Knowledge-Based Semantic Retrieval of Multimedia and Image Objects Using Collaborative Indexing

Prof. C. H. C. Leung[1], Yuanxi Li[2], W. S. Chan[3]

Department of Computer Science, Hong Kong Baptist University,
RRS702, 224 Waterloo Road, Kowloon Tong, Hong Kong
[1]clement@comp.hkbu.edu.hk; [2]yxli@comp.hkbu.edu.hk; [3]wschan@comp.hkbu.edu.hk

Abstract- **With the rapid development of multimedia technology, digital resources has become increasingly available and it constitutes a significant component of multimedia contents on the Internet. Since digital resources can be represented in various forms, formats, and dimensions, searching such information is far more challenging than text-based search. While some basic forms of multimedia retrieval are available on the Internet, these tend to be inflexible and have significant limitations. Currently, most of these multimedia retrieval systems mainly rely on text annotations. Here, we present an approach for deep concept-based multimedia information retrieval, which focuses on high-level human knowledge, perception, incorporating subtle nuances and emotional impression on the multimedia resources. We also provide a critical evaluation of the most common current Multimedia Information Retrieval approaches and propose an innovative adaptive method for multimedia information search that overcomes the current limitations. The main focus of our approach is concerned with image discovery and recovery by collaborative semantic indexing and user relevance feedback analysis. Through successive usage of our indexing model, novel image content indexing can be built from deep user knowledge incrementally and collectively by accumulating users' judgment and intelligence.**

Keywords- Collaborative Indexing; Multimedia Indexing; Image Information Retrieval; Relevance Feedback; Semantic Search

I. INTRODUCTION AND RELATED WORK

Multimedia is any combination of text, art, sound, animation, and video delivered to you by computer or other electronic or digitally manipulated means [16]. Through the rapid growth of multimedia technology, multimedia content can be created, shared and distributed easily. The amount of available digital resources is continuously increasing, promoted by a growing interest of users and by the development of new technology for the ubiquitous enjoyment of digital resources [12]. For example, the adage "A picture is worth a thousand words" refers to the notion that a complex idea can be conveyed with just a single still image [22]. Thus, unlike text-oriented web documents, it crosses the barriers of national languages and cultural backgrounds. Therefore, meaningful multimedia retrieval is important and necessary.

Multimedia information has become increasingly prevalent and it constitutes a significant component of multimedia contents on the Internet. Since multimedia information can be represented in various forms, formats, and dimensions, searching such information is far more challenging than text-based search. While some basic forms of multimedia retrieval are available on the Internet, these tend to be inflexible and have significant limitations. For example, currently, most of these image retrieval systems mainly rely on text annotations (e.g., page titles, ALT text, hyperlinks, and surrounding text, etc.). Also, [17] states that we need robust techniques to index/retrieve multimedia information in accessing huge multimedia databases, as well as semantic visual interfaces integrating the above components into unified multimedia browsing and retrieval systems. Here, we present an approach for deep concept-based multimedia information retrieval, which focuses on high-level human perception, incorporating subtle nuances and emotional impression on the multimedia resources.

The chief objective of effective multimedia information retrieval is the accurate transfer of digital information from a database to a user [1, 4, 6, 11, 12]. Since multimedia information can have different characteristics and its content can be represented in various ways and formats, effectively retrieving them from a large database can be very challenging. In an effective "concept-based" multimedia retrieval system, efficient and meaningful indexing is necessary [7, 8]. Due to current technological limitations, it is impossible to extract the semantic content of multimedia data objects automatically [13, 14].

II. RELATED WORK

In the digital world, multimedia information can be represented in various forms and different formats. Different formats can only capture limited information in different dimensions about the digital resources [18]. For example, typically, the most common image information representation approaches consist of symbolic, audio, visual, and metadata (see Table 1).

TABLE 1 EXAMPLES OF IMAGE REPRESENTATION APPROACHES

Representation	Examples
Text Annotation	page titles, ALT text, hyperlinks, and surrounding, etc.
Low-level features	Color, Shape and Texture, etc.
Hi-level Concept	Users' Comment Feedback
Metadata	Album information, XML, etc.

Present image information retrieval approaches mainly focus on shallow image content, such as color, shape and texture. Content-based image retrieval (CBIR) [23, 24, 25, 26, 27, 28, 29, 30, 31] is a technique used for retrieving similar images from an image database. CBIR operates on retrieving stored images from a collection by comparing features automatically extracted from the images themselves. The commonest current CBIR systems, whether commercial or experimental, operate at level 1. A typical system allows users to formulate queries by submitting an example of the type of image being sought, though some offer alternatives such as selection from a palette or sketch input. The system then identifies those stored images whose feature or signature values match those of the query most closely, and displays thumbnails of these images on the screen. The most challenging aspect of CBIR is to bridge the gap between low-level feature layout and high-level semantic concepts. In [15], a creative MIR approach based on personal emotion is suggested, which is closer to our proposed method. However, our proposed method is not just exclusively concerned with the human personal subjective perception, but also the general human perception of the community since we collect their perceptions through analyzing their search history and behaviour. Here, we base our innovative concept-based approach on multimedia search by adaptive collaborative indexing.

