. However, as it stands there is no commonly accepted uncopyrighted testbed.

The MIR community finds itself in the same position with respect to fragmentation and a lack of standards with digital libraries. In regards to copyright issues, it is in an even more restricted position since the music that researchers are most interested in is almost all copyrighted and the rights are heavily guarded.

2.3 The World Wide Web as a testbed

The internet allows individual users to store audio files and file indexing on computers distributed throughout the world. As such, it represents a large-scale, distributed audio testbed, ideal for a web-based music information retrieval system. The issues involved in the creation of web-based multimedia information retrieval systems have been explored thoroughly in the context of mixed media consisting of images, text and video, but have yet to be applied to audio.

In [8], an integrated visual retrieval system, supporting global visual query access to multimedia databases on the web, was described. This system required a central server and media stored in databases located at many sites accessible through the web. As such, it assumes a centrally administered repository with a Napster-like model. Such a model is useful in unifying independent testbeds run by different research centers.

Other web-based IIR systems, such as WebMARS[9], WebSeek[10](which is also a VIR system) and ImageRover[11] are multimedia search engines that incorporate content-based image information retrieval as well as text-based retrieval. They all have the benefit of accessing image files anywhere on the web, i.e., the largest possible testbed. However, all these systems require indexing of images and for the features to be stored in order for the queries to be answered quickly. Images may be removed but still appear in the index. They allow searching of various file sizes and formats, and searching for different features (different types of queries). Yet they are also quite different from each other, and use different sets of features and incorporate metadata in differing ways.

3. COMMENTS ON BENCHMARKING AND EVALUATION

3.1 Competitive benchmarking

The IIR comunity have established the Benchathlon(www.benchathlon.net), which uses a competitive approach to IR evaluation. The systems under evaluation are contestants, and follow strict guidelines in order to enter the competition. The benefits of such an approach are that it can be achieved when the ratio of the number of judges to the number of entrants is low, it spurs entrants to improve their systems by inspiring the competitive spirit, and it allows analysis of the various systems through direct comparison.

However, such competition, even when suggested only partly in jest as with the benchathlon, opposes the spirit and ethos of the methods used by TREC. One of the primary tenets employed by TREC is that their evaluation system is not meant as a contest.

TREC, although not using test subjects in real world environments, still uses extensive subjective testing and evaluation. To duplicate such a deep test would prove extremely difficult for small research groups with low financial resources. Thus automatic benchmarking has been proposed.

3.2 Automated benchmarking

One of the difficulties in MMIR benchmarking has been the lack of a common access method to retrieval systems. For instance, SQL or Structured Query Language, is a relational database query language that has been adopted as an industry standard. No such equivalent is available for content based multimedia retrieval systems. Thus, the image information retrieval community have proposed the Multimedia Retrieval Markup Language(MRML) [12]. Although proposed in the context of IIR, it could be applied to other media and thus would assist in standardising access to MIR systems.

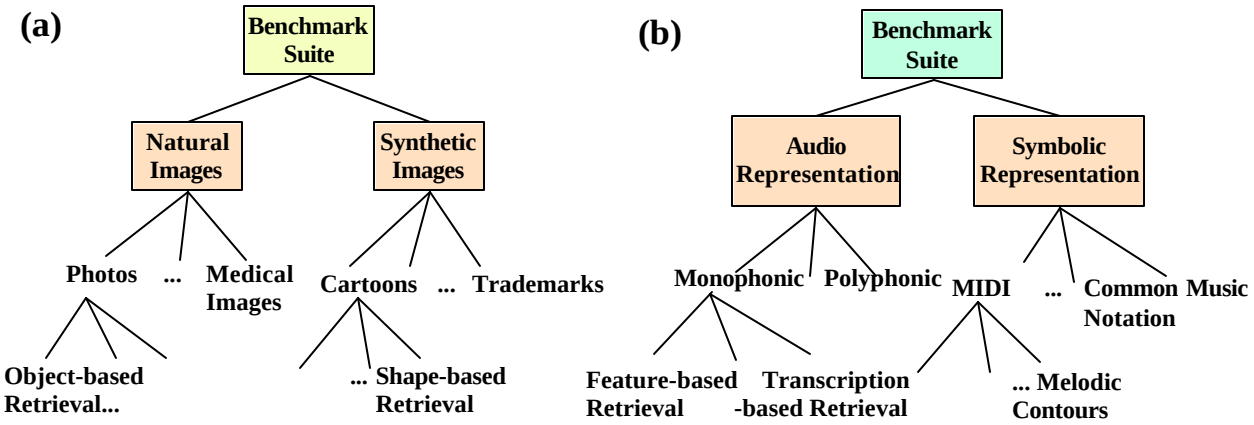

Figure 2. (a) An Extensible Image Benchmark Framework adapted from [13], and (b) an analogous musical benchmark framework.

