Scrutinizing the Video and Video Retrieval Concept

R.I. Minu, K.K. Thyagharajan

Abstract— Digital video information is some kind of proof for the events happening in our mankind. As capturing such kind of multimedia information is ease of work due to hand-held technology there are tremendous amount of such video in our Internet database. So, obviously we are in need of some technique for retrieving these data.

This paper will give an initial analysis on the structure of digital video and the overall overview of video retrieval procedure which will give effective retrieval result.

Index terms: Mpeg7, Video Retrieval, CBIR, Semantic Web

I. INTRODUCTION

Multimedia data such as image, audio and video has been widely spreading because of the extension of the application field. Structured organization of video data is the basis to build large video archives which allow efficient retrieval browsing and manipulation [1], [2]. In this paper we are giving a complete analysis of how a video visual data is organized through which we can determine the exact feature of the video content which is used for retrieval purpose. Still now there are so many proposed retrieval process has been proposed but the efficiency of multimedia data retrieval is not that much good as text based retrieval process. Among those retrieval process some of the notable process are textual based retrieval, Concept based retrieval, Content Based retrieval and Ontology based retrieval. In CBIR the feature of the image such as color, shape, texture and etc., can be used separately or combine way for retrieve purpose [3, 4, 5, 6]. In recent days we can use MPEG7 standard for extracting these feature as [7] here we have specific standard called MDS for representing these features for images or videos.

So this paper is organized in such a way that the Section 2 provides the complete analysis of the video structure, Sections 3, 4 and 5 provides the different way of determining the shot boundary, key frame and the feature of the frame and in Section 6 we emphasis why retrieval using Mpeg 7 provides good result than the other kind of retrieval.

II. VIDEO STRUCTURE ANALYSIS

Digital video information often consists of series of 25 frames per second for television broadcast and 30 frames per second for computer display. The frames are normally set of images of any desire format. In order to develop any content based manipulations on digital video information, the video information must first be structured and broken down into components [43][8]. This is shown in Figure.1

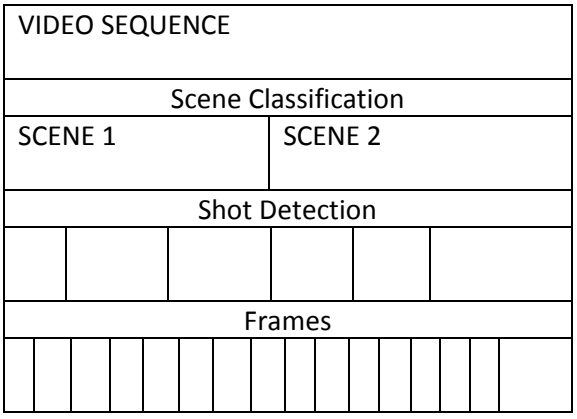

Figure.1 Graphical visualization of video structure

The basic unit of video is frames which are like an image still. A whole video can be divided into Scenes which is a large continues unit which would be taken place in one landmark. A Scene can have any number of Shot, where a shot are the video frames which are continuously recorded using single camera. Then the shot can have any number of frames. Generally CBIR of video data consists of following phases

- Video shot boundary detection
- Key frames selection
- Features extraction from the selected key frames
- Retrieval procedure

A shot boundary detection algorithm is based on the recognition of the editing effects (such as cuts, fades, dissolves, etc...) that identify the boundaries. After determining the boundaries, each shot can be represented by a key frame. A set of frames that best represent the visual content of the shot is called as the key-frames and these frames are generally used to for video search and retrieval. From the key frames we can extract both the low-level and high-level features. For the extraction of low-level feature here we analysis the usage of Mpeg-7 Descriptor tool which provide both spatial and temporal feature of a key frame. By using these features we will see what all the different types of retrieval technique used are.

So the main part of all these stages is the key frame selection stages. As the effectiveness of retrieval is only based upon the key frame selection.

Manuscript received October 21, 2011.

R.I. Minu, Research Scholar,Anna University, Trichy ,Tamilnadu, India. (E-mail: r_i_minu@yahoo.co.in)

K.K. Thyagharajan, Professor ,Information & Technology,RMK College of Engg & Technology, Tamilnadu, India (E-mail: kkthyagharajan@yahoo.com).

