Deficient Human Aspects in Current Multimedia Indexing and Retrieval (MIR) of Large Social Networks Databases

Dr. U. F. Eze,
Dept of Information Management Technology, FUTO, Nigeria

Gerald C. Ojimba
Dept of Computer Science, Anambra State University, Uli.

Abstract
An inside look at the contents of social networks databases shows a significant diversion from traditional database contents and functionality. There is also enormous evidences that Social networks are changing the way multimedia content is shared on the web, by allowing users to upload their photos, videos, and audio content, produced by any means of digital recorders such as mobile/smart-phones, and web/digital cameras. In this article, an overview of multimedia indexing and searching algorithms, following the data growth curve is presented in detail. This paper concludes with the social aspects and new, interesting views on multimedia retrieval in the large social media databases.
Keywords: multimedia, indexing, social media, algorithms social networks, databases, retrieval

INTRODUCTION

Traditionally database systems have been designed to support commercial data, consisting mainly of structured alphanumeric data. In recent years, database systems have added support for a number of nontraditional data types such as text documents, images, and maps and other spatial data. The goal is to make databases universal servers, which can store all types of data. Rather than add support for all such data types into the core database, vendors offer add-on packages that integrate with the database to provide such functionality.

PROBLEM DEFINITION:

The huge volumes of information in the social media inspired users to formulate new types of queries that pose complex questions to these heterogeneous databases. An attempt to index and retrieve multimedia content shared through the social media, using techniques of the Content Based Multimedia Retrieval community, shows clearly the inability to do so. The parameters and constraints posed from the social media aspect reformed the multimedia indexing and retrieval processes in a new problem seeking for new solutions.

The major problems of social multimedia indexing and retrieval exist due to 1. The enormous volumes of information that push existing techniques to their edges 2. The well known semantic gap between the low-level multimedia descriptors and the higher level concepts that exist in each multimedia content. These facts do not imply that previous knowledge and tools are totally useless, rather that they should be used in a different way. In general current systems have not yet had significant impact on society due to an inability to bridge the semantic gap between computers and humans.

MAJOR CHALLENGES

The following major research challenges are noteworthy of particular importance to the MIR research community. (1) Semantic search with emphasis on the detection of concepts in media with complex backgrounds; (2) Multi-modal analysis and retrieval algorithms especially towards exploiting the synergy between the various media including text and context information; (3) Experiential multimedia exploration systems toward allowing users to gain insight and explore media collections; (4) Interactive search, emergent semantics, or relevance feedback systems; and (5) Evaluation with emphasis on representative test sets and usage patterns.

HISTORICAL DEVELOPMENTS IN MIR

The earliest years of MIR were frequently based on computer vision (three excellent books: [Ballard and Brown 1982]; [Levine 1985]; [Haralick and Shapiro 1993]) algorithms focused on feature based similarity search over images, video, and audio. Influential and popular examples of these systems would be QBIC [Flickner, et al. 1995] and Virage [Bach, et al. 1996] circa mid 90s. Within a few years the basic concept of the similarity search was transferred to several Internet image search engines including Webseek [Smith and Chang 1997] and Webseer [Frankel, et al. 1996]. Significant effort was also placed into direct integration of the feature based similarity search into enterprise level databases such as Informix datablades, IBM DB2 Extenders, or Oracle Cartridges [Bliujute, et al. 1999; Egas, et al. 1999] towards making MIR more accessible to private industry. In the area of video retrieval l, the main focus in the mid 90s was toward robust shot boundary detection of which the most common approaches involved thresholding the distance between color histograms corresponding to two consecutive frames in a video [Flickner, et al. 1995]. Hanjalic, et al. [1997] proposed a method which overcame the problem of subjective user thresholds. Their approach was not dependent on any manual parameters. It gave a set of keyframes based on an objective model for the video information flow. Haas, et al. [1997] described a method to use the motion within the video to determine the shot boundary locations. Their method outperformed the histogram approaches of the period and also performed semantic classification of the video shots into categories such as zoom-in, zoom-out, pan, etc. A more recent practitioner's guide to video transition detection is given by Lienhart [2001].

Starting near the turn of the 21st century, researchers noticed that the feature based similarity search algorithms were not as intuitive nor user-friendly as they had expected. In ACM Transactions on Multimedia Computing, Communications, and Applications, Feb. 2006 One could say that systems built by research scientists were essentially systems which could only be used effectively by scientists.

