6. Liu Hong, Yang Lianhong, Wang Chao, Study on temperature control of greenhouse based on fuzzy control. Journal of Changji University, 2013.(6):77-80
7. Lan Fujun, Research on intelligent greenhouse temperature control based on fuzzy control and neural network.Anhui Agricultural Sciences, 2012.40(7):4437-4438,4441.
8. Ren Zhimiao, Present situation and Prospect of fuzzy control system,Shanxi Electronic Technology, 2011.(2):89-90.
9. Yang Jing, Zhang Lei, Design and implementation of agricultural greenhouse control system,Guangdong Agricultural Sciences, ,2011.(4):166-167.
10. Park D H, Kang B J, Cho K R, Shin C S, Cho S E, Park J W, Yang W M. A Study on Greenhouse Automatic Control System Based on Wireless Sensor Network. Wireless Personal Communications, 2011.56(1): 117-130.
11. Ren Wenttao, Xiang Quanli, Yang Yi. Implement of Fuzzy Control for Greenhouse Irrigation. IFIP International Federation for information processing,2011.344: 267-274
12. Qiu Ying, Tan Ding Zhong. Greenhouse control system based on WSN. Key Engineering Materials, 2011.486: 254-257
13. Tang Binyong, Lu Linji, Wang Wenjie, Fuzzy control theory and Application Technology, Tsinghua University press,2010.
14. Yang Weizhong, Wang Yiming, Li Haijian, Online self setting algorithm for fuzzy control parameters of greenhouse temperature, Journal of agricultural machinery, 2005,36(9):79-82.
15. Lei Jinli, Study of temperature control system of deaerator based on Fuzzy Theory , Xi'an: Xi'an Electronic and Science University, 2005

Research on a Fast Algorithm for Mining Association Rules Based on Vertically Distributed Data in Large Dense Databases

Xia Runliang, Feng Xingkai

Yellow River Institute of Hydraulic Research, Zhengzhou, china

Corresponding author is Xia Runliang

Abstract
In this paper, we prompt a new a fast algorithm for mining association rules based on vertically distributed data in large dense databases. In order to calculate item sets support, this paper puts forward the concept characteristic matrix and characteristic vector, and emerges an algorithm for mining association rules based on the characteristic matrix. As a result of drawing the advantages of CARMA(continuous association rule mining) algorithm,the algorithm needs to scan the database for only twice. Experimental results show that the algorithm is correct,and in the large dense transaction databases, VARMLDb algorithm has higher implementation efficiency.
Keywords: MULTI NODE COOPERATE. FAST ALGORITHM, MINING ASSOCIATION RULES, VERTICALLY DISTRIBUTED DATA, LARGE DENSE DATABASES

1. Introduction

With the rapid development of the information society and the arrival of the era of big data, various application systems data is being accumulated in the explosive growth by geometrically. In order to make better use of data, people eager to dig out some regular or more valuable information from large-scale data, so data mining techniques emerge [1].

Association rule mining as a key technology in the field of data mining has attracted wide attention, with far-reaching significance and practical value. Currently, domestic and foreign scholars have proposed many algorithms and made a lot of research for the association rule mining; however, frequent item sets mining still have some deficiencies, which have three main bottlenecks [2]:

(1) The speed of data processing is not high, the process of solving is slow;

(2) Mining process occupies a larger space, resulting in a large number of intermediate set of frequent item sets.

In order to solving the slower problems of computing support, the paper proposed association rule mining algorithm research based on bit operation. Firstly, the database is converted to vertical data format, and using a two-dimensional array to store binary; Secondly, pruning the candidate collection is used. Thirdly, the using the K-item sets frequent sets combined into K-item sets candidate collection; finally, using depth-first search algorithm to determine the entire frequent item sets. Experimental results show that the algorithm can effectively simplify the calculation of support and improve the efficiency of the algorithm.

The paper proposes an association rule mining algorithm based on different set in order to avoiding the waste of memory. Firstly, calculating all item sets support using bit operation. Secondly, in order to obtain higher memory utilization, the paper divides 2-item sets into several groups according diffsets. Thirdly, Generate k-item sets (k> 2) from a different grouping which can effectively reduce the time of determining whether item set is frequent.

Experimental results show that the algorithm is effective in reducing the number of frequent candidate generating sets and improve the efficiency of the algorithm. This Paper proposes two algorithms simplify the calculation of support and improve memory utilization, which improve the traditional frequent item sets mining efficiency.

