

Multimedia Indexing and Retrieval Techniques: A Review

Avinash N Bhute

Research Scholar

Dept. of Computer Technology,
VJTI, Matunga, Mumbai-19

B.B. Meshram

Professor & Head

Dept. of Computer Technology
VJTI, Matunga, Mumbai-19

Harsha A. Bhute

Assistant Professor

Dept. of Computer Engineering,
SCOE, Pune-41

ABSTRACT

Retrieval of multimedia has become a requirement for many contemporary information systems. These systems need to provide browsing, querying, navigation, and, sometimes, composition capabilities involving various forms of media. In this survey, we review techniques for text, image, audio and video retrieval. We first look at indexing and retrieval techniques for text, audio, image and video. We also discuss features visual features for video retrieval such as colour, texture, shape. The indexing techniques are discussed for these features. We also compare most popular techniques used for indexing and retrieval.

General Terms

Pattern Recognition, Computer Vision, Image/Video Processing, Video Indexing and Retrieval.

Keywords

IR, MIR, Multimedia Retrieval, Video Indexing.

1. INTRODUCTION

Although the term "search engine" is often used indiscriminately to describe crawler-based search engines, human-powered directories, and everything in between, they are not all the same. Each type of "search engine" gathers and ranks listings in radically different ways.

Crawler-based search engines such as Google, compile their listings automatically. They "crawl" or "spider" the web, and people search through their listings. These listings are what make up the search engine's index or catalog. One can think of the index as a massive electronic filing cabinet containing a copy of every web page the spider finds. Because spiders scour the web on a regular basis, any changes made to a web site may affect search engine ranking. It is also important to remember that it may take a while for a spidered page to be added to the index. Until that happens, it is not available to those searching with the search engine [11].

Directories such as Open Directory depend on human editors to compile their listings. Webmasters submit an address, title, and a brief description of their site, and then editors review the submission. The hybrid search engines will typically favor one type of listing over the other however.

Rest of the paper is organized as follow: Section 2 reviews various algorithms and approaches used for text information retrieval, section 3 we explain the various indexing and retrieval techniques for image, section 4 discusses audio indexing retrieval techniques, section 5 deals with techniques for video indexing and retrieval and finally in section 6 we conclude.

1.1 Recent Work in CBVIR

IBM's QBIC [15] system is the first commercial CBIR system and probably the best known of all CBIR systems. QBIC supports users to retrieve images by color, shape and texture. QBIC provides several query methods: Simple, Multi-feature, and Multi-pass. In the Simple method, a query is processed using only one feature. A Multi-feature query involves more than one feature, and all features have equal weights during the search. A Multi-pass query uses the output of a previous query as the basis for further refinements. Users can draw and specify color and texture patterns in desired images. In QBIC, the color similarity is computed by quadratic metric using k-element color histograms, and the average colors are used as filters to improve query efficiency. Its shape function retrieves images by shape area, circularity, eccentricity, and major axis orientation. Its texture function retrieves images by global coarseness, contrast, and directionality features. An R*-tree is used as the underlying indexing structure to index multi-dimensional features. The online demo of QBIC is at www.qbic.almaden.ibm.com.

The Photobook[16] system, developed at the Massachusetts Institute of Technology, allows users to retrieve images by color, shape and texture features. This system provides a set of matching algorithms, including Euclidean, mahalanobis, divergence, vector space angle, histogram, Fourier peak, and wavelet tree distances, as distance metrics. In its most recent version, users can define their own matching algorithms. The system includes a distinct interactive learning agent (FourEyes), which is a semi-automated tool and can generate query models based on example images provided by users. This allows users to directly address their query demands for different domains and, for each domain, users can obtain an optimal query model. The online demo can be seen at: <http://www-white.media.mit.edu/vismod/demos/photobook>

Both VisualSEEK and WebSEEK[17] were developed at Columbia University. The VisualSEEK system is an image database system. It allows users to retrieve images using color, spatial layout, and texture features. Color Set and the Wavelet Transform based texture are used to represent these features. In addition, VisualSEEK allows users to form queries by specifying color regions and their spatial locations. WebSEEK is an image catalog and search tool for the web. This system provides a prototype to catalog images and videos on the web using a combination of the text-based processing and content-based visual analysis. The online demo can be seen at: <http://disney.ctr.columbia.edu/webseek>.

