1998 International Computer Symposium
Workshop on Software Engineering and Database Systems
December 17-19, 1998, N.C.K.U., Tainan, Taiwan, R.O.C.

A MULTI-CHANNEL STORAGE ARRANGEMENT SCHEME OF PICTORIAL DATABASE FOR CONCURRENT SPATIAL MATCH RETRIEVAL

Jau-Ji Shen, and Chin-Chen Chang***

*College of Management, Chaoyang University of Technology, Taichung, Taiwan 413, R.O.C.

Tel:04-3323000 ext. 7122, Fax:04-3742369

e-mail:jjshen@mail.cyut.edu.tw

**College of Engineering, National Chung Cheng University Chiayi, Taiwan 62107,R.O.C.

Tel:04-3259100 or 05-2720412 ext. 6011, Fax:04-3277425

e-mail:ccc@cs.ccu.edu.tw

ABSTRACT

When we touch the applications related to pictorial databases then the first problem, that we should face to, is the large amount of pictorial data (pictures) to be retrieved. So, in this paper, we propose a multi-channel storage arrangement scheme to allocate the pictures into a multi-channel storage system such that the average response time for all possible spatial match queries is minimized.

Keywords: 9-DLT Representation, Parallel Spatial Match Query, Pictorial Database.

1. INTRODUCTION

The fast progress of computer hardware and software technologies stimulates the endless requests for computer applications. In the last two decades, computers were taking a fast influence of people's life. Especially for the past few years, computers supported the more powerful capability of picture-oriented applications such as multimedia applications for entertainment products, digital map of navigation supports system on an aircraft and pattern recognition in artificial intelligent applications. Many picture-oriented applications involve lots of similar pictures searching, this kind of retrieval is known as spatial match retrieval [Chang 1989, Chang, Shi and Yan 1987]. There are many researches related to the topic of spatial match retrieval such as [Chang and Buehrer 1992, Chang 1991, Chang and Lin 1990, Lee, Shan and Yang 1989].

A spatial match query is a searching request of pictures which retrieves pictures from a pictorial database by specifying some spatial relationships among the objects of a picture. The spatial relationships only interest in the bearings among objects without the distance consideration. In practice, a real picture will be denoted by a symbolic picture which extracts the spatial characteristics from the real picture. Therefore, the operational targets of spatial match query are symbolic pictures which can further be used to facilitate the similar pictures searching.

Fig.1.1 is a simple symbolic picture. In that figures A, B, C, D and E are five different icons, where an icon represents an atom that relates to an object in the real picture. The spatial match query, which is relevant to the symbolic picture in Fig.1.1, might contain some spatial specification such as A is to the north of B, and to the northwest of C, and so on. For the sake of processing convenience, Chang et al. [1987] proposed the 2D-string method, and Chang and Lin [1990] proposed a 9-DLT representation to store the symbolic picture into a list form.

A		
	DE	C
B		

Figure 1.1 A symbolic picture

Since huge amount of data processing is the major feature of any application dealing with pictorial data, hence an efficient storage processing capability to make up the timing requirement is required. In modern computer technology, the multi-channel storage device can be built easily such as multiple disk or memories system with concurrent accessing mechanism. It will incorporate some efficient storage arrangement algorithm into such powerful storage device then the throughput can be drastically increased. In this paper, we will arrange pictorial databases onto multi-channel storage devices via a concept deriving from consecutive retrieval property [Ghosh 1972] and 9-DLT representation. By using this data organization, we derive a storage arrangement algorithm that can be used to facilitate the parallel retrieving of spatial matched symbolic pictures.

In the rest of this paper, we will review the concept of the consecutive retrieval property and 9-DLT representation in Section 2. Section 3 introduces the pictorial consecutive retrieval organization. The framework of the storage arrangement algorithm of pictorial databases for parallel spatial match queries will be introduced in Section 4. Finally, we give some concluding remarks in Section 5.

2. The Consecutive Retrieval Property and the 9-DLT Representation

2.1 The Consecutive Retrieval Property

The Consecutive Retrieval (CR for short) property was first introduced by Ghosh [1972, 1973]. And, some important results have been proposed in [Al-Fedaghi and Chin 1979,Deogaun, Raghavan and Tsou 1984, Ghosh 1977].

