Proc. VIIth Digital Image Computing: Techniques and Applications, Sun C., Talbot H., Ourselin S. and Adriaansen T. (Eds.), 10-12 Dec. 2003, Sydney

Image Indexing Technique and Its Parallel Retrieval on PVM

Wichian Premchaiswadi*, Nucharee Premchaiswadi**, Theianchai Patnasirivakin*, Sutasinee Chimlek*, and Seinosuke Narita***

*Siam University, Bangkok 10163, Thailand
**Faculty of Information Technology, Dhurakijpundit University, Bangkok 10210, Thailand
***School of Science & Engineering, Waseda University, Tokyo, Japan

Abstract. The paper presents the technique for image indexing and its parallel implementation on a parallel virtual machine. The indexing method is the combination of the autocorrelogram method and color difference correlograms method. Color filtering is used for quantization of the RGB colorspace into 64 colors to reduce the complexity and calculation time. Then, both color and spatial information are employed to build the index of each image. Experimental results suggest that the method can index and retrieve images with different views of the same scene or large changes in appearance. The performance of the image retrieval system is also improved significantly when compared with both color histogram and color correlogram method. The scheme is also implemented on PVM with the load balancing technique to speed up its retrieval time. The experimental results also show that it could speed up the processing time significantly.

1 Introduction

At present, image databases are used in many application areas and their sizes become larger and larger. Therefore, the efficient image management system is required. There are two steps that are commonly used in most image database information retrieval systems as followings: indexing and searching [1]. In indexing process, each image in a database, a feature vector capturing certain essential properties of the image is computed and stored in a feature-base. Then, in searching process, given a query image, its feature vector is computed, compared to the feature vectors in the feature-base, and images most similar to the query image are returned to the user. In a content-based image retrieval system, color [2][3], the most recognizable element among image contents, is usually used for various image database applications. Color histogram, however, does not provide enough spatial information such that different images may have similar color histogram. To handle the weakness of color histograms several new features that strive to integrate both color and spatial information have been proposed [1][10-13].

In this paper, the AC/CDC (Autocorrelogram and Color Different Correlogram) method that makes use of both color and spatial information is described. In order to speed up its retrieval time, the scheme is implemented on parallel virtual machine (PVM).

2 Image Indexing Scheme

The indexing scheme is the combination of the autocorrelogram method and color difference correlograms method. It consists of three steps namely: color filtering, autocorrelogram and color different correlogram.

a. Color Filtering

The process is used for quantization of the RGB colorspace of each image into 64 colors to reduce the complexity and calculation time in the following processes, autocorrelogram and color different correlogram process.

b. Autocorrelogram

A color correlogram expresses how the spatial correlation of color changes with distance [1] as shown in Eq.1 and Eq.2.

$$\gamma_{c_i,c_j}^{(k)}(I) \cong \Pr_{p_1 \in I_{c_i}, p_2 \in I}[p_2 \in I_{c_j} \,||\, p_1 - p_2 |= k] \tag{1}$$

$$\gamma_{c_i,c_j}^{(k)}(I) \cong \frac{\Gamma_{c_i,c_j}^{(k)}(I)}{H_{c_i} x 8k} \tag{2}$$

The autocorrelogram captures spatial only correlation between identical colors and defined by Eq. 3.

$$\alpha_{c_i}^{(k)}(I) \cong \gamma_{c,c}^{(k)}(I) \tag{3}$$

c. Color Difference Correlogram

Color difference correlogram (CDC) [14] is the scheme that the authors modified from the texture description technique [4-8]. Color difference correlogram of an image is graph or table that indexed by color difference value, where the k-th entry for $Diff_i$ specifies the probability of finding color difference value $Diff_i$ of pixels at distance k from any pixels in image. The color difference value between two pixels having distance equal to k is defined by Eq. 4.

$$Diff_i = | C(p1) ñ C(p2) | \quad (4)$$
where $| p_1 ñ p_2 | = k$

The color difference correlogram is calculated by the formula in Eq. 5.

$$Cdf^{k}_{diff_i}(I) \cong \Pr_{p_1,p_2 \in I}[| C(p_1) - C(p_2) | = diff_i \,||\, p_1 - p_2 | = k]$$
$$Cdf^{k}_{diff_i}(I) = \frac{\beta^{k}_{diff_i}(I)}{n_1 x n_2 x 8k} \quad (5)$$

