Modeling and Retrieving Audiovisual Information
- A Soccer Video Retrieval System -

A. Woudstra[1], D.D. Velthausz[2*], H.J.G. de Poot[2], F. Moelaert El-Hadidy[2], W. Jonker[1], M.A.W. Houtsma[2], R.G. Heller[1], J.N.H. Heemskerk[1]

[1] KPN Research
P.O. Box 15000, 9700 CD, Groningen, the Netherlands
Phone: +31-50 -5821124, Fax: +31-50 -3122415
E-mail: {Woudstra, Jonker, Heller, Heemskerk}@research.kpn.com
[2] Telematica Instituut
P.O. Box 589, 7500 AN, Enschede, the Netherlands
Phone: +31-53-4850485, Fax: +31-53-4850400
E-mail: {Velthausz, Poot, Moelaert}@telin.nl

Abstract. This paper describes the results of an ongoing collaborative project between KPN Research and the Telematics Institute on multimedia information handling. The focus of the paper is the modelling and retrieval of audiovisual information. The paper presents a general framework for modeling multimedia information (ADMIRE) and discusses the application of this framework to the specific area of soccer video clips. The core of the paper is the integration of feature extraction and concept inference in a general framework for representing audio visual data. The work on feature extraction is built on existing feature extraction algorithms. The work on concept inference introduces a new approach to assigning semantics to collections of features in order to support concept-based retrieval, rather than feature-based retrieval. Finally, the paper describes our experiences with the implementation of the methods and techniques within the ADMIRE framework using a collection of commercially available tools. The latter is done by implementing a soccer video clip annotation and query tool.

1 Introduction

The development of the WEB has lead to an increasing research effort into methods and techniques for the realization of multimedia applications ([8], [18], [21], [32]), like video-on-demand, tele-shopping, and e-commerce. A major part of this research effort is devoted towards multimedia database management ([1], [4], [16]), driven by the expectation that multimedia database management systems will form the

* Contact person for ADMIRE framework

S. Jajodia, M.T. Özsu, and A. Dogac (Eds.): MIS'98, LNCS 1508, pp. 161–173, 1998.

cornerstone of the next generation of multimedia applications, just like relational database management systems are at the heart of current information systems.

Multimedia database research brings together researchers from different disciplines like databases, image processing, information retrieval, and artificial intelligence, and thus covers a wide variety of issues. The research in this paper concentrates on issues related to modeling of audiovisual data to support ad-hoc retrieval of video clips. The motivation for this research is the fact that the amount of video data in Web based information systems is growing so fast that techniques for efficient retrieval are mandatory. Current practice is semi-automated classification and manual annotation of video material after which the video material is stored and can be retrieved based on (keywords from) the classification and annotation scheme being used. This approach has a serious drawback: the retrieval can only be done based on a predefined classification and annotation. So one can not search on issues not covered by the classification and annotations. To overcome this problem there is a need for retrieval based on the content of the video, rather than on annotations. (An analogy can be drawn with text retrieval, where we have seen a shift from retrieval based on classification by means of indices towards full text retrieval [10]). So, what is needed are retrieval techniques that act directly on the video content. These techniques are commonly denoted as content-based retrieval techniques ([9], [11], [20], [28]).

A widely accepted first step towards content-based retrieval is feature extraction. Features are interpretation independent characteristics. Examples are pitch and noise for audio, and color histogram and shape for images. Currently some impressive results have been shown on retrieval of images based on feature extraction ([7], [12], [15], [23], [25]). The approach is purely based on features and does not incorporate any conceptual interpretation of the images. The approach is roughly a query-by-example approach in which the system is asked to retrieve images 'similar' to an example image. Quite a number of features have been identified and a lot of feature detection algorithms already exist ([3], [13], [22]).