In 1977, Ghosh tried to construct a file with the CR property for a given set of queries. This scheme proposed an opportunity such that records relevant to a query will be clustered together and known as CR organization. By using CR property, we can analyze the relationship between records and its relevant queries. In the traditional sequential process, one record is accessed at a time for the relevant query. Under this circumstances, we can build a file with CR organization and put the clustered records into contiguous memory space then the access time will be minimized. In the parallel process, several records are accessed at one time for the relevant query. In this case, we can also transform a file to its corresponding CR organization and put the clustered records into distinct memories such that the response time will be minimized. So, CR property shows a well-constructed data structure in our framework. We will illustrate this detailed concept in Section 4.

In the followings, let's introduce the CR property first. The relationship between the query set and the record set is expressed by the record-query incidence matrix. This incidence matrix is an n×m 0/1 matrix. The (i, j)th element of the matrix is 1, if the record R_i is relevant to the query Q_j, and 0 otherwise. This property can be formally stated as follows:

Definition 2.1 [Ghosh 1972]

A query set $\{Q_i\}_{i=1,2,\ldots,m}$ is defined to have the CR property with respect to a set of records $\{R_i\}_{i=1,2,\ldots,n}$ if there exists a permutation of the records in $\{R_i\}_{i=1,2,\ldots,n}$ after permuting the corresponding rows in the incidence matrix M, such that each column in M has all 1's standing at consecutive rows and if $\{R_i'\}_{i=1,2,\ldots,n}$ is the permutated set of $\{R_i\}_{i=1,2,\ldots,n}$ then $\{R_i'\}_{i=1,2,\ldots,n}$ is a CR organization with respect $\{Q_i'\}_{i=1,2,\ldots,m}$ where the notation $\{X_i\}_{i=1,2,\ldots,r}$ denotes the set $\{X_1, X_2, \ldots, X_r\}$.

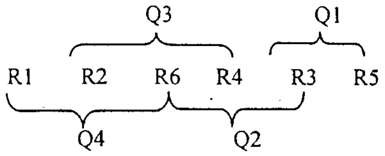
For example, a file of six records is related to four queries as follows:

	Q1	Q2	Q3	Q4
R1	0	0	0	1
R2	0	0	1	1
R3	1	1	0	0
M = R4	0	1	1	0
R5	1	0	0	0
R6	0	1	1	1

Where "R_1, R_2, R_3, R_4, R_5, R_6" is not a CR organization since there exists a column in M, such as column one, which has "1" without standing at consecutive rows. If we permute the rows in M appropriately then we can find a CR organization.

	Q1	Q2	Q3	Q4
R1	0	0	0	1
R2	0	0	1	1
R6	0	1	1	1
M = R4	0	1	1	0
R3	1	1	0	0
R5	1	0	0	0

By Observing M', We know that each column has all 1's standing at consecutive rows, thus "R_1, R_2, R_6, R_4, R_3, R_5" is a CR organization with respect to Q_1, Q_2, Q_3 and Q_4. From this example, we can see the clustering power of CR organization that can be further illustrated as follows:



By the way, there is other important result as shown in the following theorem, which can be easily concluded from Definition 2.1 and we will use this result in our framework.

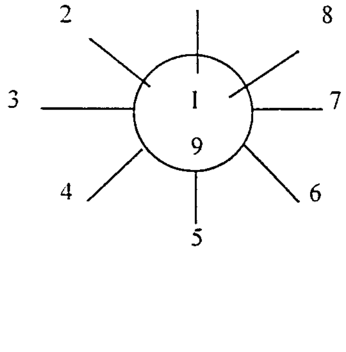
Theorem 2.1 [Ghosh 1997]

If a query set $\{Q_i\}_{i=1,2,\ldots,m}$ has CR property with respect to a set of records $\{R_i\}_{i=1,2,\ldots,mn}$ then any subset $\{Q_i'\}_{i=1,2,\ldots,k}$ of $\{Q_i\}_{i=1,2,\ldots,m}$,where k<=m, also has the CR property with respect to $\{R_i\}_{i=1,2,\ldots,n}$.

If we plant the concept of CR property to the pictorial database construction then we can employee the clustering power of CR organization to help the analysis of picture's similarity. Consequently, the concept of CR property can facilitate the spatial match retrieval of symbolic pictures in sequential and parallel manners. In order to approach our goal, we will introduce an important scheme to represent a symbolic picture.