III. ANALYSIS OF SHOT BOUNDARY DETECTION

The first phase in video content based retrieval process is Shot boundary detection. A shot is a single take camera visual content. So the boundary between shots can be either abrupt shot boundary [9] (ASB) or also gradual shot boundary (GSB). Where informer the change in shot occur in single frame we also called as shot cut or hard cut where as in GSB the effect like fade-in, fade-out, dissolve and wipe kind of concept where used. There are so many technique been [11] – [16] introduced in shot boundary as the efficiency of an video retrieval process in fully depend upon the key frame which we selected from the detected shot so the effectiveness of shot detection is need to be high.

The change between one shot to other can be identified using the feature detail of the frames the most common feature commonly used for shot boundary detection are color, Edge, Pixel intensity and so on. Among them the color plays an important role. Now most of the research are been used hybrid of all feature. The feature of the frames are compared with each other frames either using spatial feature domain or temporal feature domain (i.e) either comparing the feature frame by frame or group of frames and so on. Once determining the discontinuity the decision of shot boundary can be determine by means of thresholding value, probability detection and Heuristics combination.

IV. KEY FRAME SELECTION TECHNIQUE

Selection of key frame is one of the crucial parts of the CBVR as the further result is depending upon this step. Usually key frame selection technique is integrated with the processes of segmentation. Each time a new shot is identified, the key-frame extraction process is invoked.

Many approaches for keyframe selection is proposed [21][22][23][24][25]. By analysis these researches it seems most of the research suggest clustering technique for determining the keyframe selection. But in [25] specifies that clustering the frames within a shot and select the centroid of the clusters as key frame approach is a computationally expensive and it is also not easy to determine one cluster within a shot. So they perform principle component analysis on video sequences and derive two discriminate forms the first few retained principle components. Where the first discriminate captures the angle and distance of principle component to determine the shot. The Eigen images of shots are used as key frames. Figure 2 shows the shows the key frame selection of an video frame sequence.

Figure.2 Segmentation and key frame

V.FEATURE EXTRACTION

One of the best ways to represent an image is by its low level features. The low level features of an colored image or frame is color, shape, motion, texture, spatial location and etc.,

For each features there are so many sub techniques for example color can be represent in RGB, CMYK, HSL, HSV etc., color space for each space different kind of representation are available so for generalization we can use MPEG-7 [7] which is an ISO/IEC standard which specify a standard way of describing various kind of audiovisual information.

MPEG 7 standard consist of 3 part Description tools, Description Definition tool and System tool.

In Description tool the Descriptor (D) and Descriptor Schema (DS) of the audiovisual content will be defined. In D the low level data such as data type of the multimedia data will be given and in DS the high level data such as function description will be given. In Description Definition Tool DDL the whole details about the multimedia information will be coded in XML which pave the way for interoperability.

System Tool supports two type of representation for D and DS. One is Textual format and other one is Binary Format. We can use these formats for those platforms where XML is not feasible.

Table 1: Mpeg-7 Descriptors [7]

Color Descriptor	Color space
	Color Quantization
	Dominant Color(s)
	Scalable Color
	Color Layout
	Color-Structure Descriptor
	GoF/GoP Color
Texture Descriptor	Homogenous Texture Descriptors
	Texture Browsing
	Edge Histogram
Shape Descriptor	Region Shape
	Contour Shape
	Shape 3D
Motion Descriptor	Camera Motion
	Motion Trajectory
	Parametric Motion
	Motion Activity

Table.1 gives the complete overall details about the Mpeg-7 Descriptors which are used to determine the features of an image or a frame.

For each descriptor so many research has been done, through their results we can provide a better way of representing the feature of the frame. Initially by the year 2004 Patrick et al [36] proposed a texture analyzer which can be used for video content analysis for video coding. They suggest that they can analyze the image texture irrespective about their relevancy at a rate of 93%.

One of the great advantages of using Mpeg-7 as low level visual descriptor is that the results obtain about an image feature is expressed in XML file. So it is easy to deploy on web services. Figure 3 shows the XML schema for the given image.

VI. ANALYSIS OF RETRIEVAL PROCEDURE

In early 90's to take a picture or a video, we need a sophisticated camera and person to take video images. But now a day due to the development of technology with the hand held cameras and mobile phone we can capture the video in instance. So, now one can think about tremendous archive of video Collection which can be easy to publish in internet through YouTube kind of web services after the introduction to web3.0 technology.