The new direction was toward designing systems which would be user friendly and could bring the vast multimedia knowledge from libraries, databases, and collections to the world. To do this it was noted that the next evolution of systems would need to understand the semantics of a query, not simply the low level underlying computational features. This general problem was called "bridging the semantic gap". From a pattern recognition perspective, this roughly meant translating the easily computable low level content-based media features to high level concepts or terms which would be intuitive to the user.

Examples of bridging the semantic gap for the single concept of human faces were demonstrated by Rowley, et al. [1996] and Lew and Huijsmans [1996]. Perhaps the earliest pictorial content-based retrieval system which addressed the semantic gap problem in the query interface, indexing, and results was the ImageScape search engine [Lew 2000]. In this system, the user could make direct queries for multiple visual objects such as sky, trees, water, etc.using spatially positioned icons in a WWW index containing 10+ million images andvideos using keyframes. The system used information theory to determine the best features for minimizing uncertainty in the classification. At this point it is important to note that the feature based similarity search engines were useful in a variety of contexts [Smeulders, et al. 2000] such as searching trademark databases [Eakins, et al. 2003], finding video shots with similar visual content and motion or for DJs searching for music with similar rhythms [Foote 1999], and automatic detection of pornographic content [Forsyth and Fleck 1999; Bosson, et al. 2002].

Intuitively, the most pertinent applications are those where the basic features such as color and texture in images and video; or dominant rhythm, melody, or frequency spectrum in audio [Foote 1999] are highly correlated to the search goals of the particular application.

RECENT DEVELOPMENT

In this section we discuss representative work [Dimitrova 2003; Lew 2001; Sebe, et 2003 (CIVR)] done in content-based multimedia retrieval in the recent years. The two fundamental necessities for a multimedia information retrieval system are (1) Searching for a particular media item; and (2) Browsing and summarizing a media collection. In searching for a particular media item, the current systems have significant limitations such as an inability to understand a wide user vocabulary, understand the user' satisfaction level, nor do there exist credible representative real world test sets for evaluation nor even benchmarking measures which are clearly correlated with user satisfaction.

The prevalent research topics which have potential for improving multimedia retrieval by bridging the semantic gap are human-centered computing, new features, new media, browsing and summarization, and evaluation or benchmarking. In human-centered computing, the main idea is to satisfy the user and allow the user to make queries in their own terminology. User studies give us insight directly into the interactions between human and computer.

Experiential computing also focusses on methods for allowing the user to explore and gain insights in media collections. On a fundamental level, the notion of user satisfaction is inherently emotional. Affective computing is fascinating because it focusses on understanding the user's emotional state and intelligently reacting to it. It can also be beneficial toward measuring user satisfaction in the retrieval process.

Learning algorithms are interesting because they potentially allow the computer to understand the media collection on a semantic level. Furthermore, learning algorithms may be able to adapt and compensate for the noise and clutter in real world contexts.

New features are pertinent in that they can potentially improve the detection and recognition process or be correlated with human perception. New media types address the changing nature of the media in the collections or databases. Some of the recent new media include 3D models and biological imaging data. (i.e. towards understanding the machinery of life). As scientists, we need to objectively evaluate and benchmark the performance of the systems and take into account factors such as user satisfaction with results. Currently, there are no large international test sets for the wide problems such as searching personal media collections, so significant effort has been addressed toward developing paradigms which are effective for evaluation.

Furthermore, as collections grow from gigabyte to terabyte to petabyte sizes, high performance algorithms will be necessary toward responding to a query in an acceptable time period. Currently, the most commonly used test sets include collections involving personal photos, web images and videos, cultural heritage images, news video, and the Corel stock photography collection, which is also the most frequently mentioned collection. We are not asserting that the Corel collection is a good test set. We suspect it is popular simply because it is widely available and related loosely to real world usage. Furthermore, weIn ACM Transactions on Multimedia Computing, Communications, and Applications, Feb. 2006 think that it is only representative and suitable if the main goal of the particular retrieval system is to find professional stock photography. For the most recent research, there currently are several conferences dedicated to the field of MIR such as the ACM SIGMM Workshop on Multimedia Information Retrieval and the International Conference on Image and Video For a searchable MIR library, we suggest the community driven digital library at the Association for Multimedia Search and Retrieval Additionally, the general multimedia conferences such as ACM

Multimedia (http://www.sigmm.org) and the IEEE International Conference on Multimedia and Expo (ICME) typically have MIR related tracks.