The biggest problem in automatically benchmarking IIR systems is the lack of a common access method. The advent of MRML has solved this problem. MRML standardizes IIR access. It allows a client to log onto a database and ask for the available image collections as well as to select a certain similarity measure, and to perform queries using positive and negative examples. With such a communication protocol the automated evaluation of IIR systems is possible.

As depicted in Figure 1, MRML serves as the communication layer between the evaluated systems and the benchmark server. The multimedia digital library and the performance measures are known to all the systems. Relevance judgments may be known for initial testing, but for proper and fair evaluation, should not be known by the MMIR systems.

Of importance to the MIR community is the fact that MRML was designed to be both highly extensible and gracefully degradable. This means that it could also be extended to MIR systems. It allows for support of other media, and commands intended for images could be easily ignored by an MIR system.

Automated benchmarking can then be achieved by using MRML as the common access method to enable different MIR systems to answer the same queries, using the same testbed, and receive immediate relevance judgments.

3.3 Creation of an Evaluation Framework

As has been pointed out previously, MIR benchmarking has the problem of being a somewhat vague problem because different MIR systems are built to answer different queries and often deal with very different representations of digital music.[14,15] The video community, due to the complexity of its data, has been concerned with more low level tasks, such as segmentation and feature extraction[16]. IIR researchers, however, have carefully classified the information retrieval tasks they deal with and the related data types that are used[13]. This classification, breakdown and analysis of Information Retrieval tasks can easily crossover into MIR.

An example of this is given in Figure 2. It depicts how a benchmark suite would consist of parts that should not be fully complete and unchangeable. Indeed, with few exceptions, similarity measures for symbolic representations of music are very different from those for raw audio representations. This classification scheme, together with the work on testbeds, database access, and automated and competitive benchmarking, has led to a full evaluation framework for content-based IIR[17].

Recent work has presented a wide variety of ideas and early-stage research in MIR benchmarking and evaluation[5,18]. However, this work is yet to be formalized to the extent that it has been in the IIR community. Thus, MIR researchers should be aware that others in multimedia have devised a structure for evaluation that can deal with the variety and complexity of multimedia queries.

4. SUPPORTING MUSICAL QUERIES IN AN MMIR SYSTEM

In this section, we seek to address how best to incorporate MIR searches into a multimedia information retrieval system. We consider the complex interplay that can arise from searching across media, and the special types of issues that this creates. Lastly, we show how the creation and indexing of metadocuments can lead to an effective large-scale MIR system with the ability to retrieve multimedia documents.

Table 1. A list of hypothetical queries which use different media for the queried document, retrieved documents and/or indexing system. In each of the presented queries, musical content is in some way processed or retrieved.

Query	Input	Output	Intermediate
Find video performances of music that sounds like…	Audio	Video	Audio
Find the transcription of…	Audio	Image	Symbolic
Find all documents in the database related to the Beatles.	Text	Multimedia	Metadocument
Here is an album cover. Find music from that album.	Image	Audio	Metadocument
What does this sheet music sound like?	Image	Audio	Symbolic
Show me highlights of this sports event..	Video/Text	Video	Audio
Find the album version of the song in this video.	Video	Audio	Audio
What is the soundtrack of this movie?	Video	Audio	Audio

4.1 Cross-Media Queries

Music Information Retrieval, although having unique problems and involving an interesting and unusual mix of interdisciplinary challenges, may be classified within the general subject of multimedia information retrieval (MMIR). In a rigorous study of multimedia queries on the WWW[19], analysis of over a million queries to text based search engines suggested that, as lower estimates, approximately 3.39% of them were image or video related, whereas only 0.37% were audio related. Furthermore, the audio queries typically used more terms. This suggests that there may be a sizeable number of image, video, or text related queries where the preferred retrieved documents were music related, but not necessarily audio. This is given further credence given that *lyrics* was a popular audio related search term, and *videos* was a popular video related search term.

Even though it is difficult to determine the demand for cross-media based information retrieval systems, it is relatively easy to construct pertinent multimedia queries. Table 1 lists a variety of queries which use or retrieve music-related content. In each situation, multiple media types are used. The categorization is meant to be indicative as opposed to formal. The simplest examples involve music videos, which are often standard MIR-style audio queries with the exception that the corpus has a video as well as audio component. Other examples, such as "What does this sheet music sound like?" represent active research areas in the MIR community[20,21].