RetrievalWare [18], developed by Excalibur Technologies Corp., lets users retrieve images by Color Content, Shape Content, Texture Content, Brightness, Color Structure, and Aspect Ratio. Users can adjust the weights of these features during the search process. The online demo can be seen at: <http://vrw.excalib.com:8015/>.

The IMatch [19] system allows users to retrieve images by color, texture, and shape. IMatch supports several query methods to query similar images: Color Similarity, Color and Shape (Quick), Color and Shape (Fuzzy), and Color Distribution. Color Similarity queries for images similar to an example image based on the global color distribution. Color and Shape (Quick) queries similar images for a given image by combining shapes, textures, and colors. Color and Shape (Fuzzy) performs additional steps to identify objects in example images. Color Distribution allows users to draw color distributions, or specify the overall percentage of one color in desired images. IMatch also supports non-CBIR features to identify images: binary identical images, duplicate images that have been resized, cropped, or saved in different file formats, and images that have similar file names to the given images. Demos for IMatch can be downloaded from: www.mwllabs.de/download.htm.

Oscar D. Robles et. al.[20] propose the two new primitives for representing the content of a video in order to be used in a Content-Based Video Retrieval System. The techniques presented in the paper titled "Towards A Content-Based Video Retrieval System Using Wavelet-Based Signature" compute first a multi-resolution representation using the Haar transform. Two types of signatures are extracted afterwards, one based on multi-resolution global color histograms and the other one based on multi-resolution local color histograms. The tests performed in the experiments include the recall measure achieved with the proposed primitives.

Wang et al.[21] propose a new approach for semantics-based image retrieval. They use color-texture classification to generate the codebook which is used to segment images into regions. The content of a region is characterized by its self-saliency and the lower-level features of the region, including color and texture. The context of regions in an image describes their relationships, which are related to their relative-salencies. High-level (semantics-based) querying and query-by-example are supported on the basis of the content and context of image regions.

Bruijn and Lew[22] investigate multi-modal content-based browsing and searching methods for Peer2Peer retrieval systems. This is the first P2P system which performs analysis on the video content for browsing multimedia collections over large, distributed P2P networks. Their work targets the assumption that keyframes are more interesting when they contain people. The main contribution of in this paper toward video retrieval is the heuristic algorithm for finding a single keyframe to represent an entire video. It was designed intentionally to minimize computational complexity. The novel aspect of the keyframe selection algorithm was to create a new color space based on the theory of optimal linear classifiers. This new color space is optimized for classifying small pixel regions into the categories of skin or non-skin.

Vendrig and Worring [23] propose a system that allows Character Identification In Movies. In order to achieve this, they relate visual content to names extracted from movie scripts. They propose the WhoIsWho function for our interactive i-Notation system. WhoIsWho relates visual content to names extracted from movie scripts, working in both ways. that present extensive evaluation of character identification on six hours of movies. Employment of a user model enables evaluation of interactivity in WhoIsWho.

Denman et al [24] present the tools in a system for Creating Semantically Meaningful Summaries of Broadcast Snooker Footage. Their system parses the video sequence, identifies

relevant camera views, and tracks ball movements. They propose the three new tools appropriate for content analysis of sports, applied to footage from snooker broadcasts in particular. The first tool is a new feature for parsing a sequence based on geometry without the need for deriving 3D information. The second tool allows events to be detected where an event is characterized by an object leaving the scene at a particular location. The third tool is a mechanism for summarizing motion in a shot for use in a content based summary. As a matter of course, the paper considers a number of enabling techniques such as the removal of irrelevant objects and object tracking using a particle filter. An approach presented by Kim et al [25] extracts semantic information from basket-ball videos based on audio-visual features

A Semantic Video Retrieval Approach using audio analysis is presented by Bakker and Lew [26] in which the audio can be automatically categorized into semantic categories such as explosions, music, speech, etc. In their research literature, significant attention has been given to the visual aspect of video, however, relatively little work directly uses audio content for video retrieval. The paper gives an overview of our current research directions in semantic video retrieval using audio content. In this paper discuss the effectiveness of classifying audio into semantic categories by combining both global and local audio features based in the frequency spectrum. Furthermore, introduce two novel features called Frequency Spectrum Differentials (FSD), and Differential Swap Rate (DSR), that both model the shape of the spectrum.

