QUEEN MARY, UNIVERSITY OF LONDON
DEPARTMENT OF ELECTRONIC ENGINEERING

CLUSTERING OF VISUAL INFORMATION USING SPECTRAL METHODS

Thesis submitted to University of London
in partial fulfilment of the
requirements for the degree of
Doctor of Philosophy

UROS DAMNJANOVIC

London, 2009.

Acknowledgements

I would like to thank Prof. Ebroul Izquierdo for giving me this opportunity, assisting and advising me during my period in London. I will be ever thankful to him for this chance. I also would like to express my greatest gratitude to professor Branimir Reljin from Belgrade University for his always helpful and meticulously constructive technical advices during my studies, and for his encouragement regarding my Ph.D. in London.

Many thanks to all the friends I made at QMUL. From the swimming pool and to The Curve and Half Moon, this has certainly been a great time! Special thanks to my friends Divna Djordjevic and Ivan Damnjanovic for spotting and taking the initiative to solve technical issues in my work, directly contributing to this thesis.

I would like to thank the professors and all my friends from Belgrade University for their contributions and friendship.

I would like to thank my family from Serbia for being part of my life. Special thanks my parents Danuta and Nebojsa and my brother Nemanja for their help, love and friendship, along with the financial support and great education provided to me during my life.

Abstract of Thesis presented by Uros Damnjanovic to University of London.

CLUSTERING OF VISUAL INFORMATION USING SPECTRAL METHODS

ABSTRACT: Clustering is widely used in the computer vision community as a first step of content organisation, resulting in many different algorithms available in the literature. We propose a self adjustable clustering approach based on the Markov Random walk interpretation of pairwise data similarities. The proposed algorithm makes no assumption on the number of clusters present in the dataset, instead it recursively uncovers clusters in the data until all groups have been found. The algorithm is capable of automatically adjusting its parameters to new data, using properties of the graph-based data representation. The main advantage of the proposed approach is the algorithm's ability to cluster data with no user interaction. By applying the proposed clustering algorithm to video material, we propose solutions for different problems in the field of video analysis. A novel shot boundary detection, and key frame extraction method based on the self adjustable recursive clustering is proposed. It utilises the information about cluster properties to find different types of shot changes. By extending the shot boundary detection approach and including the temporal information, a method for scene detection and redundancy removal is created. Scenes are found by letting each frame belong to a number of different clusters instead of assigning each frame with one cluster label. The coverage of each cluster is defined to describe fuzzy clusters, in order to merge clusters that belong to the same scene. Detected scenes are then used to find similar patterns in the eigenvector space by the proposed redundancy removal algorithm. These patterns are used to detect repetitive segments, which are used to remove redundant information from the original video. Finally by introducing the context-based domain information we propose a clustering based summarisation technique for news and surveillance videos.

Contents

Chapter 1

INTRODUCTION

As we enter the 21st century the process of acquiring the right information at the right moment can be seen as the most important prerequisite of a successful and efficient everyday life. With the number of interconnected users increasing every day, growing stocks of available data lead to significant research efforts in the task of organising, storing and delivering information. Over the years a need for efficient information management has resulted in a move from text based to content based information systems. Modern multimedia documents use pictures, texts, and audio signals to carry specific information. Development of information systems able to find and retrieve documents by content is a goal of many research initiatives in the field of information retrieval. The work described in this thesis is motivated by the fact that the current state of these systems is still in its infancy, and that there are many challenges, and open questions still waiting to be answered.

1.1 Information Management Systems

From the early stages of human progress, people looked for efficient ways to communicate as the possession of useful knowledge often meant a difference between life and death. During the aeons of human evolution the means of sharing information evolved from primitive cave paintings to modern information systems. Even though means of communication changed over time the main purpose was the same: to pass information between people. Historically, communication systems started with sounds and movements and over time included signs and pictures. Early communication systems resulted in historical records on papyrus, temple and tomb walls or clay tablets. Together with technological progress, means of communication improved too, resulting in the creation of printing machines in the industrial era. The invention of printing machines automated the process of saving and passing information in the form of books and printed documents. Over a long period of time,

since the first written documents were made, the written documents were the main source of information. Except in everyday activities communication systems were also used for knowledge-sharing purposes. Either by saving important information for future generations, or by distributing available knowledge to different groups, a critical mass of educated people is created, a process which can be seen as the main initiator of human progress. Another great technological breakthrough, the invention of transistors in the middle of 20th century, marks the key point in the technological progress of information-management systems. From that point on, the technology is advancing at high speed using many different tracks, accelerating the use of radio, telegraph, television and computers. The start of the digital era meant all information could be processed using a computer, opening doors for automatic data analysis. Instead of using conventional multimedia systems such as TV, video and audio cassettes, visual data could now be processed, stored and consumed using a computer. Together with existing textual archives, new storage mediums such as magnetic tapes, floppy disks and hard drives were used for information storing. Initially, all information stored in such a way had to be manually indexed to prevent the loss of useful information. Clearly this task becomes highly impractical when working with sheer volumes of data available in modern information-management systems. Another key point in the technological progress of information technologies is the emerging of the Internet. The Internet provided better means for information sharing, than any of the existing systems. This resulted in the fast development of tools used over the network. The most important property of the Internet is the big number of users, which steers the progress of information technologies.

1.2 Problem Definition

These technological advances made available huge amount of data to every person sitting in front of a computer and surfing the web. Huge stocks of multimedia items make it more difficult for users to find right information. Searching for a specific piece of content in a traditional multimedia database is almost impossible using available browsing tools, due to the semantic-gap. The semantic-gap is a barrier between content-representation readable to computers, and actual semantic meaning perceived by humans. One way of narrowing the gap between two types of representation is to use an army of human annotators to annotate each piece of content, and then apply text-based information-management techniques. However this is neither practical nor efficient, since the amount of multimedia content is huge nowadays. Content based information systems, developed on the foundations of traditional text-base information retrieval techniques, and are used to provide

more complete understanding of data then traditional information systems. Images and videos contain much more information than related textual meta-data, due to the way humans understand and convey visual information. In order to save information present in images and videos, and use it to create semantic relations between different pieces of content, modern information systems should extract the meaning from a signal-based representation. At the current level of technology this is still not possible. The main reason for this is that the low-level based representation of media items is not sufficient to capture the high-level meaning of the content. Even though it is not possible to achieve high levels of content understanding it is possible to improve the performance of conventional information management systems using different techniques for visual content analysis. Detecting groups with similar layout in the low-level representation space improves a system capability to provide users with the right piece of content. Content based information systems are developing in many different directions nowadays. The analysis of a video material can be seen as one the most challenging research tasks, since video data connect different modalities such as images, sounds and text. The main problem in video data organisation is the huge amount of visual material that needs to be processed. Efficient organisation and retrieval of visual information depends on the capability to break down a video into its elementary units, extract some information and present the extracted information in the proper way. We focus on two main problems in this thesis: the problem of efficient information extraction from a video and the problem of information presentation to the user.

