An Efficient Content Based Image Retrieval System Using the Mesh-of-Trees Architecture*

WEI-MIN JENG AND JEN-HAO HSIAO
Department of Computer Information Science
Soochow University
Taipei, 100 Taiwan

With the fast development of the Internet and computer technology in recent years, the use of visual information on the World Wide Web has become more popular. There has been a trend in retrieving images on the basis of automatically-derived features such as color and shape. Conducting the image retrieval by query image has attracted plenty of attention from researchers in the field of information technology. The CBIR (Content Based Image Retrieval) technique refers to the image retrieval operation based on the key primitives. Despite the fact that the CBIR technique has been accepted and implemented in many application domains, extracting images of their contents is really an arduous job. The difficulty is determining how to organize these image features in some fashion without consuming too much time. In this paper, a novel MOT (Mesh Of Trees) based CBIR system is proposed. It makes use of dual image signatures to more efficiently provide better query results. Images are preprocessed and indexed by means of both color histogram and wavelet image signatures, and "similar" images are adjacently stored in a two-dimensional MOT. With superb traversing capabilities in rows and column trees, the architecture is very effective in accommodating dual image features. For higher dimensional image feature vectors and massive image databases, the experimental results show that the proposed architecture achieves its superior performance figures in comparisons of other existing systems.

Keywords: CBIR, mesh of tress, wavelet transform, color histogram, speedup

1. INTRODUCTION

Traditional information retrieval techniques, such as the keyword-based information retrieval method, have relied largely on text descriptors, which are often subjective. On the other hand, CBIR collects the most important features of images in order to characterize the contents of stored and query images. The CBIR technique [1-3] has the advantage over manual indexing of being inherently objective in the retrieval operation. Users may search an image database for desired images having similar attributes with respect to a query image. Traditional image indexing methods ignore the abundant information of an image, which can make it difficult to characterize text descriptors. Recognition of the image retrieval problem has led to increased interest in techniques now generally referred as CBIR. Researchers have devoted themselves to these issues and proposed methods to

Received April 27, 2004; revised September 1 & December 27, 2004; accepted January 19, 2005.
Communicated by Ming-Syan Chen.
* The preliminary version of this paper was presented at the 6th World Multiconference on Systemics, Cybernetics and Informatics, Orlando, Florida, July 14-18, 2002.

try to effectively manage the enormous amount of image data. Because the volume of multimedia data is often large, content-based analysis is an expensive process in terms of both time and storage. Therefore, it is impractical to compare the entire image contents while performing the retrieval operation. The major challenge for a CBIR system is to determine a set of attributers or features that can best describe the contents of an image and enable efficient similarity measurement.

CBIR is the name of a technique that has been widely used to describe the process of retrieving desired images from a large collection on the basis of features (such as color, texture, and shape) that can be automatically extracted from the images themselves. First generation CBIR systems allowed access to images through string attributes. Each image was indexed through manually annotation. However, the first generation CBIR systems suffered from two drawbacks. First, it was difficult to use subjective text descriptions to capture visual features. Second, hand-annotating images took a lot of time. The new generation of CBIR systems supports full retrieval by means of visual contents. Instead of employing annotation, it extracts an image's contents, such as its color, shape, or texture, as keys. Commercial systems like the IBM QBIC System [4] query images based on raw image data. This allows queries to use multiple features based on image contents. Other approaches have also been proposed. Examples are VisualSEEk [5] and WebSEEk [6], both of which are developed at Columbia University.

After more than a decade of intense research, various schemes have been proposed, and the above-mentioned commercial products are available. However, the technology is still not mature because of the underlying complexity. Due to the fact that the image repository is potentially huge and varies in scope, obtaining results within a reasonable response time is a challenging task when designing a successful CBIR system. The key issue in building a CBIR system, therefore, is organizing the image features appropriately in order to perform similarity search and analysis efficiently. We propose a new CBIR system that uses the MOT [7, 8] architecture to store image information for efficient indexing. With superb traversing capabilities in rows and column trees, the architecture is very effective in accommodating dual image features. For higher dimensional image feature vectors and massive image databases, the proposed architecture also achieves improved scalability and performance.