Human-centered Computing

By human-centered we mean systems which consider the behavior and needs of the human user [Jaimes and Sebe, 2006]. As noted earlier, the foundational areas of MIR were often in computing-centric fields. However, since the primary goal is to provide effective browsing and search tools for the user, it is clear that the design of the systems should be human-centric. There have been several major recent initiatives in this direction such as user understanding, experiential computing, and affective computing. One of the most fascinating studies was done on whether organization by similarity assists image browsing [Rodden 2001]. The users were asked to illustrate a set of destination guide articles for a travel website. The similarity by visual content view was compared with a text caption similarity view. In 40 of the 54 searches, users chose to use the text caption view with comments such as "it gave me a breakdown of the subject." In many cases the users began with the text caption view to ensure sufficient diversity. Also, it was noted by the users that they would want both possibilities simultaneously. In another experiment, the visual similarity view was compared with a random set. Most users were slightly more satisfied with the visual similarity view, but there was one user who preferred the random images view. Specifically, the visual similarity view was preferred in 66% of the searches. A nice description of user requirements for photoware is discussed in [Frohlich 2002] and Lim, et al. [2003]. The importance of time in user interfaces is discussed in Graham, et al. [2002]. By understanding user types [Enser and Sandom 2003; Rubin 2004, Enser, et al. 2005], it is clear that the current work has not addressed the full plurality of image and user types and that a broad evaluation is important. In specific cases there has been niche work such as the use of general purpose documentary images by generalist andIn ACM Transactions on Multimedia Computing, Communications, and Applications, Feb. 2006 specialist users [Markkula and Sormunen 2000] and the use of creative images by specialist users [Hastings 1999]. Other interesting studies have been done on the process of managing personal photograph collections [Rodden and Wood 2003]. Worring and Gevers [2001] describe a concise analysis of methodologies for interactive retrieval of color images which includes guidelines for selecting methods based on the domain and the type of search goal. Also, Worring, et al. [2004] gave useful insights into how users apply the steps of indexing, filtering, browsing, and ranking in video retrieval.

Usage mining in large multimedia databases is another emerging problem. The objective is to extract the hidden information in user behaviors on large multimedia databases. A framework for video usage mining has been presented in Mongy, et al. [2005]. The idea behind experiential computing [Jain 2003; Jain, et al. 2003] is that decision makers routinely need insights that come purely from their own experience and experimentation with media and applications. These insights come from multiple perspectives and exploration [Gong, et al. 2004]. Instead of analyzing an experience, experiential environments provide support for naturally understanding events. In the context of MIR, experiential environments provide interfaces for creatively exploring sets of data, giving multiple perspectives and allowing the user to follow his insights. Affective computing [Picard 2000; Berthouze and Kato 1998; Hanjalic and Xu 2005] seeks to provide better interaction with the user by understanding the user's emotional state and responding in a way which influences or takes into account the user's emotions.

For example, Sebe, et al. [2002] recognize emotions automatically using a Cauchy classifier on an interactive 3D wireframe model of the face. Wang, et al. [2004] examine the problem of grouping images into emotional categories. They introduce a novel feature based on line direction-length which works effectively on a set of art paintings. Salway and Graham [2003] develop a method for extracting character emotions from film which is based on a model that links character emotions to the events in their environment.

Learning and Semantics

The potential for learning in multimedia retrieval is quite compelling toward bridging the semantic gap and the recent research literature has seen significant interest in applyingclassification and learning [Therrien 1989;

Winston 1992; Haralick and Shapiro 1993]algorithms to MIR. The Karhunen-Loeve (KL) transform or principal components method [Therrien 1989] has the property of representational optimality for a linear description of the media. Another approach toward learning semantics is to determine the associations behind features and the semantic descriptions. Djeraba [2002 and 2003] examines the problem of data mining and discovering hidden associations during image indexing and consider a visual dictionary which groups together similar colors and textures. A learning approach is explored by Krishnapuram, et al. [2004] in which they introduce a fuzzy graph matching algorithm. Greenspan, et al. [2004] performs clustering on space-time regions in feature space toward creating a piece-wise GMM framework which allows for the detection of video events.

Concept Detection

One of the most important challenges and perhaps the most difficult problem in semantic understanding of media is visual concept detection in the *presence of complex backgrounds*. Many researchers have looked at classifying whole images, but the granularity is often too coarse to be useful in real world applications. Its typically necessary to find the human in the picture, not simply global features. Another limiting case is where researchers have examined the problem of detecting visual concepts in laboratory conditions where the background is simple and therefore can be easily segmented. Thus, the challenge is to detect all of the semantic content within an image such as faces, trees, animals, etc. with emphasis on the presence of complex backgrounds.