Mining association rules is one of the most active methods in data mining field. It is also one of the most important and widely used methods in this area currently. A large number of scholars have made much improvement on the basis of this algorithm. The algorithm replaces the original transaction database with shrinking Tide table to improve the search efficiency. Partition algorithm introduces the idea of parallel mining. Sampling algorithm makes compromise for mining accuracy and efficiency. DHP algorithm uses hashing to improve the efficiency of candidate item sets (especially the 2-candidateitemset) formatting process. However, these algorithms are all based on Apriori algorithm [3] which uses continuous iterative procedure to generate frequent item sets. Although these algorithms have taken some measures in pruning candidate item sets, however, a large number of useless candidate item sets will still be produced when database is too huge. In addition that Apriori algorithm needs to scan the database repeatedly. These features have become the bottleneck restricting efficiency of the algorithm. FP-Tree algorithm is a kind of association rule algorithm that does not generate candidate item sets. This algorithm has greatly improved the efficiency of generating frequent item sets, but disadvantage is that the algorithm takes too much memory. Especially, the cost is great or even it is impossible to establish FP-tree in the memory when the number of transactions is too large. Yen first proposed the graph-based association rule algorithm Direct Larger Item set Generation (DLG) [4]. DLG algorithm translates frequent item sets mining into association graph searching. This algorithm scans database only once, omitting the connection steps in apriori algorithm, reducing the size of candidate item sets, so the efficiency is improved [5]. In this paper, an improved algorithm fast algorithm based on compresses candidate item sets with the combination of node degree in association graph and the relevant characteristics of frequent item sets. And experiment is done to prove the effectiveness of the improved algorithm.

2. The mathematical model of fast algorithm

The operation and maintenance management of information communication database mainly refers to timely discovery, locating and handling of any database fault to ensure smooth and efficient operation as well as guarantee in major emergencies pertinent to database operation, complaints about database quality from customers, assessment and analysis of database quality, prediction of planning, construction, and so forth. The time consumed during fault location and judgment in the application layer of a large-scale database accounts for 93% of its total time for failure of recovery. The huge database structure and multifunctional device types also bring about large

amounts of alarm data due to such characteristics of the information communication database as topological structure densification, database device microminiaturization, communication board precision, and so forth. Therefore, the foundation of the database operation and maintenance is the effective management of the database alarms.

As an important supporting means for database operation and maintenance management, database management system directly influences the quality of service which the information communication database provides to its customers. The database management system is developing toward integrated service database management update from independent device database management, manufacturer device database management, and integrated professional database management. The centralized monitoring management function of the professional information communication database operation management will make problems exhibit a sharp full data increasing, including database faults, device alarms, and customer complaints.

As the information communication system consists of various medium interlinked database devices and operating systems implicit and complex-correlated logic is ubiquitous among database elements; that is, a certain fault point may trigger numerous alarms in the whole database. The sudden intensive alarms not only consume the resources of the database management system but also obscure the position of the database fault source point's thus severely impeding trouble shooting by the database operation and maintenance personnel. Several alarms are incorporated into a single alarm or source alarm with a large amount of information by such links as paraphrasing and explaining, eliminating and filtering, information integration, and correlating and transforming, and so forth. It aims at assisting the operation and maintenance personnel to analyze fault messages and locate faults quickly that is, mining analysis on alarm association rules.

In the research field of data mining, the research on association rules is deeper and more extensive. The focus of the research is to find frequent item sets (frequent-sets). There are numbers of typical algorithm, such as Apriori algorithm and DHP algorithm proposed by R.Agrawal et al. They have improved the production process of candidate set Ck by using Hash technology. These algorithms are database traversal algorithm. A method of database segmentation algorithm was proposed by Savasere in 1995, the algorithm has reduced the number of times of I/O in the mining process and has lightened the burden of CPU. H. Toivonen has found out the association rules from large databases by sampling method, this method costs less. These algorithms have improved mining process of association mining algorithm based on the characteristic matrix of database; it calculates the item set support by the calculation of feature vectors. It simplified the process of calculation of the support so that the efficiency of association rules mining algorithms has been improved.