Section 2 briefly introduces dual image signatures and their implementation. Section 3 describes the strategy and describes the architecture of our model and our approach to performance analysis. Section 4 presents experimental results obtained from actual runs of the proposed MOT-based system. Section 5 summarizes the results and discusses possible future works.

2. DUAL IMAGE SIGNATURES

Comparing images using their entire contents proves to be very time-consuming and requires significant amounts of storage space. According to the interactive nature of today's Internet applications, this approach proves to be infeasible for most of the CBIR systems. Intuitively, retrieving images with reduced amount of information becomes the solution for its better response time characteristic. In order to achieve better query results, finding an approximation with much less computation is usually conducted by extracting "fingerprints" of images into feature vectors. The combination of feature vectors is what we called image signature.

It is tempting for CBIR system designers to select many low-level features for the purpose of good representation and retrieval of images. Oftentimes, such high-dimensional image signature approaches suffer from the problem of excessive computational time in both indexing and retrieval operations. It seems that by using a single descriptor, it is hard to capture the contents of an image precisely. Both wavelet transformation and color histograms are utilized to form the dual image signature used in the retrieval process. In this paper, we focus on a combined color and wavelet based similarity measure for efficient and effective operations.

2.1 Wavelet Image Signature

The wavelet-based approach has attracted much attention recently, and it has been proven that it is really effective in extracting accurate shape/texture information from an image [9, 10] in a short amount of time. For example, the centroid vector of all the three-component Haar wavelet moment vectors can be used to identify texture features [11] in image retrieval. Moreover, since transformations are also used to compress image data, indexing based on compressed data increases both the storage efficiency and performance of the multimedia system. Jacobs [9] made significant use of wavelet coefficients to perform multi-resolution wavelet decomposition of images. WBIIS [12] applies a Daubechies' wavelet transform to each of the three color components. WALRUS [1, 13] employs a similarity model in which each image is first decomposed into its regions, and the similarity measure between a pair of images is then defined as the fraction of the area of the two images covered by matching regions from the images. A statistical approach to the texture retrieval problem was proposed [14] which adaptively varies two parameters of the generalized wavelet sub-band coefficient histogram Gaussian density. To expedite the wavelet based similarity analysis process, techniques [10, 15] have been implemented to improve the overall performance.

The purpose of getting a signature is to produce a "fingerprint" of an image. After image signatures are produced, a similarity metric can help us estimate the difference between the images. The basic idea behind this approach is that the energy distribution in the frequency domain identifies shape and texture. The energy magnitudes of the coefficients resulting from the computationally fast Haar wavelet transform are computed, and a fixed number of significant ones (say k) are selected according to the magnitudes of their absolute values. The quantization process takes the top k (60 in our case) and converts all the positive numbers into 1 and negative values into -1 to form a 2D matrix composed of 1s, -1s, and 0s. The quantized "fingerprint" of a reference image has to be estimated at the time of tree construction in order to compute the Euclidean distance between the database image and the reference image. The actual value of k does not affect the system performance because only the "fingerprint differences" are stored in the image signature used to measure the similarities. The image on the right in Fig. 1 is an example of a displaying wavelet image signature resulting from the left image where dots represent energies.

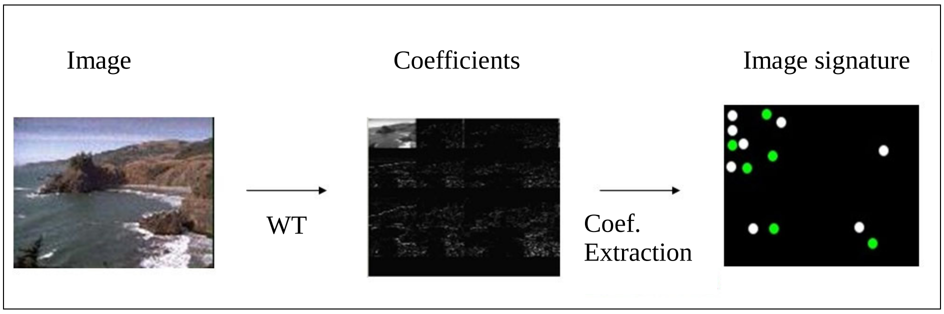


Fig. 1. The procedure of producing wavelet signature.

