E-CATALOG IMAGE METADATA MINING SYSTEM USING USER USAGE PATTERNS FOR E-BUSINESS INTELLIGENCE

Seong-Yong Hong, Savannah State University, hongs@savstate.edu
Hae-Yeon Choi, Savannah State University, choih@savstate.edu
Hye-Jin Jin, Korea University, hjjin2003@korea.ac.kr

ABSTRACT

Recently, the web sites such as e-business sites and shopping mall sites deal with a lot of image information. To find a specific image from these image sources, we usually use web search engines or image database engines which rely on keyword only retrievals or color based retrievals with limited search capabilities. This paper presents an intelligent web e-catalog image retrieval system using metadata and user log. We propose the system architecture, the texture and color based image classification and indexing techniques, and representation schemes of user usage patterns. The query can be given by providing keywords, by selecting one or more sample texture patterns, by assigning color values within positional color blocks, or by combining some or all of these factors. The system keeps track of user's preferences by generating user query logs and automatically adds more search information to subsequent user queries. To demonstrate the usefulness of the proposed system, some experimental results showing recall and precision are also explained.

Keywords: Data Mining**,** Image Retrieval System, E-Business Intelligence, Image Metadata Mining.

INTRODUCTION AND BACKGROUND

Data and Metadata Mining is becoming a mainstream technology used in business intelligence applications supporting industries such as financial services, retail, healthcare, telecommunications, and higher education, and functions of business such as marketing, manufacturing, customer experiences, customer service, and sales [1]. Many of the business problems that data mining can solve cut across industries such as customer retention and acquisition, cross-sell, and response modeling. In recent years, however, data mining products have simplified data mining considerably by automating the process-making the fruits of the technology more widely accessible. New algorithms and heuristics have been evolved to provide good results with little or no experimentation or data preparation. In addition, the availability of data mining has increased with in-database data mining capabilities [2, 3, 4].

More recently, the concept of e-business intelligence is being expanded to include data mining techniques to extract knowledge and generated predictions for business problems, thereby enabling companies to make better use of costly cooperate asset, their organizational data.

Data Mining Techniques

Data mining is the process of finding patterns and relationships in data. At its core, data mining consists of developing a model, which is typically a compact representation of patterns found using historical data, and applying that model to new data [5]. Data Mining offers different techniques that can be used depending on the type of problem to be solved. For example, a marketing manager who wants to find out which customers are likely to buy a new product can use the classification function. Similarly a supermarket manager who wants to determine which products to put next to milk and eggs, or what coupons to issue to a given customer to promote the purchase of related items can use the association function [6, 7].

Java technology, specifically as leveraged within the scalable J2EE architecture, facilitates integration with existing application such as B2B and B2C web sites, customer care centers, campaign management, as well as new applications supporting national security, fraud detection, bioinformatics and life sciences. JDM (Java Data Mining) allows users to draw on the strengths of multiple data mining vendors for solving business problems, by applying the most appropriate algorithm implementations to a given problem without having to invest resources in learning each vendor's proprietary API [8].

Image Retrieval System

With the rapid development of the Internet technology, the number of internet users and the amount of multimedia information on the Internet is ever increasing. Recently, the web sites such as e-business sites and shopping mall sites deal with lots of image information. To find a specific image from these image sources, we usually use image database engines or web

search engines. But, the feature based retrieval capabilities of these systems are quite limited, especially for the web images.

Many image retrieval systems have been developed, such as QBIC [9], Safe, VisualSEEK, Photobook, WBIIS, SIMPLIcity, Blobworld [10,11,12,13]. Some systems rely on keyword only retrievals and others support image content based retrievals. In the latter approach, they support image retrievals based on the image feature information, such as average colors, color histograms [17,18,19,20], texture patterns [15], and shape objects [13]. But, most of them are developed for image database applications [14,21,22].

This paper is an effort to develop an intelligent web image retrieval system. The approach of this paper is a little bit different. We are currently developing the agent-based image search engine which supports the content-based retrieval on web images. We are trying to support various access paths on web images with customized feedback according to internet user's preferences. To support this, we have applied both data mining [16] and web mining techniques [2].

This system supports hybrid image retrievals based on query keywords, colors, and textures. For a given web image, our system classifies it whether it is textured or non-textured and, for the textured image, assigns the appropriate texture pattern(s) to the image. The proposed system also generates positional color information. The query can be given by providing keywords, by selecting one or more sample texture patterns, by assigning colors within positional color blocks, or by combining some or all of these factors.

The proposed system is intelligent since it remembers user's preferences and adds more feature information to the given query. For example, if the user provides search keyword “shirts”, the system automatically adds texture patterns (e.g., check pattern) and color information according to user's previous preferences. As a result, the system can keeps track of user's query logs and applies data mining techniques to determine user usage patterns on colors and textures.

CONCLUSION

In this paper, we presented an intelligent e-catalog image retrieval system IM2S. We have proposed the texture and color based image indexing techniques. To support fast retrieval, we utilized the bit vector indexing for textures representing the presence of each texture pattern by 1 bit. For positional color search, we divided given images into region blocks and stored the average color of each block in system databases. The query can be given by providing keywords, by selecting one or more sample texture patterns, by assigning colors within positional color blocks, or by combining some or all of these factors. The system is designed to remember user's preferences by mining user query patterns, so that it can add more feature information automatically to subsequent user queries.