2.1 9-DLT Representation

For a symbolic picture, 9-DLT representation uses nine integers 1,2,3,4, 5,6,7,8 and 9 to denote nine spatial relationships among icons. Fig. 2.1 shows the nine spatial codes, where I indicates the reference icon, 1 stands for "north of I", 2 stands for "northwest of I" and 9 stands for "at the same location as I", and so on.



Here we can easily obtain the spatial code between two icons. For example, the icons A and B in Fig. 1.1 can be expressed as (A, B, 5) which denotes "B is to the south of A". Since if the spatial relationship between icons A and B is known, then the spatial relationship between icons B and A can be easily concluded. So, by using 9-DLT representation, each pair of icons in a symbolic picture can be expressed by a spatial code then the spatial characteristic of a symbolic picture can be easily defined. For instance, Fig.1.1 can be further represented as the set of triples {(A, B, 5), (A, C, 6), (A, D, 6), (A, E, 6), (B, C, 8), (B, D, 8), (B, E, 8), (C, D, 3), (C, E, 3), (D, E, 9)}. In summary, the 9-DLT representation of a symbolic picture is a set of lexicographic ordered triples where each triples (O_i, O_j, r_{ij}) is used to denote the spatial relationship r_{ij} between icons O_i and O_j, where r_{ij} is an integer among 1,2,..., 9, which represents the spatial code, and the lexicographic order of O_i is no greater than that of O_j.

3. Tuple-Based CR Organization

In the previous section, we introduced the CR organization and the 9-DLT representation of a symbolic picture. In this section, we will combine the CR organization and 9-DLT representation to form a new data organization, which is modified from the CR organization and called the tuple-based CR organization. By using this organization, we can construct a pictorial database such that the spatial match queries can be efficiently performed in a multi-channel system.

Before formally defining the tuple-based CR organization, let us define the spatial match query first. Let us recall the case in Fig.1.1. A spatial match query Q which retrieves the symbolic picture in Fig.1.1 can be the proper combination of triples {(A, B, 5), (A, C, 6), (D, E, 9)}. Now, we formally define the spatial match query as follows.

Definition 3.1

Let $\{P_i\}_{i=1,2,...,n}$ be a set of symbolic pictures and S_i be the corresponding 9-DLT representation of the picture P_i for i=1,2,...,n. Let $\{T_i\}_{i=1,2,...,m}$ be the set of $\cup S_i$ (i.e. the union of S_i), where each triple T_i $\{T_i\}_{i=1,2,...,m}$ is said to be a simple spatial match query of $\{P_i\}_{i=1,2,...,n}$. For a general spatial match query Q, it is defined as $\{T_{ij}\}_{j=1,2,...,k}$ where $T_{is} \neq T_{it}$ if $s \neq t$ and $\{T_{ij}\}_{j=1,2,...,k} \subset \{T_i\}_{i=1,2,...,m}$, and it can be expressed as "to search the set of pictures $\{P_i\}_{i=1,2,...,n}$ in which each picture matches the spatial characteristics T_{i1} and T_{i2} and T_{ik}".

By Definition 3.1, we can easily conclude that if A_{ij} is the pertinent picture set of T_{ij} then query $Q=\{T_{ij}\}_{j=1,2,...,k}$ has a pertinent picture set as $\{T_{ij}\}_{j=1,2,...,k}$ has a pertinent picture set as $A_{i1} \cap A_{i2} \cap ... \cap A_{ik}$.

Example 3.1

Let us consider a pictorial database, which contains six symbolic pictures {P_1, P_2, P_3, P_4, P_5, P_6}. The associated six 9-DLT representations of these six pictures are expressed as follows:

$P_1 \rightarrow S_1$={(A, B, 8), (A, C, 8), (B, C, 1)},

$P_2 \rightarrow S_2$={(A, B, 7), (A, D, 8), (B, D, 2)},

$P_3 \rightarrow S_3$={(A, B, 7), (A, C, 8), (B, C, 1)},

$P_4 \rightarrow S_4$={(A, D, 1), (B, C, 3), (B, D, 2), (C, D, 8)},

$P_5 \rightarrow S_5$={(A, B, 7), (A, D, 1), (B, D, 2)},

$P_6 \rightarrow S_6$={(A, D, 1), (B, C, 2), (B, D, 2), (C, D, 8)},

By definition, we obtain the set of triples of this database is {(A, B, 7), (A, B, 8), (A, C, 8), (A, D, 1), (A, D, 8), (B, C, 1), (B, C, 2), (B, C, 3), (B, D, 2), (C, D, 8)} = $\{T_i\}_{i=1,2,...,10}$. For a query Q={T_4, T_9, T_{10}}= {(A, D, 1), (B, D, 2), (C, D, 8)}, we know that the pertinent picture set is {T_4, T_5, T_6} $\cap$ {T_2, T_4, T_5, T_4} $\cap$ {T_4, T_6} = {T_4, T_6} thus there are two pictures T_4 and T_6 pertinent to the query Q. Now, we are ready to plant the concept of CR organization into pictorial databases. Let $\{P_i\}_{i=1,2,...,n}$, $\{S_i\}_{i=1,2,...,n}$, and $\{T_i\}_{i=1,2,...,m}$ be three sets as defined in Definition 3.1. Then we say M, an n×m matrix, is a picture-triple incidence matrix if the (i, j)th element a_{ij} of M satisfied that

$$a_{ij} = \begin{cases} 1, \text{ if } T_j \quad S_i; \\ 0, \text{ otherwise} \end{cases}$$

For instance, the picture-triple incidence matrix of the pictorial database as shown in Example 3.1 is illustrated as below.

M =

	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}
P_1	0	1	1	0	0	1	0	0	0	0
P_2	1	0	0	0	1	0	0	0	1	0
P_3	1	0	1	0	0	1	0	0	0	0
P_4	0	0	0	1	0	0	0	1	1	1
P_5	1	0	0	1	0	0	0	0	1	0
P_6	0	0	0	1	0	0	1	0	1	1

By observing Definition 2.1, we define the tuple-based CR organization and the tuple-based CR property as follows:

Definition 3.2

An organization of the symbolic picture set $\{P_i\}_{i=1,2,...,n}$ and its pertinent simple spatial query triple set $\{T_i\}_{i=1,2,...,m}$ have a picture-triple incidence matrix M. If each column of M has all 1's standing at the consecutive rows then $\{P_i\}_{i=1,2,...,n}$ is defined to be a tuple-based CR organization for $\{T_i\}_{i=1,2,...,m}$, and $\{T_i\}_{i=1,2,...,m}$ is said to have the tuple-based CR property.

Recalling to Example 3.1, since the corresponding M is not the case defined in Definition 3.2, thus {T_1, T_2, T_3, T_4, T_5, T_6} does not form a CR organization. But if we permute the rows of M properly (i.e. changing the sequence of pictures) as follows:

M =

	T_1	T_2	T_3	T_4	T_5	T_6	T_7	T_8	T_9	T_{10}
P_1	0	1	1	0	0	1	0	0	0	0
P_3	1	0	1	0	0	1	0	0	0	0
P_2	1	0	0	0	1	0	0	0	1	0
P_5	1	0	0	1	0	0	0	0	1	0
P_4	0	0	0	1	0	0	0	1	1	1
P_6	0	0	0	1	0	0	1	0	1	1

Then we obtain a new matrix M' which satisfies Definition 3.2. Thus, we say that $\{P_1, P_3, P_2, P_5, P_4, P_6\}$ is tuple-based CR organization with respect to $\{T_i\}_{i=1,2,...,10}$.

By Definition 3.1, we know that a general spatial match query is the proper combination of simple queries. That is, a general query $\{T_{ij}\}_{j=1,2,...,k}$ (see Definition 3.1) is a proper subset of $\{T_i\}_{i=1,2,...,m}$. By Theorem 2.1, the following corollary can be easily concluded.

Corollary 3.1

If $\{T_i\}_{i=1,2,...,m}$ has tuple-based CR property with respect to the relevant picture set $\{P_i\}_{i=1,2,...,m}$ then any general query $\{T_{ij}\}_{j=1,2,...,k}$ has the tuple-based CR property with respect to $\{P_i\}_{i=1,2,...,m}$.

Lemma 3.1

Let $\{P_i\}_{i=1,2,...,n}$ be the symbolic picture set and $\{T_i\}_{i=1,2,...,m}$ be the set of simple queries, and $\{Q_i\}_{i=1,2,...,l}$ be the set of general queries in which any query is a proper combination of some simple queries in $\{T_i\}_{i=1,2,...,m}$. If $\{T_i\}_{i=1,2,...,m}$ has tuple-based CR property with respect to $\{P_i\}_{i=1,2,...,n}$ then $\{Q_i\}_{i=1,2,...,l}$ has CR property with respect to $\{P_i\}_{i=1,2,...,n}$. On the other hand, if $\{P_i\}_{i=1,2,...,n}$ forms a tuple-based CR organization for $\{T_i\}_{i=1,2,...,m}$ then it is also true for $\{Q_i\}_{i=1,2,...,l}$.