III. INDEX SYSTEM STRUCTURE AND HIERARCHY

In this section, we shall introduce our adaptive indexing approach for concept-based multimedia information retrieval. Our approach concerns multimedia discovery by collaborative semantic indexing and user relevance feedback. It enables the semantic search of multimedia information resources by the collective discovery and meaningful indexing of their semantic concepts. Therefore, we make full use of high-level human knowledge, perception incorporating subtle emotional impression on the multimedia information (e.g., styles, characteristics, mood etc.) rather than its low-level feature. Through the successive use of our model, semantic properties can be discovered and incorporated by user relevance feedback and users' search behaviour. Based on user-provided knowledge, more semantic concepts can be injected to multimedia resources. Eventually, through the growth and evolution of our index hierarchy, the semantic index can be dynamically constructed, validated, built-up, and converge towards communities' expectations. Our system consists of different levels for supporting multimedia search. It includes the user level, the interface level, and the database level. The database level contains two sublevels, query level and index level.

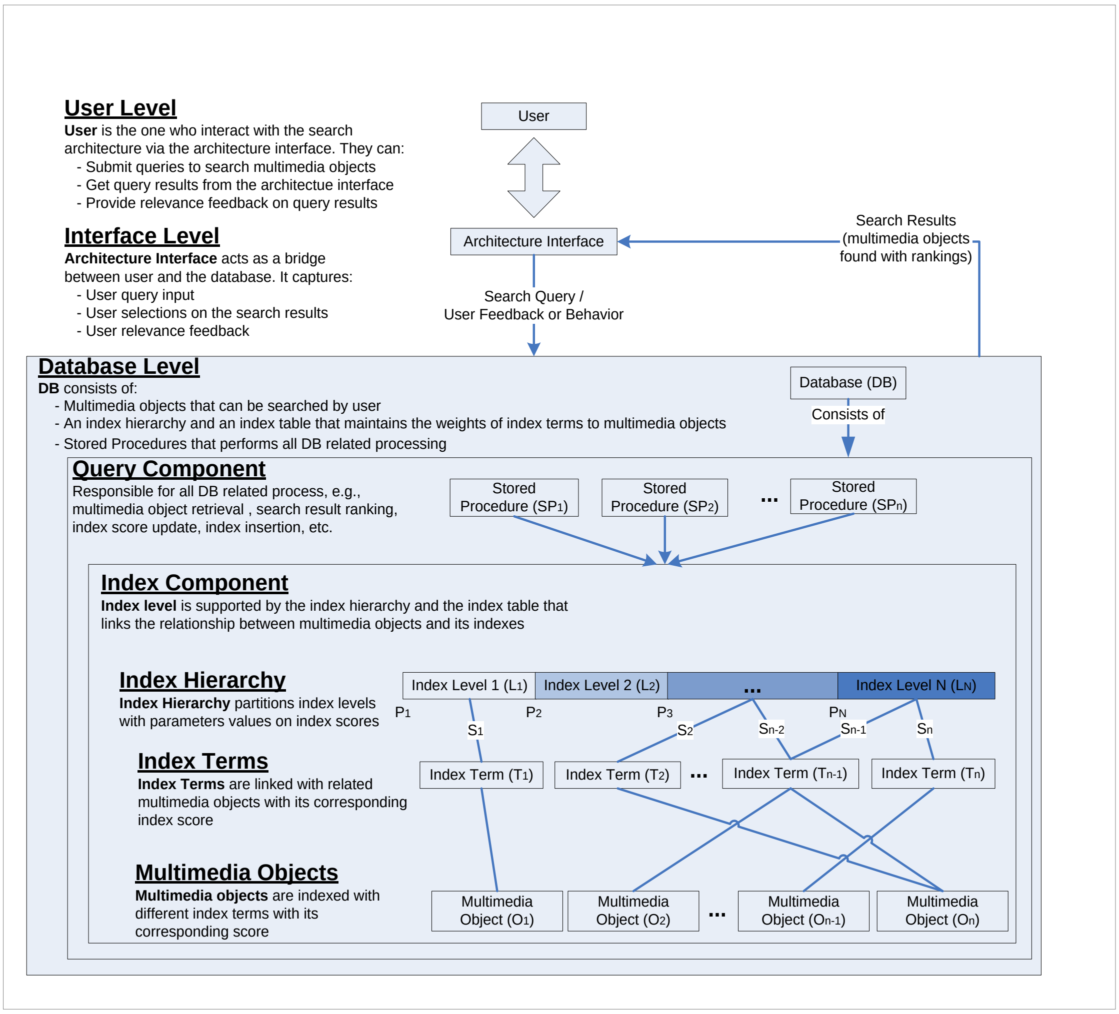

Fig. 1 A sample line graph using colors which contrast well both on screen and on a black-and-white hardcopy

A. The User and System Interface Levels

Here we use image retrieval as an example to describe how our multimedia information retrieval system works. The users are the ones who interact with the search system via the system interface. They can search image objects by submitting query with keywords. After the system has processed a query, they can receive search results from the system interface and proceed to provide feedback through the system interface. The system interface acts as a bridge between users and ultimately the system database. It captures user query input with search criteria, search result feedback and selection from the user, and employs the information to update the index (Fig. 1).