However, query-by-example is not always appropriate, especially in the case of video clips. In the Soccer domain, for example, instead of providing the system with an example video shot of a goal, a user simply wants to issue a query like 'show me all goals of this match'. Therefore, a next step is adding semantics to (collections of) features. For example, from a collection of features (e.g. whistle sound in audio, grouped players in video, ball in goal area, etc.) one can infer that a goal has happened. Some simple but effective interpretation techniques are already available, for example, using color histograms of individual video frames for shot detection in videos ([2], [27]). For real content-based retrieval a more fine-grained approach is needed, for example, in the soccer domain one must be able to infer that an object that has a round shape and a black and white color histogram is a ball. Given that a video can be seen as a sequence of frames, many techniques applied to images can be used. An additional difficulty that is introduced by videos is temporal relationships, such a tracking objects through a sequence of frames.

In the this paper we present our approach towards content-based retrieval of video clips based on an extensive modeling of the content of the videos. The main focus of the research is:

- to refine and assess the ADMIRE framework (see below) for modeling audiovisual data for content-based retrieval;
- to develop a concept inference technique and validate it in the context of soccer video clips;
- to investigate which steps to automate and implement a prototype to support that.

Our approach covers both feature extraction and semantic interpretation (we call it ‘concept extraction’). As a general representation framework, ADMIRE [29] (see Chapter 2) is used. Roughly ADMIRE distinguishes three levels: the raw data level, the feature level, and the concept level (where interpretations of data objects are given, such as player and ball). In this framework specific techniques for the representation of raw data, features, and concepts can be put, as well as specific techniques for feature extraction and concept inference [24]. Feature extraction and concept inference is given in Chapter 3. The validation of the approach is done by means of the implementation of a soccer video retrieval system [19] described in Chapter 4. Apart from validating the approach the implementation also served as a means to assess the quality of generally available tools for the manipulation of audiovisual data.

2 ADMIRE Model

Audiovisual information consists usually of three types of media: audio, video and text (e.g. subtitles). To disclose this information efficiently we first need to represent it in a model. This model should support the representation of the different types of media in a uniform way. Information should be characterized at different aggregation levels (e.g. frame, shot, coverage). Further, the model should support different types of relationships (i.e. spatial, temporal) and a wide variety of query mechanisms.

Existing, multimedia models focus mostly on a single aspect of multimedia information, like presentation (e.g. PREMO), or exchange of documents, or on a particular format (e.g HyTime). Models that do facilitate content-based information retrieval in general are for example MORE [26], VODAK [14], CORE [31], AIR [15]. These models either do not support a layered definition of information objects (e.g. MORE and VODAK) or can only represent the content of specific unstructured media types (e.g. AIR). The ADMIRE model [29] can be seen as a generalization of the existing models. It resembles the CORE model but offers more flexibility in modeling object relationships. ADMIRE emphasizes on the disclosure of all kinds of forms and types of existing digital information [30]. It uses an object-oriented modeling technique together with a layered definition of information objects. It is suitable for representing multimedia information and thus audiovisual information.

An information object in ADMIRE consists of different properties, *raw data* (a sequence of elementary data units), *format* (a representation of the raw data, e.g. MPEG-1), *attribute* (characterization of the raw data that can not be extracted from the raw data and its format in relation to the external world, e.g. creation date), *feature* (a domain-independent representation of the raw data and format, e.g. color histogram), and *concept* (a domain-dependent and format-independent semantic interpretation of the other property types); *relations* refer to other information objects. Properties are modeled in a three-layer hierarchy, see Fig. 1. The lowest layer is the data layer. It contains the raw data, format and attribute properties. All the properties in this layer are stored as they convey information that cannot be determined differently. The feature and concept layers contain information that can be determined (respectively via feature extraction and concept extraction) using the properties of the lower layers and extraction algorithms (for features) and (domain) knowledge rules for concepts. As opposed to features, concepts are context dependent, different concepts can be inferred from the same lower layers. For example, a 'thumb pointing upwards' has a different interpretation for scuba-divers than for e.g. stock brokers. The context only influences the usefulness of particular features.

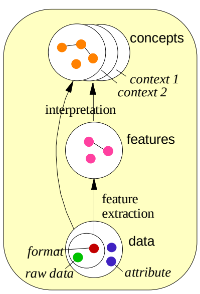

Fig. 1. Information object operations.