1.3 Contributions of the Thesis

1.3.1 Proposed Self-Adjustable Clustering Approach

Clustering of visual information is the first step towards the semantic based indexing of multimedia items [1]. The main goal of the proposed work is to cluster information and extract meaning from clusters using a low level data representation. We propose a novel clustering approach which is based on well known Normalised Cuts originally used in image segmentation. The main contribution of the proposed approach is its capability to automatically select parameters which are usually set by the user. The basic step in all spectral clustering algorithms is the creation of the similarity matrix. The structure of the similarity distribution is defined by the parameter σ. We propose a novel way for automatic σ selection using random walk interpretation of clusters. By interpreting the resulting random walk for different values, σ is automatically set to the value that successfully finds clusters in a data set. Another parameter which usually has to be known in advance, the number of

clusters, increases the need for human interaction. In our approach this problem is solved by defining stopping criteria based on the geometrical interpretation of eigenvectors. The geometrical interpretation of eigenvectors is based on the fact that the structure of an eigenvector depends directly on the structure of a data set. By analysing the structure of eigenvectors we give a new interpretation of clusters and create a framework for adjustable recursive clustering of data. Together with proposed solutions we give an intuitive explanation of the proposed approach. The adjustable recursive clustering is used in different problems of visual information analysis.

1.3.2 Clustering of Video Data

It is extended to find shot boundaries in a video by localising the clustering in a sliding window, and introducing a temporal dimension to the process of eigenvectors interpretation. A method for representation of visually informative segments is used as a first step of the video summarisation. Another application of the adjustable recursive clustering is the problem of scene detection and redundancy removal. By assigning each frame to a number of different clusters and by using a proposed cluster merging procedure a video is broke down into scenes which are used for fast video browsing. The cluster merging procedure based on temporal properties of clusters is used to build scenes by putting together different clusters. The proposed geometrical eigenvector model is used to find similar temporal patterns in scenes and remove redundant information. This is used in the summary creation where only selected segments from scenes are presented to the user. Finally context-based domain information is used to cluster a video and to build a summary using clusters as basic building blocks.

1.4 Structure

The thesis is organised as follows. Content based retrieval systems are described in the following chapter focusing on open issues and problems. The overview is a result of the literature reading, and investigation of the state of the art techniques in the area of: content based retrieval, data clustering and content based video retrieval. The state of the art overview is followed with an overview of spectral clustering algorithms. In the chapter III we describe the proposed recursive clustering approach. The adjustable choice of input parameters and eigenvector based stopping criteria are explained with main focus being on practical applications. Experiments on both synthetic and real world examples give more detail about proposed approach. They are also used to evaluate the performance of different clustering algorithms. In the

chapter V a shot detection method based on the proposed clustering approach is presented. The proposed method for scene detection and redundancy removal is presented in the chapter VI. It is shown how scene and redundancy detection can be used for Rushes summarisation and present results achieved in 2008 Trecvid. In chapter VII contextual information is introduced to summarise different types of video material. It is shown how the summarisation procedure depends on an interpretation of clusters found in a video. In chapter VIII we give directions for future work and conclude the thesis.

Chapter 8

CONCLUSION

8.1 Main Accomplishments

In this thesis we presented a novel clustering approach based on Normalised Cuts, a well known spectral clustering method. Spectral clustering emerged in recent years as a promising clustering tool which was mainly applied to the problem of image segmentation. Our approach was based on the fact that a real world problems algorithm should be able to produce proper results with no human interaction. We showed in chapter 3 solutions for two important problems in data clustering. The automatic choice of the parameter σ used in building the similarity, based on the Markov random walk interpretation, not only produced better results when compared to different clustering methods proposed in the literature, but also gave an intuitive explanation on why such a choice is made. By selecting σ that introduced the most information to the system, a graph representing data is connected maximising the ratio between connections inside clusters and connections between different clusters. Another important issue when clustering data is to automatically select the number of clusters. We proposed a solution for this problem based on the geometric interpretation of the eigenvector structure. By applying the recursive clustering defined with the proposed stopping criteria, a data set is clustered with no need for the number of clusters. The stopping criteria is defined to distinguish between different cases when there are more than one cluster in a data set and when all points belong to the same clusters. In order to compare the proposed approach to different clustering methods available in the literature experiments are done on 2D synthetic data sets, gene expression data and on a set of images.

By introducing a temporal dimension to the clustering problem we applied the proposed clustering algorithm to a problem of shot detection. We introduced a sliding window in which frames are clustered in order to localise the clustering. The proposed stopping criteria was extended so it used the temporal information in the

decision process. We showed how different types of transitions can be found using modified eigenvector interpretations, followed by the process of key frame extraction. We classified frames into two groups. One group contained frames which were part of a static scene and which are understood to hold visually important information. The second group of frames are frames that are part of a transition or a change in the video content. Both groups are found by looking at the eigenvector temporal structure.

The next logical step in applying clustering to the content based video retrieval, is to detect scenes and redundant information in the task of video summarisation. Our approach is based on the fact that similar frames which are close in time should be put in the same scene. A scene is defined as a basic semantic video block used to build a summary. By clustering frames defining the coverage of each cluster, a number of different clusterings are produced. By merging clusters with overlapping coverage regions, scenes are found as unions of overlapping clusters. In order to select segments which are representing each scene redundant material needs to be removed from the summary. Our approach for redundancy removal is based on the fact that regions of similar visual layout results in the areas of similar patterns in an eigenvector space. By detecting these patterns and selecting only one segment to represent sets of repetitive intervals, a summary representing the original video is built. Such a summary is made of segments representing each scene with the level of detail depending on the time available for the summary. Such a summarisation approach is evaluated in the 2008 Trecvid resulting in a high ranking in terms of redundancy removal and pleasantness and rhythm of the summary.