Association rules can be described with mathematical model. A set of letters I={i1,i2,…,im} is called the set of data items, the data items collection is called item-set, D is a collection of all the affairs, business T is a subset of I, i.e. $T \subseteq I$, each transaction consists of an only TID mark. Association rules have the following form: $X \Rightarrow Y$, and $X \subseteq I, Y \subseteq I$, $X \cap Y = \varnothing$, X is the conditions of rules, Y is the results of rules. If a record contains X, Y, the record will met rules $X \Rightarrow Y$. For $X \subseteq I$, if the record number of X which is included in D is s, then the support of X: support (X)=s

Confidence can be described as strength of rules, defined as:

$$confidence(X \Rightarrow Y) = \frac{\sup port(X \cup Y)}{\sup port(X)} \quad (1)$$

Association rules mining algorithm is usually divided into the following two steps:

(1) Find frequent item sets (frequent item-sets) which are a set of data items that its support is greater than a given value.

(2) Generate candidate association rules by using the frequent item sets, and verify the credibility of the rule.

Step (1) is the key to improve the efficiency of the algorithm. First of all, there are the definitions as follows:

Definition 1: the attribute domain r of relational database is BOOL variables, matrix M is the vector matrix of a database:

$$M_{i\times j} = (m_{ij})_{i\times j} = \left[\overrightarrow{m_1}, \overrightarrow{m_2}, ..., \overrightarrow{m_j}\right] \quad (2)$$

$$\begin{cases} m_{ij} = 1 & attibute \quad J \subseteq I \\ m_{ij} = 0 & attibute \quad J \not\subset I \end{cases}$$

Which I is the number of records in database, J is the number of BOOL variables, each column vector $\overrightarrow{m_j}(j = 1,2,...,J)$ which corresponds to the attribute is the feature vector of attribute.

Definition 2: the inner product is described as k n-dimensional characteristic vectors in K elements function:

$$\left\langle \overline{R_1}, \overline{R_2}, ..., \overline{R_k}, \right\rangle = \sum_{i=1}^{k} r_{1i} \times r_{2i} \times .... \times r_{ki} \quad (3)$$

Definition 3: If 1-frequent item sets L1 of the database D is not empty, then the matrix composed by order of the feature vector of element in L, is defined as the characteristic matrix of database.

Theorem 1 support (X) of element X in k frequency set of database D is equal to k elements inner product of feature vector which k attributes form X, that is:

$$\sup pose(X) = \left\langle \overrightarrow{x_1}, \overrightarrow{x_2}, ..., \overrightarrow{x_k} \right\rangle \quad (4)$$

The proof of theorem l: If there is k frequent item sets in record U of X, then $U(x_1), U(x_2), ..., U(x_k)$ is real at the same time, so $U(x_1) \times U(x_2) \times ... \times U(x_k) = 1$; Otherwise it is zero, $n = \left\langle \overrightarrow{x_1}, \overrightarrow{x_2}, ..., \overrightarrow{x_k} \right\rangle$ is the number of k property set in D. Theorem 1 has been proved.

Theorem 2 property set A belongs to the k- frequent item sets Lk, if X⊆A, then support(X) will notless than the stated threshold of k project set.

The proof of theorem2: According to X ⊆ A ⊆ I (l is the attribute collection in database D), with the functional dependency A → X, then record U, V are in D. If U(A) = V(A), then U(X)=V(X). So if A meets the conditions, X will also meet them, otherwise it won't bring into existence.

So support(X) > support(A)

sup port(X) > supportmin because of support (A) > supportmin.

Step1: Detection of Scale-Space Extreme. The database is convolved with Gaussian filters at different scales, and then the difference of successive Gaussian-blurred databases is taken. Key points are then taken as maxima/minima of the Difference of Gaussians that occur at multiple scales.

$$D(x, y, \sigma) = L(x, y, k\sigma) - L(x, y, \sigma) \quad (5)$$

Where $L(x, y, \sigma)$ is the convolution of the original $I(x, y)$ with the Gaussian blur $G(x, y, \sigma)$ at scale $k\sigma$, k is a constant multiplicative factor and we have:

$$L(x, y, k\sigma) = G(x, y, k\sigma) * I(x, y) \quad (6)$$

Step 2: Accurate Key points Localization. Key points are identified as local minima/maxima of the across scales. This is done by comparing each pixel in the databases to its eight neighbors at the same scale and nine corresponding neighboring pixels in each of the neighboring scales. Then using the (3) to calculate the D(X) of these key points and reserve the extreme value that is greater than 0.03 as the candidate key.

$$D(X) = D + \frac{\partial D^T}{\partial X} X + \frac{1}{2} X^T \frac{\partial^2 D^T}{\partial X^2} X \quad (7)$$

Where D and its derivatives are evaluated at the candidate key point and $X = (x, y, \sigma)$ is the offset from this point. Finally, in order to increase stability, the Hessian matrix is used to remove the unstable edge response points.