2.2 Color Image Signature

It is well-known that the wavelet-based image signature fails to deal with the rotation and shifting of images. The color histogram algorithm, on the contrary, is robust in handling such variances. We wish to emphasize the importance of incorporating at least two image signatures to improve the effectiveness without sacrificing the system efficiency when using the proposed MOT-based system. Traditionally, the color histogram has been used to describe the low-level color properties of images. It can be represented as three independent color distributions in each primary color. Though it doesn't include spatial information, it records the distribution of colors in an image. In addition, color histograms are fairly insensitive to variations caused by rotation or scaling. According to Swain [16], good results have been obtained using 64 bins and the histogram intersection approach. Fig. 2 shows an example of a color histogram on the right side in the blue component for the input image. The similarity between the reference image and the database image is determined based on the Euclidean distance of the corresponding color histogram elements. We perform this operation of calculating color histogram vectors and store only the differences as signatures in the pre-processing phase of database construction.

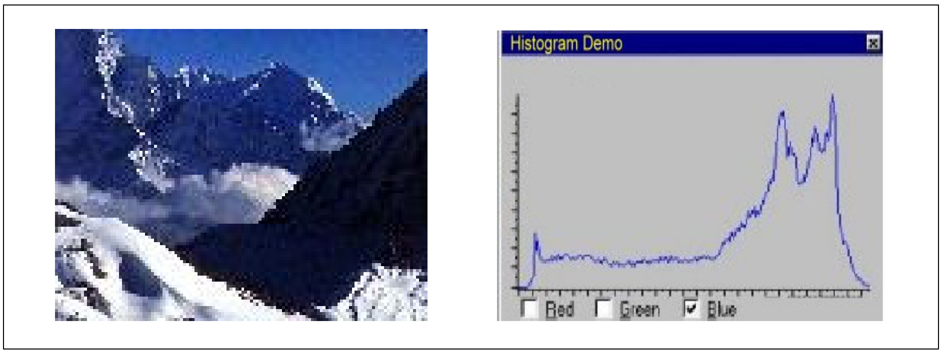


Fig. 2. Color histogram of a sample image.

Even though we agree that human perception is subjective, we can still generalize some principle as to what kind of images people would expect. If the input image is made up of a few objects and a simple background, it is highly probable that the user will want to find an object similar to the query image in the database. At this point, the query mode can be set to shape match. An image with complex objects, such as a scenic photo or abstract painting, may be queried. It may be concluded that users will care more about the whole image than about part of it. In this case, color match is a better choice. Hybrid match is the default mode and provides another choice. Furthermore, users can adjust the weights themselves, depending on which part of the image they care about most.

3. MOT-BASED ARCHITECTURE

Oftentimes, direct search of matched images is impractical owing to the fact that the size of the image database. Different mechanisms [17, 18] have been explored in the context of data structure models, where ordinary tree searching strategies are applied. In light of the need for multiple image features to achieve better image retrieval, we have developed a novel way to manage the abundant image information by using the MOT architecture.

3.1 Organization of the MOT-Based Architecture

A two-dimensional $N \times N$ mesh of trees consists of an $N \times N$ grid of nodes obtained by adding nodes and wires to form a complete binary tree on each row and each column [19]. The trees built on rows and columns are called row trees and column trees, respectively. We organize the representative characteristics of images in the mesh of trees as follows. First, we compute both wavelet and color image signatures of images as explained in section 2. The leaf nodes are ordered according to the magnitudes of the Euclidean similarity distances of the color and wavelet signatures for the row and column trees, respectively. The internal nodes of the trees are formed so as to contain the range information of the underlying sub-trees. This organization facilitates image retrieval by making it possible to conduct binary tree searches to narrow down the results. Images that have the similar shape or color attributes can be found in the same/nearby row/column trees based on minor differences in the signatures. Similarly, two images with very different contents will stay far apart in the proposed trees due to the drastically different image characteristics. Fig. 3 shows the architecture. The leaf node in Fig. 3 (a) represents the database image and stores the related image information. In Fig. 3 (b), row-trees R1, R2, R3, and R4 each store images with similar color distributions. Fig. 3 (c) indicates that the C1, C2, C3, and C4 column trees each store images with similar shape and texture. All rows and column trees share the same leaf nodes; therefore, all the trees in the mesh are connected, and topologically nearby nodes naturally have similar contents.