One of the most interesting examples, "Show me highlights of this sports event," involves neither musical content in the query statement or in the retrieved documents. However, musical content can be an important feature in the audio stream. If critical moments in a sporting event are accompanied by music (such as a popular song played on the loudspeakers in baseball stadia after each run is scored), then identifying this in audio content will often prove much easier than identifying visual clues in the video stream. The use of audio identifiers to retrieve relevant video-based information, is well-known to researchers in [20]Video Information Retrieval[22-24], but the specific use of musical content in the audio-visual stream represents a novel area of research.

4.2 Use of metadocuments in an MMIR Indexing Scheme

An inherent problem with retrieving multiple media types in a feature-based index is that the appropriate features are different for each media. Image classification schemes often use the discrete cosine transform or the Gabor transform, whereas text feature extraction uses lexical analysis, and audio feature extraction often uses windowed harmonic content. Thus similarity indexing based on features would not enable retrieval of images related to audio, or vice versa. Such a problem necessitates the use of metadata as a means of hyperlinking related multimedia.

Through the combined use of metadata, metadocuments and hyperlinks, an effective cross-media indexing scheme may be devised. This has also been considered in IIR where the text metadata often associated with images on the web may be used for retrieval of web pages with relevant visual content[9].

For instance, an image may be entered as the query, features extracted and the closest match found. the related metadocuments for this image are then searched to find related media. thus, a scanned image of an album cover can be used to find songs off the album, lyrics, similar images, a video interview with the album producer, and so on. this has the additional benefit that it creates both an appropriate indexing and an appropriate browsing scheme for multimedia.

Furthermore, the system need be no more complex than the sum of its parts. All media retrieval can be performed using the same multidimensional feature set indexing and retrieval scheme (see [25] for details of an appropriate multidimensional search method). Then metadocuments can be searched using any appropriate keyword and text based scheme.

5. CONCLUSION

In this work, we considered the approaches of the Multimedia Information Retrieval community to the problems of multimedia digital library creation and information retrieval benchmarking and evaluation. For all multimedia resources, there are problems concerning copyrights. Some researchers in the video community have taken the somewhat risky approach of assuming that all use will comply with regulations regarding the Fair Use of Copyrighted Materials. In IIR, several commercially available data sets are frequently used, thus there remains the problems of standardization and expense.

It therefore seems that the current proposals for MDL testbeds, secure copyrighted databases or contributed uncopyrighted testbeds, are preferable. A third approach, which is not specific to just images or video, is the use of the web as a multimedia corpus. This concept has obvious benefits because of its use of a sufficiently large and varied collection. However, this would allow for only certain types of MIR systems, i.e., web-based, and has additional issues concerning the ever-changing nature of the testbed, speed of access, and lack of metadata.

In regards to benchmarking and evaluation, both video and imaging researchers have made great strides. The video community has been very active in TREC, and thus can serve as a guide to how TREC can assist in MMIR evaluation, and also of how MIR benchmarking could mirror the TREC approach if the music retrieval community chose not to participate in TREC.

In image retrieval, they have considered an approach to benchmarking that is distinctly different from the TREC approach. First, they have referred to the evaluation as a contest, whereas TREC emphasizes its noncompetitive nature. Furthermore, the IIR researchers do not have the resources to manage large data collections, create and refine topic statements, pool individual results, judge retrieved documents, and evaluate results. Thus, they have automated the process through the use of MRML, a standardized access scheme for IIR systems. It is clear that the approach followed by IIR researchers deserves consideration by the MIR community. Certainly, those interested in having a common access method should consider the extension of MRML to musical queries.

Finally, we considered how MMIR systems may be linked to allow cross-media queries. Video, images and music retrieval systems all often use feature extraction in the summarization of content. Only metadata is required to link them. Given that research into retrieval is advancing throughout the multimedia community, and that many topic statements might require different media for the query, the retrieved documents or for intermediate stages, it is clear that the convergence of Multimedia Information Retrieval systems will be an active area of research in the future.

References

1 R.J. McNab, L.A. Smith, D. Bainbridge et al., *The New Zealand Digital Library MELody inDEX*, D-Lib Magazine **May** (1997).

2 G. Marchionini and G. Geisler, *The Open Video Digital Library*, D-Lib Magazine **8** (12) (2002).

3 M. Beck and T. Moore. *The Internet2 Distributed Storage Infrastructure Project: An Architecture for Internet Content Channels*. 3rd International WWW Caching Workshop, Manchester, England, June 15-17 1998.