Finally by introducing a context in the task of video summarisation it is possible to extend the proposed clustering approach to find clusters for a specific application. For two different video types, news and surveillance, we showed how recursive clustering can be used in combination with contextual information to build both static and dynamic summaries. Since news and surveillance videos both have a specific structure and different purposes, a summary of such videos is built taking into account different properties of video. For news videos the main purpose is to present the most important news in a pleasant way so a user gets interested for the original content of a video. A news summary is built by detecting story segments in a video and by looking at the dominant cluster of each story segment using it for summary building. On the other hand surveillance videos are made of short and rare events which are of interest to the user. In order to build a surveillance summary the difference between energy levels in consecutive frames is used to find intervals of strong change in a video and then use these intervals in the clustering and summary building process. By clustering only rare events detected in the event

detection process the set of static and dynamic summaries are built. We presented both approaches together with experimental evaluation done for two types of video material. Interesting theoretical and practical issues based on spectral recursive clustering have been tackled and solved, such that the contributions presented in this work are manifold, as enumerated in the following:

1. The automatic choice of parameter σ is only free parameter used in spectral clustering. We proposed a way to automatically select this parameter using the Markov random walk interpretation of the clustering problem. Also the proposed approach gives an intuitive explanation why such a choice is made.

2. The stopping criterion is defined based on the second smallest eigenvector structure. Based on the expected eigenvector structure the decision is made whether the clustering should be continued or stopped. In order to initialise parameters used in the decision process, a learning phase may be introduced. Depending on data that is clustered, the user may create a set of clusters and use them to initialise parameters so they represent data structure better. Since the basic step in spectral clustering is the creation of the similarity matrix, which maps distance value to an unit interval, eigenvectors will have the same structure independently of data. This is the reason why the set of parameters used in the stopping criteria may be initialise with one type of data and used in various applications.

3. Shot boundaries are found using clustering of frames in the interval defined by a sliding window. The eigenvector structure model is extended to take into the account the temporal dimension of the problem, and different types of shot boundaries are found.

4. By clustering frames of a video and by defining the coverage area of each cluster a set of different clusterings of a single video is created. We defined a process of merging together clusters which have overlapping areas.

5. Once the scenes are found each scene is used as a building block for a final summary. By looking at similar patterns in an eigenvector space, we proposed a way to find repetitive intervals within each scene. Each repetition is then used as a source of a number of segments that are used to build a summary. Scene detection and redundancy removal was tested on 2008 Trecvid and resulted in the highest ranking in redundancy removal and the third highest result in terms of pleasantness and rhythm.

6. By introducing a context in the summarisation process, we proposed the summarisation of two different video types. Surveillance and news video have

different structures and purposes and it is common for both video types that contextual information can be used to lead the clustering process towards the final goal. For news video different importance measures are defined and used to define stories over the time line. By clustering shots each story is assigned its cluster label and then used to build a summary. In the surveillance domain we used the energy difference between consecutive frames to quantify the amount of change in a frame, and then used this information to find frames holding rare events. By clustering events and analysing the cluster structure two types of summaries are built.

8.2 Future Work

With respect to the future work a challenging task is to improve the ability of the proposed clustering approach to adjust itself to new data. In terms of selecting the optimal σ it would mean the creation of a stable model for choosing sigma based on two critical σ values. The stopping criteria will be extended so the decision process depends on parameters which are automatically learnt from positive and negative examples rather then using a fixed thresholds.

For the shot boundary detection problem an interesting direction for future work is to define a distance between two frames as a commute distance defined in Markov random walks [147]. The commute distance is also highly connected to the graph based methods for electrical circuit solutions [148]. Defining a random walk over set of frames and calculating the commute distance between each pair of frames will be the basis of the improved shot boundary approach.

For video summarisation using scene detection and redundancy removal the main challenge is to improve the algorithm capability to detect important events in the scene. By extracting different low level and mid level features such as face detection and motion activity estimation and by combining these features with features that are already used in our approach, we will try to improve the algorithm performance in terms of ground truth inclusions.

Summarisation of news and surveillance videos can be improved by introducing more contextual information. For the news videos face detection can be used to find anchor person shots and other shots with visually less important information. The results of speech recognition could be used as an additional source of information in video analysis and may be used to define a set of textual annotation for news videos. Textual information can be then combined with textual captions detected in the shots, and used to create the semantic relations between different stories in the news. For the surveillance videos the main challenge is to identify interesting events

properly. By describing events more completely and by using these descriptions to cluster events more intuitively better summaries may be created. Finally it is worth noting that the proposed clustering approach can be used on any type of data. Since the proposed clustering algorithm chooses parameters independently on data type a minimal level of human interaction is needed.

Bibliography

[1] U. Damnjanovic and E. Izquierdo, "Asymmetric and normalized cuts for image clustering and segmentation," *Neural Network Applications in Electrical Engineering, 2006. NEUREL 2006. 8th Seminar on*, pp. 5–9, Sept. 2006.

[2] U. Damnjanovic, E. Izquierdo, and M. Grzegorzek, "Shot boundary detection using spectral clustering," in *Proceedings of European Signal Processing Conference*, 2007.

[3] U. Damnjanovic, V. Fernandez, E. Izquierdo, and J. M. Martinez, "Event detection and clustering for surveillance video summarization," in *WIAMIS '08: Proceedings of the 2008 Ninth International Workshop on Image Analysis for Multimedia Interactive Services*, (Washington, DC, USA), pp. 63–66, IEEE Computer Society, 2008.

[4] U. Damnjanovic, T. Piatrik, D. Djordjevic, E. Izquierdo, and J. M. Martinez, "Video summarisation for surveillance and news domain," in *2nd International Conference on Semantics and Digital Media Technologies (SAMT2007)*, pp. 99–112, Springer Berlin / Heidelberg, 2007.