When a user begins a query, the number of rows/columns of retrieved images can be precisely determined according to the signature range specified. Depending on the circumstances of the query, various forms of searches can be arranged to accommodate the image query. For example, shape similarity search is the better choice if the target image

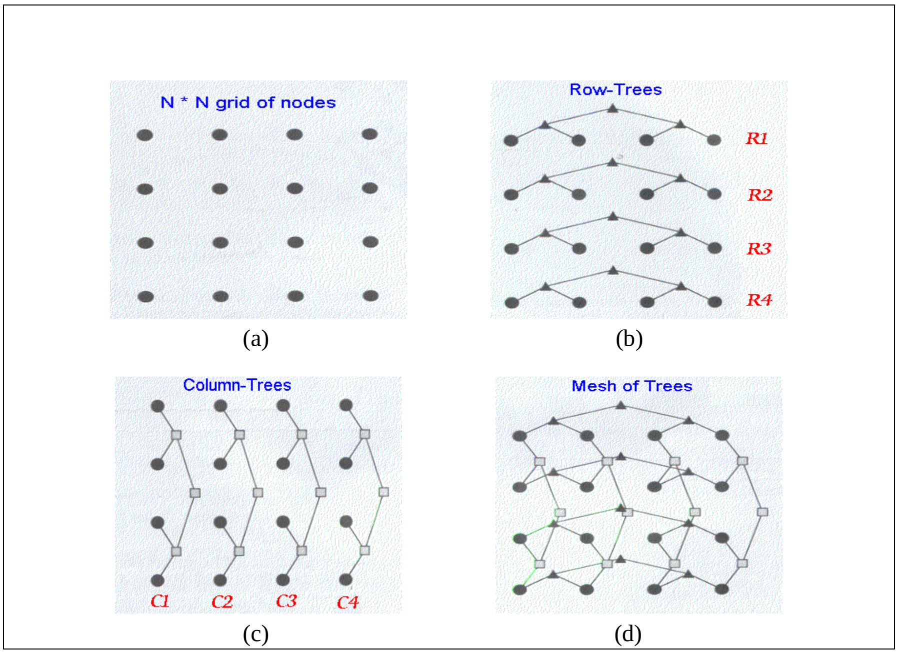


Fig. 3. The two-dimensional mesh-of-trees.

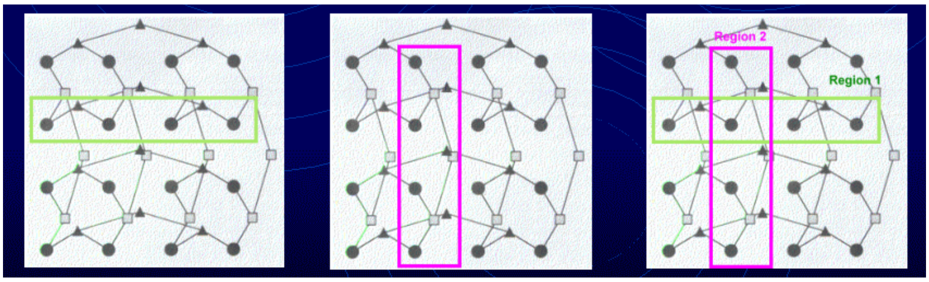


Fig. 4. Color-only, shape-only, and hybrid searching strategies.

is made up of a few objects and a simple background, while hybrid search combines the advantages of the two different techniques. The retrieval process can be completed in logarithmic time because of the nature of complete binary trees that we utilize in the design. Fig. 4 shows the three basic searching strategies.

3.2 Intersections between Row and Column Trees

The retrieval process starts with extraction dual image signatures from the query image for the matching of row/column trees. Each node in MOT stores dual signature values because both of the image features are taken into account. Each signature determines a region consisting of one or several trees. In Fig. 5, regions 1 and 2 are selected according to wavelet coefficients and color distribution information, respectively. The intersectional node(s) of regions 1 and 2 are considered to be the images that are most similar to query image due to the fact that they satisfy both of the constraints.