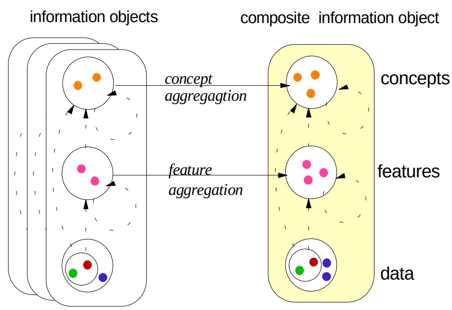

Fig. 2. Property aggregation operations.

Information in the ADMIRE model can be modeled at multiple levels of granularity, e.g. an individual frame or a whole movie. Through the use of composite relations logical structures between the information objects can be modeled. For example, a video can subsequently be decomposed into a sequence of scenes, shots, and eventually frames. The composite relationships facilitate aggregation of property values, see Fig. 2. In addition to the layout information conveyed in the information object's properties, i.e. the format, the layout structure of a composite information object is represented by the spatial and temporal relations.

An example of a composite information object for audiovisual information is a TV-coverage. Within a TV-coverage we distinguish the following information object classes: *scene* (a sequence of semantic coherent shots), *shot* (continuous recording of audiovisual material), *video shot* (a successive collection of frames, see Fig. 4, *audio track* (shot's audio part), *text title* (e.g. actual score, play time, name of player that received a yellow card), and *frame* (a 2-dimensional image sample). This is presented in Fig. 3.

To model any subpart of an information object we introduce the term 'pseudo' as these objects are not retrievable, only via their accompanying (basic or composite) information objects, see [29]. The data layer of a pseudo information object refers to a subset of the raw data and format of the accompanying retrievable information object. For example, (a pseudo frame object) a rectangle box (e.g. indicating a ball or player) within a frame, see Fig. 4, and (a pseudo video shot object) correlated successive boxes within a video shot, e.g. indicating the ball (labeled with ‘1’ in Fig. 4). Similarly, an audio-track can have accompanying pseudo objects, e.g. that part of the track that indicates the reporter’s voice. Notice that a pseudo video shot object is a composite object, i.e. composed of pseudo frame objects, see [30] for detailed description. Although these pseudo information objects are not directly retrievable they are very helpful during the inference of property values of retrievable information objects and provide a flexible way to model relations, see [29].

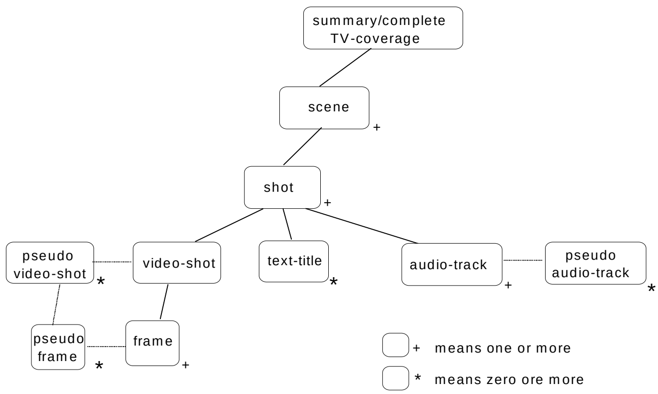

Fig. 3. Hierarchy of information objects within a summary- or complete TV-coverage**.**

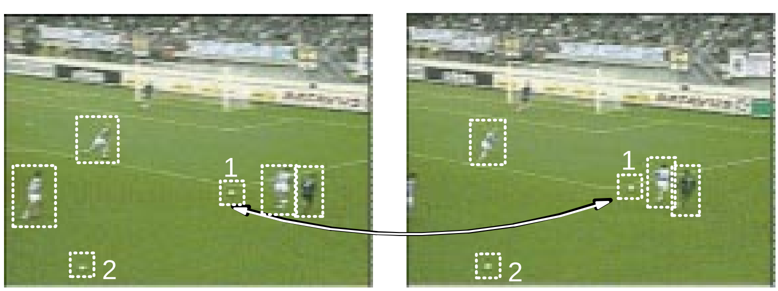

Fig. 4. Example of a video shot information object (IO) consisting of two successive frame IOs. The frame information objects may contain multiple pseudo frame IOs. Corresponding pseudo frame objects, e.g. indicated by ‘1’, form a pseudo video-shot IO.