<VisualDescriptor xsi:type="ColorLayoutType">
<YDCCoeff>48</YDCCoeff>
<CbDCCoeff>22</CbDCCoeff>
<CrDCCoeff>20</CrDCCoeff>
<YACCoeff63>19 20 16 12 14 20 15 16 14 16 15 16 16 9 19 17 13 19 16 15 16 16 14 15 18 15 16 14 14 15 13 17 16 15 15 16 15 14 17 18 16 18 15 16 13 15 16 15 16 16 17 16 15 17 15 15 15 15 15 15 16 16 16</YACCoeff63>
<CbACCoeff63>31 9 19 21 16 5 18 14 16 16 14 13 10 23 16 17 18 18 15 15 16 16 14 15 14 16 16 13 17 16 13 17 15 15 16 16 15 15 16 18 15 13 15 15 14 17 16 15 16 15 15 17 16 16 15 16 15 15 15 16 16 16 15</CbACCoeff63>
<CrACCoeff63>1 17 11 15 13 24 15 17 18 15 16 20 16 15 13 16 15 11 16 17 14 15 16 17 13 14 13 19 16 15 18 13 15 16 16 15 16 17 16 14 17 16 16 15 18 15 15 16 16 16 15 14 17 14 16 15 15 16 16 16 16 16 15</CrACCoeff63>
</VisualDescriptor>
<VisualDescriptor xsi:type="ScalableColorType" numOfBitplanesDiscarded="0" numOfCoeff="256">
<Coeff>-71 -37 -55 46 38 14 22 31 31 16 31 30 18 11 29 28 2 1 0 2 3 5 0 2 12 3 -3 2 -4 5 -14 4 0 2 0 1 0 0 1 2 9 5 1 3 1 2 4 7 7 3 3 4 7 3 3 2 2 11 0 -2 15 0 -3 1 0 0 0 0 0 0 1 2 6 2 1 2 2 2 1 0 1 0 1 0 0 1 1 7 4 2 1 3 3 2 2 2 3 0 1 0 0 2 2 6 1 2 1 3 3 1 2 1 0 -1 -1 3 0 1 2 3 1 1 0 -14 0 -1 -1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 -1 1 0 -1 0 0 1 -1 -2 2 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 -1 0 0 0 -1 1 1 0 -1 0 0 1 -2 0 1 0 0 0 0 1 1 0 -1 0 0 0 0 0 0 2 -1 0 -1 0 0 1 -1 -1 -1 1 -1 1 1 -2 -1 -2 0 0 0 1 0 1 0 -1 0 2 1 0 0 -1 1 -1 1 0 -3 0 0 1 -1 -1 0 0 -3 1 0 3 1 -3 0 0 0 1 0 3</Coeff>
</VisualDescriptor>
<VisualDescriptor xsi:type="EdgeHistogramType">
<BinCounts>0 3 0 0 0 0 3 0 0 0 0 3 1 4 1 5 1 2 2 2 0 0 0 0 0 0 0 2 1 1 0 2 3 6 4 4 1 6 3 1 0 0 0 0 0 0 0 1 1 1 0 6 3 4 2 4 1 6 4 5 0 0 0 0 0 0 0 0 0 0 1 3 5 3 2 4 1 7 4 2</BinCounts>
</VisualDescriptor>
<VisualDescriptor xsi:type="DominantColorType">
<SpatialCoherency>0</SpatialCoherency>
</VisualDescriptor>
</Image>
</MultimediaContent>
</Description>
</Mpeg7>

Figure 3: Schema for the image

Even though the concept of video retrieval was there before these inventions. Now it is a must required technology because of the growing rate of multimedia data's on web.

Initially the retrieval concept start by the key-word based search then comes the CBIR – Content Based image Retrieval a boom to the field of image retrieval where they use the low level feature of an image for retrieval. Thus in key-word based search there is a need of a manual annotation of the image which is a tedious job. With CBIR it extracts the low level feature of an image for retrieval purpose. The overall operation of CBIR is given in Figure 4.

So if we are dealing with the video file than the low level feature of the key frame of all the video in the database would be stored in feature database. So the queried image can be compared with the feature database content through relevant feedback we can ranked our similarity retrieval and concern output can obtained.