B. The System Database Level

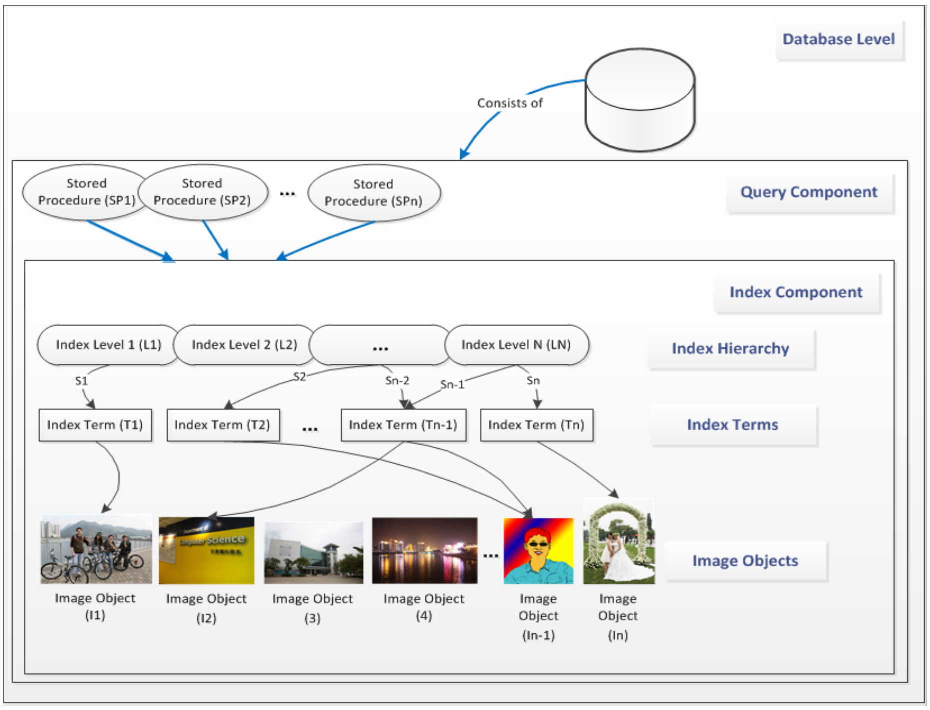

Fig. 2 The system database level

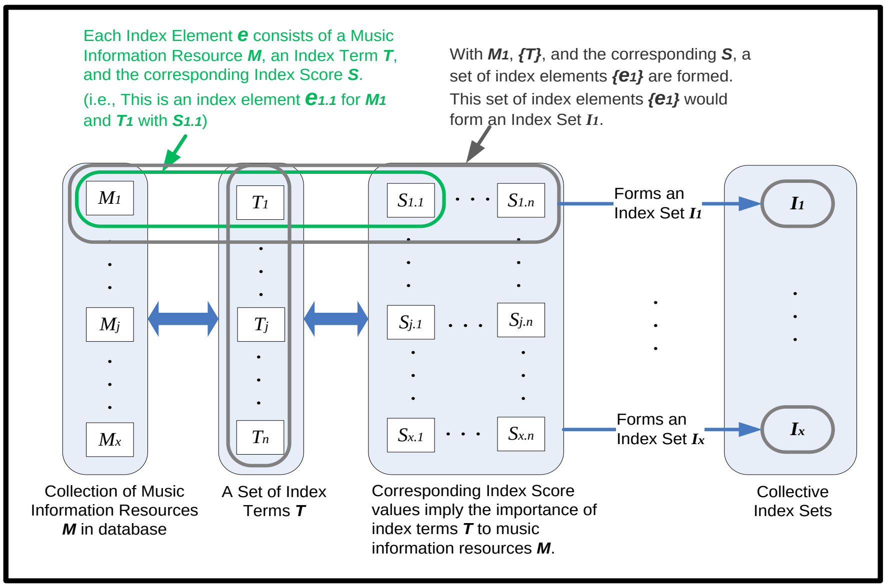

Fig. 3 Index Structure: Multimedia Information, Index & Index Element

The system database level is the core part of the system. It consists of a query level and an index level (Fig. 2). The query level consists of stored procedures that are responsible for all database related processes, such as index insertion, index score updating, object ranking and retrieval for the entire system. The index level is an index table organized as a hierarchy, which links multimedia objects and index terms with the corresponding index score. The index score indicates the importance of the index term that is related to the multimedia object. The following sections will focus on discussing the system database level.