A system for Recognizing Objects in Video Sequences is presented by Visser et al [27]. They use the Kalman filter to obtain segmented blobs from the video, classify the blobs using the probability ration test, and apply several different temporal methods, which results in sequential classification methods over the video sequence containing the blob.

An Automated Scene Matching Algorithm is presented by Schaffalitzky and Zisserman [28]. Their goal is to match images of the same 3D scene in a movie. They describe progress in matching shots which are images of the same 3D scene in a film. The problem is hard because the camera viewpoint may change substantially between shots, with consequent changes in the imaged appearance of the scene due to foreshortening, scale changes and partial occlusion. They demonstrate that wide baseline matching techniques can be successfully employed for this task by matching key frames between shots. The wide baseline method represents each frame by a set of viewpoint invariant local feature vectors. The local spatial support of the features means that segmentation of the frame (e.g. into foreground/background) is not required, and partial occlusion is tolerated. Results of matching shots for a number of different scene types are illustrated on a commercial film.

Ruiz-del-Solar and Navarrete [29] present a Content-Based Face Retrieval System that uses self-organizing maps (SOMs) and user feedback. SOMs were also employed by Oh et al, Hussain et al, and Huang et al for visual clustering. A ranking algorithm using dynamic clustering for content-based image retrieval is proposed by Park et al. A learning method using the AdaBoost algorithm and a k-nearest neighbor approach is proposed by Pickering et al for video retrieval.

An overview of challenges for Content-Based Navigation of Digital Video is presented by Smeaton[30]. The author presents the different ways in which video content can be used directly to support the navigation within large video libraries and lists the challenges that still remain to be addressed in this

area. Some of these content-based mechanisms have been developed and implemented on video already and we use our own Físchlár system to illustrate many of these. Others remain beyond our current technological capabilities but by sketching out the possibilities and illustrating with examples where possible, as we do in this paper, we help to define what challenges still remain to be addressed in the area of content-based video navigation.

An insight into the problems and challenges of Retrieval of Archival Moving Imagery Via The Internet is presented by Enser and Sandom[31]. The authors conclude that the combination of limited CBIR functionality and lack of adherence to cataloging standards seriously limits the Internet’s potential for providing enhanced access to film and video-based cultural resources. They reports on research which sought to gain an informed view on this potential. Working in collaboration with eleven representative film collections, evidence was gathered relating to subject access facilities and to client requirements in the form of 1,270 requests. The work revealed that the formulation and satisfaction of requests for archival footage places a heavy reliance on human intellectual input for which CBIR techniques offer little prospect of being an effective substitute. The conclusion is drawn that the combination of limited CBIR functionality and lack of adherence to cataloging standards seriously limits the Internet’s potential for providing enhanced access to film and video-based cultural resources.

Burke[32] describes a research project which applies Personal Construct Theory to Individual User Perceptions of Photographs. This work presents a librarian viewpoint toward content-based image retrieval. A user-centric system for visualization and layout for content-based image retrieval and browsing is proposed by Tian et al.[33]

2. TEXT RETRIEVAL AND INDEXING

The basic document retrieval process is shown in Figure 1. As shown on the right side of the figure, the documents are processed off-line to obtain document representations. These representations are stored together with documents themselves. During retrieval (left side of the figure), the user issues a query that is processed (online) to obtain its representation. Then the query representation is compared with the document representations. Documents deemed relevant by the system are retrieved and presented to the user, who evaluates the returned documents and decides which ones are actually relevant to the information need. A good IR

system should then allow the user to provide relevance feedback to the system. The system uses this information to modify query, query representation, and/or document representations. Another retrieval is done based on the modified query and document representations. If necessary, the retrieval-feedback process is iterated a few times. Note that not all IR systems support the user relevance feedback process.