3 Extraction and Inference

Previous approaches for content-based information retrieval focused on manual attachment or automated properties extraction and identification of information objects. We propose a combination of these. First, we compare these approaches. Manual attachment is good for information reduction, but lacks consistency [33] and details and is labor intensive. Automatic property extraction requires massive common sense knowledge bases (like Cyc [34]), which are slow, cumbersome [17] and fail unique determination of concepts [5]. Semi-automatic systems combine the best of manual and automatic extraction. Humans can give semantic descriptions, annotations [12], while computers are more precise and consistent in measurements and can propagate annotations [6]. Since multimedia information extraction involve technologies from various disciplines (e.g. image process, pattern recognition, AI, neural networks) and most of these technologies are continuously changing and improving, we need an approach that can easily adapt to the future needs. In a fourth approach, sophisticated manual, human resources are used to 'simulate' existing or future algorithms, introducing restrictions to a subset of key-frames and requiring human labeling as error prone as the algorithms. This approach supports the simulation of future capabilities, e.g. when new algorithms become available.

In the soccer prototype the identification of (pseudo) information objects is partly automated. For example, via pair wise comparison of two successive frames, correlated pseudo frame objects (and thus pseudo video shot information objects) can be found, see Fig. 4. Given a pseudo video shot information, additional features are extracted, i.e. motion features. Further, the identification of concepts like goal events, involving the combination of multiple pseudo shot information objects, can be done. These are inferred from the features and attribute data. . This is discussed in the next section. A similar hierarchy exists for the audio track. Another example is the detection of scene breaks by comparing successive shot information objects. These examples illustrate that concept inference is a bottom up approach, see ([30], [24]).

3.1 Property Aggregation

In an information object, features are extracted from data and concepts are inferred from features and attribute data. The properties of composite information object, as presented in Fig. 2, can be determined using the inferred properties of the underlying information objects. This operation is called property aggregation. There are two different ways to derive concepts of a composite information object via features aggregation and concept aggregation. See Fig. 5 for an example.

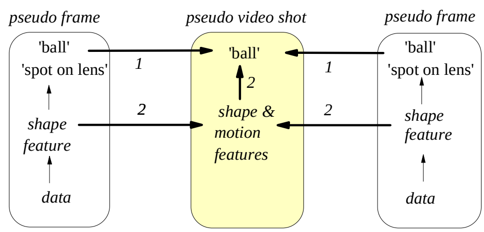

Fig. 5. Example of feature (2) and concept aggregation (1) for two pseudo frames using motion feature information and knowledge that a moving object can not be a fixed spot on the lens.

3.2 Inference Rules for a Soccer Example

Concepts become pronounced at a sufficiently high level of abstraction in the information object hierarchy. Nevertheless, they can not be inferred with absolute certainty. Hence, we incorporate uncertainty in our inference mechanism, discussed in [24]. Take for example the inference of the concept *ball*. Using the shape feature in a frame information object to identify the concept *ball*, may not only identify the ball but also a spot on the camera for example. If the inference takes place on the video shot level, spatial and temporal features, e.g. motion can be combined and inference happens with a higher certainty. The inference of, e.g. the concept *ball*, should thus be delayed. In this manner a *ball* and *spot* can be distinguished, as shown in Fig. 4 where '1' indicates the ball and '2' a spot.

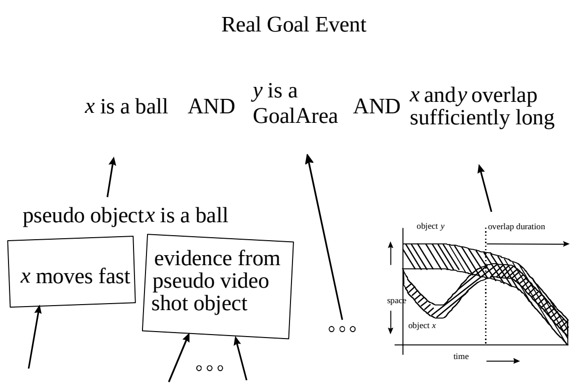

Fig. 6. Schema of the inference of the concept Goal shot.