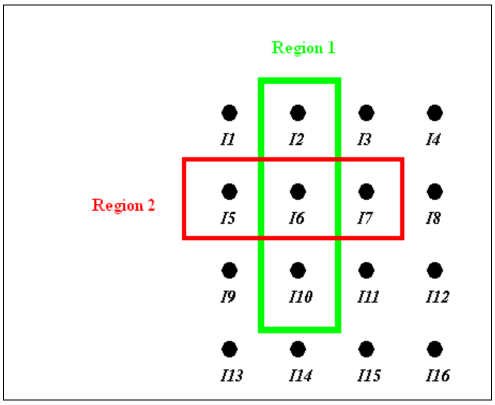


Fig. 5. Intersection between regions.

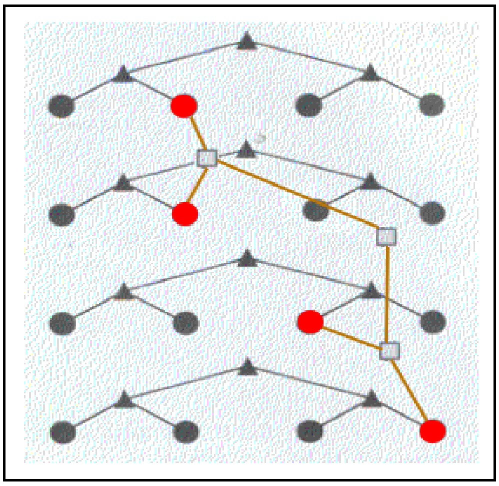


Fig. 6. Oblique MOT.

Note that row and column intersections do not always appear as show in Fig. 5. On the contrary, a row- or column-tree can be oblique as displayed in Fig. 6. Provided that the MOT approach first groups row trees based on the color distributions, the chances of getting ideal orthogonal columns for those images that are similar in shape are slim. As far as the performance is concerned, oblique row or column trees in one dimension will have no adverse impact on the resulting tree traversal.

3.3 Disjoint Row and Column Trees

A part from the ideal case described n the previous section, the query image is not likely to appear in the database. Consequently, no intersection between two trees is obtained through our MOT-based database search. When these two trees are disjoint, we conclude that there is no image in the database that is identical to the query image. As shown in Fig. 7, we can still treat the union of two image sets as a sub-optimal solution even though there is no intersection between them. The images in regions 1 and 2 are then estimated and sorted based on the hybrid similarity metric that we mentioned in a previous section. It then returns the nodes which have the smallest computed distances from the query image.

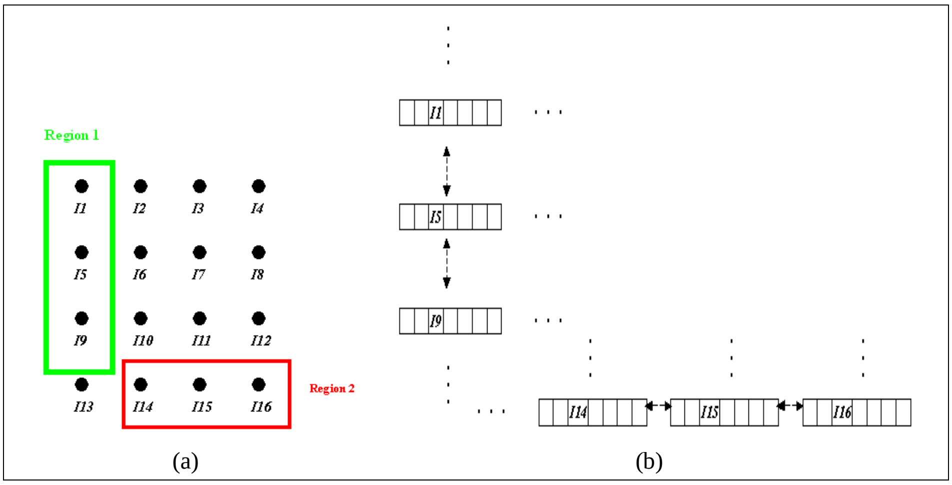


Fig. 7. Disjoint row and column trees.