[5] S. U. Naci, U. Damnjanovic, B. Mansencal, J. Benois-Pineau, C. Kaes, M. Corvaglia, E. Rossi, and N. Aginako, "The cost292 experimental framework for rushes summarization task in trecvid 2008," in *TVS 08: Proceedings of the 2nd ACM TRECVid Video Summarization Workshop*, (New York, NY, USA), pp. 40–44, ACM, 2008.

[6] P. H. Schiller, "The central visual system," *Vision Research*, vol. 26, no. 9, pp. 1351 – 1386, 1986. Twenty-Fifth Anniversary Issue of Vision Research.

[7] M. Dastani, "The role of visual perception in data visualization," *Journal of Visual Languages and Computing*, vol. 13, pp. 601–622, 2002.

[8] D. H. Hubel, *Eye, Brain, and Vision (Scientific American Library)*. W H Freeman & Co (Sd), 1988.

[9] D. A. Norman, "Affordance, conventions, and design," *interactions*, vol. 6, no. 3, pp. 38–43, 1999.

[10] J. J. Gibson, *The Ecological Approach to Visual Perception*. Lawrence Erlbaum Associates, 1979.

[11] D. Chang, L. Dooley, and J. E. Tuovinen, "Gestalt theory in visual screen design: a new look at an old subject," in *CRPIT '02: Proceedings of the Seventh world conference on computers in education conference on Computers in education: Australian topics*, (Darlinghurst, Australia, Australia), pp. 5–12, Australian Computer Society, Inc., 2002.

[12] T. Gevers and A. Smeulders, "Pictoseek: Combining color and shape invariant features for image retrieval," *IEEE Transactions on Image Processing*, vol. 9, pp. 102–119, January 2000.

[13] S. D. Servetto, Y. Rui, K. Ramchandran, and T. S. Huang, "A region-based representation of images in mars," *J. VLSI Signal Process. Syst.*, vol. 20, no. 1-2, pp. 137–150, 1998.

[14] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 12, pp. 1349–1380, 2000.

[15] G. Pass and R. Zabih, "Histogram refinement for content-based image retrieval," in *IEEE Workshop on Applications of Computer Vision*, pp. 96–102, 1996.

[16] Y. Tao and W. I. Grosky, "Spatial color indexing using rotation, translation, and scale invariant anglograms," *Multimedia Tools Appl.*, vol. 15, no. 3, pp. 247–268, 2001.

[17] J. Huang, S. R. Kumar, M. Mitra, W.-J. Zhu, and R. Zabih, "Image indexing using color correlograms," in *CVPR '97: Proceedings of the 1997 Conference on Computer Vision and Pattern Recognition (CVPR '97)*, (Washington, DC, USA), p. 762, IEEE Computer Society, 1997.

[18] S. Dickinson, A. Pentland, and S. Stevenson, "Viewpoint-invariant indexing for content-based image retrieval," in *Proc. IEEE Int. Workshop on Content-Based Access of Image and Video Database*, pp. 20–30, 1998.

[19] G. D. Finalyson, "Color in perspective," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 18, no. 10, pp. 1034–1038, 1996.

[20] G. Gimel'Farb and A. Jain, "On retrieving textured images from an image database," *Pattern recognition*, vol. 29, pp. 1461–1483, September 1996.

[21] H. Tamura, S. Mori, and T. Yamawaki, "Textural features corresponding to visual perception," *Systems, Man and Cybernetics, IEEE Transactions on*, vol. 8, pp. 460–473, June 1978.

[22] A. Pinheiro, E. Izquierdo, M. Ghanhari, "Shape Matching using a Curvature Based Polygonal Approximation in Scale-Space", *IEEE Proceedings ICIP*, September 2000, Vol. 2, Pages 538-541.

[23] A. Dorado, E. Izquierdo, "Semi-Automatic Image Annotation Using Frequent Keyword Mining", *IEEE Proceedings 7th International Conference on Information Visualisation* (IV 2003), Pages 532-535.

[24] Y. Wang, E. Izquierdo, "High-Capacity Data Hiding in MPEG-2 Compressed Video", 9th International Workshop on Systems, Signals and Image Processing (IWSSIP 2002), World Scientific, Pages 212-218.

[25] A. Dorado, E. Izquierdo, "Fuzzy Color Signatures", IEEE Proceedings International Conference on Image Processing (ICIP 2002). Rochester, New York, 22-25 September 2002, Volume 1, Pages 433-436.

[26] P. Salembier and T. Sikora, *Introduction to MPEG-7: Multimedia Content Description Interface*. New York, NY, USA: John Wiley & Sons, Inc., 2002.

[27] P. Vinicius Koerich Borges, J. Mayer, E. Izquierdo, "Robust and Transparent Color Modulation for Text Data Hiding", *IEEE Transactions on Multimedia*, Volume 10, Issue 8, Pages 1479-1489

[28] T. Zgaljic, N. Sprljan, E. Izquierdo, "Bitstream Syntax Description based Adaptation of Scalable Video", *2nd European Workshop EWIMT 2005*, London, England, 30 November - 1 December 2005

[29] P. K. Borges, J. Mayer, E. Izquierdo, "Efficient visual fire detection applied for video retrieval", *European Signal Processing Conference (EUSIPCO-2008)*, Lausanne, Switzerland, 25-29 August 2008, Pages 1-5

[30] E. Izquierdo, M. Ernst, "Motion/Disparity analysis for 3D-Video-Conference Applications", *International Workshop on Stereoscopy and 3-Dimensional Imaging, Santorin*i, Greece, September 1995

[31] S Kay, E. Izquierdo, "Robust Content Based Image Watermarking", *3rd Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2001)*, Tampere, Finland, 16-17 May 2001, Pages 53-56.

[32] X.-Y. Wei and C.-W. Ngo, “Ontology-enriched semantic space for video search,” in *MULTIMEDIA ’07: Proceedings of the 15th international conference on Multimedia*, (New York, NY, USA), pp. 981–990, ACM, 2007.