Object motion can often be told from camera motion (pan, tilt, roll, zoom), after identification of some fixed points, e.g. spots on the soccer field. Having defined a distance measure, e.g. between objects' boundaries, spatio-temporal phenomena like: player *A* getting the ball *B* from player *C* may be inspected. For the soccer prototype a set of key events and key objects where defined for every information (pseudo)object layer in the hierarchy, e.g. at the pseudo frame level and pseudo video shot level *ball* is a supported concept (i.e., inference rules exist for it). At the pseudo video shot level the event *goal shot* is also supported as interaction between the *ball* and the *goal* object. At the audio track level *cheering audience* is a supported pseudo object, etc. An example of the combination of evidence for a *goal shot* is shown in Fig. 6.

4 Implementation

This chapter describes the implementation of the Soccer Video Retrieval System (SVRS). The implementation forms a validation of the information modeling approach described in the previous chapters. In addition, it serves as a means to assess the quality of generally available tools and algorithms for the manipulation of video data. The architecture and the mapping of the information model onto the architecture is described. This architecture provides for storing videos at the raw data, feature and concept levels. In order to extract this data from the raw video material and store it in the multimedia database an *annotation module* was implemented. Retrieving this pre-annotated information is done by means of a *query module*.

4.1 SVRS Architecture

The SVRS is implemented as a client/server application using the Informix Universal Server (IUS) as the object-relational database platform. The clients are implemented on Windows NT 4.0 machines using Delphi and IUS query tools. Communication between the clients and server takes place using ODBC. All Soccer data is stored at the server side. Raw video material is stored in Quicktime format. The clients contain several applications for manipulating the data. The Universal Server is extended with a number of software libraries called DataBlades that provide data storage and management functionality. The SVRS uses DataBlades for video (Informix), image (Excalibur), audio information retrieval (AIR by Musclefish) and text (Excalibur).

The architecture revealed two weaknesses of the currently available multimedia database technology. First, there is the impedance mismatch between the data models at client and server sides. The ADMIRE information model perfectly maps upon an object oriented data model. At the client side an object oriented implementation could be used, while at the server side, the object-relational IUS lacks true object oriented characteristics such as inheritance and encapsulation. It therefore forced us to implement all data types in a relational table format. A second limitation is the lack of support for continuous video streams, which requires either the transmission of complete video clips before playout or the development of add-on streaming functionality. A solution for this problem is a streaming DataBlade within IUS which is currently not available.

4.2 SVRS Annotation Module

As stated in the introduction of this paper, we were interested in tools that support the automatic annotation processes. In our implementation, the annotation process can be seen as a ‘human supervised approach’. This differs from the initial approach we proposed in section 0 for simplicity reasons. Although the user is supported by several tools (e.g., feature extraction) still a number of decisions and corrections have to be made manually.

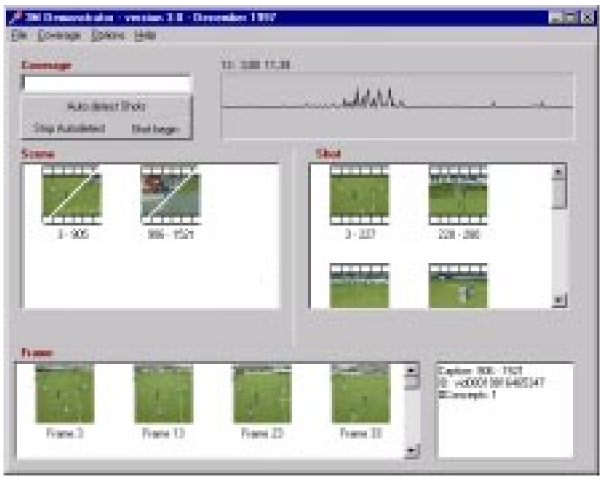

Fig. 7. User interface of the SVRS annotation module, the upper right window shows the color histogram of the current frame.

