Mining Techniques for Integrated Multimedia Repositories: A Review

Naveen Aggarwal[1], Dr. Nupur Prakash[2] and Dr. Sanjeev Sofat[3]

Abstract - ***The multimedia databases and the need to intuitively handle their content, which meets the user's requirements with the available content based video indexing and retrieval technology, are the main focus of the research in the field of multimedia and computer vision. The researchers mainly focus on the problem of bridging the "semantic gap" between a user's need for meaningful retrieval and the current technology for computational analysis and description of the media content. It takes into account both the high complexity of the real-world implementation and user's need for conceptual video retrieval and browsing. In this paper, the initial work done in this area during last 15 years has been categorized in three generations. The key technologies in each generation are reviewed and characterized based on the standard parameters. It is found that in first and second generation all techniques are semantic less techniques, but in third generation, techniques based on semantics have been evolved. But still most of the techniques are in their infancy and require lots of research for their use in daily applications. In last a solution is proposed for a general purpose multimedia mining application which caters to the needs of the different types of domains.***

1. INTRODUCTION:

The development of various multimedia compression standards in last decade has made the widespread exchange of multimedia information a reality. Due to significant increase in desktop computer performance and a decrease in the cost of storage media, extraordinary growth of multimedia information in private and commercial databases has been seen. Further its ubiquity throughout the World Wide Web, presents new research challenges in computing, data storage, retrieval and multimedia communications. Intuitive handling of this vast multimedia information is the demand of users. Keeping this in mind, multimedia and computer vision researchers are focusing on the development of content based multimedia indexing and retrieval. However, evolution of functional multimedia management system is hindered by the "semantic gap"; a discontinuity between simplicity of content description that can be currently computed automatically and the richness of semantics in user's queries posed for media search and retrieval [1]. The availability of cost effective means for obtaining digital video has led to the easy storage of digital video data, which can be widely distributed over networks or storage media such as CDROM or DVD. Unfortunately, these collections are often not catalogued and are accessible only by the sequential scanning of the sequences. To make the use of large video databases more feasible, user should be able to automatically index, search and retrieve relevant material. Content-Based Video Indexing and Retrieval (CBVIR) has been the focus of the research community during last 15 years. The main idea behind this concept is to access information and interact with large collections of videos for referring and interacting with its content, rather than its form. Although there has been a lot of effort put in this research area, the outcomes have not been very encouraging. The discontinuity between the available content descriptions like color layout or motion activity and the user's need for rich semantics in user queries makes user approval of automated content retrieval systems very difficult. Thus, in order to develop a meaningful CBVIR system one has to involve multidisciplinary knowledge ranging from image and video signal processing to semiotic theories and video production techniques. Signal processing and computer vision methodologies achieved astonishing results in extracting structural and perceptual features from the video data. Algorithms from database system theory and other computer science disciplines enabled efficient, adaptive and intelligent indexing and retrieval of data with various structure and content. Furthermore, fields like computational linguistics and even semiotics have engaged with problems of natural language and even visual media semantics. However, this knowledge is scattered and needs a way to fuse into one system that will enable content-based retrieval of videos in a way natural for users. Multimedia Mining has evolved immensely in last decade. This paper classifies the evolution of Multimedia Mining in three generations. The performance analysis of algorithms of shot detection is analyzed and the paper is concluded with some comments on the future directions.

[1]University Institute of Engg. & Technology, Panjab University, Chandigarh, H.No 230, Punjab Engineering College Campus, Sector-12, Chandigarh-160012, India. Phone: +91-9814865455
[2]School of Information Technology, Guru Gobind Singh Indraprastha University, Delhi
[3]Computer Sc. & Engg. Deptt., Punjab Engg. College, Chandigarh.
E-Mail: [1]navagg@gmail.com

2. FIRST GENERATION

In the first generation of visual retrieval systems, feature descriptors of video data are expressed manually. Representation of these features provides high level of image ABSTRACTion and model visual content at a conceptual level. These features identify significant entities contained in the image or video (an object, a person, etc.), object parts (eyes in the face, boat in the lake, etc.) or the scene represented and concepts associated to it (a landscape, a storm, etc.). Features can be represented in schemes like relational models and object oriented models and can be queried using query languages like SQL. In Figure 1, main process of first generation is shown. Classification and indexing depends upon how accurate the image features are annotated manually. Cost of annotation is typically very high and the whole process suffers from subjectivity of descriptions, in that the annotator is a different person from the one who issues the query. Search engines like Google and Yahoo uses semantics of web to provide high level descriptions of video data.