Even though the retrieval process using CBIR is seems to be good in theory the result shown is not up to the satisfaction the main reason behind this is the Semantic gap. The concept of Semantic gap is shown in Figure 5. The way how we represent an image and the way how a machine represent an image is completely different those difference is said to be the semantic gap that is the gap between a low level and high level description of an image. So the reduction of this semantic gap is the main goal research area for those researchers who rely on CBIR. For this we can go for the concept of Ontology based image retrieval, this concept reduces the gap between high levels semantic with low level credential.

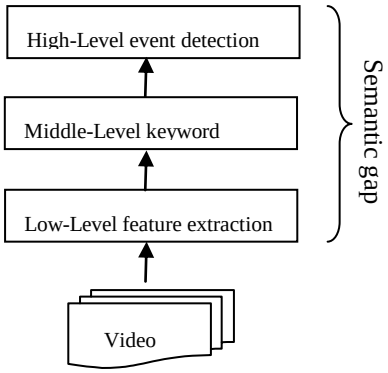

Figure 5: Semantic gap

The high level semantic can be provided by the Ontology technique [40],[41],[42]. There are different kind of ontology to the required system we can select the appropriate Ontology technique. The efficiency of the retrieval system depends literally on the kind of mapping technique we use to map the determined low level and high level features.

This mapping can be performed by three methods[38] [39] as listed below:

- Manual annotation
- Semi-Automatic annotation
- Automatic annotation

Table 2 shows the different kind of approach used for annotation

Table 2: Kind of annotation approach

KIND OF ANNOTATION	APPROACH
Manual Annotation	User Annotation
Semi-Automatic Annotation	Bayer Probability
	Co-Occurrence Model
	Graph Based Link
Automatic Annotation	Decision Tree
	Bootstrapping
	Latent Semantic Analysis
	Hidden Markov Model

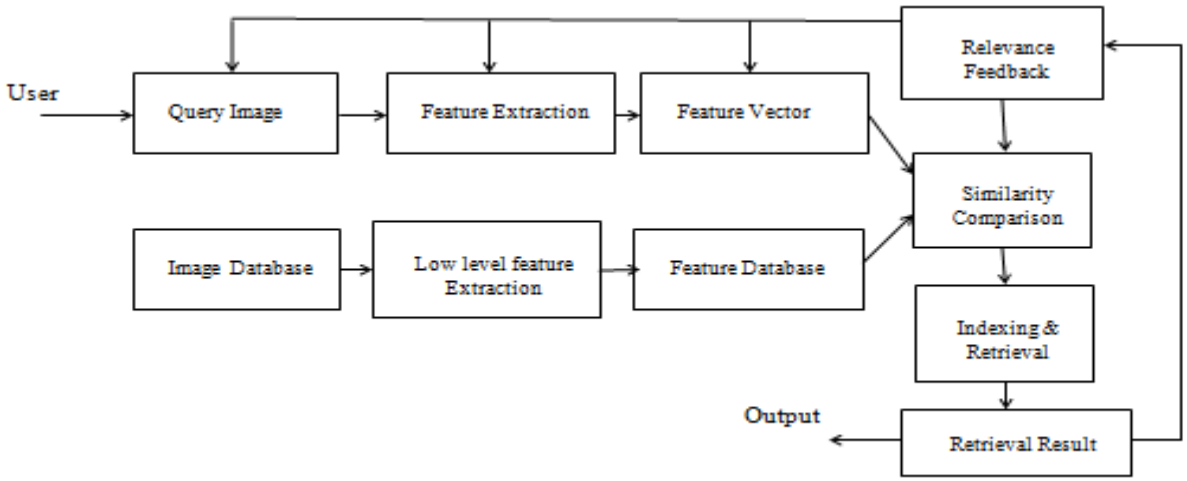

Figure 4: Basic block diagram of CBIR

VII. CONCLUSION

This paper provide a overall analysis on how a video stream look like and what are all the techniques required for retrieval of video frames by comparing and enhancing the concepts used for image retrieval. Through this analysis we conclude this paper by promoting our future work [42, 43, 44] on developing a retrieval technique which will provide a semantic meaning to it. So to reduce the gap between high level and low level semantic we are proposing a system of using Ontology technique in it. Once we use MPEG 7 for feature description and Ontology for retrieval purpose we can provide an effective image search engine for Semantic Web as efficient as document search engine which always provide you relevant information for a given query.