Step 3: Orientation Assignment. In this step, according to the scale of the key point and the location information, the direction of the coefficient is calculated using the following formula:

$$D(X) = D + \frac{\partial D^T}{\partial X} X + \frac{1}{2} X^T \frac{\partial^2 D^T}{\partial X^2} X \quad (8)$$

$$m(x, y) = \sqrt{\begin{matrix} (L(x+1), y) - L(x-1, y)^2 \\ + (L(x, y+1) - L(x, y-1))^2 \end{matrix}} \quad (9)$$

$$\theta(x, y) = \tan^{-1} \frac{L(x, y+1) - L(x, y-1)}{(L(x+1), y) - L(x-1, y)}. \quad (10)$$

This algorithm is based on the direct data instead of a database management system (DBMS). The first is to traverse the database, make the database vector. Identify records in the database by computing the vector product of vectors. By Theorem 1, we can obtain support degree of candidate elements. The calculation of support degree steps is simplified, the speed of discovery the frequent item sets can be improved; it leads to improve of algorithm of mining association rules. Because data mining in large scale database (Very large database), database vectorization can only make the data more compact, at the same time, it requires large memory, so this algorithm to compute K frequent item set L, is based on the reduced dimension feature matrix, The algorithm occupied a smaller memory. If the database is too large, we can also use the database segmentation vectorization.

3. The improved algorithm of mining association rules based on vertically distributed data

To describe conveniently, we name the improved algorithm as fast algorithm. Before introducing the specific algorithm, we dig out some useful theorems in mining association graph, detailed as follows:

Theorem 1: In association graph, if node degree is smaller than k-1 when we generate k-item sets, this node can not be extended to k-frequent item sets. The node associated with its edge can be removed from the association graph

Proof: Suppose that the node can be extended to k-frequent item sets, considering all the 2-subsets of frequent item sets which contain the node, according

to the law that any subset of frequent item sets is frequent item sets, all the k-1 2-subsets are frequent item sets. Then there are k-1 edges between this node and the other k-1 nodes, contradiction arising because the degree of this node is smaller than k-1. So the former part of the theorem is established. Then we prove why we can delete the node and its connected edges from the association graph. We just need to prove that the node showing in higher frequent item sets is impossible. Suppose the node can be extended to higher frequent item sets, denoted by (k+m)(m>0), thus there are at least k+m-1 edges connected with the node. We have already known the number of edges connected to this node is not more than k-1. Therefore, it is impossible that the node appears in the (k+m)-frequent item sets (m>0). So we can remove the node from the association graph, with no impacts to the future generation of frequent item sets

Theorem 2: In association graph, when we make extensions to known k-frequent item sets, if we find out at least one edge does not exist from one item in k frequent item sets to the node to be extended, we should not make this extension. That is because the extended (k+1)-item set will not be a frequent item set

Proof: Suppose the extended (k+1)-item set would be a frequent item set, then the (k+1)-item set would have k 2-subsets containing the extended node, and all the 2-subsets would be frequent item sets. Then, it is inevitable that there will be edges between the k nodes and the extended node, as is shown in Fig. (1).

Here we have a specific example to demonstrate the procedure of the algorithm. When we compute the support of the 3-frequent item sets in Fig. (l), the degree of h is smaller than 2. Thus we delete h and its connected edge, and we get figure (2). In Fig. (2), edge {I2, I3} can be extended to {I2, I3, I5} and {I2, I3, I4}, but the latter extension {I2, I3, I4} is not necessary. That is because there is no edge between I2 and I4. Thus item set {I2, I3, I4} is not a frequent item set. Edge {I3, I4} can be extended to {I3, I4, I5} which is proved to be another 3-frequent item set. Finally we find out two frequent item sets {I2, I3, I5} and {I3, I4, I5}. Now we know we need not extend this edge to 4 frequent item sets, because the degree of I2, and I4 is smaller than 3. Then we can delete them from association graph, after that we get Fig. (3) which contains only two nodes, and mining procedure is over.

