IJCCCE, VOL.7, NO.2, 2007

PROPOSED DATABASE IMAGE INDEX KEY EXTRACTING ALGORITHM BASED ON IMAGE CONTENTS (IIKE)

*Dr. Abdul mon'em S. Rahma & * Dr. Emad k. Jbbar

Receipted: 1 / 10 / 2007
Accepted:10 /2/ 2008

Abstract

The images have large amount of information and they are important object in modern database system, so there is a great need for efficient image indexing key as access and retrieval tool. In database systems the index key is an attribute which has unique value for all tuples of the relation by " name" , " numbers" or any codes, this key aids to identify the tuples. In archive system a vast amount of digital images it is very necessary to find unique image index key to store it corresponding to its image and can use it to retrieval documents. So in the proposed algorithm we generate image index key which depends on extracting the key from the contents of the image itself, so such an image index key will contain the properties and characteristics of image, it is Unique, because its value depends on the value of image pixels. So it is a good method for image indexing , image verification , image ordering and checking the authenticity of retrieved image to prevent any juggle perhaps happening with the image by recomputing index keyword of retrieved image and comparing it with index keyword which is stored in database. Using shrinking method With Image Index Key Extracting algorithm (IIKE) reduce storage space of stored images.

الخلاصة

ان الصور تمتلك كميات هائلة من المعلومات وهي من الاشياء المهمة في انظمة قواعد البيانات الحديثة لذا برزت حاجة كبيرة لمفتاح فهرست, للوصول الى هذه الصور واسترجاعها. ان مفاتيح الفهرست في انظمة قواعد البيانات هي عبارة عن حقل او اكثر من حقل تمتلك قييمة وحيدة لكل قيود علاقات معينة ويكون هذا المفتاح اما ارقام او اسم او أي رمز يساعد على تعريف القيود والوصول لها. في انظمة الارشفة الحديثة توجد كميات كبيرة من الصور الرقمية التي تحتاج الى مفتاح فهرست خاص لكل منها يخزن مقابل الصورة (الوثيقة) في قاعدة البيانات لتسهيل عملية استرجاع الوثيقة والتعامل معها وان الطرق الاعتيادية في تحديد مفاتيح الفهرست تعتمد على اختيار حقل او اكثر من حقول قاعدة البيانات ولكن الخوارزمية المقترحة تعتمد في توليد مفتاح الفهرست على استخلاص هذا المفتاح من مكونات الصوره (الوثيقة) نفسها وهذا يعني ان المفتاح يحتوي على خصائص وصفات الصورة ويكون وحيد لان قيمة المفتاح تعتمد على قيمة النقاط المكونة للصورة. ان مثل هذا المفتاح يعتبر طريقة جديدة لفهرست الصورة , التحقق المنطقي , الترتيب وتدقيق صحة وسلامة الصورة (الوثيقة) عند استرجاعها من أي تلاعب قد يحدث بالصورة وذلك من خلال اعادة احتساب المفتاح للصورة المسترجعة ومقارنته مع المفتاح المخزون بالقاعدة فاذا لم يتطابقا فان الوثيقة بها تلاعب . لقد تم استخدام خوارزمية التصغير للصورة اولا ثم استخلاص المفتاح حسب الخوارزمية IIKE وذلك مساحة الخزن للصورة وتقليل طول مفتاح الفهرست الناتج.

* Dept. of Computer Sciences University of Technology, Baghdad-Iraq

1-Introuduction

Most of modern database systems use images since they show a large amount of information and are easy to store without data entry. The amount of digital images available for the office document is rapidly growing. (They include biomedical purposes, scientific experiments charts, personal and official information documents). There is a great need for efficient image indexing and access tools in order to fully utilize this massive digital resource. Image retrieval is a research area dedicated to address this issue and substantial research efforts have been made. [1]

The earlier image retrieval systems have all taken keyword or text-based approaches for indexing and retrieval of image data. Because image annotation is a tedious process, it is practically impossible to annotate all images on database without index. Furthermore, it is also difficult to make exactly the same image annotations retrieval. To address those limitations, Content-Based Image Retrieval (CBIR) approaches have been studied in the last decade. These approaches work with description based on properties that are inherent in the images themselves such as color, texture, and shape and utilize them for retrieval purposes. Since visual features are automatically extracted from images, automated indexing of image databases becomes possible. [2]

The advantages of a database system over traditional, paper - based methods of record keeping are perhaps easier to see in these cases. Here are some of them:

***ompactness*:** There is no need for possibly voluminous paper files.

Speed: The machine can retrieve and update faster than a human.

Less drudgery: Much of the sheer tedium of maintaining files by hand is eliminated mechanical tasks are always better done by machines.

Currency: Accurate, update information is available on demand at any time.