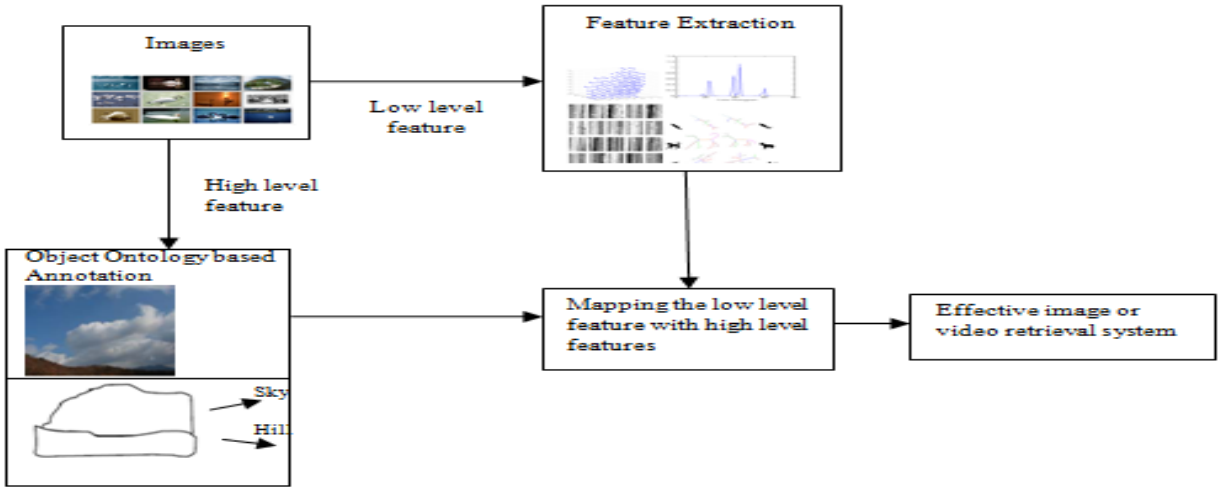

Figure 6: Overall Procedure of retrieval procedure

REFERENCES

1. D. Zhong, S.-F. Chang, "Video Object Model and Segmentation for Content-Based Video Indexing",IEEE Intern. Conf. on Circuits and Systems, June, 1997, Hong Kong.
2. S. Chang, W. Chen, H. Meng, H. Sundaram, D. Zhong, (1998, September), "A fully automated content based video search engine supporting spatio-temporal queries", IEEE Trans. Circuis and Systems for Video Technology, 8(5), 602-615.
3. M.Swani and D.Ballard,"Color indexing",International Journal of Computer Vision,vol. 7,no.1,00 11-32, 1991.
4. Jing Huang, S.R.Kumar,M.Mitra,Wei-Jing ZhuR.Zbih,"Image indexing using color correlograms",IEEE Processdings on Computer vision and Pattern Recognition, pp.762-768,1997
5. R.Mehtotra and J.E.Gray, "Similar shape retrieval in shape data mangement",IEEE Computer, vol.28,pp.57 -62,Sept. 1995
6. B.S.Manjunath and W.Y.Ma."Texture features for browsing and retrieval of image data",IEEE Transaction on Pattern Analysis and Machine Intelligence,vol.18,pp.837-842,Aug.1996
7. ISO/IEC JTC1/SC29/WG11N6828 Palma de Mallorca, "MPEG-7 Overview (version 10)" October 2004
8. Ivica Dimitrovski,Suzana Loskovska,Gorgi Kakasevki and Ivan Chorbev"Video content Based Retrieval System"IEEE The international Conference on "Computer as a Tool" Pg.978 -983,Sep 07
9. Hong lu and Yap-Peng Tan"'An effective Post-Refinement Method for Shot Boundary Detection",IEEE Transactions on Circuits and systems for Video Technology,Vol.15,No.11,Nov05,pp.1407 - 1420.
10. W. Wolf, "Key frame selection by motion analysis," in Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing, vol. 2, 1996, pp. 1228–1231.
11. M. Yeung, B.-L. Yeo, and B. Liu, "Extracting story units from long programs for video browsing and navigation," in Proc. IEEE Int. Conf. Multimedia Computing and Systems, 1996, pp. 296–305.
12. X. B. Gao and X. O. Tang, "Unsupervised video-shot segmentation and model-free anchorperson detection for news video story parsing," IEEE Trans. Circuits Syst. Video Technol., vol. 12, no. 9, pp. 765–776, Sep. 2002.
13. H. J. Zhang, Y. H. Gong, S. W. Smoliar, and S. Y. Tan, "Automatic parsing of news video," in Proc. Int. Conf. Multimedia Computing and Systems, 1994, pp. 45–54.
14. Y.-P. Tan, D. D. Saur, S. R. Kulkarni, and P. J. Ramadge, "Rapid estimation of camera motion from MPEG video with application to video annotation," IEEE Trans. Circuits Syst. Video Technol., vol. 10, no. 1, pp. 133–146, Feb. 2000.
15. R. Lienhart, "Comparison of Automatic Shot Boundary Detection Algorithms", Proc. SPIE Vol. 3656, p.290-301, Storage and Retrieval for Image and Video Databases VII, Editor(s), 1998
16. J. S. Boresczky and L. A. Rowe, "A Comparison of Video Shot Boundary Detection Techniques. Storage & Retrieval for Image and Video Databases IV", SPIE 2670, pp. 170-179, 1996.
17. S. Jun, S. Park, "An Automatic Cut Detection Algorithm Using Median Filter and Neural Network", ITC-CSCC 2000, Vol. 2, pp 1049-1052.
18. D. Zhong, "Segmentation, Index and Summarization of Digital Video Content," Doctoral Dissertation, Columbia University, 2001.
19. Houng Lu,Yap-Peng Tan,"An effective post refinement method for shot boundary detection"IEEE Transactions on circuits and systems for video technology,Vol 15,No.11,Nov 05
20. W. Wolf, "Key frame selection by motion analysis", Proceedings of IEEE International Conference Acoustic, Speech, and Signal Processing, 1996
21. M. Yeung and B. Liu, "Efficient matching and clustering of video shots", Proceedings of IEEE International Conference Image Processing, 1995
22. Girgensohn and J.Borecky "Time constrained key frame selection technique", Proceedings of IEEE International Conference Multimedia Computing and Systems, 1999.
23. Pascal Kelm, Sebastian Schmiedeke, and Thomas Sikora "FEATURE-BASED VIDEO KEY FRAME EXTRACTION FOR LOW QUALITY VIDEO SEQUENCES", Proceeding of IEEE WIAMIS 2009 conference