The level of automation that can be achieved with currently available feature extraction technology in a real-world application is still very basic. The annotation module *automatically* divides videos into frames and supports automatic shot detection based on differences in color histograms of successive frames. This is based on functionality offered by the image DataBlade for extracting and storing a color, shape and texture index of each image. Automatically detected shots can be manually corrected. In addition, some very limited automated detection and tracking of pseudo frame information objects is supported, e.g. the tracking of players and the ball, as illustrated in Fig. 4. Finally, the audio DataBlade is used for indexing and storing sound features of audio fragments. All additional feature and concept annotation has to be done manually. The feature and concept annotations are used for comparison and inference to support retrieval (see below).

4.3 SVRS Query Module

The query module was built to retrieve previously annotated information from the database. Fig. 8 shows a screen-dump of the query module. The query module supports standard functions like viewing video clips (including fast forwarding, rewinding, pause and stop) as well as querying video clips based on annotated (or inferred) concepts.

Fig. 8. User interface of the SVRS query module.

In Fig. 8 the example of searching for exciting moments with the help of audio features is displayed. Exciting moments in the soccer game were manually selected and used to define the distinguishing characteristics of the audio information features such as duration, pitch, tone, and loudness. These patterns were used to automatically search for similar audio tracks in the video clips using the functionality of the AIR DataBlade. The query module returns an ordered list of exciting moments that can directly be displayed on the screen by simple mouse clicks.

The current prototype implementation offers basic support for inference. Some simple rules like a moving round shape with black and white color histogram is a ball are implemented as well as some limited inference on audio as explained above. Improving this inference is one of the major topics we currently work on.

5 Conclusions

In this paper we presented an integrated approach towards the modeling and retrieval of audiovisual information. The approach includes an overall modeling framework (ADMIRE), a formalism for concept representation and inference, as well as an experimental implementation environment with tools supporting modeling, annotation, and retrieval of audiovisual information.

In the introduction we listed the main research issues of the project. As far as the assessment of ADMIRE is concerned we conclude that ADMIRE is a very useful framework for structuring the overall representation of audiovisual information. Especially, the explicit distinction between feature level and concept level offers a strong modeling advantage. Explicit concept modeling and inference allows querying at the conceptual level. The latter is an important improvement over existing approaches, which lack the notion of concept inference and as a result only offer querying at the feature level (mostly by means of query-by-example).

As far as the refinement of ADMIRE is concerned we recall that ADMIRE is a modeling framework. In order to use ADMIRE in the context of a specific application the framework has to be populated by adequate representations at the data, feature, and concept level. In our research we focused on refining ADMIRE by providing a representation formalism for concept modeling and inference. We based our formalism on knowledge representations from artificial intelligence and presented a logic based formalism augmented with an uncertainty model. The formalism supports multi-modal inference, i.e. combining information from different media (e.g. audio and video) to infer context dependent concepts.

Although our experiments with inference are in an early stage, we already conclude that fully automated inference with an acceptable degree of certainty for inferred concepts is limited to very specific and well defined concepts here. In addition, we of course face the same problems as people working on knowledge representation in general.

One of the current shortcomings of the inference model is that it lacks adequate representation of spatio-temporal information. At the moment the way to represent the temporal information to explicitly model dependencies between concepts. We will look for an appropriate temporal logic to have explicit time modeling.

An additional focus of our research is the investigation of the level of automation that can be achieved with currently available techniques. We distinguish two areas: annotation and database support. As far as annotation is concerned, we conclude that fully automated annotation is not feasible, not even at the feature level. Actually only for simple features like color histograms fully automated annotation is feasible, however, for more completes features like shape or object recognition the only feasible way is a human supervised annotation process. At the concept level the situation is even worse, although concept inference can be of help for very simple concepts, also here human supervision is indispensable.