We consider a set of multimedia objects $\{O_j\}$, where the characteristics or semantic contents of each object O_j cannot be extracted automatically. Typically, a multimedia object would be a song, represented in digital formats such as mp3, midi, wave, or an image. For every O_j, it links with an index set I_j that consists of a number of elements: $I_j = \{e_{j1}, e_{j2}, \ldots, e_{jMj}\}$. Each index element e is a triple, such that $e_{jk} = (t_{jk}, s_{jk}, o_j)$, where t_{jk} is an index term ID, s_{jk} is the score associated with t_{jk}, and O_j is the object ID. The higher the score s_{jk} is, the more important the index term t_{jk} is to the object O_j. The relationship of t_{jk}, s_{jk}, and O_j can be represented as:

IndexTerm (index_id, index_term)

ImageObject (object_id, object_name, object_description, …)

IndexTable (index_id, object_id, score, …)

Each index term is uniquely identified by the primary key, index_id. Similarly, each object is uniquely identified by the primary key, object_id. In the index table, each item is uniquely identified by the composite key, index_id and object_id (Fig. 3).

IV. INDEX EVOLUTION AND GROWTH

From the index score values, an index hierarchy may be established, which consists of the index sets of all the objects stored in the database (Figs. 4 and 5). By partitioning the value of score s_{jk}, it can be divided into N levels $L_1, L_2, \ldots, L_N$ using a set of parameters $P_1, P_2, \ldots, P_N$. For example, for $P_1 = 0$, $P_2 = 10$, $P_3 = 20$, and $P_4 = 30$, there would be four levels of the index set. The index score is directly affected by user search behaviour, such as result selection and relevance feedback. By the continuous use of the search system, user search behaviour can be collected and analyzed.

For a given index element ***E*** which consists of an index term $\boldsymbol{T_x}$ for a multimedia information resource $\boldsymbol{M_y}$ with index score value ***X***, the index level of this index element can be determined by the following algorithm:

If $P_i \leq X < P_{i+1}$, where $i = 1, 2, \ldots, N$-1.
Then, E would be placed in level Li of the index hierarchy.
Otherwise, E would be placed in level N if $P_N \leq X$.

Fig. 4 The structure of the index hierarchy. In this index hierarchy, the higher the level is, the more significant the associated index terms are.

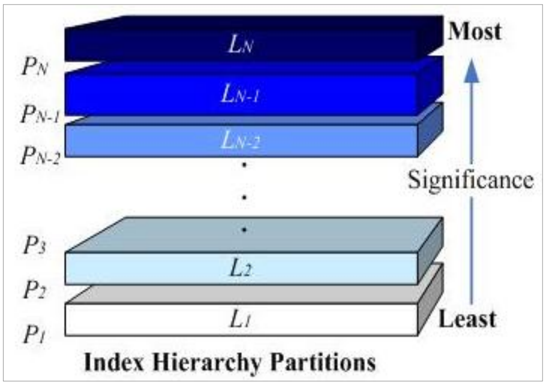

Fig. 5 Index hierarchy structure

Consider the situation of a user input search query $Q(T_1, T_2)$, and suppose N multimedia objects $O_1, O_2, \ldots, O_n$ are returned in the query result and ordered by the corresponding score $S_1, S_2, \ldots, S_n$ in descending order. The related index scores on T_1 and T_2 for the desired object O_x would be increased when the user selects O_x in the query result list, or when the user provides positive feedback on O_x. These two cases would increase the related index scores on T_1 and T_2 for the desired object O_x by a predefined value. In contrast, the related index scores on T_1 and T_2 for the desired object O_x would be decreased when the user provides negative feedback on O_x. Furthermore, the related index scores on T_1 and T_2 for all objects $O_1, O_2, \ldots, O_n$ in the query result would be decreased when the user does not select any object on the query result list. These two cases would decrease the related index scores by a predefined value.

In the index growth approach, consider an object K that is indexed with a term T_1. K can be searched by a user query which contains T_1 and many objects may be returned in the query result since many objects are indexed with T_1. Among these returned objects, a user can distinguish objects by adding another index term T_2 to K. Thus, the user can search the desired object by entering both index terms in the search query. Consider the searching of the picture “Starry Night”, painted by Vincent Willem van Gogh, and we assume that the image object is indexed with the term “Vincent Willem van Gogh” initially. Users can search this image by the term “Vincent Willem van Gogh”, but sometimes, some user query would be more specific, with both search terms “Vincent Willem van Gogh” and “Starry Night” used. The same object would be returned in the result when searching by the term “Vincent Willem van Gogh”, since the term “Starry Night” is not indexed yet. Eventually, the user would select the object “Starry Night” and this suggests that a new index term, “Starry Night”, may be useful for indexing this data object. Thus, the new index term would be included in the lowest level of the index hierarchy for this object. For every query that specifies both terms, “Vincent Willem van Gogh” and “Starry Night”, the user on selecting this image object will cause an increase in the score of the index terms for that object. Thus, the score of the index would be gradually increased and the new index term would be properly installed. Through progressive usage, the indexing of image objects would be enriched, and such attributes as “De sterrennacht”, “Saint-Remy”, “oil painting”, “Nuit etoilee” or “New York, The Museum of Modern Art” may be indexed for this image object (Fig. 6). The index evolution and growth of the system is very efficient.