In the mid 90s, there was a great deal of success in the special case of detecting the locations of human faces in grayscale images with complex backgrounds. Lew and Huijsmans [1996] used Shannon's information theory to minimize the uncertainty in the face detection process. Rowley, et al. [1996] applied several neural networks toward detecting faces. Both methods had the limitation of searching for whole faces which prompted later component based model approaches which combined separate detectors for the eyes and nose regions. For the case of near frontal face views in high quality photographs, the early systems generally performed near 95% accuracy with minimal false positives. Non-frontal views and low quality or older images from cultural heritage collections are still considered to be very difficult. An early example of designing a simple detector for city pictures was demonstrated by Vailaya, et al. [1998]. They used a nearest neighbor classifier in conjunction with edge histograms. In more recent work, Schneiderman and Kanade [2004] proposed a system for component based face detection using the statistics of parts. Chua, et al. [2002] used the gradient energy directly from the video representation to detect faces based on the high contrast areas such as the eyes, nose, and mouth. They also compared a rules based classifier with a neural network and found that the neural network gave superior accuracy. For a good overview, Yang, et al.

[2002] did a comprehensive survey on the area of face detection. Detecting a wider set of concepts other than human faces turned out to be fairly difficult. In the context of image search over the Internet, Lew [2000] showed a system for detecting sky, trees, mountains, grass, and faces in images with complex backgrounds. Fan, et al. [2004] used multi-level annotation of natural scenes using dominant image components and semantic concepts. Li and Wang [2003] used a statistical modeling approach toward converting images to keywords

Relevance Feedback

Beyond the one-shot queries in the early similarity based search systems, the next generation of systems attempted to integrate continuous feedback from the user toward learning more about the user query. The interactive process of asking the user a sequential set of questions after each round of results was called *relevance feedback* due to the similarity with older pure text approaches. Relevance feedback can be considered a special case of *emergent semantics*. Other names have included query refinement, interactive search, and active learning from the computer vision literature. The fundamental idea behind relevance feedback is to show the user a list of candidate images, ask the user to decide whether each image is relevant or irrelevant, and modify the parameter space, semantic space, feature space, or classification space to reflect the relevant and irrelevant examples. In the simplest relevance feedback method from Rocchio [Rocchio 1971], the idea is to move the query point toward the relevant examples and away from the irrelevant examples. In principle, one general view is to view relevance feedback as a particular type of pattern classification in which the positive and negative examples are found from the relevant and irrelevant labels, respectively. Therefore, it is possible to apply any learning algorithm into the relevance feedback loop. One of the major problems in relevance feedback is how to address the small training set. A typical user may only want to label 50 images when the algorithm really needs 5000 examples instead. If we compare the simple Rocchio algorithm to more sophisticated learning algorithms such as neural networks, its clear that one reason the

Rocchio algorithm is popular is that it requires very few examples. However, one challenging limitation of the Rocchio algorithm is that there is a single query point which would refer to a single cluster of results. In the discussion below we briefly describe some of the recent innovations in relevance feedback. describe the first WWW image search engine which focused on relevance feedback based improvement of the results. In their initial system, where they used relevance feedback to guide the feature selection process, it was found that the positive examples

New Features & Similarity Measures

Research did not only proceed along the lines of improved search algorithms, but also toward creating new features and similarity measures based on color, texture, and shape. One of the recent interesting additions to the set of features are from the MPEG-7 standard [Pereira and Koenen 2001]. The new color features [Lew 2001, Gevers2001] such as the NF, rgb, and m color spaces have specific benefits in areas such as lighting invariance, intuitiveness, and perceptual uniformity. A quantitative comparison of influential color models is performed in Sebe and Lew [2001]. In texture understanding, Ojala, et al. [1996] found that combining relatively simple texture histograms outperformed traditional texture models such as Gaussian or Markov features. Jafari-Khouzani and Soltanian-Zadeh [2005] proposed a new texture feature based on the Radon transform orientation which has the significant advantage of being rotationally invariant. Insight into the MPEG-7 texture descriptors has been given by Wu, et al. [2001].

New Media

In the early years of MIR, most research focused on content-based image retrieval. Recently, there has been a surge of interest in a wide variety of media. An excellent example, “life records”, which encompasses all types of media simultaneously is being actively promoted by Bell [2004]. He is investigating the issues and challenges in processing life records - all the text, audio, video, and media related to a person's life. Beyond text, audio, images, and video, there has been significant recent interest in new media such as 3D models. Another fascinating area is peering into biological databases consisting of imagery from the atomic through the visible light range. Applications can range from understanding the machinery of life to fast identification of dangerous bacteria or viruses.