Where $\beta^{k}_{diff_i}(I)$ is the number of pixels having distance of the center equal to k and having color difference equal to $Diff_i$

3 Image Retrieval

The process is used to retrieve target images from the image database. The technique of distance measure [1] is employed as the criteria for selecting the target images in the process. The paper use d_1 distance measure for comparing histogram, autocorrelogram, color difference correlogram and AC/CDC Eq. 6 to 9.

$$d_1(h) = | I - I' |_{h,d_1} \cong \sum_{i \in [m]} \frac{| h_{c_i}(I) - h_{c_i}(I') |}{1 + h_{c_i}(I) + h_{c_i}(I')} \quad (6)$$

$$d_1(\gamma) = | I - I' |_{\gamma,d_1} \cong \sum_{i,j \in [m], k \in [d]} \frac{| h_{c_i}(I) - h_{c_i}(I') |}{1 + h_{c_i}(I) + h_{c_i}(I')} \quad (7)$$

$$d_1(CDC) = | I - I' |_{CDC, d_1} \cong \sum_{diff_i \in [diff], k \in [d]} \frac{| CDC^{k}_{diff_i}(I) - CDC^{k}_{diff_i}(I') |}{1 + CDC^{k}_{diff_i}(I) + CDC^{k}_{diff_i}(I')} \quad (8)$$

$$d_1(\gamma, CDC) = d_1(\gamma) + d_1(CDC) \quad (9)$$

4 Performance Measures

The performance of the scheme is measured by using parameters described in [1] as follows.

1) r-measure is the sums up of the rank of correct answer of all queries and average r-measure is the r-measure divided by the number of queries q as shown in the formula below.

$$r - measure = \sum_{i=1}^{q} rank(Q_i^{'}) \quad (10)$$

$$Avg \ \text{r - measure} = \frac{\text{r - measure}}{\text{q}} \quad (11)$$

2) p_1-measure is the sum of the precision at the recall equal to 1and the average p_1-measure is the p1-measure divided by q as shown below.

$$p_1 - measure = \frac{1}{\sum_{i=1}^{q} rank(Q_i^{'})} \quad (12)$$

$$Avg \ \text{p}_1 \text{ - measure} = \frac{\text{p}_1 \text{ - measure}}{\text{q}} \quad (13)$$

5 The Implementation on PVM

The master/slave technique is employed in the implementation. The master program is in charge of the task scheduling and allocation of resources for slaves. While each slave is in charge of processing the assigned task and sends the result to the master. The basic configuration is shown in Fig.1. In order to balance tasks on each slave machine efficiently, the dynamic task scheduling is employed. The capacity of each machine is defined by using its CPU speed. The parameter can be obtained by building the process to check for the CPU speed of each machine. The configuration of the proposed scheme is shown in Fig. 2.

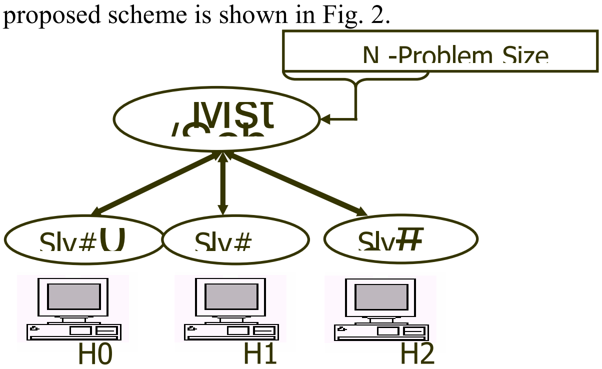


Fig. 1. Master/Slave configuration.