We are planning to develop an integrated e-catalog image search engine by combining the image crawling agent and automatic image caption extractor module. There should be further researches to improve the performance of the feature extractor modules and the matching modules on massive amount of web images.

REFERENCES

1. M.S. Chen, J. Han, and P.S. Yu. (1996). Data Mining: An Overview from a Database Perspective, *IEEE Trans. Knowledge and Data Engineering*, pp. 8:866-883.
2. O. Zaiane and J. Han. (1998). WebML: Querying the World-Wide Web for Resources and Knowledge,' *in Proc. Int'l Workshop on Web Information and Data Management (WIDM'98)*, pp. 9-12.
3. J. Han and Y. Fu. (1994). Dynamic Generation and Refinement of Concept Hierarchies for Knowledge Discovery in Database, *in Proc. AAAI'94 Workshop on Knowledge in Databases (KDD'94)*, Seattle, WA, pp.157-168.
4. R. Agrawal and R. Srikant. (1995). Mining Sequential Patterns, *in Proc. of the Int'l Conference on Data Engineering (ICDE)*, Taipei, Taiwan, pp.3-14.
5. R. Agrawal and R. Srikant. (1994). Fast Algorithms for Mining Association Rules, *in Proc. VLDB, Santiago, Chile*, pp.487-499.
6. D.-H. Lee, D.-Y. Seo, N.-H. Kim, J.-Y. Lee. (1998). Discovery and Application of User Access Patterns in the World Wide Web, *in Proc of the 4th World Congress on Expert Systems*, pp.321-327.
7. R. Agrawal, T. Imielinski, and A. Swami. (1993). Mining Association Rules Between Sets of Items in Large Databases. *In Proceedings of the 12th ACM SIGMOD International Conference on Management of Data*, pp.207-216.
8. Mark Hornick. (2006). Java™ Specification Request 247: Java™ Data Mining(JDM)2.0, *JSR-247 Expert Group*. Available: http://www.jcp.org/en/jsr/detail?id=247
9. Niblack, W., Barber, R., Equitz, W., Flickner, M., Glasman, E., Petkovic, D., Yanker, P, & Faloutsos, C. (1993). The QBIC Project: Querying images by content using color, texture, and shape, *in Proc. SPIE Storage and Retrieval for Image and Video Database*, pp.173-187. Available: http://Iibra.ucdavis.edu/cgi-bin/QbicStable/
10. V. N. Gudivada and V. V. Raghavan. (1995). Content-Based Image Retrieval Systems, *IEEE*

Computer, 28(9), pp.18-22.

11. C. Faloutsos, R. Barber, M. Flickner, J. Hafner, W. Niblack, D. Petkovic, and W. Equiz. (1994). Efficient and Effective Querying by Image Content, *Journal of Intelligent Information System(JIIS), 3(3)*, pp.231-262.
12. Smith, J. R. & Chang, S.-F. (1996). VisualSEEk: A Fully Automated Content-Based Image Query System, *ACM Multimedia 96*, pp. 87-98.
13. Li, Jia., Wang, James Z., & Wiederhold, Gio. (2000). IRM: Integrated region matching for image retrieval, *Proc. ACM Multimedia*, Los Angeles, ACM, pp. 147-156.
14. Jia. Wang, et.al. (1997). Color Clustering Techniques for Color Content-B Image Retrieval from Image Database, *in Proc. of the International Conference on Multimedia Computing and Systems*, pp. 442-449.
15. J.Li, J.Z. Wang, G. Wiederhold. (2000). Classification of texture and non-textured images using region segmentation, *in Proc. of the Seventh International Conference on image Processing, Vancouver, BC, Canada*, pp. 754-757.
16. J. Han, Y. Huang, N. Cercone and Y. Fu. (1996). Intelligent Query Answering by Knowledge Discovery Techniques, *IEEE Transactions on Knowledge and Data Engineering, 8(3)*, pp.373-390.
17. Stricker, M. & Orengo, M. (1995). Similarity of color images, *Proc. SPIE on Storage and Retrieval for Image and Video, Databases, Vol. 2420*, San Jose, USA, pp. 381-392.
18. Swain, M. J. & Ballard, S. H. (1991). Color Indexing, *Int. Journal of Computer Vision, Vol.7, No. 1*, pp. 11-32.
19. Rickman, R. & Stonham, J. (1996). Content-based image retrieval using color tuple histograms, *SPIE proceedings*, 2670: 2-7.
20. Smith, J. & Chang, S.-F. (1996). Tools and techniques for color image retrieval, *SPIE proceedings*, 2670: 1630-1639.
21. Ogle, V.E. and Stonebraker, M., (1995). Chabot: Retrieval from a Relational Database of Images, *IEEE Computer*, Sept, pp.40-48.
22. Smith, J. R. & Chang, S.-F. (1995). Single Color extraction and image query, *in Proc. ICIP, vol. 3*, pp. 528-531.