2.1 Temporal Video Extraction

In first generation, main concern was to effectively extract the data temporally from video sources. Further video sources may be in compressed or uncompressed domain. All temporal video parsing techniques that exploit information in uncompressed domain lack efficiency. The reason for that is in the nature of the approach. In the feature extraction part the majority of uncompressed analysis techniques must initially decode the video stream and afterwards apply some processing on the vast pixel data, which additionally slows down the processing time. Thus, algorithms that base their analysis on pixel data require substantial processing time. Block-based algorithms [2, 3] and methods based on histogram comparison [4, 5] achieved considerable improvement in both processing requirements and sensitivity to camera and object motion, but far from the efficiency of the compressed domain analysis.

Gargi et al[5] compared the approaches of Arman et al.[6], Patel and Sethi [7], Yeo and Liu [8] and Shen and Delp [9] using different parameters such as: classification performance (recall and precision), full data use, ease of implementation, source effects. Ten MPEG video sequences containing more than 30,000 frames connected with 172 cuts and 38 gradual transitions are used as an evaluation database. It is found that the algorithm of Yeo and Liu and those of Shen and Delp perform best when detecting cuts. Although none of the approach recognizes gradual transitions particularly well, the best performance is achieved by Shen and Delp[9]. The reason for the poor gradual transition detection of the algorithms is their design.

The algorithms are designed by keeping the ideal behavior of transitions in mind. The gradual transitions ideally remain linear in space and time. But the actual frame differences do not follow this ideal pattern smoothly for the entire transition due to the presence of noise. Frame differences are also affected by the degradations due to different sampling and bit rates. Another interesting conclusion is that performance decreases significantly if we don't process all frame types (e.g., like in the first two methods). The algorithm of Yeo and Liu is found to be easiest for implementation as it specifies the parameter values and even some performance analysis is already carried out by the authors. The dependence of the two best performing algorithms on bitrate variations is investigated and shown by Gargi et al [5] that they are robust to bitrate changes except at very low rates. It is found that software encoder implementations also affect the performance. Yeo and Liu compared the effect of different software encoders. Although the techniques that works well for both uncompressed and compressed videos can be considered. But these techniques lack efficiency in either of domain. On the other hand, algorithms that access compressed domain features without additional processing and thus having the similar efficiency, underperformed in the accuracy and robustness criteria.

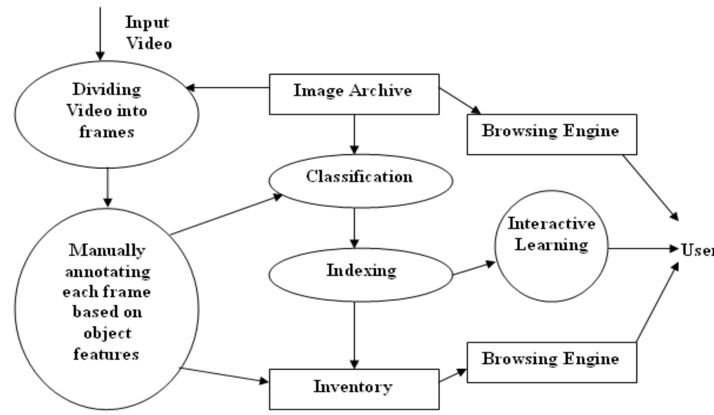

Figure 1: First Generation Video Retrieval Systems

3. SECOND GENERATION

Apart from the descriptors of first generation systems, the second generation systems also describes the perceptual features like color, textures, shape, spatial relationships, etc. These features are numeric descriptors of a video and can be obtained by fully automated objective measurements of the visual contents. So retrieval of content based data can be supported by combination of these features. Most of the techniques used to extract visual primitives from image frames have come from field of computer vision and pattern recognition. Therefore image processing, computer vision and pattern recognition subsystems are integral part of the architecture and operations of the second generation systems. Retrieval is based on similarity models that somehow replicate the way in which humans assess similarity between different objects. Apart from these parameters, Videos can be considered as a source of multi-planar visual information. Each plane communicates different attributes of information. These include the way in which the frames are linked together by using editing effects (cut, fades, dissolves, mattes, etc.), and high level information embedded in the frame sequence (the characters, the story content, the story message). Text embedded in the video frames and the other sensory data like speech and sound can be employed to extract useful data.