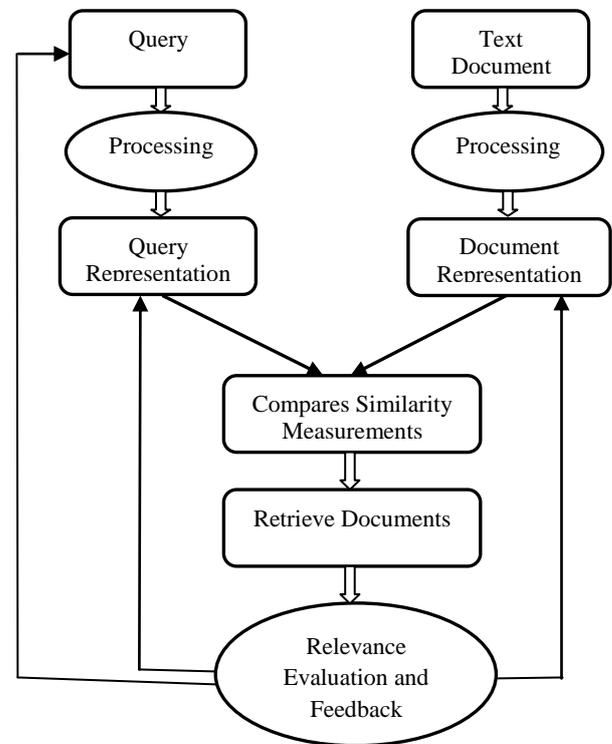


Fig 1: Basic Text/ Document Retrieval Model.

Following Table 1 briefly review various retrieval approaches with its advantages and limitations.

Table 1. Comparison of Text Indexing and retrieval Approaches

Algorithm	Approach	Advantage	Limitation
Basic Boolean	Documents are indexed by sets of keywords. Queries are also represented by a set of keywords joined by logical (Boolean) operators that supply relationships between the Query terms.	Simple to Implement. Ideal for small text data patterns and databases.	It is difficult to formulate Boolean queries and the retrieval results are very sensitive to query formulation. Query term weights are normally not used as queries are often very short.

Vector Space Based Indexing and Retrieval	<p>The vector space model assumes that there is a fixed set of index terms to represent documents and queries.</p> <p>A document D_i and a query Q_j are represented as</p> $D_i = [T_{i1}, T_{i2} \dots \dots T_{iK}, \dots T_{iN}]$ $Q_j = [Q_{j1}, Q_{j2} \dots \dots Q_{jK}, \dots Q_{jN}]$ <p>where T_{ik} is the weight of term k in document i, Q_{jk} is the weight of query k in query j, and N is the total number of terms used in documents and queries.</p> <p>Term weights T_{ik} and Q_{jk} can be binary (i.e., either 1 or 0), or tf.idf or weights obtained by other means.</p>	When subsequent queries similar to the queries used to modify the documents are issued, performance is improved.	Treats terms as unrelated and it only works well with short documents and queries.
Probabilistic Based Indexing & Retrieval	<p>It is based on the following four parameters:</p> <p>P(rel): the probability of relevance of a document P(nonrel): the probability of nonrelevance of a document a1: the cost associated with the retrieval of a non relevant document. a2: the cost associated with the non retrieval of a relevant document.</p>	<p>Since the retrieval of a nonrelevant document carries a loss of $a_1 P(\text{nonrel})$ and the rejection of a relevant document carries a loss of $a_2 P(\text{rel})$, the total loss caused by a given retrieval process will be minimized if a document is retrieved whenever</p> $a_2 P(\text{rel}) \geq a_1 P(\text{nonrel})$	Retrieval effectiveness and efficiency are low because not all relevant items may be retrieved and whole document space has to be searched.
Cluster-Based Indexing and Retrieval	<p>Document clustering grouping similar documents into clusters are used. During document retrieval, the query vector is compared with the centroids of clusters. After the cluster with highest similarity to the query vector is identified, either all documents in the cluster are retrieved or most similar documents are retrieved.</p>	Effective Retrieval	High computation, memory and time requirement.