Proof:

By definition 3.1, any Q in $\{Q_i\}_{i=1,2,...,l}$ can be express as $\{T_{ij}\}_{j=1,2,...,k}$ and $\{T_{ij}\}_{j=1,2,...,k} \subset \{T_i\}_{i=1,2,...,m}$. Let A_1, A_2, ..., A_k be the relevant picture set of T_{i1}, T_{i2}, ..., T_{ik} respectively. By Corollary 3.1, we know that any query $\{T_{ij}\}_{j=1,2,...,k}$ has tuple-based CR property with respect to $\{P_i\}_{i=1,2,...,n}$, therefore all pictures in A_1, A_2, ..., A_k are located at consecutive rows of the associated incidence matrix.

Since the relevant picture set of Q is obtained from $A_1 \cap A_2 \cap ... \cap A_k$. In the following, we shall show that intersection pictures are located at the consecutive rows of the associated incidence matrix. The are two cases after intersection. One is $A_1 \cap A_2 \cap ... \cap A_k = \varnothing$ while the other is $A_1 \cap A_2 \cap ... \cap A_k = \{P_{j1}, P_{j2}, ..., P_{jb}\} \subset \{P_i\}_{i=1,2,...,n}$. We have to show that P_{j1}, P_{j2}, ..., P_{jb} are located consecutively .

Suppose that a picture P_y in P_{j1}, P_{j2}, ..., P_{jb} is not located consecutively with its neighbor(s). Then, there is a contradiction that is $\{P_i\}_{i=1,2,...,m}$ without CR property can be easily concluded. It is because that its neighbor(s) should belong to each of A_1, A_2, ..., A_k which will be destroy the fact of Corollary 3.1. For any general query in $\{Q_i\}_{i=1,2,...,m}$, the above discussion holds. Thus, we have the proof.

Q.E.D.

As we showed in Example 3.1 $\{P_1, P_3, P_2, P_5, P_4, P_6\}$ is a tuple-based CR organization with respect to its relevant simple queries $\{T_i\}_{i=1,2,...,m}$. If we consider a general query $\{Q_4, Q_9, Q_{10}\}$, then its pertinent picture set $\{P_4, P_6\}$ is a tuple-based CR organization as well. Based on Lemma 3.1, we see that if a picture set has a tuple-based CR organization with respect to the set of its relevant simple queries, then this set of pictures also forms a CR organization for all possible general queries composed from the relevant simple queries.

But, how to examine and construct the CR organization of picture set $\{P_i\}_{i=1,2,...,m}$, with respect to all possible simple queries? For this problem, the interested reader is encouraged to consult [Al-Fedaghi and Chin 1979].

4. The Multi-channel Arrangement Scheme of Pictorial Database and the Processing Framework

In the past ten years, the techniques of parallel computers have been maturely developed and many architecture were proposed [Duncan 1990, Feitelson and Rudolph 1990]. Moreover, many real parallel machines have been built and operated such as Cedar, HM^2P and HEP [Dinning 1989]. In this paper, we will consider the general multiprocessor machine, say GMM for short, which can be depicted generally as Fig.4.1.

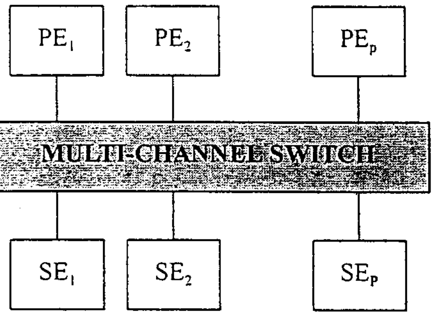


Figure 4.1: A general architecture of multiprocessor machine. PE stands for processing element. SE stands for processing element.