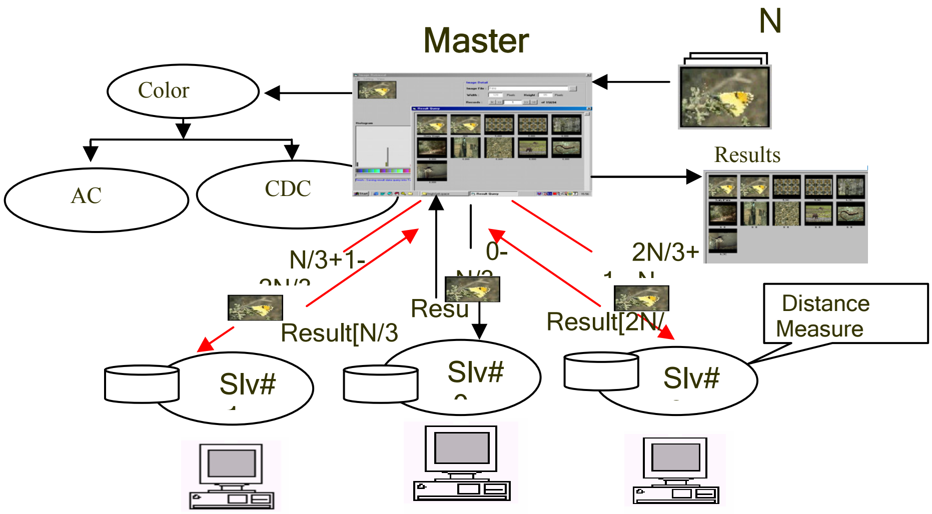


Fig. 2. Configuration for the implementation.

6 Experimental Results

The proposed scheme is implemented by using Visual Basic version 6.0 and Microsoft Access Database. The experimental image database is a heterogeneous image database. It consists of 15,694 images in varieties formats such as JPEG, BMP and GIF and varieties sizes. The number of color in each image is quantized into 64 colors. Thus, color difference of 64 values are computed, set {0, 1, .., 63}. Figure 3 shows the sample query of the proposed system. The result of the query is shown in Fig. 4.

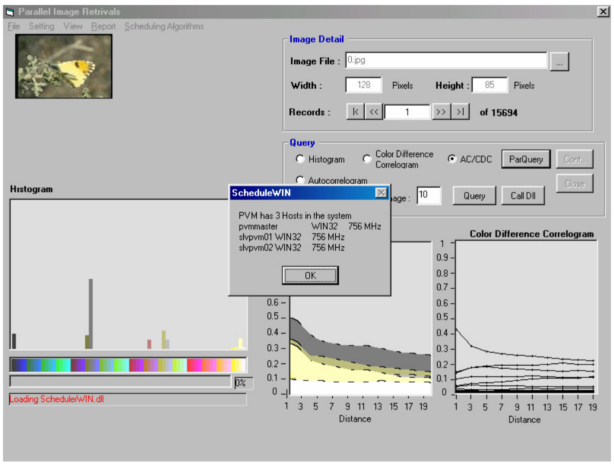


Fig. 3. Query for the proposed scheme.

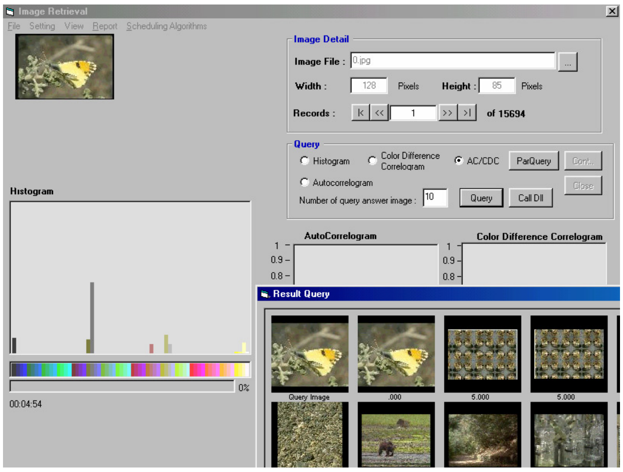


Fig. 4. Result of the query in Fig.3.

The system is tested with 100 queries and some of the results are shown in Table 1.

From the experimental results, the AC/CDC scheme performs better performance than the other compared schemes color histogram (H), autocorrelogram (A) and color difference correlogram (C) as shown in Table 2. The scheme can also retrieve images with different views of the same scene, large changes in appearance.

For the parallel processing, the scheme is tested up to three PEs and the results are compared with that of a sequential program on a single host. In the experiment, each host is already installed PVM software and connected by the Ethernet network 10/100 T-Base. Specification of each individual host is shown in Table 3.

For the parallel implementation, the experiments are conducted by using many sample query images and the average processing times and speed up factor (ratio of a sequential processing time to a parallel processing time) are shown in Table 4.

From Table 4, it can be seen that all speed up factors exceed a number of hosts in a virtual machine which seem to be impossible in a practical.