[33] A. Vailaya, A. Member, M. A. T. Figueiredo, A. K. Jain, H. jiang Zhang, and S. Member, “Image classification for content-based indexing,” *IEEE Transactions on Image Processing*, vol. 10, pp. 117–130, 2001.

[34] Y. Chen, J. Z. Wang, and R. Krovetz, “Clue: Cluster-based retrieval of images by unsupervised learning,” *IEEE Transactions on Image Processing*, vol. 14, p. 2005, 2005.

[35] E. Izquierdo, et al., "Advanced Content-Based Semantic Scene Analysis and Information Retrieval", *4th European Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2003)*, World Scientific Publishing, London, England, 9-11 April 2003, Pages 519-528.

[36] Janko Calic, Ebroul Izquierdo, “Temporal Segmentation of MPEG Video Streams”, *EURASIP Journal on Applied Signal Processing*, Issue 6, Pages 561—565, 2002/6/26.

[37] Q. Zhang, E. Izquierdo, “A Multi-Feature Optimization Approach to Object-Based Image Classification”, *5th International Conference Image and Video Retrieval*, Tempe, Arizona, 13-15 July 2006, Pages 310—319.

[38] Krishna Chandramouli, Ebroul Izquierdo, “Image Classification using Self-Organising Feature Map and Particle Swarm Optimisation”, *7th International Workshop WIAMIS 2006*, Korea, Pages 313-316.

[39] E. Izquierdo, M. Ghanbari, "Key Components for an Advanced Segmentation System", *IEEE Transactions on Multimedia*, Vol. 4, Issue 1.

[40] T Kliegr, K. Chandramouli, J. Nemrava, V. Svatek, E. Izquierdo, “Combining Image Captions and Visual Analysis for Image Concept Classification”, *9th International Workshop on Multimedia Data Mining”*, Las Vegas, NV, 24-27 August 2008, Pages 817.

[41] J.-P. Schober, T. Hermes, and O. Herzog, “Picturefinder: Description logics for semantic image retrieval,” *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*, pp. 1571–1574, July 2005.

[42] Y. Rui, M. Ortega, T. S. Huang, and S. Mehrotra, "Information retrieval beyond the text document," *Library Trends*, vol. 48, pp. 455–474, 1999.

[43] Janko Calic, Ebroul Izquierdo, "A Multiresolution Technique for Video Indexing and Retrieval", *International Conference on Image Processing Rochester*, NY, September, 2002, Volume 1, Pages 952—955.

[44] N. Ramzan, S. Wan, E. Izquierdo, "Joint Source-Channel Coding for Wavelet-Based Scalable Video Transmission Using an Adaptive Turbo Code", *EURASIP Journal on Image and Video Processing*, Volume 2007.

[45] T. Zgaljic, N. Sprljan, E. Izquierdo, "Bit-stream allocation methods for scalable video coding supporting wireless communications", *Signal Processing: Image Communicatio*, Volume 22, Issue 3, Pages 298-316

[46] J. Calic, S. Sav, E. Izquierdo, S. Marlow, N. O'Connor, "Temporal Video Segmentation For Real-Time Key Frame Extraction", *27th IEEE International Conference on Acoustics, Speech, and Signal Processing*, Orlando, FL, 13-17 May 2002, Volume 4, Pages 3632-3635

[47] J. R. Cózar, N. Guil, E. L Zapata, E. Izquierdo, "Logotype detection to support semantic-based video annotation", *Journal Signal Processing: Image Communication*, Volume 22, Issue 7-8, Pages 669-679.

[48] O. Marques and B. Furht, *Content-Based Image and Video Retrieval*. Norwell, MA, USA: Kluwer Academic Publishers, 2002.

[49] M. S. Lew, N. Sebe, and J. P. Eakins, "Challenges of image and video retrieval," in *in Lecture Notes in Computer Science*, pp. 1–6, Springer-Verlag, 2002.

[50] K. S. Lee, W. B. Croft, and J. Allan, "A cluster-based resampling method for pseudo-relevance feedback," in *SIGIR '08: Proceedings of the 31st annual international ACM SIGIR conference on Research and development in information retrieval*, (New York, NY, USA), pp. 235–242, ACM, 2008.

[51] I. Ahmad and W. I. Grosky, "Spatial similarity-based retrievals and image indexing by hierarchical decomposition," in *Proceedings of the International Database Engineering and Application*, pp. 269–278, 1997.

[52] Y. A. Aslandogan, C. Thier, C. T. Yu, C. Liu, and K. R. Nair, "Design, implementation and evaluation of score a system for content based retrieval of pictures," in *In Eleventh International Conference on Data Engineering*, pp. 280–287, 1995.

[53] S. Chang and S. Liu, "Picture indexing and abstraction techniques for pictorial databases," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 6, pp. 475–484, July 1984.

[54] C.-C. Chang and T.-C. Wu, "Retrieving the most similar symbolic pictures from pictorial databases," *Inf. Process. Manage.*, vol. 28, no. 5, pp. 581–588, 1992.

[55] S. K. Chang, Q. Y. Shi, and C. W. Yan, "Iconic indexing by 2-d strings," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 9, no. 3, pp. 413–428, 1987.

[56] P. Huang and Y. Jean, "Using 2d c+-strings as spatial knowledge representation for image database systems," *Pattern Recognition*, vol. 27, pp. 1249–1257, September 1994.

[57] A. Pentland, R. Picard, and S. Sclaroff, "Photobook: Content-based manipulation of image databases," *International Journal of Computer Vision*, vol. 18, no. 3, pp. 233–254, 1994.

[58] Y. Tao and W. I. Grosky, "Object-based image retrieval using point feature maps," in *Proceedings of The 8th IFIP 2.6 Working Conference on Database Semantics (DS8), Rotozur, New Zealand*, pp. 5–8, 1999.

[59] A. K. Jain and R. C. Dubes, *Algorithms for clustering data*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1988.

[60] L. Fisher and J. W. V. Ness, "Admissible clustering procedures," *Biometrika*, vol. 58, no. 1, pp. 91–104, 1971.

[61] M. Fiedler, "Algebraic connectivity of graphs," *Czechoslovak Mathematical Journal*, vol. 23, no. 98, pp. 298–305, 1973.