Main concern of research on second-generation systems is to extract video structure automatically [10]. In second generation, video indexing is performed by temporally segmenting the videos in unit like shots and scenes. Different image processing and computer vision techniques are used on index frames to generate low level feature descriptors. A metadata database of index frames and their corresponding feature descriptors can be created for later retrieval. When user makes a query, query is transformed into the structurally same low-level feature descriptor and the search engine finds the closest match from a metadata base. The system learns from relevance feedback from users during retrieval process and adapts the feature descriptor in order to achieve more consistent results in terms of perceptual similarity.

Even though some efficient results are reported in literature, there is a problem of bridging the gap between the systems and users. Similarity of perceptual properties is generally of little use in most practical cases of retrieval by content, if not combined with similarity of high-level information.

3.1 Low Level Features

Indexing of images and video contents using low level visual features is touch upon by several content based information retrieval systems. (e.g. WEBSEER [11], QBIC [12], and VisualSeek [13]). Three main task performed by these systems are:

1. Automatic extraction of visual feature descriptors
2. Indexing the extracted descriptors for fast access
3. Querying and matching descriptors for the retrieval of the visual data.

Apart from these features, learning of system is considered as an important aspect. Relevance feedback from users is used to refine the queries and learn through examples that the user may be looking for [14]. More recently, there has been focused effort on automatically producing certain semantic labels that could contribute significantly to retrieving visual data. For example, recent work has focused on portrait vs. landscape detection, indoor vs. outdoor classification, city vs. landscape classification, sunset vs. forest classification [15], [16], and other attempts to answer basic questions of who, what, when, and where about the visual content. Most of the approaches rely on traditional machine learning techniques to produce semantic labels, and some degree of success has been reached for various constrained and sometimes skewed test sets. However, these efforts represent only a small initial step towards achieving the real understanding of the visual content.

3.2 Bridging Gap

Numerous papers [17][18][19][20] have explored the problem of semantic gap and give valuable insights into the current state of the art. Wang et al [17] generate a code book by using the color-texture classification. Different regions of image are segmented using class definitions in the code book. The Entropy of a region contents describes its perceptual importance. Denman et al [18] present the system for creating semantically meaningful summaries of broadcast Snooker footage. Different tools in system are used for parsing the video sequence, identifying relevant camera views, and tracking ball movements. A system for recognizing objects in video sequences is presented by Visser et al [19]. 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 result in sequential classification methods over the video sequence containing the blob. An automated scene matching algorithm is presented by Schaffalitzky and Zisserman [20]. Main process is base on template matching of given scene in 3-D movies. Ruiz-del-Solar and Navarrete [21] present a system that uses self-organizing maps (SOMs) and user feedback for content-based face retrieval. A ranking algorithm using dynamic clustering for content-based image retrieval is proposed by Park et al [22].

3.3 Shot Detection

Shot Detection technique divides the continuous video, based on recognition of different frame sequences. One shot is characterized when camera remains still i.e. background is still and foreground is changing. When shot changes, background changes significantly. Earlier work in this category is evaluated in [5]. Earlier work is based on recognizing shot boundaries using color histograms, average pixel difference and histogram similarity matrix. Mihai Datcu et al used the motion compensation vector to estimate the motion of different objects [23]. Error in the motion estimation is used to differentiate the key frames from videos. Timo Volkmer described moving query window technique for TREC-12 video techniques [24]. Separate decision stages for abrupt and gradual transitions are applied during a single pass through the video clip. Similarly Masaru Sugano et al performed the shot boundary determination from I picture sequence of MPEG coded video [25]. Both pixel differences and histograms methods are used to overcome problems when either one of them is used. In this same approach is extended to detect shot boundaries in one frame unit. Jordi Mas et al used the color histogram differences and temporal color variations for video shot boundary detection [26]. The technique is able to differentiate abrupt shot boundaries by analysis of color histogram differences and smooth boundaries by temporal color variation. Five different algorithms for shot boundary detection have been analyzed using test data set of different types of videos ranges from news, sports, televisions and movies.