3. IMAGE INDEXING & RETRIEVAL

This section briefly describes seven main approaches, which are based on structured attributes, object-recognition, text, and low-level image features. The first approach, attribute-based, uses traditional database management systems for image indexing and retrieval. The second approach color and third based on shape are not mature yet as it relies on automatic object recognition. Text-based image retrieval uses traditional IR for image indexing and retrieval. The low-level content-based approach to image indexing and retrieval requires the

extraction of low level image features. The common features used are color, object shape, and texture. As most images are stored in compressed form, it would be advantageous if we could derive image features directly from compressed image data. Table 2 looks into a number of image indexing and retrieval techniques based on compressed image data. An individual feature will not be able to describe an image adequately. For example, it is not possible to distinguish a red car from a red apple based on color alone. Therefore, a combination of features is required for effective image indexing and retrieval.

Table 2. Comparison of Image Indexing and Retrieval Approaches

Algorithm	Approach	Advantage	Limitation
Text-Based Indexing and Retrieval Techniques	Manual, Automated Annotation, Attributes based indexing and retrieval	Easy to implement, low memory and computational requirement.	Most text-based retrieval systems require manual annotation of images.
Color-Based Indexing and Retrieval Techniques	Histogram, Spatial relation among Pixels, Color distribution, etc	Histogram is easy to compute and effective in characterizing both the global and local distribution of colors in an image.	The retrieval performance may be poor.
Image Retrieval Based on	Major and Minor axis, rectangle,	Object can be recognized	Accurate shape detection

Shape	Eccentricity and moment invariants, histogram of edges detected, elastic template matching, Region based shape representation, Rotational Normalization.	and retrieved.	will require human intervention because methods like segmentation are very difficult to completely automate.
Sequential Attribute Tree	The SMAT is a multi-tier tree structure, where each layer corresponds to an indexing attribute.	Multiple features are used for indexing.	
Texture Based Methods	Coarseness, Contrast, Directionality, Line likeness, Regularity, Roughness	Ideal for natural images and satellite images.	Poor performance on images other than natural images.
Indexing and Retrieval Based on DCT Coefficients	DCT Coefficients	Performs well with JPEG images	The choice of windows affects the performance dramatically, as the objects of interest may be located anywhere in images.
Indexing and Retrieval Based on Wavelet Coefficients	In a wavelet transform, a signal is decomposed into a number of the chosen base function and its variations, called wavelets. The result of the wavelet compression is a sequence of wavelet coefficients.	Effective performance if integrated with clustering techniques.	

4. AUDIO INDEXING & RETRIEVAL

Improvements in speech recognition technology and computing power have enabled the development of usable indexes for vast spoken audio repositories (e.g. [9]). A standard technique is to use speech recognition to transcribe the audio and then to build an index using this transcription. However, this approach suffers from the fact that a speech recognizer has a limited vocabulary so the system cannot retrieve out of vocabulary (OOV) queries. A popular technique to confront this problem is to use phoneme rather than word recognition. Here, a phoneme recognition system is used to transcribe the spoken audio. Word queries are then converted to phoneme sequences and searched for in the transcriptions. Following Table 3 summarizes various techniques.

5. VIDEO INDEXING & RETRIEVAL

With the development of multimedia data types and available bandwidth there is huge demand of video retrieval systems, as users shift from text based retrieval systems to content based retrieval systems. Content based Video Indexing and Retrieval (CBVIR), in the application of image retrieval problem, that is, the problem of searching for digital videos in large databases. "Content-based" means that the search will analyze the actual content of the video. The term 'Content' in this context might refer colours, shapes, textures. Without the ability to examine video content, searches must rely on images provided by the user [10].