[62] F. R. K. Chung, *Spectral Graph Theory (CBMS Regional Conference Series in Mathematics, No. 92) (Cbms Regional Conference Series in Mathematics)*. American Mathematical Society, February 1997.

[63] Y.-C. Wei and C.-K. Cheng, "Towards efficient hierarchical designs by ratio cut partitioning," *Computer-Aided Design, 1989. ICCAD-89. Digest of Technical Papers., 1989 IEEE International Conference on*, pp. 298–301, Nov 1989.

[64] R. Kannan, S. Vempala, and A. Vetta, "On clusterings: good, bad and spectral," in *Journal of the ACM*, pp. 367–377, 2000.

[65] C. Ding, X. He, H. Zha, M. Gu, and H. Simon, "A min-max cut algorithm for graph partitioning and data clustering," *Data Mining, 2001. ICDM 2001, Proceedings IEEE International Conference on*, pp. 107–114, 2001.

[66] J. Calic, E. Izquierdo, "Towards Real-Time Shot Detection in the MPEG Compressed Domain", *3rd Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2001)*, Finland, 16-17 May 2001, Pages 1-5.

[67] A. Pothen, H. D. Simon, and K.-P. Liou, "Partitioning sparse matrices with eigenvectors of graphs," *SIAM J. Matrix Anal. Appl.*, vol. 11, no. 3, pp. 430–452, 1990.

[68] J. Shi and J. Malik, "Normalized cuts and image segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 8, pp. 888–905, 2000.

[69] M. Maila and J. Shi, "A random walks view of spectral segmentation," in *AI and STATISTICS (AISTATS) 2001*, 2001.

[70] F. R. Bach and M. I. Jordan, "Learning spectral clustering," in *Advances in Neural Information Processing Systems 16*, MIT Press, 2003.

[71] M. Meila and L. Xu, "Multiway cuts and spectral clustering," Tech. Rep. 442, University of Washington, September 2004.

[72] X. Zheng and X. Lin, "Automatic determination of intrinsic cluster number family in spectral clustering using random walk on graph," vol. 5, pp. 3471–3474 Vol. 5, 2004.

[73] M. Meila and J. Shi, "Learning segmentation by random walks," in *NIPS*, pp. 873–879, 2000.

[74] M. Meila, S. Shortreed, and X. Liang, "Regularized spectral learning," Tech. Rep. 465, University of Washington, 2005.

[75] G. Sanguinetti, J. Laidler, and N. D. Lawrence, "Automatic determination of the number of clusters using spectral algorithms," *IEEE Workshop on Machine Learning for Signal Processing*, pp. 55–60, 2005.

[76] M. Meila and V. Deepack, "A comparison of spectral clustering algorithms," Tech. Rep. 03-05-01, University of Washington, 2001.

[77] A. Y. Ng, M. I. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," in *Advances in Neural Information Processing Systems 14*, pp. 849–856, MIT Press, 2001.

[78] U. von Luxburg, "A tutorial on spectral clustering," *Statistics and Computing*, vol. 17, pp. 395–416, December 2007.

[79] S. X. Yu and J. Shi, "Multiclass spectral clustering," in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pp. 313–319 vol.1, 2003.

[80] C. J. Alpert, A. B. Kahng, and S. zen Yao, "Spectral partitioning: The more eigenvectors, the better," in *Proc. ACM/IEEE Design Automation Conf*, pp. 195–200, 1995.

[81] C. Chennubhotla and A. D. Jepson, "Half-lives of eigenflows for spectral clustering," in *In NIPS*, pp. 689–696, MIT Press, 2002.

[82] T. Xiang and S. Gong, "Spectral clustering with eigenvector selection," *Pattern Recogn.*, vol. 41, no. 3, pp. 1012–1029, 2008.

[83] D. M. Cvetkovic, M. Doob, and H. Sachs, *Spectra of Graphs: Theory and Applications, 3rd Revised and Enlarged Edition*. Vch Verlagsgesellschaft Mbh, December 1998.

[84] D. Cvetkovic, P. Rowlinson, and S. Simic, *Eigenspaces of Graphs (Encyclopedia of Mathematics and its Applications)*. Cambridge University Press, January 1997.

[85] B. Mohar, "Laplace eigenvalues of graphs—a survey," *Discrete Math.*, vol. 109, no. 1-3, pp. 171–183, 1992.

[86] R. Merris, "Laplacian graph eigenvectors," *Linear Algebra and its Applications*, vol. 278, pp. 221–236, July 1998.

[87] R. Marris, "A survey of graph laplacians," *Linear and Multilinear Algebra*, vol. 39, pp. 19–31, 1995.

[88] P. Rowlinson, D. Cvetkovic, and S. Simic, "Graphs with least eigenvalue -2: a new proof of the 31 forbidden subgraphs theorem," *Designs, Codes and Cryptography*, vol. 34, pp. 229–240, 2005.

[89] W. Hesseler and S. Eickeler, "Mpeg-2 compressed-domain algorithms for video analysis," *EURASIP Journal on Applied Signal Processing*, vol. 2006, pp. 186–186, 2006.

[90] J. Calic and E. Izquierdo, "Towards real time shot detection in the mpeg compressed domain," in *Proceedings of WIAMIS 2001*, 2001.

[91] B. Yeo and B. Liu, “Rapid scene analysis on compressed video,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 5, pp. 533–544, December 1995.

[92] M. Hanke, E. Izquierdo, R. März, "On Asymptotics in Case of Linear Index-2 Differential-Algebraic Equations", *SIAM Journal on Numerical Analysis 1998*, Volume 35, Issue 4, Pages 1326-1346

[93] S. Wan, E. Izquierdo, “Rate-distortion optimized motion-compensated prediction for packet loss resilient video coding”, *Image Processing, IEEE Transactions on*, Volume 16, Issue 5, Pages 1327-1338

[94] K. Chandramouli, T. Kliegr, J. Nemrava, E. Izquierdo, “Query Refinement and User Relevance Feedback for Contextualized Image Retrieval”, *5th International Conference on Visual Information Engineering* (VIE 2008). Xi'an, China, 29 July - 1 August 2008, Pages 453—458

[95] Q. Zhang, E. Izquierdo, “Combining Low-Level Features for Semantic Inference in Image Retrieval”, J. on Advances in Signal Processing, 2007.