The aspect of particular interest is how to combine the data from different imaging methods such as electron microscopes, MRI, X-ray, etc. Each imaging method uses a fundamentally different technique however the underlying content is the same. For example, Haas, et al. [2004] used a genetic algorithm learning approach combined with additional knowledge sources to search through virus databases and video collections.

Toward supporting imprecise queries in bio-databases, Chen, et al. [2002] used fuzzy equivalence classes to support query relaxation in biological imagery collections.

Browsing and Summarization

There have been a wide variety of innovative ways of browsing and summarizing multimedia information. Spierenburg and Huijsmans [1997] proposed a method for converting an image database into a movie. The intuition was that one could cluster a sufficiently large image database so that visually similar images would be in the same cluster. After the cluster process, one can order the clusters by the inter-cluster similarity, arrange the images in sequential order and then convert to a video. This allows a user to have a gestalt understanding of a large image database in minutes. Sundaram, et al. [2002] took a similar approach toward summarizing video. They introduced the idea of a video skim which is a shortened video composed of informative scenes from the original video. The fundamental idea is for the user to be able to receive an abstract of the story but in video format. Snoek, et al. [2005] propose several methods for summarizing video such as grouping by categories and browsing by category and in time. Chiu, et al. [2005] created a system for texturing a 3D city with relevant frames from video shots. The user would then be able to fly through the 3D city and browse all of the videos in a directory.

High Performance Indexing

In the early multimedia database systems, the multimedia items such as images or video were frequently simply files in a directory or entries in an SQL database table. From a computational efficiency perspective, both options exhibited poor performance because most filesystems use linear search within directories and most databases could only perform efficient operations on fixed size elements. Thus, as the size of the multimedia databases or collections grew from hundreds to thousands to millions of variable sized items, the computers could not respond in an acceptable time period. Even as the typical SQL database systems began to implement higher performance table searches, the search keys had to be exact such as in text search. Audio, images, and video were stored as blobs which could not be indexed effectively. Therefore, researchers [Egas, et al. 1999; Lew 2000] turned to similarity based databases which used tree-based indexes to achieve logarithmic performance. Even in the case of multimedia oriented databases such as the Informix database, it was still necessary to create custom datablades to handle efficient similarity searching such as k-d trees [Egas, et al. 1999]. In general the k-d tree methods had linear worst case performance and logarithmic average case performance in the context of feature based similarity searches. A recent improvement to the k-d tree method is to integrate entropy based balancing [Scott and Shyu 2003].

FUTURE DIRECTIONS

Despite the considerable progress of academic research in multimedia information retrieval, there has been relatively little impact of MIR research into commercial applications with some niche exceptions such as video segmentation. One example of an attempt to merge academic and commercial interests would be Riya (www.riya.com).

Their goal is to have a commercial product that uses the academic research in face detection and recognition and allows the users to search through their own photo collection or through the Internet for particular persons.

The potential landscape of multimedia information retrieval is quite wide and diverse. Below are some potential areas for additional MIR research challenges.

Human Centered Methods.

We should focus as much as possible on the user, who may want to explore instead of search for media. It has been noted that decision makers need to explore an area to acquire valuable insight, thus experiential systems which stress the exploration aspect are strongly encouraged. Studies on the needs of the user are also highly encouraged toward giving us understanding of their patterns and desires. New interactive devices (e.g., force, olfactory, and facial expression detectors) have largely been overlooked and should be tested to provide new possibilities, such as human emotional state detection and tracking.

Conclusion

Discovering more effective means of human-human computer-mediated interaction is increasingly important as our world becomes more wired or wirelessly connected. In a multimodal collaboration environment many questions remain: How do people find one another? How does an individual discover meetings/collaborations? What are the most effective multimedia interfaces in these environments for different purposes, individuals, and groups? Multimodal processing has many potential roles ranging from transcribing and summarizing meetings to correlating voices, names, and faces, to tracking individual (or group) attention and intention across media.

REFERENCES

2004, M.S. LEW, N. SEBE, C. DJERABA, Eds. ACM, New York, 77-83.

BALLARD, D.H. AND BROWN, C.M. 1982. *Computer Vision*. Prentice Hall, New Jersey, USA.

BAKKER, E.M. AND LEW, M.S. 2002. Semantic Video Retrieval Using Audio Analysis. *In Proceedings of the 1st International Conference on Image and Video Retrieval*, London, July 2002, M.S. LEW, N. SEBE, AND J.P. EAKINS, Eds. Springer-Verlag, London, 262-270.

BATTELLE, J. 2005. *The Search: How Google and Its Rivals Rewrote the Rules of Business and Transformed Our Culture*. Portfolio Hardcover, USA.