6.2 Analysis of the results

There is a need to find out that why the speed up factor of the experimental results exceeds a number of processors used.

As described in the previous section that in finding the target images of each query, the scheme needs to process in 3 steps as follows:

a) loading all feature vector recorded in the images database
b) calculate a distance measure, and
c) ranking all N results.

The loading time and the total processing time of different number of images in the database are shown in Table 5.

From Table 5, it can be seen that the loading time of data from image database is used more than 85% of the total processing time.

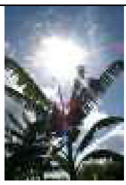
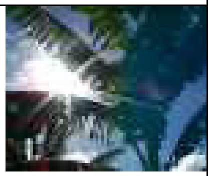
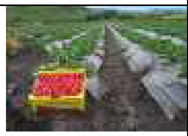
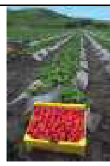
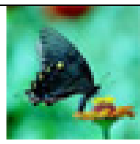
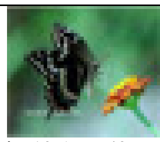
Query	**Result**	**Query**	**Result**
	 Hist 74 Auto 15 CDC 29 AC/CDC 4		 Hist 4 Auto 2 CDC 1 AC/CDC 1
	 Hist 28 Auto 2 CDC 3 AC/CDC 1		 Hist 5 Auto 3 CDC 7 AC/CDC 1
	 Hist 8 Auto 8 CDC 1 AC/CDC 1		 Hist 32 Auto 4 CDC 1 AC/CDC 1
	 Hist 32 Auto 3 CDC 1 AC/CDC 1		 Hist 7 Auto 4 CDC 1 AC/CDC 1
	 Hist 29 Auto 16 CDC 86 AC/CDC 9		 Hist 19 Auto 43 CDC 28 AC/CDC 5

Table 1. Some of the experimental results.

Method	Hist	Auto	CDC	AC/CDC
Avg. r-measure	20	7	9	2
Avg. p1-measure	0.56	0.70	0.58	0.80

Table 2. Performances of various methods

Hostname	CPU	Memory
Master	AMD THUNDER BIRD 755 MHz	SDRAM 128MB
Slvpvm01	ASUS ATHLON 550 MHz	EDO 64 MB
Slvpvm02	ASUS CELERON 400 MHz	EDO 64 MB
Slvpvm03	Pentium MMX 233 MHz	EDO 32 MB

Table .3 Host specifications

Number of Hosts	Time(sec)	Speed Up
1	214	1.0
2	94	2.28
3	65	3.29

Table 4. Experimental results

Number of images	Loading time (sec)	Total processing time (sec)	Percentage of loading time
15694	189	214	88.32
7847	82	94	87.23
5232	54	65	83.07

Table 5. Loading and total processing time.

This testing has measured a loading time of various number vector data records and plotted and as shown in Fig. 5. It can be seen from Fig. 5 that the loading time is not linear especially when a number of records exceed 8,000 records.