For a large database, we usually consider each SE in Fig.4.1 as a disk. In order to speedup the spatial match query, we will show a parallel execution manner of query and propose a disk storage arrangement of pictures with respect to the designated parallel queries. Let's show the parallel query first. Again, let $\{T_{ij}\}_{j=1,2,...,k}$ be a general query and $\{P_i\}_{i=1,2,...,n}$ be the pertinent picture set. For convenience, let $\{T_{i1}, T_{i2}, ..., T_{ik}\} = \{T_1', T_2', ..., T_k'\}$. Let a'_{ij} $1 \leqq i \leqq n$ and $1 \leqq j \leqq k$, be an element of the picture-triple incidence matrix of $\{P_i\}_{i=1,2,...,n}$ and $\{T_i'\}_{i=1,2,...,k}$ then we design the parallel query on GMM as the following figure.

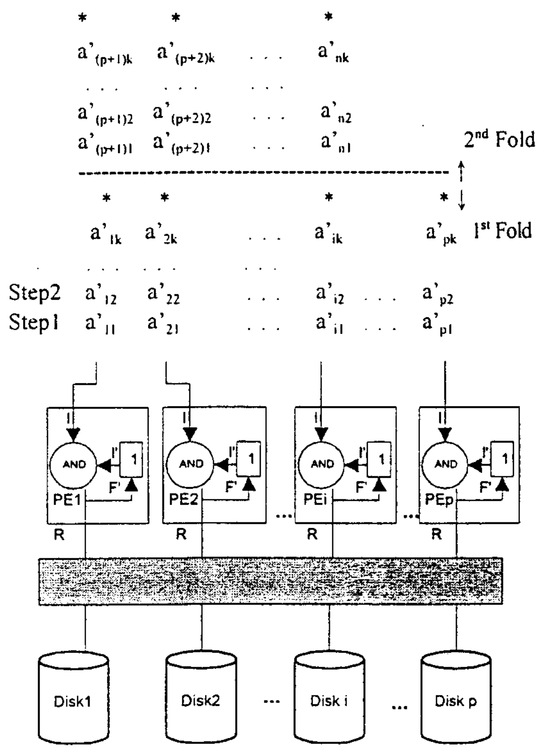


Figure 4.2 The illustration of a parallel spatial match query.

In Fig.4.2, it is a case with p<n (the total number of pictures)<2p, so there are two fold inputs to be processed by each PE. Function of each PE can be defined as follows:

```
Initialize F=1; R = null;
While I≠* do
    F = I AND I'
    R = null;
End while;
R=I';
F=1;
```

After O(k) "AND" operations, we obtain the final result of R in each PE. If R is 1 then a picture retrieval action should be initiated. If R is 0 then it does nothing. For a sequential query, the number of "AND" operations should be O(nk), thus the parallel query has significantly improved its preprocessing time. But, there is a more important problem is, how to arrange the picture onto disks such that the requested pictures are just located at disk i when PE_i has R=1.

For a system with p disks and p controllers, if a spatial match query Q with b pertinent pictures then the minimal number of disk accesses is equal to [b/p]. (Here [x] represents the smallest integer $\geq$ x). Since one disk access can examine p pictures, therefore if we arrange all pictures carefully for any possible spatial match query then the optimality of disk accesses can be expected.

Based on the tuple-based CR organization specified in Section 3, we propose the following algorithm to achieve the optimal arrangement of the set of pictures $\{P_i\}_{i=1,2,\ldots,n}$ into p disks with respect to all possible queries.

Algorithm A

Input: M_{n*m} is the picture-triple incidence matrix for $\{P_i\}_{i=1,2,\ldots,n}$ and $\{T_i\}_{i=1,2,\ldots,m}$. p is the number of disk.

Output: The assignment of all pictures in $\{P_i\}_{i=1,2,\ldots,n}$ into the disks.

Step1: Rearrange M_{n*m} into M'_{n*m} which has the tuple-based CR organization as $\{PR_i\}_{i=1,2,\ldots,n}$ [Al-Fedaghi and Chin 1979], where each $PR_i \in \{P_i\}_{i=1,2,\ldots,n}$ stands for picture location at row i.

Setp2: Store Picture PR_i in Disk (i mod p), if i mod p≠0. Store Picture PR_i in Disk p, if otherwise.

Theorem 4.1

If picture set $\{P_j\}_{j=1,2,\ldots,n}$ has a CR organization with respect to its relevant simple query set $\{T_i\}_{i=1,2,\ldots,m}$ then Algorithm A produced the optimal arrangement of pictures in $\{P_i\}_{i=1,2,\ldots,n}$ into p disks such that the average number of disk accesses for all possible parallel general queries on GMM is minimized.