BELL, G. 2004. A New Relevance for Multimedia When We Record Everything Personal. In *Proceedings of the 12th annual ACM international conference on Multimedia*, ACM, New York, 1. 1105.

BERTHOUZE, N.B. AND KATO, T. 1998. Towards a comprehensive integration of subjective parameters in database browsing. In *Advanced Database Systems for Integration of Media and User Environments*, Y.

KAMBAYASHI, A. MAKINOUCHI, S. UEMURA, K. TANAKA, AND Y. MASUNAGA, Eds, World Scientific: Singapore, 227–232.

BOSSON, A., CAWLEY, G.C., CHAN, Y., AND HARVEY, R. 2002. Non-retrieval: Blocking Pornographic Images. *In Proceedings of the 1st International Conference on Image and Video Retrieval*, London, July 2002, M.S. LEW, N. SEBE, AND J.P. EAKINS, Eds. Springer-Verlag, London, 50-60.

CHANG, S.-F.,CHEN,W., AND SUNDARAM,H. 1998. Semantic visual templates: Linking visual features to semantics. In *Proceedings of the IEEE International Conference on Image Processing*, IEEE Computer Society Press, Los Alamitos, Calif. 531–535.

CHEN, Y., ZHOU, X.S., AND HUANG, T.S. 2001. One-class SVM for Learning in Image Retrieval, In *Proceedings of IEEE International Conference on Image Processing*, Thessaloniki, Greece, October, 815-818.

CHUA, T.S., ZHAO, Y., AND KANKANHALLI, M.S. 2002. Detection of human faces in a compressed domain for video stratification, *The Visual Computer 18(2)*, 121-133.

COOPER, M., FOOTE, J., GIRGENSOHN, A., AND WILCOX, L. 2005. Temporal event clustering for digital photo collections. *ACM Transactions on Multimedia Computing, Communications, and Applications* 1(3). 269-288.

DIMITROVA, N., AGNIHOTRI, L., AND WEI, G. 2000. Video Classification Based on HMM Using Text and

Faces. *European Signal Processing Conference*, Tampere, Finland.

DIMITROVA, N., ZHANG, H. J., SHAHRARAY, B., SEZAN, I., HUANG, T., AND ZAKHOR, A. 2002. Applications of video-content analysis and retrieval. *IEEE Multimedia 9(3)*, 42-55.

2003, E.M. BAKKER, T.S. HUANG, M.S. LEW, N. SEBE, AND X. ZHOU, Eds. Springer-Verlag, London, 28-38. *In ACM Transactions on Multimedia Computing, Communications, and Applications, Feb. 2006*

EGAS, R., HUIJSMANS, N., LEW, M.S., AND SEBE, N. 1999. Adapting k-d Trees to Visual Retrieval. *In Proceedings of the International Conference on Visual Information Systems*, Amsterdam, June 1999, A. SMEULDERS AND R. JAIN, Eds., 533-540.

ENSER, P.G.B., SANDOM, C.J. AND LEWIS, P.H. 2005. Automatic annotation of images from the practitioner perspective. *In Proceedings of the 4th International Conference on Image and Video Retrieval*, Singapore, July 2005, W. LEOW, M.S. LEW, T.-S. CHUA, W.-Y. MA, E.M. BAKKER, AND L.

CHAISORN, Eds. Springer-Verlag, London, 497-506.

FLICKNER, M. SAWHNEY, H. NIBLACK, W. ASHLEY, J. QIAN HUANG DOM, B. GORKANI, M.

HAFNER, J. LEE, D. PETKOVIC, D. STEELE, D. YANKER, P. 1995. Query by image and video content: the QBIC system, *IEEE Computer*, September, 23-32.

FOOTE, J. 1999. An Overview of Audio Information Retrieval. *ACM Multimedia Systems 7(1)*, 42-51.

FOOTE, J. 2000. Automatic audio segmentation using a measure of audio novelty. In *Proceedings of the IEEE International Conference on Multimedia and Expo*. IEEE, Computer Society Press, Los Alamitos, CA, 452–455.

FORSYTH, D.A., AND FLECK, M.M. 1999. Automatic Detection of Human Nudes, *International Journal of Computer Vision 32(1)*, 63-77.

FRANKEL, C., SWAIN, M.J., AND ATHITSOS, V. 1996. WebSeer: An Image Search Engine for the World Wide Web. *University of Chicago Technical Report 96-14*, University of Chicago, USA. Piecewise GMM. *IEEE Transactions on Pattern Analysis and Machine Intelligence 26(3)*, 384-396.