24. R. Datta, D. Joshi, J. Li, and J. Z. Wang. Image retrieval: Ideas, inuences, and trends of the new age. In ACM Computing Surveys, 2008Letter Symbols for Quantities, ANSI Standard Y10.5-1968.
25. M. Worring and G. Schreiber. Semantic image and video indexing in broad domains. IEEE Transactions on Multimedia, 9(5):909{911, August 2007
26. Shneiderman B and Kang H (2000) Direct Annotation: A Drag-and-Drop Strategy for Labeling Photos. In: Proc. International Conference Information Visualisation (IV2000). London, England.
27. B.S.Manjunath,Jens-Rainer Ohm,Vinod V.Vaudevan and Akio Yamada "Color and Texture Descriptors" IEEE Transaction on Circuits and Systems for Video technology, VOL. 11, NO. 6, JUNE 2001
28. Kuen-Long Lee,Ling-Hwei Chen "An efficient computation method for the texture browsing descriptor of MPEG-7",Science Direct Image and Vision Computation 23(2005)
29. Yong Man Ro, Munchurl Kim, Ho Kyung Kang, B.S. Manjunath, and Jinwoong Kim," MPEG-7 Homogeneous Texture Descriptor", ETRI Journal, Volume 23, Number 2, June 2001
30. Dong Kwon Park Chee Sun Won and Soo-Jun Park. Efficient use of mpeg-7 edge histogram descriptor. ETRI Journal, 24(2), 2002.
31. Bober, M. Price, W. Atkinson, J. ," The contour shape descriptor for MPEG-7 and its applications", International Conference on **Consumer Electronics, 2000. ICCE. 2000 Digest of Technical Papers.**
32. Carla Zibreira and Fernando Pereira." Image Description and Retrieval Using MPEG-7 Shape Descriptors ",Springer Publication Book on **Research and Advanced Technology for Digital Libraries**, 2009
33. Mokhtarian, F., Bober, M.," Curvature Scale Space Representation: Theory, Applications, and MPEG-7 Standardization",Computational Image and Vision Vol25,2003
34. Patrick Ndjiki-Nya, Oleg Novychny and Thomas Wiegand "VIDEO CONTENT ANALYSIS USING MPEG-7 DESCRIPTORS","Merging Mpeg-7 descriptors for image content analysis" and "A Mpeg-7 aided segmentation tool for content based video coding",IEEE Conference ICASSP,2004
35. Mojsilovic, B. Rogowitz, Capturing image semantics with low-level descriptors, Proceedings of the ICIP, September 2001, pp. 18–21
36. Ying Liu, Dengsheng Zhang, Guojun Lu,Wei-Ying Ma ," A survey of content-based image retrieval with high-level semantics", Science Direct, Pattern Reorganization ,40,2007
37. Jianjiang Lu,Qi Xiao,Tianzhong Zhao, Yafei Zhang and Yanhui Li,"Image Annotation Based On Bi-Coded Chromosome Genetic Algorithm for Feature Selection", IEEE proceesing,International Conference on Information and Automation,2008
38. Lei Shi,Guochang Gu,Haibo Liu,Jing Shen,"A Semantic annotation algorithm based on image regional object ontology",2008 IEEE International Conference on Computer Science and Software Engineering
39. Hui Hui Wang,Dzulkifli Mohamad,N.A.Ismail,"Image Retrieval:Techniques,Challenge and Trend",World Academy of Science,Engineering and Technology 2009
40. Lei Wang,Latifur Khan,Li Lie and Weili Wu,"Automatic Image Annotation and Retrieval using Weighted Feature Selection", Proceedings of the IEEE Sixth International Symposium on Multimedia Software Engineering (ISMSE'04)
41. Thyagharajan K. K. & R. Harikrishnan," Contentbased Bandwidth Aware Hierarchical Video Summarization",International Journal of Computer Science, System Engineering and Information technology, Volume 3 Number 1 January-June 2010, Pages: 197-203
42. R.I.Minu and K.K.Thyagharajan "Automatic image classification using SVM Classifier"CiiT International Journal of Data Mining Knowledge Engineering,July issue,2011
43. G.Nagarajan, Prof.K.K.Thyagharajan,"A Survey on the Ethical Implications of Semantic Web Technology": Journal of Advanced Reasearch in Computer Engineering,Vol.4No.1 jan-june 2010
44. R.I.Minu,Prof.K.K.Thyagharajan,"Evolution of Semantic Web and Its Ontology": Second Conference on Digital Convergence 2009.