Proof:

By Step 1 of Algorithm A, the CR-organized set $\{PR_i\}_{i=1,2,\ldots,n}$ of $\{P_i\}_{i=1,2,\ldots,n}$ is constructed. Step 2 assigns all pictures as follows:

Disk 1: PR_1, PR_{p+1}, PR_{2p+1}, ...
Disk 2: PR_2, PR_{p+2}, PR_{2p+2}, ...
.
.
Disk i: PR_i, PR_{p+i}, PR_{2p+i}, ...
.
.
Disk p: PR_p, PR_{2p}, PR_{3p}, ...

For any spatial match query Q, if Q has b pertinent pictures then the optimal number of disk accesses is [b/p]. On the other hand, all pictures pertinent to Q should be evenly dispersed to different disks then the optimality can be achieved.

According to the previous illustration of Step 2 of Algorithm A, we find that PR_1, PR_2, ... and PR_n are evenly assigned to different disks. By Lemma 3.1, we have that any possible general query Q composed from $\{T_i\}_{i=1,2,\ldots,m}$ has CR property with respect to $\{PR_i\}_{i=1,2,\ldots,n}$. Thus, the b pertinent pictures of Q should be located at consecutive order in $\{PR_i\}_{i=1,2,\ldots,n}$, hence we obtain the following result.

Result A : All b pertinent pictures of Q are assigned into p disks with no more than [b/p] pictures on each disk.

By the designated parallel query Q on GMM as shown in Fig.4.2, if Q has b pertinent pictures then we have the number of R's being set with 1 is equal to b. In other words, there has b retrieval actions to be initiated. We should prove that these b times initiations of retrieval actions have to be performed by different PE's. Again, by Lemma 3.1, we know that any Q has CR property with respect to $\{PR_i\}_{i=1,2,\ldots,n}$. Thus, we have the Result B.

Result B: All R's with value 1 shall be happened at consecutive PE's thus the retrieval action of

Q would be initiated by consecutive PE's with no more than [b/p] retrieval actions to be initiated at the same PE.

By Results A and B, we have proved that Algorithm A produces an optimal disk arrangement of pictures in $\{P_i\}_{i=1,2,\ldots,n}$ such that the all possible queries composed from $\{T_i\}_{i=1,2,\ldots,m}$ have the minimal number of disk accesses under parallel retrieving on GMM.

Q.E.D.

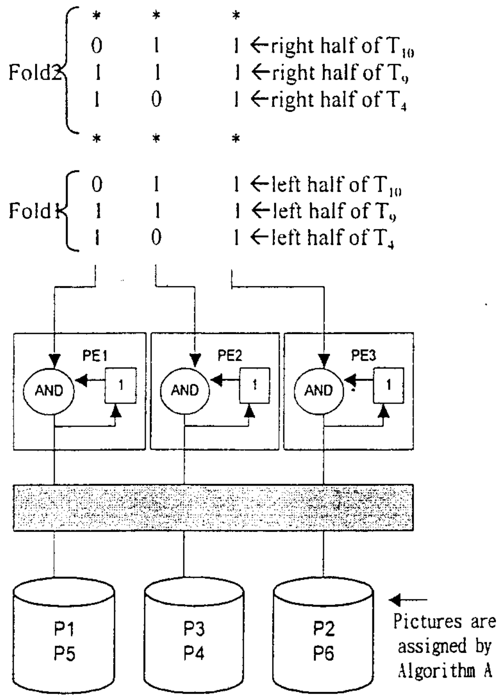


Figure 4.3 An example of triple GMM execution.

Let's recall Example 3.1 $\{P_1, P_3, P_2, P_5, P_4, P_6\}$ forms a CR organization for $\{T_1, T_2, \ldots, P_{10}\}$. Suppose that there are 3 disks with 3 controllers in our processing framework. Let the considered general query Q be $\{T_4, T_9, T_{10}\}$. Then the whole job is performed as figure 4.3. We have the column vector of T_4 is (0, 0, 0, 1, 1, 1), the column vector of T_9 is (0, 0, 1, 1, 1, 1) and the column vector of T_{10} is (0, 0, 0, 0, 1, 1) with respect to $\{P_1, P_3, P_2, P_5, P_4, P_6\}$.

In Fig.4.3, we will find that, PE_2 and PE_3 should be initiated to retrieve P_4 and P_6 after six "AND" operations. By Algorithm A, P_4 and P_6 are assigned into DISK2 and DISK3, thus it requires just one disk access in this example.