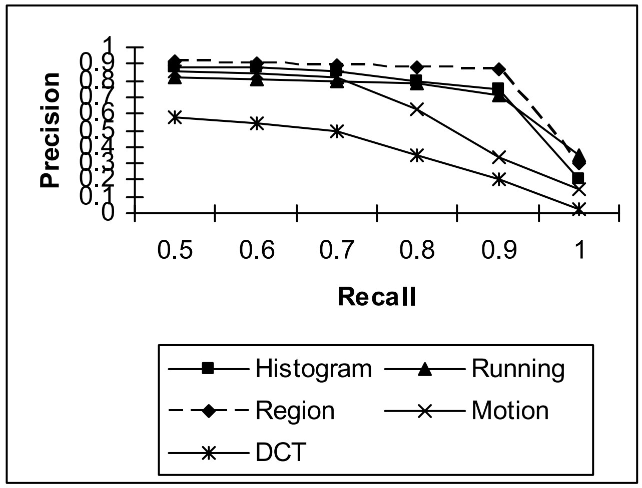

Figure 2: Comparison of Shot Detection Algorithms

As shown in Figure 2, different algorithms have been compared on a common platform from TRECVID 2006 [39] using different set of videos such as BBC News Videos, Sports Video, MTV Videos etc. Two main parameters precision and recall are calculated for each of the algorithm. From graph, it is clear that histogram based and region based algorithms outperforms other in shot detection. Region based algorithm also show some stability in detecting the gradual transitions.

4. THIRD GENERATION

The third generation retrieval systems seek for more high level information from images, audio and video content. Moving videos are not perceived as collection of shots by humans. Spectators even do not realize any editing in video based on the gradual transitions of one shot into another. So our aim is not only to extract the syntactic content (perceptual features) but also the semantic contents. Semantic content involves (1) the rhythm of the sequence (which is induced by the editing), (2) the scenes (which are obtained from shots), (3) the story (including the characters, their roles, actions and their logical relations), and (4) the feelings (which depends on the combination of perceptual facts like colors, objects, music, sounds, etc. and from the meaning of the story). Semantic contents and the perceived feelings of the user are used to support semantic-based retrieval automatically, with no or minimal manual intervention [27]. As human beings are more concerned with the narrative and discourse structure of video contents, so retrieval of video is generally meaningful only if performed at high levels of representation. Image sequence classification must also be based on semantically meaningful categories of information.

4.1 Object Detection

Most challenging problem in visual information retrieval is recognizing and detecting the objects in the moving videos. Several papers present the advances in this area [28][29][30][31]. Matas et al [28] present a method to recognize the objects even when viewing range is wide and illumination is not constant. Their method is robust to occlusion and background clutter in the frames. Sebe and Lew [29] have found that Gaussian noise distribution assumption is invalid. They proposed to use other metrics close to real noise distribution such as Cauchy metric. Further they explained how to create a maximum likelihood metric based on the real noise distributions and found that it consistently outperformed other analytic metrics. A hierarchical shape descriptor for object-based retrieval is proposed by Leung and Chan [30]. Brucale, et al [31] propose a class of topological-geometrical shape descriptors called size functions. Main limitation of this approach is hand drawn shape descriptors, which makes whole idea very much subjective to quality of hand drawn shapes. Based on the idea that the distribution of colors in an image provides a useful clue for image retrieval and object recognition, Berens and Finlayson [32] propose an efficient coding of three dimensional color distributions for image retrieval. Main focus in object recognition area is shifting to make whole process less computational intensive for real time queries.

4.2 Story Segmentation

Story Segmentation is another emerging area in which video is segmented based on different stories. So whereas Shot detection is semantic less technique, story segmentation is semantic based technique. Lekha Chaisorn et al presents a story segmentation technique as shown in Figure 3 using various low level and high level features[33]. It also makes use of various object level features.

4.3 Frameworks

A variety of different frameworks has been proposed to store retrieve and mine the raw videos. Jung Hwan Oh et al proposed such a framework, in which first data is characterized into different segments then each segment is clustered based on low level features and finally data is grouped using motion vectors and multi level hierarchical techniques [34].