4 J. A. Montalvo. *A MIDI Track for Music IR Evaluation*. JCDL Workshop on the Creation of Standardized Test Collections, Tasks, and Metrics for Music Information Retreival (MIR) and Music Digital Library (MDL) Evaluation, Portland, Oregon, July 18 2002.

5 J. S. Downie. *Toward the Scientific Evaluation of Music Information Retrieval Systems*. Fourth International Conference on Music Information Retrieval (ISMIR 2003), Washington, D.C., USA, October 26-30 2003.

6 H. Muller, M. Muller, S. Marchand-Maillet et al. *Automated Benchmarking In Content-Based Image Retrieval*. 2001 IEEE International Conference on Multimedia and Expo (ICME2001), Tokyo, Japan, August 2001.

7 "Annotated groundtruth database," Department of Computer Science and Engineering, University of Washington, 1999, www.cs.washington.edu/research/imagedatabase.

8 D. Murthy and A. Zhang. *WebView: A Multimedia Database Resource Integration and Search System over Web*. WebNet 97: World Conference of the WWW, Internet and Intranet, Toronto, Canada, October 1997.

9 M. Ortega-Binderberger, S. Mehrotra, K. Chakrabarti et al. *WebMARS: A Multimedia Search Engine for Full Document Retrieval and Cross Media Browsing*. Multimedia Information Systems Workshop, Chicago, IL, October 2000.

10 S.-F. Chang, *Visually Searching the Web for Content*, IEEE Multimedia Magazine **4** (3), 12 (1997).

11 S. Sclaroff, L. Taycher, and M. La Cascia. *ImageRover: A Content-Based Image Browser for the World Wide Web*. Workshop on Content-Based Access of Image and Video Libraries (CBAIVL '97), Puerto Rico, June 1997.

12 H. Muller, M. Muller, S. Marchand-Maillet et al. *MRML: A communication protocol for content-based image retrieval*. International Conference on Visual

Information Systems (Visual2000), Lyon, France, November 2–4 2000.

[13] C. H. C. Leung and H. H. S. Ip. *Benchmarking for Content-Based Visual Information Search*. Advances in Visual Information Systems (Visual 2000), Lyon, France, November 2-4 2000.

[14] J. D. Reiss and M. B. Sandler. *Beyond Recall and Precision: A Full Framework for MIR System Evaluation*. 3rd Annual International Symposium on Music Information Retrieval, Paris, France, October 17 2002.

[15] J. D. Reiss and M. D.. Sandler. *Benchmarking Music Information Retrieval Systems*. JCDL Workshop on the Creation of Standardized Test Collections, Tasks, and Metrics for Music Information Retreival (MIR) and Music Digital Library (MDL) Evaluation, Portland, Oregon, July 18 2002.

[16] A. F. Smeaton. *An Overview of the TREC Video Track*. 10th TREC Conference, Gaithersburg, Md, 13-16 November 2001.

[17] H. Müller, W. Müller, S. Marchand-Maillet et al., *A framework for benchmarking in visual information retrieval*, International Journal on Multimedia Tools and Applications **21** (2), 55 (2003).

[18] J. S. Downie. *Panel on Music Information Retrieval Evaluation Frameworks.* 3rd International Conference on Music Information Retrieval (ISMIR), Paris, France2002.

[19] B. J. Jansen, A. Goodrum, and Spink. A., *Searching for multimedia: video, audio, and image Web queries*, World Wide Web Journal **3** (4), 249 (2000).

[20] G. S. Choudhury, T. DiLauro, M. Droettboom et al. *Optical music recognition system within a large-scale digitization project*. First International Conference on Music Information Retrieval (ISMIR 2000), Plymouth, Massachusetts, October 23-25 2000.

[21] J. R. McPherson. *Introducing feedback into an optical music recognition system.* Proceedings of the Third International Conference on Music Information Retrieval: ISMIR 20022002.

[22] S.-F. Chang. *Searching and Filtering of Audio-Visual Information: Technologies, Standards, and Applications.* IEEE International Conference of Information Technology, Las Vegas, NV, March 2000.

[23] H. Sundaram and S.-F. Chang. *Video Scene Segmentation using Audio and Video Features*. ICME 2000, New York, New York, July 28-Aug 2 2000.

[24] N. V. Patel and I. K. Sethi. *Audio characterization for video indexing*. Proceedings of the SPIE on Storage and Retrieval for Image and Video Databases, San Jose, CA, February 1-2 1996.

[25] J. D. Reiss, J.-J. Aucouturier, and M. B. Sandler. *Efficient Multidimensional Searching Routines for Music Information Retrieval*. 2nd Annual International Symposium on Music Information Retrieval, Bloomington, Indiana, USA, October 15-17 2001.