[96] E. Izquierdo, V. Guerra Ones, "An Ill-Posed Operator for Secure Image Authentication", *IEEE Transactions on Circuits and Systems for Video Technology*, Volume 13, Issue 8, Pages 842-852

[97] N. E O'Connor, T. Adamek, V. Mezaris, E. Izquierdo, C. Bennström, J. R Casas, "Region and Object Segmentation Algorithms in the Qimera Segmentation Platform", *3rd International Workshop on Content-Based Multimedia Indexing*, France, 22-24 September 2003, Pages 1-8

[98] M. Mrak, N. Sprljan, T. Zgaljic, N. Ramzan, S. Wan, E. Izquierdo, "Performance evidence of software proposal for Wavelet Video Coding Exploration group", *Conference 76th MPEG Meeting*, 2006/4/7.

[99] M. Yeung, B. Yeo, and B. Liu, “Segmentation of video by clustering and graph analysis,” *Computer Vision and Image Understanding*, vol. 71, pp. 94–109, July 1998.

[100] W. Heng, K. Ngan, and M. Lee, “Comparison of mpeg domain elements for low-level shot boundary detection,” *Real Time Imaging*, vol. 8, pp. 341–358, October 1999.

[101] L. Cheong, “Scene-based shot change detection and comparative evaluation,” *Computer Vision and Image Understanding*, vol. 79, pp. 224–235, August 2000.

[102] C. Huang and B. Liao, “A robust scene-change detection method for video segmentation,” *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 11, pp. 1281–1288, December 2001.

[103] E. Izquierdo, S. Kruse, "Image Analysis for 3D Modeling, Rendering, and Virtual View Generation", *Elsevier Journal Computer Vision and Image Understanding*, 1998, Volume 71, Issue 2, Pages 231-253.

[104] N. Sprljan, M. Mrak, C. Abhayaratne, E. Izquierdo, "A Scalable Coding Framework for Efficient Video Adaptation", *6th International Workshop on Image Analysis for Multimedia Interactive Services (WIAMIS 2005)*, Montreux, Switzerland, 2005, 13-15 April 2005, Pages 1-4

[105] P. Wilkins, E. Izquierdo, et al., "K-Space at TRECVid 2006", *4th TRECVID Workshop*, Gaithersburg, Maryland, 13-14 November 2006

[106] E. Izquierdo, J-R. Ohm, "Image-based rendering and 3D modeling: a complete framework", *Signal Processing: Image Communication*, Volume 15, Issue 10, 2000, Pages 817-858

[107] K. Chandramouli, E. Izquierdo, "Image Classification using Chaotic Particle Swarm Optimization", *13th IEEE International Conference on Image Processing (ICIP 2006)*, Atlanta, Georgia, 8-11 October 2006, Pages 3001-3004.

[108] A. Divakaran and I. Otsuka, “A video-browsing-enhanced personal video recorder,” in *ICIAPW ’07: Proceedings of the 14th International Conference of Image Analysis and Processing - Workshops*, (Washington, DC, USA), pp. 137–142, IEEE Computer Society, 2007.

[109] W. Yong, Z. Tong, and D. Tretter, “Real time video motion analysis toward semantic understanding,” in *In SPIE Video Communications and Image Processing (VCIP)*, 2005.

[110] N. Adami, S. Benini, and R. Leonardi, “An overview of video shot clustering and summarization techniques for mobile applications,” in *MobiMedia ’06: Proceedings of the 2nd international conference on Mobile multimedia communications*, (New York, NY, USA), pp. 1–6, ACM, 2006.

[111] O. Bao and L. Guan, “Scene change detection using dc coefficients,” in *Proceedings of the International Conference on Image Processing*, pp. II: 421–424, 2002.

[112] H. Kang, “A hierarchical approach to scene segmentation,” in *IEEE Workshop on Content-Based Access of Image and Video Libraries*, p. 65, 2001.

[113] A. Girgensohn, "A fast layout algorithm for visual video summaries," in *ICME '03: Proceedings of the 2003 International Conference on Multimedia and Expo*, (Washington, DC, USA), pp. 77–80, IEEE Computer Society, 2003.

[114] G. Ciocca and R. Schettini, "An innovative algorithm for key-frame extraction in video summarization," *Journal of Real-Time Image Processing*, vol. 1, pp. 69–88, 2006.

[115] B. Luo, X. Tang, J. Liu, and H. Zhang, "Video caption detection and extraction using temporal information," in *Proceedings of the International Conference on Image Processing*, pp. I: 297–300, 2003.

[116] A. G. Money and H. Agius, "Video summarisation: A conceptual framework and survey of the state of the art," *J. Vis. Comun. Image Represent.*, vol. 19, no. 2, pp. 121–143, 2008.

[117] A. Ekin, A. Tekalp, and R. Mehrotra, "Automatic soccer video analysis and summarisation," *IEEE Transactions on Image Processing*, vol. 12, pp. 796–807, July 2003.

[118] A. Jaimes, T. Echigo, M. Teraguchi, and F. Satoh, "Learning personalized video highlights from detailed mpeg-7 metadata," *Image Processing. 2002. Proceedings. 2002 International Conference on*, vol. 1, pp. I–133–I–136 vol.1, 2002.

[119] S. Lu, M. Lyu, and I. King, "Video summarization by spatial-temporal graph optimization," *Circuits and Systems, 2004. ISCAS '04. Proceedings of the 2004 International Symposium on*, vol. 2, pp. II–197–200 Vol.2, May 2004.

[120] A. Hanjalic and L.-Q. Xu, "Affective video content representation and modeling," *Multimedia, IEEE Transactions on*, vol. 7, no. 1, pp. 143–154, 2005.

[121] E. Izquierdo, "Stereo Image Analysis for Multi-viewpoint Telepresence Applications", *Elsevier Journal Signal Processing: Image Communication*, Volume 11, Issue 3, Pages 231-254.

[122] A. Hanjalic and H. Zhang, "An integrated scheme for automated video abstraction based on unsupervised cluster-validity analysis," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, pp. 1280–1289, December 1999.