4. EXPERIMENTAL RESULTS

For demonstrating its performance and effectiveness superiority of the proposed MOT-based system, both single-node SQL system and WALRUS are selected as the two representative systems to compare with. Efficacy measures such as retrieval time, recall rate, and precision rate are used in our query-by-example experiments. All the experiments were performed on a machine equipped with an AMD 800 CPU, Sun Ultra box, and 512 MB of RAM, and Windows 2000 Professional was the run time operating system environment.

Table 1. Performance comparison between SQL server and MOT-based systems.

System	Image Size	Database Size	Platform	Search Strategy	Retrieval Time(secs)
Single Node SQL Server System	128 * 128	10,000	AMD 800	Sequential search	80
MOT-Based CBIR	128 * 128	10,000	AMD 800	MOT	1.35

In Table 1, the system performance is compared for sequential search on SQL servers. It reports the average executing times of the different methods. For a problem size of

10,000 images, we measured the speed of the system by running random queries 1000 times. This running time included the time needed to extract dual image signatures from query images, pelmet the MOT-based search strategy, and find the top 30 similar images.

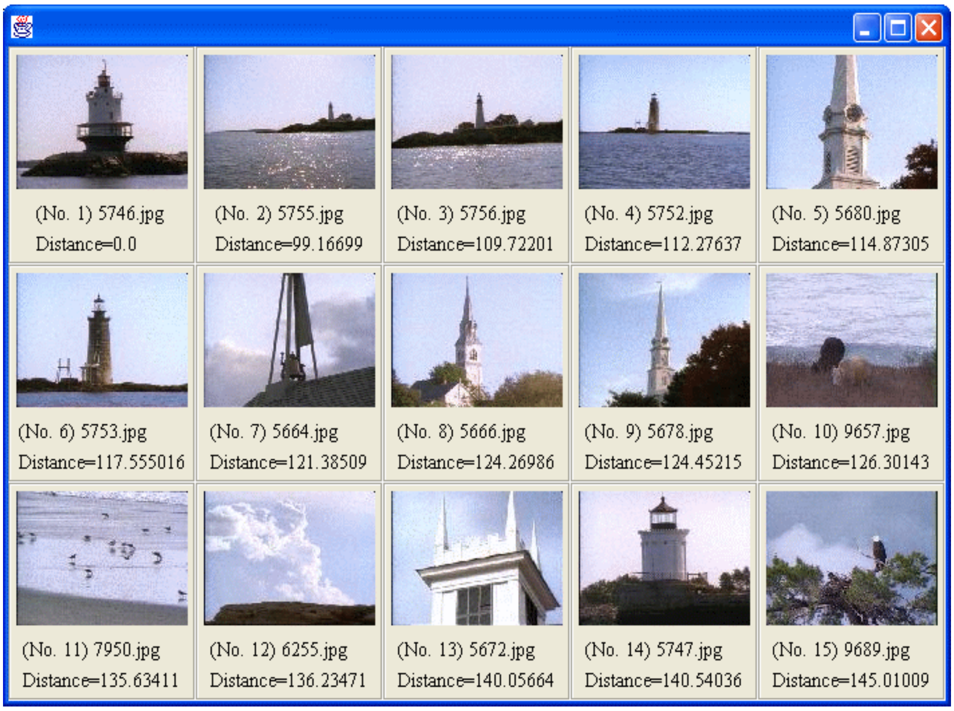


Fig. 8. Query using 5746.jpg.

To overcome the problem of characterizing shapes under spatial transformations caused by region features extracted through wavelet transform for a system like Walrus, our proposed MOT-based system additionally includes invariant color-based features to improve the accuracy. Table 2 shows the efficacy of the proposed CBIR system. As the number of matching images retrieved increases, the recall rate that can be achieved increases compared with the rates obtained [21] with the Walrus system. More than half of the relevant images can be retrieved when the number is 10. When the number exceeds 20, about 90 percent of the relevant images can be recalled.

The reduction in the image retrieval time results from use of the proposed architecture as opposed to traditional approaches such as R*-tree implementations. Ideally, logarithmic time complexity can be achieved with this MOT-based architecture due to its binary tree organization. Table 3 shows the performance superiority achieved compared with the performance of the WALRUS system [21] when the experiment was conducted on a Sun Ultra 2 machine.