To ensure that newly added multimedia objects have a reasonable chance of being discovered and processed, our system incorporates a degree of variation in the return of search results. After considerable usage, the high scored multimedia objects always rank near the top of the search results. Generally, users tend to be interested in the highly ranked objects. Consequently, the high scored objects always have greater chance of increasing the score while the chance for the newly added objects appearing in the result list would be reduced. Thus, the newly added objects would be ranked very low and nearly “hidden”. In order to optimize the search results, our search system would introduce small degrees of perturbations so as to allow constructive variations in the result. When considering a large number of objects returned by a query result, the users may not reach their desired object since the target object always rank very low and nearly hidden. By exploiting the random variations introduced through Genetic Algorithms, those “hidden” objects would have a chance of being promoted to a higher ranking position and discovered eventually. For example, for a returned page containing thirty relevant results, we make it contain twenty top-ranked objects and ten newly-added objects.

Our model enables users to indicate relevance explicitly using a binary vote relevance system. The binary vote relevance feedback indicates that a multimedia information resource is either relevant or irrelevant for a specific query. Once a user submits a query, our system will return a list of query results to the user. In the spirit of Web 2.0 [19-21] with collaborative user involvements, our system allows users to provide their relevance feedback for the multimedia information resources of the query results. Their feedback can be either positive or negative. When positive relevance feedback is received, the related index scores would increase. Similarly, the related index scores would decrease when negative relevance feedback is received.

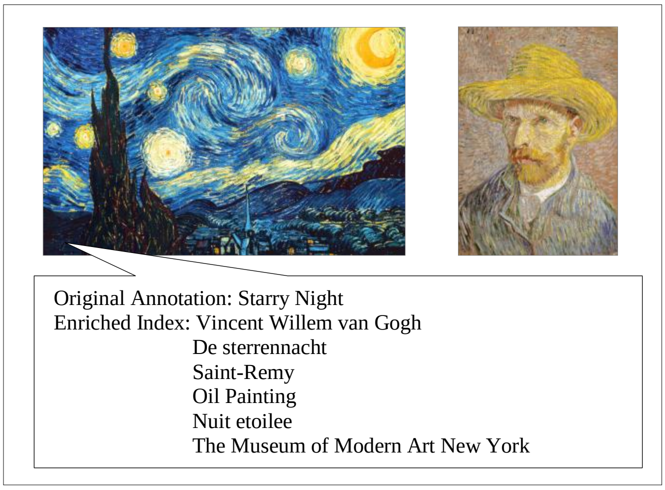

Fig. 6 Automatic image index enrichment

Although the idea of exploiting user’s feedback to rate relevance seems promising, it is not easy to convince a community of users to spend their time to explicitly rate resources. Therefore, our model also collects implicit relevance feedback from them. The implicit feedback is inferred from user behaviour and their history, such as noting which resource(s) that users do and do not select for viewing, and the duration of time spent in viewing a resource. All such information can be collected automatically and would reflect user satisfaction and expectation of the query result. When users click on a resource in the answer vector, we can infer that the selected resource may be relevant to the user query. Our system will treat it as a kind of positive feedback from the user implicitly. On the contrary, when users do not select any resource in the answer vector, we can infer that they may think that the resources in the answer vector are irrelevant to their input query or they are not interested in those resources. Our system will treat it as a kind of negative feedback from the user implicitly. The index score update approach is similar to the explicit user feedback; the related index score would increase when receiving a positive user feedback, and the related index score would decrease when receiving a negative user feedback.

V. EXPERIMENTAL RESULTS

We have carried out experiments to measure the effectiveness of our approach using general multimedia objects.

A. Index Growth

Initially, a hundred multimedia objects are stored in the database. Each object is initially indexed with one index term. The index table is grown through sixty queries, input by the user. Throughout the queries, the user add new index term(s) to the query results. Fig. 7 shows the relationship between the number of queries and the number of index terms. As the number of queries increases, the number of index terms also increases as expected. At the initial state, the number of index terms grows faster since there is more scope for users to add new ones. The number of index terms keeps on increasing while the scope for user to add new ones is reduced. Thus, it grows slower and the number of index terms tends towards an equilibrium level. Fig. 8 shows the index growth rate. We observe that the index growth rate is the highest at the first ten queries, and then, it drops rapidly. As the number of queries increases, it drops continuously and more gradually. The drop slows down until it reaches an equilibrium level.