Description of FAST Algorithm:

1. Scan the database and create transaction matrix
2. Find out all the 2-frequent item sets
3. Create association graph
4. Find out all the k-frequent item sets (k>2), and compute the degree of the nodes. If the degree is smaller than k-1, delete the node and its connected edge, applying theorem 2 when extending the (k-1)-item sets.
5. Execute k++. If the number of nodes in association graph is smaller than k (theorem 1), repeat step 5, else the algorithm is over.

4. The experiment and data analysis

The following is the illustration of the new algorithm with an example, the threshold is set to 2, find all frequent item sets. Database D is shown in Table 1:

Table 1. The Database D

TID	Items
100	ACD
200	BCE
300	ABCE
400	BE

Step 1: the vector matrix in Table 1:

$$M_{4\times5}=\begin{bmatrix}1&0&1&1&0\\0&1&1&0&1\\1&1&1&0&1\\0&1&0&0&1\end{bmatrix}$$

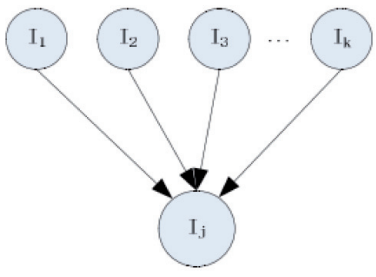

Figure 1. The Condition of Theorem 2

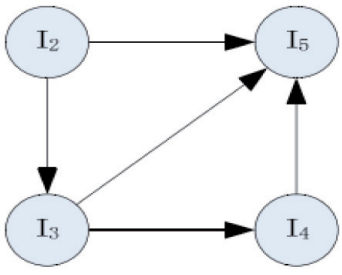

Figure 2. The Association Graph for Sample

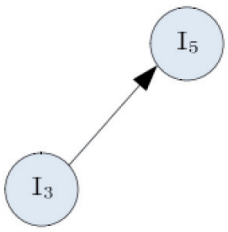

Figure 3. The Association Graph for Sample 2

Let $C_1 = \{A, B, C, D, E\}$

$$S(C_1) = (1 \quad 1 \quad 1 \quad 1) \begin{bmatrix} 1 & 0 & 1 & 1 & 0 \\ 0 & 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \end{bmatrix} = [2 \quad 3 \quad 3 \quad 1 \quad 3]$$

Step 2: According to the theorem 2 and L1, construct candidate $C_2 = \{AB, AC, BC, BE, CE\}$. In order to express the operations of seeking support in matrix form, we define operation $\otimes$: A, B are both N dimensional column vector.

We make the comparison between DLG and fast algorithm through experiments. In Fig. (4), we set the number of transactions for 1000, and we test two algorithms with different support. In Fig. (4), we set the support for 2%, we also test two algorithms with different number of transactions. As the Fig. (4) and Fig. (5) show, under the conditions of different number of transactions and support, the fast algorithm has a better performance.