Table 2. Retrieval accuracy of query results.

Query image/ number of retrieved images	10		15		20	
	Recall	Precision	Recall	Precision	Recall	Precision
1394.jpg	0.66	0.8	0.66	0.53	0.91	0.55
6076.jpg	0.73	0.8	0.73	0.53	0.90	0.5
5746.jpg	0.56	0.9	0.69	0.73	0.94	0.75
7348.jpg	0.56	0.9	0.81	0.8	0.87	0.7
7311.jpg	0.7	0.7	0.8	0.53	0.9	0.45

Table 3. The Performance of MOT versus WALRUS on an Ultra 2.

System	Image Size	Database Size	Platform	Search Strategy	Retrieval Time(secs)
WALRUS (Duke)	128 * 128	10,000	Sun Ultra 2	R* -tree	5-20
MOT-Based CBIR	128 * 128	10,000	Sun Ultra 2	MOT	1.82

5. CONCLUSION AND FUTURE WORK

CBIR systems are characterized by the growing sizes of the image datasets, hence efforts have been made to take on the challenge. When one deals with multimedia data, algorithms tend to become computationally intractable. This paper has described a query image based method for selecting a calibrated set of results from within a potentially huge image database. The dual image signature composed of shape and color weighted features is used to extract the characteristics of images to overcome the problem addressed in section 2.2. In order to take into account the retrieval effectiveness, a novel MOT-based search strategy has been proposed, which expedites similarity analysis due to its logarithmic complexity. The automatic procedure can be used to objectively retrieve the most similar images in accordance with the parameter settings. Both direct and oblique cases of tree formation have been addressed to reflect actual traversal scenarios encountered. Results show that the proposed architecture indeed offers improved retrieval accuracy and performance.

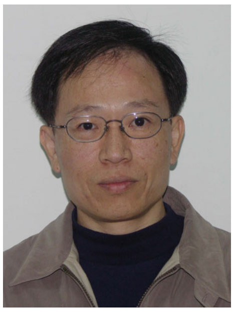
Though we tested our approach on a single system, it is possible to extend the system to a distributed environment because of the recursive decomposition property of MOT. When the $2N$ row and column roots and wires incident to them are removed from the $N \times N$ mesh of trees, we are left with four disjoint copies of the $(N/2 \times N/2)$ mesh of trees. This makes it easy for us to design an efficient layout for a distributed environment. Merging and splitting of the image database will not be a problem because of the intrinsic property of the MOT architecture. It is highly probable that matching images will reside in the same or adjacent PCs due to the way in which the data is arranged. Queries can be made through the web-based interface in the front-end, which provides access for client workstations outside the cluster.

Nevertheless, color and shape are by no means the only differentiating measures for judging similarity between images. In light of the fact that there may be more than two representative characteristics of an image, the proposed MOT based scheme can be further extended for higher dimensions. Higher dimensional MOT will possess an interesting and powerful structure such that the proposed scheme can be extended in nearly the same algorithmic details.