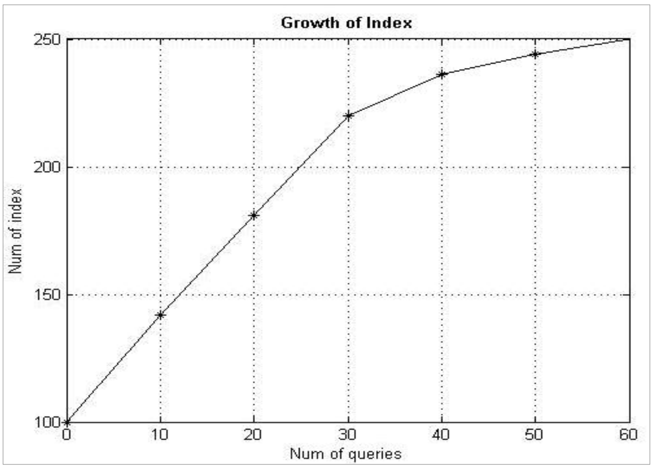

Fig. 7 Growth of index

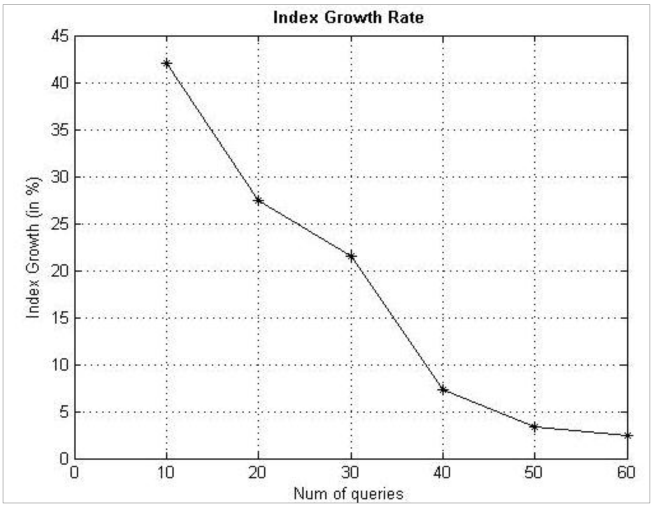

Fig. 8 Index growth rate

B. The Effect of Indexing

In this experiment, a total of thirty queries are used. These queries are performed before indexing with fifteen queries performed before usage, and an identical set of queries performed after the usage. Initially, all the multimedia objects are minimally indexed with one index term. As users interact with the system, the index table grows. The multimedia data objects become not only searchable with its initial index term, it is also searchable by other index term(s) that are added by users. The new index terms that are meaningful will gradually be promoted to the most significant level of the index hierarchy.

Fig. 9 shows the recall rate. In the first 5 queries of our standard query set, we use the initial index terms for searching the desired multimedia object. From the experimental results of these queries, all the targeted multimedia objects can be found. By comparing performance before and after indexing through user queries, the targeted objects are ranked significantly higher (Fig. 10). Then, we use new index terms as the input queries for the remaining queries in the query set. Although none of these can be found with the original index terms, they can be found after collaborative indexing through sixty queries. Assuming a cutoff rank of ten, the improvement in search result rankings after indexing is shown in Fig. 11.