GUO, G., ZHANG, H.J., AND LI, S.Z. 2001. Boosting for Content-Based Audio Classification and Retrieval: An Evaluation, In *Proceedings of the IEEE Conference on Multimedia and Expo*, Tokyo, Japan, August.

HAAS, M., LEW, M.S. AND HUIJSMANS, D.P. 1997. A New Method for Key Frame based Video Content Representation. In *Image Databases and Multimedia Search*, A. SMEULDERS AND R. JAIN, Eds., World Scientific. 191-200.

HAAS, M., RIJSDAM, J. AND LEW, M. 2004. Relevance feedback: perceptual learning and retrieval in biocomputing, photos, and video, In *Proceedings of the 6th ACM SIGMM international workshop on Multimedia information retrieval*, New York, October, 151-156.

HANJALIC, A., LAGENDIJK, R.L., AND BIEMOND, J. 1997. A New Method for Key Frame based Video Content Representation. In *Image Databases and Multimedia Search*, A. SMEULDERS AND R. JAIN, Eds., World Scientific. 97-107.

HANJALIC, A. AND XU, L-Q. 2005. Affective Video Content Representation and Modeling, *IEEE*

HARALICK, R.M. AND SHAPIRO, L.G. 1993. *Computer and Robot Vision*. Addison-Wesley, New York, USA.

HUIJSMANS, D.P. AND SEBE, N. 2005. How to Complete Performance Graphs in Content-Based Image Retrieval: Add Generality and Normalize Scope. *IEEE Transactions on Pattern Analysis and Machine Intelligence 27(2)*, 245-251.

JAIMES A., SEBE N. 2006, Multimodal Human-computer Interaction: A survey, *Computer Vision and Image Understanding,* in press.

JAIN, R. 2003. A Game Experience in Every Application: Experiential Computing. *Communications of the ACM 46(7)*, 48-54.

JAIN, R., KIM, P., AND LI, Z. 2003. Experiential Meeting System. In *Proceedings of the 2003 ACM SIGMM Workshop on Experiential Telepresence*, Berkeley, USA, 1-12.

KRISHNAPURAM, R., MEDASANI, S., JUNG, S.H., CHOI, Y.S., AND BALASUBRAMANIAM, R. 2004. Content-Based Image Retrieval Based on a Fuzzy Approach. *IEEE Transactions on Knowledge and Data Engineering 16(10)*, 1185-1199.

LEVINE, M. 1985. *Vision in Man and Machine*, Mcgraw Hill, Columbus.

LEW, M.S. AND HUIJSMANS, N. 1996. Information Theory and Face Detection. In *Proceedings of the International Conference on Pattern Recogntion*, Vienna, Austria, 601-605.

LEW, M.S. 2000. Next Generation Web Searches for Visual Content. *IEEE Computer*, November, 46-53.

LEW, M.S. 2001. *Principles of Visual Information Retrieval*. Springer, London, UK.

LEW, M.S. AND DENTENEER, D. 2001. Fisher Keys for Content Based Retrieval. *Image and Vision Computing 19*, 561-566.