Table 3. Comparison of Audio Indexing and Retrieval Approaches

Algorithm	Approach	Advantage	Limitation
Forward and Viterbi algorithms Based on Hidden Markov model	An HMM consists of a number of states, linked by a number of possible transitions Associated with each state are a number of symbols, each with a certain occurrence probability associated with each transition. When a state is entered, a symbol is generated. Which symbol to be generated at each state is determined by the occurrence probabilities.	Forward Algorithms words well with isolated words and Viterbi algorithms with continuous speech.	The challenge of speech recognition is to model audio/speech variations mathematically.
Techniques Based on Artificial Neural Networks	An ANN consists of many neurons interconnected by links with weights. Speech recognition with ANNs	Suitable Speech Recognition	Extensive training is required for effective retrieval.

	consists of training and recognition stages.		
The Virtual-Node (Vn) Algorithm	It exploits the multi-resolution capabilities of the hierarchical tree structure.	Fast and effective retrieval	High memory requirements.
Robust search Algorithm	Tree Indexed on query block boundaries aligned with those of data block boundaries.	Use of only a small number of transform coefficients to characterize the data. query blocks could be at any part of data, not necessarily aligned with data block boundaries	
Feature vector based techniques	Pitch, ZCR, etc features are extracted and indexed	Good retrieval results	Requires more time to compare query with database.

Video segmentation is first step towards the content based video retrieval aiming to segment moving objects in video sequences. Video segmentation initially segments the first image frame as the image frame into some moving objects and then it tracks the evolution of the moving objects in the subsequent image frames. After segmenting objects in each image frame, these segmented objects have many applications, such as surveillance, object manipulation, scene composition, and video retrieval[10]. Video is created by taking a set of shots and composing them together using specified composition operators.

Due to limitation of space in this paper we discuss Texture, Color and Shape features used for indexing.

5.1 Texture Feature

Texture can be defined as the visual patterns that have properties of homogeneity that do not result from the presence of only a single color or intensity. These consists of six statistical features, including coarseness, contrast, directionality, line-likeness, regularity and roughness to describe various texture properties.

5.2 Color Feature

Color is one of the most widely used visual features in multimedia context and image / video retrieval, in particular.

To support communication over the Internet, the data should compress well and be suitable for heterogeneous environment with a variety of the user platforms and viewing devices, large scatter of the user's machine power, and changing viewing conditions. The CBIR systems are not aware usually of the difference in original, encoded, and perceived colors, e.g., differences between the colorimetric and device color data.

5.3 Shape Feature

An approach to shape recognition which has been prevalent for many years, the “statistical pattern recognition” approach. We make a set of measurements which independently characterize some aspect of the shape. Hopefully, we have a large collection of examples, so we may then characterize the shape statistically. Suppose, for example, the mission of the project is to distinguish between sharks and sting rays. Measurements may include properties of a region such as area, perimeter, aspect ratio, eigen values, convex discrepancy, and various central moments. Though the computation of such features can be challenging, we do not discuss the actual computational process here, but rather refer the reader to texts on computational geometry. Before we can consider the use of simple geometric features, we must discuss briefly how such features might be used, which is, in turn, a pattern recognition problem. Simple region growing algorithm is to segment a black-and-white image in regions.

Following Table 4 summarizes the video indexing and retrieval methods.

Table 4. Comparison of Video Indexing and Retrieval Approaches

Algorithm	Approach	Advantage	Limitation
Indexing and Retrieval Based on r Frames of Video Shots	The most common way of creating a shot index is to use a representative frame to represent a shot. An r frame captures the main contents of the shot. Features of this frame are extracted and indexed based on color, shape, and/or texture, as in the case of image retrieval.	Single frame represent a shot, reducing complex computations.	Treats video as a collection still images and ignores temporal or motion information contained in the video
Indexing and Retrieval Based on Motion Information	Motion content, motion uniformity, motion panning, motion titling, etc. is captured from video.	Extension of r frame method and takes care of motion information	Retrieval performance depends on window size.