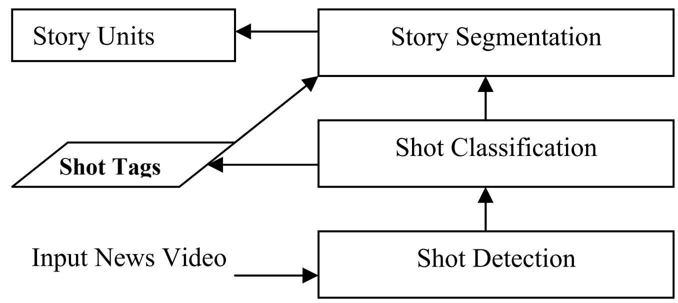

Figure 3: Story Segmentation

It is found that this technique is useful in limited domain and very time consuming. Mihai Datcu et al give an innovative concept for image information mining [23]. It represents the image in four steps. First step is to extract image features using different algorithms. In second step these are grouped together and further data is reduced by parametric modeling of clusters and finally supervised learning is used to represent the knowledge. William Perrizo et al present the significance of P-tree in mining any multimedia data [35]. In this paper, properties of P-trees are presented to take P-tree as Data Mining ready structure. The various implementation issues and scalability of P-tree for large size data are discussed. Xin Huang et al proposed a multimedia data mining framework. It incorporates Multiple Instance Learning into the user relevance feedback in a seamless way. Main aim is to discover the concept patterns of users (especially where the user's most interested region) and how to map the local feature vector of that region to the high-level concept pattern of users [36]. This underlying mapping can be progressively discovered through the feedback and learning procedure. Ankur Teredesai et al proposed the auto-annotation problem as a multi-relational association rule mining problem where the relations exist between image-based features, and textual annotations [37]. Their approach combined low-level image features, such as color, orientation, intensity, etc. and the corresponding text annotations to generate association rules across multiple tables using multi-relational association mining.

5. CONCLUSION

It is clear that Multimedia mining has evolved phenomenally in the last ten years as seen from the progress from first generation to third generation. Whereas, in first and second generation all techniques were semantic less, in third generation, techniques based on semantics have been evolved. But still most of the techniques are in their infancy and require lots of research for their use in daily applications. Multimedia

Mining is being used in different application domain such as Medical diagnosis, News Video Analysis, Sports Video Analysis, Movie Indexing and Satellite image Analysis etc. A general purpose system is still not available which could cater to the need of different domains. Further work need to be done in following areas

1. Multimedia Repositories are supposed to have data in both compressed and uncompressed form. Therefore a representation technique is required which will be able to represent both compressed and uncompressed data in common format which can further be used for mining.
2. Various Multimedia databases provide techniques to organize the data based on its general features. However effective organization should be content based. So a content based organization approach for multimedia data is required.
3. High level features are although based on the semantic of data but they can not be used directly for querying. So an approach is required to relate the semantic of High level feature with the internal representation of data. High level features are usually represented either in First Order Logic or by using Association Rules. Both techniques need to be analyzed in variety of different domains.
4. Further the problem of multi-resolution, scalability and hierarchical video description need to be tackled to gain efficiency in task of video indexing.
5. Optimization of different techniques for temporal extraction of data from compressed and uncompressed domain is required. Already proposed different techniques need to be analyzed and tested with the Benchmark videos proposed by TRECVID.
6. Overall Mining process of Multimedia repositories is bound to be computational and storage expensive. Hence different optimizations can be proposed for limited domains.

Thus a large scale system that merges the semantic text based retrieval approach to multimedia database with content based feature analysis and investigates the significant links between them appears to be the next milestone of research in the field of multimedia management systems [38].

REFERENCES

[1] J. Calic , N. Campbell , A. Calway , M. Mirmehdi , B. T. Thomas, T. Burghardt , S. Hannuna, C. Kong , S. Porter , N. Canagarajah , D. Bull, "Towards Intelligent Content Based Retrieval of Wildlife Videos", Proc. to the 6th International Workshop on Image Analysis for Multimedia Interactive Services, WIAMIS'2005, EPFL, Montreux, Switzerland, April 2005.

[2] Shu-Ching Chen, Mei-Ling Shyu, Chengcui Zhang, and R.L. Kashyap, "Video Scene Change Detection Method Using Unsupervised Segmentation and Object Tracking," Proceedings of the IEEE International Conference on Multimedia and Expo (ICME), August, 2001, Waseda University, Tokyo, Japan.