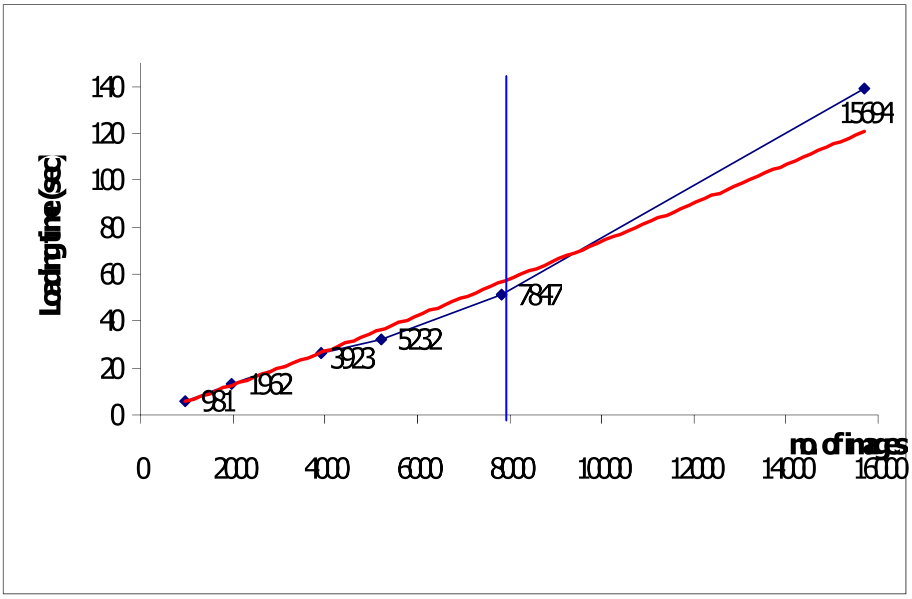


Fig. 5. Loading time.

```
Sub Load ()
RecSet.MoveFirst();
for(i=0;i<begin*10;i++)
          RecSet.MoveNext();
time(&start);
          for(i=0;i<(end-begin)*10;i++){
             for(j=0;j<64;j++){
                 RecData[ i ][ j ] = RecSet.FieldAt[ j ]
            }
                    RecSet.MoveNext();
          }
time(&stop);
```

Fig. 6. Loading program.

The sub program that is used for loading these images is shown in Fig. 6. In the experiment, the images of 15,694 images are used. One image needs 10 records to store all of its feature vector information. Therefore, the total number of feature vectors stored in the database is 156,940 records. The loading time is very significant for the parallel processing.

When the scheme is implemented on PVM, the number of records needed to be loaded is decreased approximately by the factors of a number of PE used. Therefore, it could be concluded from the experimental results that the loading time made the speed up factor more than the expected value.

7 Conclusion

The scheme for image indexing and its parallel implementation of image retrieval technique are proposed. From the experimental results, it can be concluded that a parallel image retrieval system by using AC/CDC on a PVM system is efficient and can be applied in a practical use.

References

1. J. Huang, S. Kumar, M. Mitra, W.J.Zhu, and R. Zabih: ìImage indexing using color correlogram,î *Proc. of IEEE Conf. on Computer Vision and Pattern Recognition*, pp.762-768, 1997.
2. John R. Smith and Shih-Fu Chang: ìSingle color extraction and image query,î Proc. IEEE Int. Conf. on Image Proc., pp.528-531, 1995.

3. Michael Swain and Dana Ballard: ìColor indexing,î *International Journal of Computer Vision*, 7(1), pp.11-32, 1991.
4. Robert M. Haralick, K. Shanmugam, and Itsíhak Dinstein: ìTexture feature for image classification,î *IEEE Trans. on Sys, Man, and Cyb*, SMC-3(6), pp. 610-621, 1973.
5. B. Manjunath and W. Ma: ìTexture feature for browsing and retrieval of image data,î *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 18(8), August 1996.
6. Randy Crane: ìA simplified approach to Image Processing,î New Jersey, Prentice-Hall, 1997.
7. I. Pitas: ìDigital Image Processing Algorithms and Applications,î New York: WILEY INTERSCIENCE, 2000.
8. S. Nepal, M V. Ramakrishna, J A. Thom: ìImage Retrieval Using Pixel Descriptor and its Applications to Region Extractionî, Technical report, Royal Melbourne Institute of Technology, 1997.
9. Stephan Bres and Jean-Michel Jolion: ìDetection of interest points for image indexing,î *3rd Int. Conf. on Visual Information Systems, Visual 99, Springer, Lecture Notes in Computer Science*, 1614, pp.427-434, June 1999.
10. Greg Pass, Ramin Zabih, and Justin Miller: ìComparing images using color coherence vectors,î In *Proceedings of ACM Multimedia 96*, Boston MA USA, pp. 65-73, 1996.
11. Greg Pass and Ramin Zabih: ìHistogram Refinement for Content-Based Image Retrieval,î *Workshop on Applications of Computer Vision*, 1996.
12. Markus Stricker and Alexander Dimai: ìColor Indexing with Weak Spatial Constraints,î *SPIE Conference*, San Jose, CA February 1996.
13. W.Hsu, T.S.Chua and H.K.Pung: ìAn Integrated Color-spatial Approach to Content-based Image Retrieval,î *The ACM Multimedia Conference*, pp. 305-313, 1995.
14. W.Premchaiswadi, N.Premchaiswadi, S. Chimlek and S.Narita, ìImage Indexing and Retrieval using Autocorelogram and Color Difference Corelograms (AC/CDC),î *International Conference on Fundamentals, Communications and Computer Sciences, ICFS 2002*, Tokyo, Japan, March, 2002.