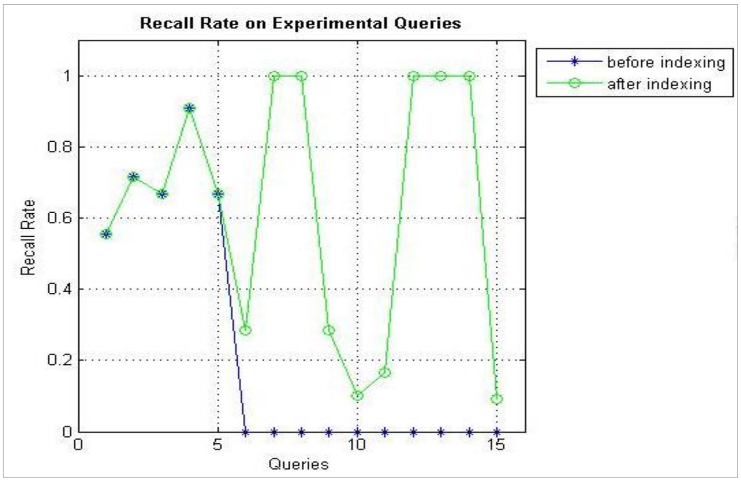

Fig. 9 Recall rate on the experimental queries

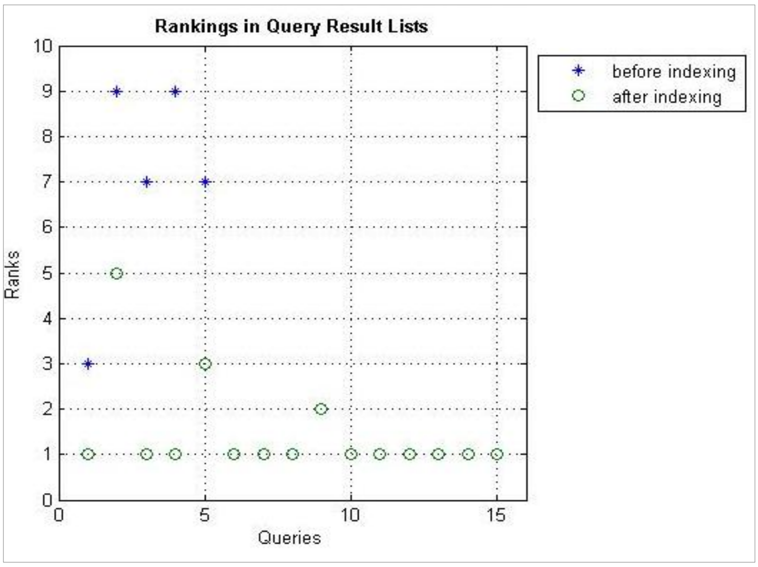

Fig. 10 Rankings in query result lists

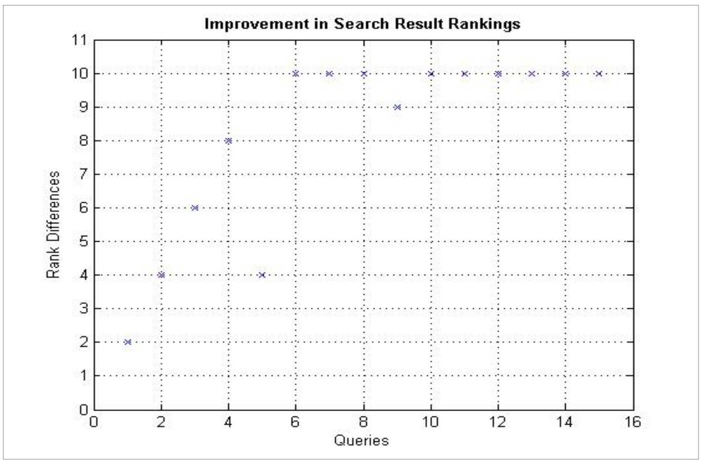

Fig. 11 Improvement in search results based on cutoff at 10

We thus see that the effect of collaborative indexing is quite pronounced and it consistently produces significant improvement in search performance. While it takes time for the index to evolve to an optimal performance level, search benefits will start to accrue as soon as the index starts to grow, and search efficiency tends to grow fastest at the initial stages of usage.

VI. CONCLUSIONS

Digital multimedia has been increasing at an accelerated pace, and a mismatch of creation and indexing rates necessitates a radically new approach to solve the multimedia object retrieval problem. Individual indexing of multimedia objects is labour intensive and has a number of disadvantages. In the context of ROC analysis [5], these are (i) false positives – i.e. wrongly adding an index term through personal subjectivity which is not shared by the wider community, and (ii) false negatives – i.e. the omission of relevant properties due to the inability to cover and capture all the nuances and deep semantics of the multimedia.

Our system is able to overcome (i) by having a resilient, community validated structure which allows personal subjectivity to be filtered off through a robust scoring system, and (ii) by exploiting collective assessment and perception of multimedia objects through continuous usage by the community. By capturing, analyzing and interpreting user response and query behaviour, the patterns of searching and finding multimedia objects may be established. Within the present paradigm, the semantic index may be dynamically constructed, validated, and built-up, and the performance of the system will tend to increase as time progresses. Our system also incorporates a high degree of robustness and fault-tolerance whereby inappropriate index terms will be gradually eliminated from the index, while appropriate ones will be reinforced. By incorporating genetic variations into the design, our system will allow multimedia objects which may otherwise be hidden to be discovered

ACKNOWLEDGMENT

This research is supported by HKBU Research Grant FRG2/11-12/161.

REFERENCES

[1] Crawford, T.: "Music Information Retrieval and the Future of Musicology", Technical Report for the Online Chopin variorum Edition project, (2005).

[2] Bischoff, K., Firan, C. S., Nejdl, W., and Paiu, R.: "Can All Tags be Used for Search?" In CIKM '08: Proceedings of the 17th ACM Conference on Information and Knowledge Management, pages 193-202, New York, NY, USA, ACM, (2008).

[3] Dannenberg, R. B., Birmingham, W. P., Pardo, B., Hu, N., Meek, C., and Tzanetakis, G.: "A Comparative Evaluation of Search Techniques for Query-By-Humming using the Musart Testbed", Journal of the American Society for Information Science and Technology, 58(5): 687-701, 2007.

[4] Downie, J. S.: "Music Information Retrieval", Annual Review of Information Science and Technology, 37:295-340, 2003.

[5] Fawcett, T.: "An Introduction to ROC Analysis", Pattern Recognition Letters, 27: 861-874, 2006.

[6] Fingerhut, M.: "Music Information Retrieval, or How to Search for (and maybe find) Music and Do Away with Incipits", In Proceedings of the IAML-IASA Congress, (2004).

[7] Go'mez, J. and Vicedo, J. L.: "Next-generation Multimedia Database Retrieval", IEEE Multimedia, 14(3):106-107, 2007.

[8] Goth, G.: "Multimedia Search: Ready or Not?", Distributed Systems Online, IEEE, 5(7). 2004.

[9] Knees, P. et al.: "A Music Search Engine Built Upon Audio-based and Web-based Similarity Measures," In Proceedings of the ACM SIGIR 2007, Amsterdam (2007).