As far as the database support is concerned, we have to conclude that commercially available systems have a long way to go before they can offer integrated support for multimedia data. The main issue being the integration problem. Multimedia data management brings together techniques from various disciplines. What is needed therefore is an open data base management kernel that allows plug-in of the various techniques. This is understood by leading providers of database technology like Oracle, IBM, and Informix. However, their current products are based on traditional database management kernels with add-ons to support multimedia. As a result the kernels do not offer the level of openness to allow tight integration of the various techniques which is an absolute prerequisite to support multi-modal retrieval.

Finally, we mention our future work. The main emphasis will be on concept inference and generalization of results. The work on concept inference will focus on modeling context dependent inference, and spatio-temporal reasoning. The work on generalization will focus on the applicability of the approach in the soccer domain to other (non-sports) domains.

References

1. S. Adali., K.S. Candan, Y. Papakonstantinou, and V.S. Subrahmania, *Query Caching and Optimization in Distributed Media Systems*. In Proc. of SIGMOD'96, Canada, pp137-148.
2. P. Aigraine, P. Joly and V. Longueville, *Medium Knowledge-Based Macro-Segmentation of Video into Sequences.* In Intelligent Multimedia Information Retrieval M. T. Maybury (Ed.) AAAI/MIT Press, 1997, pp 159-173
3. A. Akutsu, Y. Tonomura, Y. Ohba, and H. Hashimoto, *Video Indexing Using Motion Vectors.* In Proc. of SPIE Visual Comm. and Image Processing '92, Boston, Massachusetts, pp343-350, 1992
4. C. Baumgarten, and K. Meyer-Wegener, *Towards a Scalable Networked Retrieval System for Searching Multimedia Databases.* SIGIR'97 Networked Information Retrieval workshop, Philadelphia, July, 1997
5. C.H.C. Leung, J. Hibler, and N. Mwara, Content-based Retrieval in Multimedia Databases, In *Computer Graphics*, Vol. 28 No. 1, February 1994, pp 24-28
6. R.W. Picard and T.P. Minka, Visual texture for annotation, In *Multimedia Systems, Springer Verlag*, Vol. 3, No. 1, 1995, pp 3-13
7. J. C. Fuller Bach, A. Gupta, A. Hampapur, B. Horowitz , R. Humphrey, R Jain and C. Shu, *Virage image search engine: An open framework for image management.* In Proc. of the Symposium on Storage & Retrieval for Image and Video Databases VI, IS&T/SPIE, San Jose, California, Feb. 1996
8. M.G. Brown, J.T. Foote, G.J.F Jones, K. Sparck Jones and S.J. Young, *Automatic Content-based Retrieval of Broadcast News*. In Proceedings of ACM Multimedia, pp 35-43, San Francisco, 1995
9. S.F. Chang, W. Chen, H. Meng, H. Sundaram, and d. Zhong, VideoQ: *An automatic content based video search system using visual cues.* In Proc. of ACM Multimedia '97, Seattle, Washington, Nov. 1997
10. M.W. Davis and W.C. Ogden. *QUILT: Implementing a Large Scale Cross-Language Text Retrieval System.* In Proc. of SIGIR'97, Philadelphia, USA, pp. 91-98, July 1997.
11. M.E. Davis, *Media streams: representing video for retrieval and repurposing,* PhD. Thesis, MIT, 1995
12. M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Stelle and P. Yanker. *Query by Image and Video content: The QBIC System*, IEEE Computer, September 1995, pp23-32
13. R.C. Gonzalez, & Woods, R.E. (1992). *Digital Image Processing*. Addison-Wesley
14. Gu J., and E.J. Neuhold, *A data model for multimedia information retrieval*. Proc. of 1st Int. Conf. on Multimedia Modeling, Singapore, 1993, p. 113-127
15. V.N. Gudivada, V.V Raghavan, and K. Vanapipat, *A Unified Approach to data Modelling and Retrieval for a Class of Image database applications.* In Multimedia Database Systems, Issues and Research Directions, Subrahmanian, V.S., S. Jajodia (Eds.), Springer Verlag, pp37-78, 1996
16. W.I. Grosky, *Managing Multimedia Information Database Systems*. In Communications of the ACM, Vol. 40., No. 12, pp 72-80
17. V. Burrill, Group I Report: Authoring Systems, In *Multimedia Systems and Applications*, J.L. Encarnação, J.D. Foley (Eds.), Springer-Verlag, 1994, pp 11-23
18. A.G. Hauptmann, and M. Witbrock. *Informedia: News-on-Demand Multimedia Information Acquisition and Retrieval*. In Intelligent Multimedia Retrieval. Maybary (Ed.), AAAI/MIT Press 1997, pp 215-239