Sometimes, the picture set $\{P_i\}_{i=1,2,\ldots,n}$ may not exist a tuple-based CR organization with respect to its pertinent simple query set $\{T_i\}_{i=1,2,\ldots,m}$. If it is the case then we can add some redundant pictures, say (N-n) redundant pictures, into the set $\{P_i\}_{i=1,2,\ldots,n}$ such that $\{PR_i\}_{i=1,2,\ldots,N}$, N>n, forms a consecutive retrieval with redundancy(CRWR for short) organization, where there exist some distinct i and j such that $PR_i = PR_j$.

We can easily verify that if $\{PR_i\}_{i=1,2,\ldots,N}$ takes place of $\{PR_i\}_{i=1,2,\ldots,n}$ in Algorithm A then the optimality is still kept. The investment for keeping the optimality of retrieval time is the addition of redundant pictures.

We may wish to find a CRWR organization which has minimal redundancy. This problem is called the consecutive retrieval with minimal redundancy. However, this problem has been shown to be NP-complete in [Kou 1977]. In 1973, Ghosh had developed an algorithm for finding a CR organization with bounded redundancy. In 1984, Deogun, Raghavan and Tson proposed a more efficient approximation greedy algorithm to find the CRWR organization.

5. Concluding Remarks

In this paper, we present the concept of tuple-based CR organization. Based on this concept, we proposed a processing framework GMM such that the spatial match query can be performed not only concurrently but also efficiently. On this GMM framework, we also proposed an algorithm that can achieve the optimal arrangement of symbolic pictures into disks of GMM machines with minimal number of disk access times for all possible spatial match queries.

6. References

1. Al-Fedaghi and Chin, Y. H. (1979): "Algorithmic Approach to the Consecutive Retrieval Property", International Journal of Computer and Information Science, Vol.8, No.4, 1979, pp.279-301.

2. Chang, C. C. and Buehrer, D. J. (1992): "De-clustering Image Database", Future Database '92, Advanced Database Research and Development Series, Vol.3, World Scientific Publishing Company, 1992, pp.142-145.

3. Chang, C. C. (1991): "Spatial Match Retrieval of Symbolic Pictures", Journal of Information Science and Engineering, 1991,pp.405-422.

4. Chang, C. C. and Lin, J. C. (1990): "A Fast Spatial Match Accessing Scheme for Symbolic Pictures", Journal of Electrical Engineering, Vol.33, No.3, 1990, pp.129-137.

5. Chang, S. K., Shi, Q. Y. and Yan, C. W. (1987): "Iconic Indexing by 2D strings", IEEE Transactions on Pattern Recognition and Machine Intelligence, Vol.PAMI-9, No.3, 1987, pp.413-428.

6. Deogun, J. S., Raghavan, V. V. and Tsou, T. K. W. (1984): "Organization of Clustered Files for Consecutive Retrieval", ACM Transactions on Database Systems, Vol.9, No.4, December 1984, pp.646-671.

7. Dinning A. (1989): "A Survey of Synchronization Methods for Parallel Computers", IEEE Computer, Vol.22, No. 7, July 1989, pp.66-77.

8. Duncan R. (1990): "A Survey of Parallel Computer Architectures", IEEE Computer, Vol.23, No.2, Feb. 1990, pp.5-16.

9. Feotelson D. G. and Rudolph L. (1990): "Distributed Hierarchical Control for Parallel Processing", IEEE Computer, Vol.23, No. 5, May 1990, pp.65-77.

10. Ghosh, S. P. (1972): "File Organization: The Consecutive Retrieval Property", Communication of ACM, Vol.15, 1972, pp.802-808.

11. Ghosh. S. P. (1973): "On the Theory of Consecutive Storage Retrieval Records", Information Sciences, Vol.6, 1973, pp.1-9.

12. Ghosh, S. P. (1977): "Database Organization for Data Management", Academic Press, N. Y., 1977.

13. Kou, L. T. (1977): "Polynomial-complete Information Retrieval Problems", SIAM Journal of Computing, Vol.6, No1, March 1977 pp.67-65.

14. Lee, S. Y., Shan, M. K. and Yang, W. P. (1989): "Similarity Retrieval of Iconic Image Database", Pattern Recognition, Vol.22, No.6, 1989, pp.675-682.