REFERENCES

1. A. P. Berman and L. G. Shapiro, "A flexible image database system for content-based retrieval," *Computer Vision and Image Understanding*, Vol. 75, 1999, pp. 175-195.
2. A. Chadha, B. Soetarman, and J. S. Vitter, "CAMEL: concept annotated iMagE libraries," in *Proceedings of SPIE Electronic Imaging 2001: Storage and Retrieval for Image and Video Databases*, Vol. 4315, 2001, pp. 62-73.
3. A. Pentland, R. W. Pichard, and S. Sclaroff, "Photobook: content-based manipulation of image databases," *International Journal of Computer Vision*, Vol. 18, 1996, pp. 233-254.
4. M. Filickner, H. Sawhney, W. Niblack, J. Ashley, W. Huang, B. Dom, M. Gorkani, J. Hafine, D. Lee, D. Petkovic, D. Steele, and P. Yanker, "Query by image and video content – QBIC system," *IEEE Computer*, Vol. 28, 1995, pp. 23-32.
5. J. R. Smith and S. F. Chang, "VisualSEEK: a fully automated content-based image query system," in *Proceedings of 4th ACM International Conference on Multimedia*, 1997, pp. 87-98.
6. J. R. Smith and S. F. Chang, "WebSEEK: a content-based image and video search engine for the world-wide web," *IEEE Multimedia*, Vol. 4, 1997, pp. 12-20.
7. T. Leighton, "New lower bound techniques for VLSI," in *Proceedings of 22nd Annual Symposium on Foundations of Computer Science*, 1981, pp. 1-12.
8. D. Muller and F. Preparata, "Bounds to complexities of networks for sorting and of switching," *Journal of the ACM*, Vol. 22, 1975, pp. 195-201.
9. C. E. Jacobs, A. Finkelstein, and D. H. Salesin, "Fast multiresolution image querying," in *Proceedings of SIGGRAPH 95*, 1995, pp. 277-286.
10. E. Albuz, E. Kocalar, and A. A. Khokhar, "Scalable image indexing and retrieval using vector wavelets," *IEEE Transactions on Knowledge and Data Enginereing*, Vol. 13, 2001, pp. 851-861.
11. R. Zhang and Z. Zhang, "A clustering based approach to efficient image retrieval," in *Proceedings of 14th IEEE International Conference on Tools with Artificial Intelligence*, 2002, pp. 339-346.
12. J. Z. Wang, G. Wiederhold, O. Firschein, and S. X. Wei, "Content-based image indexing and searching using Daubechies' wavelets," in *Proceedings of 4th Forum on Research and Technology Advances in Digital Libraries*, Vol. 1, 1997, pp. 311-328.
13. A. Natsev, R. Rastogi, and K. Shim, "WALRUS: a similarity retrieval algorithm for image databases," in *Proceedings of ACM SIGMOD International Conference on Management of Data*, 1999, pp. 395-406.
14. M. N. Do and M. Vetterli, "Wavelet-based texture retrieval using generalized Gaus-

sian density and Kullback-Leibler distance," *IEEE Transactions on Image Processing*, Vol. 11, 2002, pp. 146-158.
15. J. Z. Wang, J. Li, and G. Wiederhold, "SIMPLIcity: semantics-sensitive integrated matching for picture libraries," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 23, 2001, pp. 947-963.
16. M. J. Swain and D. H. Ballard, "Color indexing," *International Journal of Computer Vision*, Vol. 7, 1991, pp. 11-32.
17. H. Samet, *The Design and Analysis of Spatial Data Structures*, Addison-Wesley, Reading, MA, 1990.
18. R. F. Sproull, "Refinements to nearest-neighbor searching in k-dimensional trees," *Algorithmica*, Vol. 6, 1991, pp. 579-589.
19. F. T. Leighton, *Introduction to Parallel Algorithms and Architectures*, Morgan Kaufman Press, 1992.
20. A. Berman and L. Shapiro, "Efficient image retrieval with multiple distance measures," in *Proceedings of SPIE Conference on Storage and Retrieval for Image and Video Databases*, 1997, pp. 12-21.
21. A. Natsev, R. Rastogi, and K. Shim, "WALRUS: a similarity retrieval algorithm for image databases," *IEEE Transactions on Knowledge and Data Engineering*, Vol. 16, 2004, pp. 301-316.



Wei-Min Jeng (鄭爲民) received the B.S. degree in Management Science from National Chiao Tung University, Taiwan, in 1985, the M.S. degree in Computer Science from University of Texas at Arlington, in 1989, and the Ph.D. degree in Computer Science from University of Houston, in 1999. He is an Assistant Professor of the Department of Computer Information Science of Soochow University, Taipei, Taiwan. He was with Decision Consultant Inc., Dallas, Texas, from 1989 to 1990. He later joined Positron Corporation, Houston, Texas, from 1990 to 1993. His area of research include medical imaging and high performance computing.

Jen-Hao Hsiao (蕭人豪) received the M.S. degree in Computer Science from Soochow University, Taipei, Taiwan, in 2002. He is currently pursuing the Ph.D. degree at the Department of Electrical Engineering, National Taiwan University, Taipei, Taiwan. His areas of research include content-based multimedia retrieval, image processing, and digital rights management.