Indexing and Retrieval based on Object	Object Segmentation	Object-based video indexing and retrieval can be carried out easily when video is compressed using the MPEG-4 object-based coding standard	Design and implement object tracking and segmentation techniques are difficult to implement
Indexing and Retrieval Based on Metadata	Video Meta data is used for indexing	Suitable for conventional DBMSs	
Indexing and Retrieval Based on Annotation	Manual or Automated Annotation	Performs well if transcripts are available	This approach is very challenging because speech and no speech are normally mixed in soundtracks, and there is background music and/or noise in speech signals, leading to a lower recognition rate
Fast Search Algorithm for Short Video Clips from Large Video Database Using a Novel Histogram Feature	This algorithm is based on the adjacent pixel intensity difference quantization (APIDQ) algorithm	Robust against Gaussian noise	

6. CONCLUSION

Despite the considerable progress of academic research in multimedia retrieval, there has been relatively little impact of content based video retrieval research on commercial applications with some niche exceptions such as video segmentation. Choosing features that reflect real human interest remains an open issue. One promising approach is to use meta learning to automatically select or combine appropriate features. Another possibility is to develop an interactive user interface based on visually interpreting the data using a selected measure to assist the selection process. Extensive experiments comparing the results of features with actual human interest could be used as another method of analysis. Since user interactions are indispensable in the determination of features, it is desirable to develop new theories, methods, and tools to facilitate the user's involvement.

7. REFERENCES

- [1] Dufour, Y. Estève, P. Deléglise, and F. Béchet, "Local and global models for spontaneous speech segment detection and characterization," in ASRU 2009, Merano, Italy, 2009.
- [2] Bazillon, Y. Estève, and D. Luzzati, "Manual vs assisted transcription of prepared and spontaneous speech," in LREC 2008, Marrakech, Morocco, 2008.
- [3] S. Kankanhalli and Y. Rui, "Application Potential of Multimedia Information Retrieval", Proc. IEE, April 2008.
- [4] Datta, D Joshi, J Li, and J. Wang, "Image Retrieval: Ideas, Influences, and Trends of the New Age", ACM Computing Surveys, VOL 40, No. 2, April 2008.
- [5] P. Sinha and Ramesh Jain, "Concept Annotation and Search Space Decrement of Digital Photos using Optical Context Information", In Proceedings of SPIE, Multimedia content Access: Algorithms and System, January 2008.
- [6] C. F. Wong and C. H. C. Leung. Automatic semantic annotation of real-world web images. IEEE Transactions on Pattern Analysis and Machine Intelligence, 30(11):1933{1944, November 2008.
- [7] Utz Westermann and Ramesh Jain, "Towards a Common Event Model for Multimedia Applications", in IEEE Multimedia, January 2007.
- [8] A. Scherp and R. Jain, "Towards an ecosystem for semantics", In Proceedings of Workshop on Many faces of Multimedia Semantics, at ACM Multimedia 2007, pp. 3-12, Sept. 2007. H. Yang, A. Dasdan, R.-L. Hsiao, and D. S. Parker. Map-Reduce-Merge: Simplified relational data processing on large clusters. SIGMOD, 2007.
- [9] Hampapur, A. Borger, S. Brown, L. Carlson, C. Connell, J. Lu, M. Senior, A. Reddy, V. Shu, C. Tian, Y." S3: The IBM Smart Surveillance System: From Transactional Systems to Observational Systems," in Proc. Acoustics, Speech and Signal Processing, 2007. ICASSP April 2007.
- [10] V. Patel, B. B. Meshram, "Retrieving and Summarizing Images from PDF Documents", International Conference on Soft computing and Intelligent Systems(ICSCSI-07), Jabalpur, India, 27-29 December 2007.
- [11] B Liu, A. Gupta, and R. Jain, "MEDSMAN: a live multimedia stream querying system", Int. Journal of Multimedia Tools and Applications, 2007.
- [12] S. Datta, C. Giannella, and H. Kargupta, " K-means clustering over a large, dynamic network", In Siam Conference of Data Mining, 2006.M. Lew, N. Sebe, C Djerba, and R. Jain, "Content-based Multimedia Information Retrieval: State of the Art and Challenges", ACM TOMCAPP vol.2, No. 1, pp. 1-19, 2006.
- [13] Milind Naphade , John R. Smith , Jelena Tesic , Shih-Fu Chang , Winston Hsu , Lyndon Kennedy , Alexander Hauptmann , Jon Curtis, "Large-Scale Concept Ontology for Multimedia," IEEE Multimedia, April 2006.