[3] Yazdi, M.,Zaccarin, A., Scene break detection and classification using a block-wise difference method,in Proceedings of International Conference on Image Processing, vol 3,Page(s):394 - 397,2001

[4] J. Calic, S. Sav, E. Izquierdo, S. Marlow, N. Murphy and N.E. O'Connor, "Temporal Video Segmentation for Real-Time Key Frame Extraction", ICASSP 2002, Orlando, Florida, USA

[5] U.Gargi, R.Kasturi, S.Antani, "Performance Characterization and Comparison of Video Indexing Algorithms" in proceeding of conference on computer vision and pattern recognition (CVPR), 1998

[6] F.Arman, A.Hsu, M-Y Chiu, "Image Processing on Compressed data for large Video databases" in proceedings of First ACM international conference on Multimedia, 1993

[7] ", published in Pattern Recognition N.Patel, I.K.Sethi, "Video Shot detection and Characterization for video databases vol 30, 1997

[8] B.Yeo, B.Liu, "Rapid Scene analysis on compressed video", IEEE Transactions on circuits and systems for video technology, vol 5, no. 6, December, 1995

[9] K.Shen, E.Delp, "A fast algorithm for video parsing using MPEG compressed video sequences", in proceedings of International conference on Image Processing (ICIP '96), 1996

[10] J.Boreczky, L.A.Rowe, "Comparison of video shot boundary detection techniques"in proceedings IS & T/SPIE International Symposium on Electronic Imaging, San Jose 1996

[11] C. Frankel, M. Swain & V. Athitsos, "WebSeer: an image search engine for the world-wide web" Technical Report 94-14, Computer Science Department, University of Chicago, August, 1996

[12] M. Flickner, H. Sawhney, W. Niblack, J. Ashley, Q. Huang, B. Dom, M. Gorkani, J. Hafner, D. Lee, D. Petkovic, D. Steele, and P. Yonker. Query by image and video content: The QBIC system, in Computer, Vol 28, pages 23–32, 1995.

[13] J. R. Smith and S.-F. Chang. "VisualSEEK: a fully automated content-based image query system." In proceedings of ACM Multimedia '96, November, 1996.

[14] Selim Aksoy, Ozge Cavus, "A Relevance Feedback Technique for Multimodal Retrieval of News Videos," in EUROCON, Belgrade, Serbia & Montenegro, November 21-24, 2005.

[15] M.SZummer, R.Picard, "Indoor-Outdoor Image classification" IEEE Workshop on Content-based access of image and video databases in conjunction with ICCV'98, Bombay, India , 1998

[16] A. Vailaya, A.Jain, H.J.Zhang, "On Image Classification City vs Landscape", IEEE Workshop on Content-Based access of image and video libraries", Santa Barbara, CA, June, 1998

[17] Wang,W., Song, Y., Zhang.A "Semantics-based image retrieval by region saliency", in International Conference on Image and Video Retrieval, Lecture Notes in Computer Science, vol. 2383, Springer 2002.