Protection: The data can be better protected against unintentional loss and unlawful access.

Centralized control: The database provides the enterprise with centralized control of its data.

2- Database Indexing

In the early stage of data processing, there was no established conception of indexing. Most of the data files were accessed sequentially. This was pretty slow and inefficient particularly when the data file is big enough. To get rid of this problem, the concept of indexing came to the picture. In the initial stage, a number is used to be given against each record by the system in a file created on disk. We could specify these numbers to access any record randomly. Then came the concept of Indexed Sequential Access Method where instead of assigning separate numbers against each record, a field or a combination of fields were used as a key. There are two kinds of indexing, one where the key value in a particular file is unique and the other where the key value could be a duplicate. The search method is called Binary Search where to find a particular key value, the whole file is divided into two halves and the part, which contains the particular key value we are searching, is taken and then divided into two halves again. The part here which contains the key value is again taken and divided into two halves. This process continues until it finds the match to the key value. [3]

Nowadays an old file system is hardly used in maintaining computer records. Instead, a database system has been developed. The latest development in this field is Relational Database System, which contains tables to store data. In our problem of dealing with multimedia databases, which would contain images, the concept of indexing is very important. We are trying to develop a more sophisticated method of indexing where there won't be any clear-cut definition of index against the images, but indexes would be defined based on content of image itself, that implies generating unique key.

3- The Database Indexing Key

A key is an attribute which has a unique value for all tuples of the relation. That is given a tuple with particular key, no other n-tuple belonging to that relation has that key. The purpose of the key in the relational database is to identify the tuple containing, that is, an individual and his record are corrected by the value of the key in the tuple and of the key value of the attribute for the individual in real word. For example given the name of an individual we can find his tuple if "name" and "numbers " are a key component for the relation. Conversely, having located records which fulfill certain of our needs we can identify the individuals these records correspond to by means of the key found in the tuple.

The key relates data about an individual to the individual, it may also be used to determine the sequence of records in the file. To do this we place ordering requirement upon the key and the code which replaces it on the physical medium so that, given two keys, we can determine which is the larger. This in turn, determines which should precede the other in the file. Other criteria can be used to determine order. For instance, for a transaction file, the transaction may be placed in the file according to its date and time of arrival. This may be regardless of the transaction key. Even if the transaction contains no key, the arrival

time may be inherent criterion that is said, it may be part of the data comprising that record. So it can say that a key serves two purposes: [3]

- It identifies a record so that the system can return the quested record to the user.
- It is the means for ordering records.

4- Image Representation

At the core of any description of natural images is the choice of a suitable image representation. There are, of course, many different image representations to choose from. The choice should be made based on their effectiveness in revealing statistical regularities in natural images. The simplest representation, for example, is an intensity-based approach, where the representations is simply the original intensity values. An $n \times n$ grayscale image is considered as a collection of $n * n$ independent samples of intensity values. Similarly, an $n * n$ RGB color image is represented as a collection of n^2 independent 3D vectors. [1]

In order to be in a form suitable for computer processing we must quantize the image. An image function f (x, y) must be digitized both in space and amplitude. Digitization of the spatial coordinates (x, y) will be referred to as image sampling while amplitude digitization will be called gray – level quantization. Suppose a continuous image f (x, y) is arranged in the form of an N * N array as follows: [4]

$$f(x,y) = \begin{matrix} f(0,0) & f(0,1) \ldots & f(0,N-1) \\ f(1,0) & f(1,1) \ldots & f(1,N-1) \\ f(2,0) & \ldots\ldots\ldots & \\ & \ldots\ldots\ldots & \\ f(N-1,0) \ldots & & f(N-1,N-1) \end{matrix}$$

The right side is representing what is commonly called digital image, while each element of the array is refereed to as an image element, picture element, and pixel.

The above digitization process requires that a decision be made on a value for N as well as on the number of discrete gray levels allowed for each pixel , it is common practice in digital image processing to let these quantities be integer powers of two that is N = 2 and G = 2 ** m
where G denotes the number of gray. The number of bit (b) required to store a digitized image is computed as follows b = N * N * m
For example
512 x 512 images with 32 gray levels requires 1,310,720 bit of storage
Since b = N * N * m
N = 512
m = 5 from G = 2 **m, 32=2**m
Then b = 512 * 512 * 5
= 1,310,720

The value represents the minimum number of bytes needed for each value of N and m. When no overlap is allowed in our example we use m = 5 that means only one pixel is stored in byte , even though this leaves three unused bits in the byte. Now we ask at this point how many samples and gray levels are required for a good approximation of S. The resolution of an image is strongly dependent on both N and m. the more these parameters are increased , the closer the digitized array will approximate the original image.
So from b = N * N * m we see:

The storage increases rapidly as a function of N and m.
It is interesting to consider the effect that variations in N and m have on image quality. As might be expected a good image is difficult to define because quality requirements vary with application.[5]

If N is reduced and m is kept fixed the image equality deteriorates rapidly for the value of N. The effect produced by reducing the number of bits is used to represent the gray levels in an image and keep N fixed. The 256 to 64- levels are of acceptable quality. The 32 level images have some mild smooth background area above the subject right shoulder. This effect is considerably more pronounced in the image displayed in 16- levels and increases sharply for remaining images. The number of samples and gray levels required to produce faithful reproduction of an original, image depends on the image itself as basis for comparison. So the effects produced on image quality by varying N and m independently. From the above notes we find using one of image compression algorithms is the best way to reduce image size to store shrunk image in database.

5- <u>Proposed Algorithm for Image Index Key Extracting (IIKE)</u>

In an archive systems with an amount of digital images that are rapidly growing, there is a great need for efficient image indexing key as access tool in order to fully utilize this massive digital resource, and to retrieve stored database images. The traditional index key in database like document numbers, document number and date , sequence number, codes and others is just codes to image coding since it dose not extract from the image contents.

The proposed algorithms depend on extracting image index key based on the image contents itself, so such as image index key will:

- Contain the properties and characteristic of image.
- Be unique
- Its value depends on the value of image pixels.

So it is a good method for image indexing, image verification, image ordering and check on the confidence of retrieval image.

The following steps represent Proposed Algorithm Image Index Key Extracting (IIKE)

Step1 scanning the official document and store it in BMP format.

Step2 shrinking the store image of *step 1*.

Step3 using proposed algorithm with image in step 2 to generate (IIKE).

Step4 storing *image index* keyword in database with its image corresponds.

Now let us have the following pixel values Fig (1) of any image.
How will the image index key be extracted?

220	210	210	180	210	220
230	180	190	200	200	230
225	185	140	180	190	210
180	220	190	140	160	200
210	190	180	160	170	180
220	200	210	180	190	210
190	210	220	190	200	240
190	200	210	190	210	220

Fig (1) Image pixels value

Notes: *keyword* means the image index key extracting (IIKE)

From the pixel values of Fig(1) one can generate the image index keyword as follows:

1- Divide the image size for four square matrix, by completing every sub sample with Zeros row or column if needed

2- Compute value of Keyword

Value (Keyword) $=\Sigma p_1(x, y) \| \Sigma p_2(x, y) \| \Sigma p_3(x, y) \| \Sigma p_4(x, y)$

$=2380 \| 2320 \| 2430 \| 2340$

3- Convert value (keyword) to Hexadecimal numbers

=94C || 910 || 97E || 924

Then Keyword is 94C91097E924.

This keyword will be unique to this image and any change or juggle in image will change the pixel values and that implies changing keyword value, so when retrieving an image, recomputed keyword value (retrieved image) and if recomputed keyword value is not identical, store keyword value that means the image is not true.

Algorithm: Image Index Key Extracting (**IIKE**)
Input: Database D, Images
Output: Image Index Key
Begin
Image Scanning;
Image Shrinking;
For k=1 to 4
  Compute Value(Keyword[k]) = Value(Keyword[k]) $+\Sigma p_k(x, y)$
  Convert $(\text{Value(Keyword[k])})_{10}$ to $(\text{Keyword[k]})_{16}$
Next k
Keyword = Keyword || $(\text{Keyword[k]})_{16}$
End

Fig (2) Image Index Key Extracting Algorithm

Example

Suppose you have the following official document with 1034 * 738 (height * width) and G=256 RGB color image.

1. ***Generate image index key***
2. ***Check the retrieval image***

The Solution

1. Generate image index key Fig (5)

Step1 Scan the image document will need 6,104,736 bits to store When
b = N x N x m then
1034 * 738 * 8= 6,104,736 bits

Step2 Use one of shrinking algorithms. The image document will need 676,692

Fig (3) Origin document

bits to store When b = N * N * m then 344 *x* 246 *x* 8 = 676,692 bits

Step3 Use the shrinking image document in ***step2*** to generate index keyword by using IIKE algorithm.

Fig (4) shrink document

Image Index key will be (**C7B07C36F8B0EB4E2EAC**).

Step4 Store index keyword (**C7B07C36F8B0EB4E2EAC**) in database record with its corresponding image.