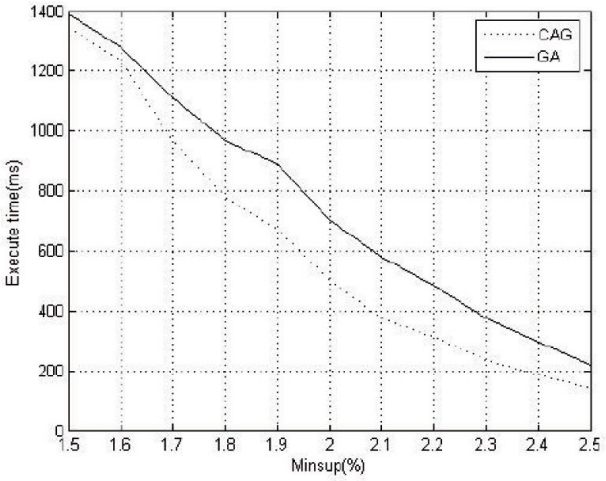

Fig. (4). The Comparison under Different Support

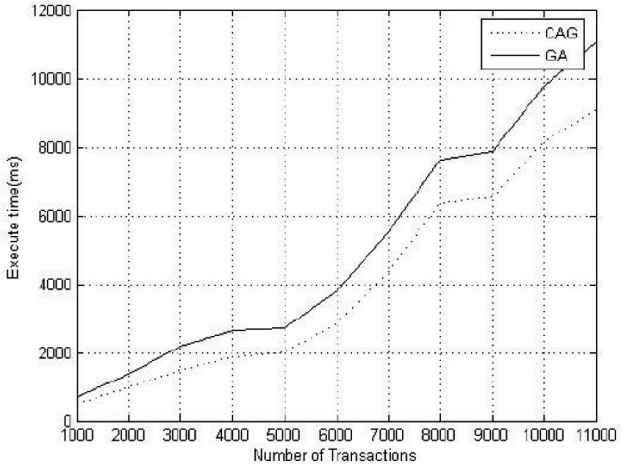

Fig. (5). The Comparison under Different Transactions

5. Conclusions

In this paper, the association rule mining algorithm based on DLG is put forward, some possible improvement is proposed. On the one hand, we propose a method to compress the association graph by digging out the potential relationship between node degree and frequent item sets; on the other hand, according to the law that any subset of frequent item set is a frequent item set, we restrict the extension of the candidate item sets. However, fast algorithm also has its disadvantages. When most transactions contain nearly all the items or the support is fairly low, fast algorithm can not improve the efficiency obviously. That is because under this condition, the association graph is close to be a complete graph and the condition to compress the association graph is hard to satisfy.

Acknowledgment

This work is supported by the National Natural Science Foundation of China(No.51309113), National Science and Technology Support Program (No.2013BAB05B01-05),and Ministry of water resources public benefit projects (No.201401033).

References

1. Szathmary, L, Napoli, A, Valtchev, P.(2007). Towards rare itmeset mining. In International Conference on Tools with Artificial Intelligence, Washington, DC. , pp. 305-312.
2. Chia-Wen Liao, Yeng-Horng Perng, Tsung-Lung Chiang, (2009). Discovery of unapparent association rules based on extracted probability, Journal Decision Support Systems Volume 47 Issue 4, November.
3. Yun Sing,(2008).Kohl Russel Pears Rare Association Rule Mining via Transaction Clustering. Seventh Australasian Data Mining Conference (AusDM 2008), Glenelg, Australia.
4. S.L. Hershberger, D.G. Fisher,(2005). Measures of Association Encyclopedia of Statistics in Behavioral Science, John Wiley&Sons.
5. Carlos Ordonez and Kai Zhao(2011). Evaluating association rules and decision trees to predict multiple target attributes, Intelligent data Analysis, vol. 15, pp. 173-192.

6. M.N. Do, M. Vetterli,(2005). The contourlet transform: An efficient directional multi-resolution database representation. IEEE Transcactions on Database Processing, vol. 12, pp. 2091-2106.
7. Jin Zhengshu, Hu Guang.(2013).Application of inquiry optimized in the Distributed database system. Computer Applications and Software. pp. 58-60.
8. Panos Vassiliadis, Timos Sellis.(2009). A Survey on Logical Models for OLAP Databases. ACM SIGMOD Record, vol. 4, pp.64-69.

A Robust Zero-watermarking Algorithm Against Geometric Attacks with Perceptual Hashing

Baoru Han[1], Jingbing Li[2], Mengxing Huang[2]

1 Department of Electric Engineering, Hainan Software Profession Institute, Qionghai, Hainan, 571400, China
2 College of Information Science and Technology, Hainan University, Haikou, Hainan, 570228, China
Email: 6183191@163.com

Corresponding author is Jingbing Li

Abstract
Taking into account the information security problem in medical image, the paper put forward a new robust zero-watermarking algorithm with three-dimensional (3D) discrete Fourier transform (DFT) and perceptual hashing. The medical volume data was made by 3D DFT. The zero-watermarking algorithm selected the real part of 3D DFT coefficients (4*4*4) to generate the watermarking extraction key sequence (64-bit). Meanwhile, it employed Legendre chaotic neural network scrambling to achieve the two-encryption, and boosted the robustness of the algorithm, so that it can better resist conventional attacks and geometric attacks. The zero-watermarking algorithm can solve the contradiction between the watermarking imperceptibility and robustness. The experimental results show that the proposed algorithm has superduper robustness.
Keywords: ZERO-WATERMARKING ALGORITHM, 3D DFT, PERCEPTUAL HASHING.

1. Introduction

Today, the digital information system plays a more and more important role in the medical industry. The digitizing system not only provides convenience for the transmission and storage of medical information, but also promotes the diagnosis mode of remote medical system. At the same time, it also brings a lot of risk, which is widely concerned by people. Currently,