- [14] Keiji Yanai, Kobus Barnard, "Finding Visual Concepts by Web Image Mining", in proc. Of WWW 2006, May 23–26, 2006, Edinburgh, Scotland.
- [15] M. Flickher, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovicand D. Steele, and P. Yanker. "Query by image and video content: The QBIC system" In IEEE Computer, volume 38, pages 23-31, 1995.
- [16] A. Pentland, R. W. Picard and S. Sclaroff," Photobook: Content-based manipulation of image databases",SPIE storage and Retrieval Image and video database II,No.2185,Feb 6-10,1994,San Jose.
- [17] R. Smith and S.-F. Chang, "VisualSEEK: a Fully Automated Content-Based Image Query System, Proceedings", ACM Multimedia '96 Conference, Boston, MA, November 1996.
- [18] Avi Rappoport URL <http://searchtools.com/tools/retrievalware.html> The report is availableon<http://www.dtic.mil/cgi-bin/GetTRDoc?AD=ADA252509&Location=U2&doc=GetTRDoc.pdf>
- [19] Stringa, Elena, Paul Meylemans, et al, "Image Retrieval byExample:Techniques and Demonstrations." proceedings of ESARDA (European Safeguards Research and Development Association) Symposium on Safeguards and Nuclear Material Management, Bruges (Belgium). 2001.
- [20] Shengjiu Wang, "A Robust CBIR Approach Using Local Color Histograms", technical report ,department of computer science, University of Alberta, Canada , 2001.
- [21] M. Bressan D. Guillamet J. Vitrià" Using an ICA representation of local color Histograms for object recognition, Pattern Recognition ,Volume 36, Issue 3, March 2003, Pages 691-701
- [22] J. Han and K. Ma, "Fuzzy Color Histogram and Its Use in Color Image Retrieval",IEEE Trans. On Image Processing, vol. 11, pp. 944 – 952, Aug. 2002.
- [23] Y. Rui, T. S. Huang and S. Chang, "Image Retrieval: Current Techniques, Promising Directions and Open Issues ", Journal of Visual Communication and Image Representation, vol. 10, pp. 39-62, March 1999.
- [24] A. Bovik, Handbook of Image and Video Processing, 2nd Edition, Elsevier Academic Press, ISBN 0-12-119792-1, pp. 993-1013, 2005.
- [25] Robert M Haralick, K Shanmugam, Its'hak Dinstein (1973). "Textural Features for Image Classification". IEEE Transactions on Systems, Man, and Cybernetics SMC-3 (6): 610–621.
- [26] B. Jähne, H. Scharr, and S. Körkel. Principles of filter design. In Handbook of Computer Vision and Applications. Academic Press, 1999.
- [27] H. Farid and E. P. Simoncelli, Optimally Rotation-Equivariant Directional Derivative Kernels, Int'l Conf Computer Analysis of Images and Patterns, pp. 207--214, Sep 1997.
- [28] H. Farid and E. P. Simoncelli, Differentiation of discrete multi-dimensional signals, IEEE Trans Image Processing, vol.13(4), pp. 496--508, Apr 2004.
- [29] R. C. Gonzalez and R. E. Woods. Digital Image Processing. Addison-Wesley, Reading, MA, USA, 3rd edition, 1992.
- [30] J. Smith and S. Chang. Tools and techniques for color image retrieval. In Proc. of the SPIE conference on the Storage and Retrieval for Image and Video Databases IV, pages 426–437, San Jose,CA, USA, 1996.
- [31] M. A. Stricker and A. Dimai. Color indexing with weak spatial constraints. In Proc. of the SPIE conference on the Storage and Retrieval for Image and Video Databases IV, pages 29–40, San Diego,CA, USA, February 1996.
- [32] W. Hsu, T. S. Chua, and H. K. Pung. An integrated color-spatial approach to content-based imageretrieval. In Proc. of the ACM Multimedia 95, pages 305–313, 1995.
- [33] J. Li, Y. Tian, and W. Gao, "Exploring inter-frame correlation analysis and wavelet-domain modeling for real-time caption detection in streaming video," San Jose, CA, United states, 2008, p. The Society for Imaging Science and Technology (IS and T); The International Society for Optical Engineering (SPIE).