[18] Denman, H., Rea, N., Kokaram, A "Content based analysis for video from snooker broadcasts" in International Conference on Image and Video Retrieval, Lecture Notes in Computer Science, vol. 2383, Springer, 2002
[19] Visser, R., Sebe, N., Bakker, E, "Object recognition for video retrieval", in International Conference on Image and Video Retrieval, Lecture Notes in Computer Science, vol. 2383, Springer, 2002
[20] Schaffalitzky, F., Zisserman.A "Automated scene matching in movies. In: International Conference on Image and Video Retrieval, Lecture Notes in Computer Science, vol. 2383, Springer, 2002
[21] Ruiz-del-Solar, J., Navarrete, P, "FACERET: An interactive face retrieval system based on self-organizing maps" in International Conference on Image and Video Retrieval, Lecture Notes in Computer Science, vol. 2383, Springer, 2002
[22] Park, G., Baek, Y., Lee, H.K, "A ranking algorithm using dynamic clustering for content-based image retrieval", in International Conference on Image and Video Retrieval, Lecture Notes in Computer Science, vol. 2383, Springer, 2002
[23] Mihai Datcu and Klaus Seidel, "An Innovative Concept For Image Information Mining", in proceeding of International workshop on Multimedia Data Mining with ACM SIGKDD, 2002
[24] Timo Volkmer, S.M.M.Tahaghoghi, Hugh E. Williams, James A. Thom, "The Moving Query Window for Shot Boundary Detection at TREC-12", Notebook Papers TRECVID2003, Gaithersburg, Maryland, NIST, October 2003.
[25] Masaru Sugano, Keiichiro Hoashi, Kazunori Matsumoto, Fumiaki Sugaya, Yasuyuki Nakajima, "Shot Boundary Determination on MPEG Compressed Domain and Story Segmentation Experiments for TRECVID 2003" ,
[26] Jordi Mas, Gabriel Fernandez, "Video Shot Boundary Detection based on Color Histogram", Notebook Papers TRECVID2003, Gaithersburg, Maryland, NIST, October 2003.
[27] Osmar R. Zaiane, Jiawei Han, Ze-Nian Li, Jean Hou, "Mining Multimedia Data", MDM/KDD2001, Second International Workshop on Multimedia Data Mining in conjunction with Seventh ACM SIGKDD International Conference on Knowledge Discovery & Data Mining , San Francisco, CA, USA, 2001
[28] S. Obdrˇz´alek, Matas, J, "Local affine frames for image retrieval", in International Conference on Image and Video Retrieval, Lecture Notes in Computer Science, vol. 2383, Springer, 2002
[29] Sebe, N., Lew, M, "Robust shape matching", in International Conference on Image and Video Retrieval, Lecture Notes in Computer Science, vol. 2383, Springer, 2002
[30] Leung, M.W., Chan, K.L, "Object-based image retrieval using hierarchical shape descriptor. ", in International Conference on Image and Video Retrieval, Lecture Notes in Computer Science, vol. 2383, Springer, 2002
[31] Brucale, A., d'Amico, M., Ferri, M., Gualandri, M., Lovato, A., "Size functions for image retrieval: A demonstrator on randomly generated curves", in International Conference on Image and Video Retrieval, Lecture Notes in Computer Science, vol. 2383, Springer 2002
[32] Berens, J., Finlayson, G, "An efficient coding of three dimensional colour distributions for image retrieval", in International Conference on Image and Video Retrieval, Lecture Notes in Computer Science, vol. 2383, Springer, 2002
[33] Lekha Chaisorn, Tat-Seng Chua, Chun-Keat Koh, Yunlong Zhao, Huaxin Xu, Huamin Feng, Qi Tian, " A Two-Level Multi-Modal Approach for Story Segmentation of Large News Video Corpus", Notebook Papers TRECVID2003, Gaithersburg, Maryland, NIST, October 2003.
[34] JungHwan Oh and Babitha Bandi, "Multimedia Data Mining Framework For Raw Video Sequences", in proceeding of International workshop on Multimedia Data Mining with ACM SIGKDD, 2002
[35] William Perrizo, William Jockheck, Amal Perera, Dongmei Ren, Weihua Wu, Yi Zhang, "Multimedia Data Mining Using P-Trees", in proceeding of International workshop on Multimedia Data Mining with ACM SIGKDD, 2002
[36] Xin Huang, Shu-Ching Chen, Mei-Ling Shyu and Chengcui Zhang, "User Concept Pattern Discovery Using Relevance Feedback And Multiple Instance Learning For Content-Based Image Retrieval", in proceeding of International workshop on Multimedia Data Mining with ACM SIGKDD, 2002
[37] Ankur Teredesai, Juveria Kanodia, Muhammad Ahmad, Roger Gaborski, "CoMMA: A Framework for Integrated Multimedia Mining using Multi-relational Associations", in proceeding of International workshop on Multimedia Data Mining with ACM SIGKDD, 2004
[38] Osmar R. Zaïane, Simeon J. Simoff, Chabane Djeraba, "Mining Multimedia and Complex Data", KDD Workshop MDM/KDD 2002, PAKDD Workshop KDMCD 2002, Revised Papers, ISBN 3-540-20305-2, Springer 2003
[39] Evaluation tools, campaigns and TRECVID Proceedings of the 8th ACM International Workshop on Multimedia Information Retrieval, http://www-nlpir.nist.gov/projects/t01v/trecvid.tools/shot.boundary.evaluation/sbeval/, 2006