2. Check the retrieved image

Let some changes happen on the store image in database like "جيد جدا" be "امتياز"

The Ministry of Higher Education and Scientific Research
University of Technology
Office of Higher Education

وزارة التعليم العالي والبحث العلمي
الجامعة التكنولوجية
قسم الدراسات العليا

No.:
Ref.:

العدد: د ع / ٩٩٦
التاريخ: ٢٠٠٥/١٢/٢٨

- أمـــــر جامعـــــي -

منح شهادة دكتوراه

بناءا على إكمال الطالب عماد كاظم جبار متطلبات الدراسات العليا/الدكتوراه بنجاح. واستنادا إلى الصلاحيات المخولة لنا من قبل مجلس الجامعة وبناءا على ما جاء بمحضر مجلس قسم علوم الحاسبات بجلسته الخامسة المنعقدة بتاريخ ٢٠٠٥/١٢/١٨ .

تقرر منحه درجة دكتوراه فلسفة في علوم الحاسبات / تخصص حاسبات وبتقدير (امتياز) مع تمتعه بكافة الحقوق والامتيازات التي تخوله إياه هذه الدرجة.

أ.د. قحطان خلف الخزرجي
رئيس الجامعة

نسخة منه إلى/
- قسم علوم الحاسبات .
- قسم الدراسات العليا .

Baghdad - University of Technology - P.O.Box (35010) Tel.:7196023, Tel- Mohammed - Baghdad

Fig (6) Forger document

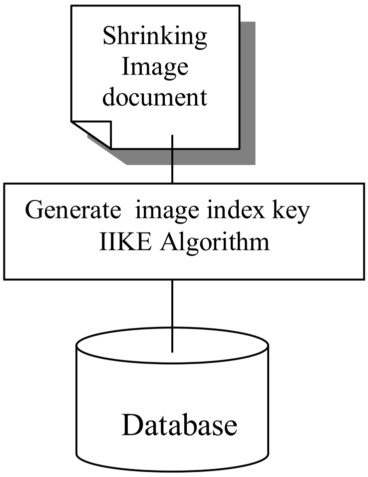

Fig (5) Generate image index key

According to this method, after retrieving the stored image by its index keyword, recomputed the index keyword for image Fig(7) and will be (**C7BO7C36F8B1653E2EC9**). If recomputed index keyword is not identical to database index keyword storage, the image is not true. Else the retrieved image is true so the confidence of retrieved image is 100%. On the other hand, if the index keyword is any numbers or codes not extracted from image contents one can not be sure of the confidence of the image store.

Fig (7) Shinning Forger document

Fig (8) Check retrieval image Shrink Forger document

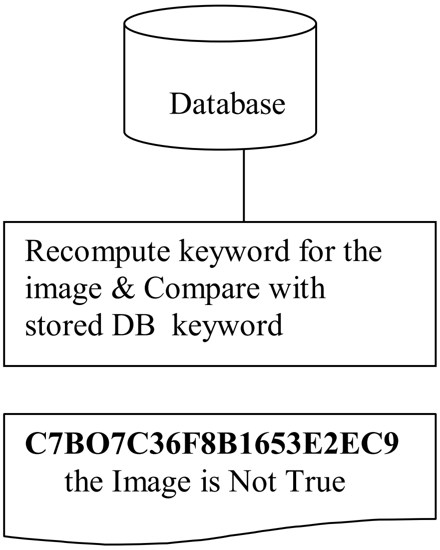

Fig (9) forger document retrieval

9nts of other image even it similar with some of information because it will be difference by other information like stamp location, signature shape, picture, …. So the contents of document image be unique key

7- Conclusion

The images in database systems are rapidly increasing and it is easy to change. Retrieved record from database needs to index keyword and the ordinary index keyword "name" or "numbers" haven't any ability to check the authenticity of retrieved image because it just codes image, but this our method involves generating index keyword by extracting its value from contents of image itself and that aids to:

- check the authenticity of retrieved image
- use keyword image to retrieve image and sort database image as any index keyword
- prevent any juggle perhaps from happening with the image by recomputing index keyword of retrieved image and comparing it with index keyword which is stored in database.
- Use shrinking method to reduce storage space of image.

References

[1] Mingjing Li , Zheng Chen, Hong-Jiang Zhang "**Statistical Correlation Analysis in Image Retrieval",** Microsoft .Research China {mjli, zhengc, hjzhang} @microsoft.com

[2] Sagarmay Deb , **"Content-based Image Retrieval Based On Emergence Index",** A dissertation submitted to The Department of Mathematics and Computing Faculty of Sciences the University of Southern Queensland For The degree of Doctor of Philosophy, 2003

[3] Date C. J. **"An Introduction to Database System "** , Pearson Education , USA ,2004

[4] Gonzalez Rafael C. **" Digital Image processing** ", Addison Wesley Publishing, USA, 1987 Scott V. " **Computer Vision and image processing " ,** Prentice Hall, 1988

[5] Scott V. **" Computer Vision and image processing " ,** Prentice Hall, 1988