[10] Musicpedia: The Open Music Encyclopedia. http://www.musicpedia.org

[11] Lesaffre, M.: "Music Information Retrieval: Conceptual Framework, Annotation and User Behaviour", PhD thesis, Faculty of Arts and Philosophy, Department of Art, Music and Theatre Sciences, Ghent University, (2006).

[12] Orio, N.: "Music Retrieval: A Tutorial and Review", Foundations and Trends in Information Retrieval, 1(1):1-90. Now Publishers Inc. (2006).

[13] Snoek, C. G. M., Worring, M., Gemert, J. C. van, Beusebroek, J. M., and Smeulders, A. W. M.: "The Challenge Problem for Automated Detection of 101 Semantic Concepts in Multimedia", In Multimedia '06: Proceedings of the 14th annual ACM international conference on Multimedia, pages 421-430, New York, NY, USA, ACM, (2006).

[14] Yang, B. and Hurson A. R.: "Ad Hoc Image Retrieval using Hierarchical Semantic-based Index", In AINA '05: Proceedings of the 19th international conference on advanced information networking and applications, pages 629-634, Washington, DC, USA, IEEE Computer Society, (2005).

[15] Yang, Y.H. et al.: "Personalized Music Emotion Recognition," In Proceedings of the SIGIR 2009, Boston, Massachusetts (2009).

[16] Vaughan, T.: "What is Multimedia?", Multimedia: Making It Work (7th edition), McGraw-Hill Professional, pp.1-17, 2006.

[17] Sebe, N. and Tian, Q.: "Personalized multimedia retrieval: the new trend?", In Proceedings of the international Workshop on Workshop on Multimedia information Retrieval (Augsburg, Bavaria, Germany, September 24 - 29, 2007). MIR '07. ACM, New York, NY, 299-306 (2007).

[18] Futrelle, J. and Downie, S.: "Interdisciplinary Communities and Research Issues in Music Information Retrieval", In Proceedings of

ISMIR 2002, Paris (2002).

[19] Boll, S.: “Multitube - where web 2.0 and multimedia could meet”, IEEE Multimedia, 14(1), 9-13, 2007.

[20] Giannakidou, E., Kompatsiaris, I., Vakali, A.: “Semsoc: Semantic, social and content-based clustering in multimedia collaborative tagging systems”, In: ICSC 2008: Proceedings of the 2008 IEEE International Conference on Semantic Computing, Washington, DC, USA, pp. 128–135. IEEE Computer Society, Los Alamitos (2008).

[21] Yesilada, Y., Harper, S.: “Web 2.0 and the semantic web: hindrance or opportunity?”, In Proceedings of W4a - International Cross-disciplinary Conference on Web Accessibility 2007. SIGACCESS Access. Comput. (90), 19-31, 2008.

[22] Wikipedia: http://en.wikipedia.org/wiki/A_picture_is_worth_a_thousand_words

[23] Ritendra Datta, Dhiraj Joshi, Jia Li and James Wang, “Image Retrieval: Ideas, Influences, and Trends of the New Age”, Proceedings of the 7th ACM SIGMM international workshop on Multimedia information retrieval, November 10-11, 2005, Hilton, Singapore.

[24] C. Carson, S. Belongie, H. Greenspan, and J. Malik, “Blobworld: Image Segmentation Using Expectation-Maximization and Its Application to Image Querying,” in IEEE Trans. On PAMI, vol. 24, iss. 8, pp. 1026-1038, 2002.

[25] Y. Chen and J. Z. Wang, “A Region-Based Fuzzy Feature Matching Approach to Content-Based Image Retrieval,” in IEEE Trans. on PAMI, vol. 24, iss. 9, pp. 1252-1267, 2002.

[26] A. Natsev, R. Rastogi, and K. Shim, “WALRUS: A Similarity Retrieval Algorithm for Image Databases,” in Proc. ACM SIGMOD Int. Conf. Management of Data, pp. 395-406, 1999.

[27] J. Li, J.Z. Wang, and G. Wiederhold, “IRM: Integrated Region Matching for Image Retrieval,” in Proc. of the 8th ACM Int. Conf. on Multimedia, pp. 147-156, Oct. 2000.

[28] V. Mezaris, I. Kompatsiaris, and M. G. Strintzis, “Region-based Image Retrieval Using an Object Ontology and Relevance Feedback,” in Eurasip Journal on Applied Signal Processing, vol. 2004, iss. 6, pp. 886-901, 2004.

[29] W.Y. Ma and B.S. Manjunath, “NETRA: A Toolbox for Navigating Large Image Databases,” in Proc. IEEE Int. Conf. on Image Processing, vol. I, Santa Barbara, CA, pp. 568-571, Oct. 1997.

[30] W. Niblack et al., “The QBIC Project: Querying Images by Content Using Color, Texture, and Shape,” in Proc. SPIE, vol. 1908, San Jose, CA, pp. 173-187, Feb. 1993.

[31] P.S. Hiremath and J. Pujari, “Content Based Image Retrieval Using Color, Texture and Shape Features,” In proceeding of Advanced Computing and Communications, 2007. ADCOM 2007.