[123] M. Sugano, Y. Nakajima, and H. Yanagihara, "Automated mpeg audio-video summarization and description," *Image Processing. 2002. Proceedings. 2002 International Conference on*, vol. 1, pp. I–956–I–959 vol.1, 2002.

[124] R. Lienhart, S. Pfeiffer, and W. Effelsberg, "Video abstracting," *Commun. ACM*, vol. 40, no. 12, pp. 54–62, 1997.

[125] M. Fayzullin, V. S. Subrahmanian, M. Albanese, and A. Picariello, "The priority curve algorithm for video summarization," in *MMDB '04: Proceedings of the 2nd ACM international workshop on Multimedia databases*, (New York, NY, USA), pp. 28–35, ACM, 2004.

[126] A. G. Papernick, N.and Hauptmann, "Summarisation of broadcast news video through link analysis of named entities," in *In Proceedings of the AAAI workshop on link analysis*, 2005.

[127] D. Zhong, R. Kumar, and S.-F. Chang, "Real-time personalized sports video filtering and summarization," in *MULTIMEDIA '01: Proceedings of the ninth ACM international conference on Multimedia*, (New York, NY, USA), pp. 623–625, ACM, 2001.

[128] D. Farin, W. Effelsberg, and P. de With, "Robust clustering-based video-summarization with integration of domain-knowledge," *IEEE International Conference on Multimedia and Expo, 2002. ICME '02.*, vol. 1, pp. 89–92, 2002.

[129] M. S. Drew and J. Au, "Clustering of compressed illumination-invariant chromaticity signatures for efficient video summarization," *Image and Vision Computing Journal*, vol. 21, pp. 705–716, 2003.

[130] C. Lin and B. L. Tseng, "Optimizing user expectations for video semantic filtering and abstraction," in *In Proceedings of the IEEE ISCAS Conference*, 2005.

[131] N. Babaguchi, Y. Kawai, and T. Kitahashi, "Generation of personalized abstract of sports video," *Multimedia and Expo, 2001. ICME 2001. IEEE International Conference on*, pp. 619–622, Aug. 2001.

[132] M. Oren, C. Papageorgiou, P. Sinha, E. Osuna, and T. Poggio, "Pedestrian detection using wavelet templates," in *In Proc. Computer Vision and Pattern Recognition*, pp. 193–199, 1997.

[133] S. Shipman, R. Radhakrishan, and A. Divakaran, "Architecture for video summarisation services over home networks and the internet," in *In Proc. IEEE International Conference on Consumer Electronics (ICCE),*, pp. 201–202, 2006.

[134] K. Jones, "Automatic summarising: The state of the art," *Inf. Process. Manage.*, vol. 43, no. 6, pp. 1449–1481, 2007.

[135] S. Pfeiffer, R. Lienhart, S. Fischer, and W. Effelsberg, "Abstracting digital movies automatically," *Journal of Visual Communication and Image Representation*, vol. 7, pp. 345–353, December 1996.

[136] C. Kim and J. Hwang, "Object-based video abstraction for video surveillance systems," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, pp. 1128–1138, December 2002.

[137] P. Fonseca and F. Pereira, "Automatic video summarization based on mpeg-7 descriptions," *Signal processing and Image communication*, vol. 19, pp. 685–699, September 2004.

[138] A. Anjulan and C. Canagarajah, "A novel video mining system," in *IEEE International Conference on Image Processing*, pp. I: 185–188, 2007.

[139] M. Meilva, "Comparing clusterings: an axiomatic view," in *ICML '05: Proceedings of the 22nd international conference on Machine learning*, (New York, NY, USA), pp. 577–584, ACM, 2005.

[140] M. Meila, "Comparing clusterings—an information based distance," *J. Multivar. Anal.*, vol. 98, no. 5, pp. 873–895, 2007.

[141] M. M. Breunig, H.-P. Kriegel, R. T. Ng, and J. Sander, "Lof: identifying density-based local outliers," *SIGMOD Rec.*, vol. 29, no. 2, pp. 93–104, 2000.

[142] A. Chaudhary, E. S. Szalay, and A. W. Moore, "Very fast outlier detection in large multidimensional data sets," in *In DMKD02*, ACM Press, 2002.

[143] H. Xiong, G. P, and M. Steinbach, "Enhancing data analysis with noise removal," *IEEE Transactions on Knowledge and Data Engineering*, vol. 18, 2006.

[144] L. Ertoz, M. Steinbach, and V. Kumar, "Finding clusters of different sizes, shapes, and densities in noisy, high dimensional data," in *in Proceedings of Second SIAM International Conference on Data Mining*, 2003.

[145] S. Mika, B. Schölkopf, A. Smola, and Muller, "Kernel pca and de-noising in feature spaces," (Cambridge, MA, USA), pp. 536–542, MIT Press, 1999.

[146] T. N. Tran, R. Wehrens, and L. M. Buydens, "Knn-kernel density-based clustering for high-dimensional multivariate data," *Computational Statistics and Data Analysis*, vol. 51, no. 2, pp. 513 – 525, 2006.

[147] R. Behmo, N. Paragios, and V. Prinet, "Graph commute times for image representation," in *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, 2008.

[148] H. Qiu and E. R. Hancock, "Commute times for graph spectral clustering," *Computer Analysis of Images and Patterns*, pp. 128–136, 2005.

[149] M. Meila and J. Shi, "Learning segmentation by random walks," in *NIPS*, pp. 873–879, 2000.

[150] A. E. Raftery, K. Y. Yeung, K. Y. Yeung, C. Fraley, C. Fraley, A. Murua, A. Murua, W. L. Ruzzo, and W. L. Ruzzo, "Model based clustering and data transformation for gene expression data," *Bioinformatics*, vol. 17, 2001.

[151] Y. Yusoff, W. Christmas, and J. Kittler, "Video shot cut detection using adaptive thresholding," in *BMVC00*, 2000.

[152] P. Over, A. F. Smeaton, and G. Awad, "The trecvid 2008 bbc rushes summarization evaluation," in *TVS '08: Proceedings of the 2nd ACM TRECVid Video Summarization Workshop*, (New York, NY, USA), pp. 1–20, ACM, 2008.