AUTHORS PROFILE

R.I.Minu (MISTE '11) was born in India December 16, 1982. She received her BE degree in Electronic & Communication Engineering from Bharathidasan University 2004. She received her ME degree in Computer Science Engineering from Anna University 2007.She is at present a PhD Scholar in Information and Communication Engineering from Anna University of Technology, Trichy. Her research areas are Image retrieval, Artificial Intelligent, Machine Leaning and Semantic Web.

Dr. K.K. Thyagharajan has received his B.E., degree in Electrical and Electronics Engineering from PSG College of Technology (Madras University). He received his M.E., degree in Applied Electronics from Coimbatore Institute of Technology and Post Graduate Diploma in Computer Applications from Bharathiar University. He has received his Ph.D., (Multimedia Streaming) degree in Information and Communication Engineering from College of Engineering Guindy, Anna University. He has written 5 books in Computing. His book "Flash MX 2004" published by McGraw Hill (INDIA) has been recommended as text / reference book by many universities. He has published more than 30 papers in National and International Journals and Conferences. He is a grant recipient of Tamil Nadu State Council for Science and Technology. His biography has been published in the 25th Anniversary Edition of Marquis Who's Who in the World Directory. He has been invited as chairperson and delivered special lectures in many National and International conferences and workshops. He is reviewer for many International Journals and Conferences. His current interests are Multimedia Networks, Mobile Computing, Web services, Data Mining, e-learning, Image Processing, Microprocessors and Microcontrollers. He has guided 10 M.E. projects and now 9 students are doing Ph.D. under him in the area of Multimedia, Image Processing and Data Mining. He is a life member of Computer Society of India and Chairman of the ISTE chapter of RMK College of Engineering and Technology.