MARKKULA, M. AND SORMUNEN, E. 2000. End-user Searching Challenges Indexing Practices in the Digital Newspaper Photo Archive. *Information Retrieval 1(4)*, 259-285.
OJALA, T., PIETIKAINEN, M., AND HARWOOD, D. 1996. Comparative study of texture measures with classification based on feature distributions, *Pattern Recognition 29(1)*, 51-59.
SEBE, AND X. ZHOU, Eds. Springer-Verlag, London, 260-270.
ROCCHIO, 1971. Relevance Feedback in Information Retrieval. In *The Smart Retrieval System: Experiments in Automatic Document Processing*, G. Salton, Ed. Prentice Hall, Englewoods Cliffs.
ROWE, L.A. AND JAIN, R. 2005. ACM SIGMM retreat report on future directions in multimedia research. *ACM Transactions on Multimedia Computing, Communications, and Application 1(1),* 3-13.
ROWLEY, H., BALUJA, S., AND KANADE, K. 1996. Human Face Detection in Visual Scenes. *Advances in Neural Information Processing Systems 8 (Proceedings of NIPS)*, Denver, USA, November, 875-881.
RUBIN, R. 2004. *Foundations of Library and Information Science*. Neal-Schuman Publishers, New York.
RUI, Y. AND HUANG, T.S. 2001. Relevance Feedback Techniques in Image Retrieval. In *Principles of Visual Information Retrieval*, M.S. LEW, Ed. Springer-Verlag, London, 219-258.
SCHNEIDERMAN, H. AND KANADE, T. 2004. Object Detection Using the Statistics of Parts, *International Journal of Computer Vision* 56(3), 151-177.
SCOTT, G.J. AND SHYU, C.R. 2003. EBS k-d Tree: An Entropy Balanced Statistical k-d Tree for Image
Databases with Ground-Truth Labels. *In Proceedings of the 2nd International Conference on Image and Video Retrieval*, Urbana, July 2003, E.M. BAKKER, T.S. HUANG, M.S. LEW, N. SEBE, AND X. ZHOU, Eds. Springer-Verlag, London, 467-476.
SEBE, N., LEW, M.S., AND HUIJSMANS, D.P. 2000. Toward Improved Ranking Metrics. *IEEE Transactions on Pattern Analysis and Machine Intelligence 22(10)*, 1132-1143.
SEBE, N., AND LEW, M.S. 2001. Color Based Retrieval, *Pattern Recognition Letters 22(2)*, 223-230.
SEBE, N., AND LEW, M.S. 2002. Robust Shape Matching. *In Proceedings of the 1st International Conference on Image and Video Retrieval*, London, July 2002, M.S. LEW, N. SEBE, AND J.P. EAKINS, Eds. Springer- Verlag, London, 17-28.
SEBE, N., COHEN, I., GARG, A., LEW, M.S., AND HUANG, T.S. 2002. Emotion recognition using a Cauchy naive Bayes classifier. In *Proceedings of International Conference on Pattern Recognition*, Quebec, August, 17-20.
SEBE, N., TIAN, Q., LOUPIAS, E., LEW, M.S., AND HUANG, T.S. 2003. Evaluation of Salient Point Techniques. *Image and Vision Computing 21(13-14),* 1087-1095.
SEBE, N., LEW, M.S., ZHOU, X., AND HUANG, T.S. 2003. The State of the Art in Image and Video Retrieval. *In Proceedings of the 2nd International Conference on Image and Video Retrieval*, Urbana, July
SNOEK, C.G.M., WORRING, M., VAN GEMERT, J., GEUSEBROEK, J.M., KOELMA, D., NGUYEN, G.P., TIAN, Q., SEBE, N., LEW, M.S., LOUPIAS, E., AND HUANG, T.S. 2001. Image Retrieval using Waveletbased Salient Points. Journal of Electronic Imaging 10(4), 835-849.
TIAN, Q., MOGHADDAM, B., AND HUANG, T.S. 2002. Visualization, Estimation and User-Modeling. *In Proceedings of the 1st International Conference on Image and Video Retrieval*, London, July 2002, M.S.
LEW, N. SEBE, AND J.P. EAKINS, Eds. Springer-Verlag, London, 7-16.
UCHIHASHI, S., FOOTE, J., GIRGENSOHN, A., AND BORECZKY, J. 1999. Video Manga: generating semantically meaningful video summaries. In *Proceedings of the seventh ACM international conference on Multimedia*, Orlando, USA, 383-392.
VAILAYA, A., JAIN, A., AND ZHANG, H. 1998. On Image Classification: City vs Landscape. In Proceedings of *Workshop on Content-based Access of Image and Video Libraries*, 3-8.
WANG, W., YU, Y., AND ZHANG, J. 2004. Image Emotional Classification: Static vs. Dynamic. In *Proceedings of IEEE International Conference on Systems, Man, and Cybernetics*, October, 6407-6411.

The IISTE is a pioneer in the Open-Access hosting service and academic event management. The aim of the firm is Accelerating Global Knowledge Sharing.

More information about the firm can be found on the homepage:
http://www.iiste.org

CALL FOR JOURNAL PAPERS

There are more than 30 peer-reviewed academic journals hosted under the hosting platform.

Prospective authors of journals can find the submission instruction on the following page: http://www.iiste.org/journals/ All the journals articles are available online to the readers all over the world without financial, legal, or technical barriers other than those inseparable from gaining access to the internet itself. Paper version of the journals is also available upon request of readers and authors.

MORE RESOURCES

Book publication information: http://www.iiste.org/book/

Academic conference: http://www.iiste.org/conference/upcoming-conferences-call-for-paper/

IISTE Knowledge Sharing Partners

EBSCO, Index Copernicus, Ulrich's Periodicals Directory, JournalTOCS, PKP Open Archives Harvester, Bielefeld Academic Search Engine, Elektronische Zeitschriftenbibliothek EZB, Open J-Gate, OCLC WorldCat, Universe Digtial Library , NewJour, Google Scholar