19. J.N.H. Heemskerk, R.G. Heller, W. Jonker, E.J. Sommer & A. Woudstra (1997). *Multi Media Manifestation Part II: Implementation of a Soccer VDS*. KPN Research, R&D-RA-97-1036
20. R. Lienhart, S. Pfeiffer, and W. Effelsberg, *Video Abstracting*. In Communications of the ACM, Vol. 40, No.12, December 1997, pp55-62
21. I. Mani, D. House, M.T. Maybury and M. Green, *Towards Content-Based Browsing of Broadcast News Video*, In Intelligent Multimedia Information Retrieval, M. T. Maybury (Ed.) AAAI/MIT Press, 1997, pp 241-258
22. J.R. Parker, *Algorithms for Image Processing and Computer Vision*. John Wiley & Sons, Inc. 1997
23. A. Pentland, R. Picard, and S. Sclaroff, Photobook: *Tools for Content Based Manipulation of Image Databases*. In International Journal on Computer Vision, Vol. 18, No. 3, 1996, pp 233-154.
24. H.J.G. de Poot, F. Moelaert El-Hadidy and D.D. Velthausz (1997). *Multi Media Manifestation Part I: Modelling a Soccer VDS using ADMIRE*. Telematica Instituut, internal report IRS/97020.
25. R.K. Srihari, *Automatic Indexing and Content-Based Retrieval of Captioned Images*. In Special Issue of Computer, IEEE, September 1995 Content-Based Image Retrieval Systems, pp 49-56
26. K. Tsuda, and K. Yamamoto, M. Hirakawa, and T. Ichikawa, *MORE: An object-oriented data model with a facility for changing object structures*. IEEE on Knowledge and Data Engineering, vol. 3-4, `91, p. 444-460
27. H.J. Zhang, A. Kankanhalli, and S. Smoliar, *Automatic Partitioning of Full Motion Video*. In Multimedia System Vol. 1 No 1., pp 1-28, 1993.
28. H.J. Zhang, C.Y. Low, S.W. Smoliar and J.H. Wu, *Video parsing, retrieval and browsing: an integrated and content-based solution*. In Proc. 3rd ACM Int. Multimedia Conf., San Francisco, USA, ‘95, p. 15-24
29. D.D. Velthausz, Bal, C.M.R., Eertink, E.H., *A Multimedia Information Object Model for Information Disclosure*. MMM'96 proc. of the Third International Conference on MultiMedia Modelling, Toulouse, France, 12-15 November 1996,pp 289-304
30. D.D. Velthausz, *Cost-effective networked based multimedia information retrieval*. PhD. thesis to be published in 1998.
31. J.K. Wu, A.D. Narasimhalu, B.M. Mehtre, A.P. Lam and Y.J. Gao, *CORE: a content-based retrieval engine for multimedia information systems*. In Multimedia Systems, Vol. 3, No. 1, 1995, p. 25-41
32. E.A. Fox, R.M. Akscyn, R.K. Furuta, and J.L. Leggett (Eds), *Special issue of Communication of the ACM on Digital Libararies*, April 1995.
33. Narisimhalu, Special section on content-based retrieval. In *Multimedia Systems*, Vol. 3., No. 1, 1995, pp 1-2
34. D.B. Lenat, and R.V. Guha, Building Large Knowledge-Based Systems, Representation and Inference in the Cyc Project, Addison-Wesley Publishing Company, 1990