

## A Novel Algorithm for Grid Assembly based Porous Structure Modeling

<sup>1</sup>Wei Lou, <sup>2</sup>Yuan Yao, <sup>1</sup>Xiaohu Huang, <sup>1</sup>Min Cheng and <sup>2</sup>Qingxi Hu

<sup>1</sup>College of Mechanical Engineering and Automation, Shanghai University, Shanghai, 200444, China

<sup>2</sup>Shanghai Key Laboratory of Mechanical Automation and Robotics, Shanghai, 200444, China

**Abstract:** This study presents a novel algorithm for assembling cell pore structure to enhance the connectivity of porous medium used in the medical science. Firstly based on sample learning, the designed cell pore structure is assembled and thus the parametric pore model can be established. Then the model is optimized by using random decision forests as evaluator and KD tree as the nearest neighbor searching area in the high dimensional space. Finally the parametric model can be transformed to solid model for evaluating the robustness of the proposed algorithm with the aid of the second development platform of UG. The test verifies that the proposed method can assemble and optimize the established cell pore model and thus significantly improve the correlation among cell models and successfully solve the difficult problem that the connectivity among cell models can't easily be controlled.

**Keywords:** Algorithm for grid assembly, porous structure modeling, sample learning

### INTRODUCTION

Motivated by replacing the defected bone with substitute made of artificial material (plastic, metal, ceramics, etc), the researchers have investigated many proposals to repair the bulk defected bone tissue. Microstructure modeling of internal bone structure is the first step of the methods and plays an important role in the process of manufacturing artificial bone by Rapid Prototyping (RP) technology. The internal microstructure of activity bone has many small pores, thus has large surface area and pore volume. The proposed modeling in this study not only reflects the size and arrangement of the actual pores in bone but also takes into account of the shape and structure of pores and the connectivity among pores and thus can significantly reveal the traits of microstructure of internal bone, thus ensuring the biological activity of artificial bone structure. Utilizing the method of forward modeling, porous structure model is obtained by Boolean Operation between solid model and pore structure in this study. Thus this process comprises two steps: modeling of solid part and modeling of pore structure. The thinking method for designing pore structure is to construct integral feature of pore structure by assembling local features from the porous medium. Firstly, based on local fractals structure of porous medium from the nature, the pore structure can be established. A six-neighbor cell pore model is designed as pore cell model (Fig. 1a) the cell model includes 7 ellipsoids which the centered one intersected

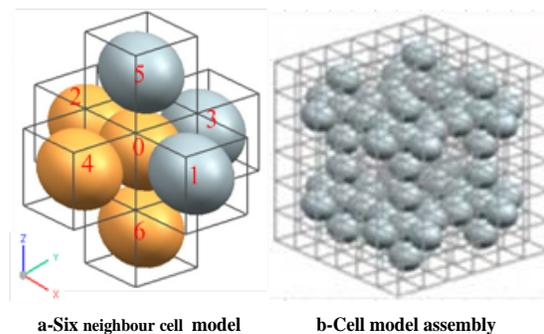


Fig. 1: Model diagram

with the other six ones two at a time and the feature of the model can be described using direction vector in the long axis direction  $l = \{10, 1..16\}$  (Cheng and Yuan, 2012). The size of included angle determines the size of intersection area between two pore structures.

The area thus expresses the connectivity of porous medium. Considering the fact that microstructure of the bone and spatial location are randomly distributed, a semi-supervised learning technique is adopted to construct the sample library. Statistically speaking, the library can guarantee the local cell model stay in a random status. Thus intersection area among 7 ellipsoids in one cell which characterize the porosity and connectivity of porous medium can be controlled within a certain range. However, this can't ensure that the connectivity among cell models is good enough (Fig. 1b). Aiming at this problem, an algorithm for grid

assembly used in cell pore structure is proposed in this study. Thus local pore model of porous medium can be assembled into integral porous model and improve the connectivity among cell pores. Finally, parameterized model is changed into pore solid model with the aid of the second development platform of UG and porous medium model is obtained by Boolean Operation.

### RELEVANT RESARCH

Porous medium comprises solid matter part and pore part. Many researchers in geology and medical field have greatly investigated porous medium. In order to effectively predict the flow behavior of various liquid in porous medium and taking into consideration of correct information about pore structure, microscopic porous medium model has to be built. The scholars have proposed several porous mediums and from the simple ones to difficult ones, they are hollow billet model, accumulated model of spheroidal particle, grid model, numerical core model and network model. With the development of computer technology and high-resolution instrument, porous medium can more accurately reflect the traits of actual porous medium (Wang and Ning, 2012). Thus quantitatively analyzing the features of pore structure and then obtaining appropriate porous medium model are very important and is also an arduous task.

So far, forward construction method and reverse construction method have been employed to obtain the pore structure. During the designing process of the product, the reverse reconstruction method for porous medium consists of physical experiment technique and numerical reconstruction technique on the basis of slice analysis. Yanlong *et al.* (2009) proposed the method of physical experiment using a real 2D micro-CT image and MPS to reconstruct the 3D structures of porous media. Forward construction method is the conventional design technique and is used in many applications ranging from the design of screw and nut to the design of aircraft and steamship. Wang *et al.*

(2011) employed volume vector and compact volume as the notation of volume object for complex internal structure. The vector notation permits quick random access and can obtain the real time visualization of any section of any complex volume object, but the resolution of notation of structure is low and so can't precisely express the engineering porous structure. So most scholars only pay attention to the reverse design technique for porous structure modeling and the research of forward design method is scarce. In this study, the forward design method is employed. The complex internal pore structure of porous medium is abstracted as the network model with certain topological structure, thus forming the unit pore models, based on assembly of which the entire porous medium structure can be obtained.

### ALGORITHM FOR GRID ASSEMBLY

Any sizeable porous structure can be assembled with the trained associated models. In order to adapt to different unite structures, the assemble procedures can be designed as a framework that can extend. Shown in the following Fig. 2, the model's size, feature size and evaluating criteria can be determined through the specifications of the design. Then, the strategy is drafted and the pore models are assembled. Finally, the porous model is built with the aid of CAD modeling system.

The assemble process includes two stages. The first is primary assembly step. And according to the design specifications, the grids are primarily filled, using RDF as evaluator. The second stage is optimization process for the grid assembly. Using RDF as evaluator and KD tree as the most recently seeking area in the high dimensional space, the designed cell models are assembled and optimized.

In order to effectively handle space data, building index mechanism for quickly accessing data has become necessary. In this study, KD tress is adopted as index. Then the geometric construction, in Fig. 3

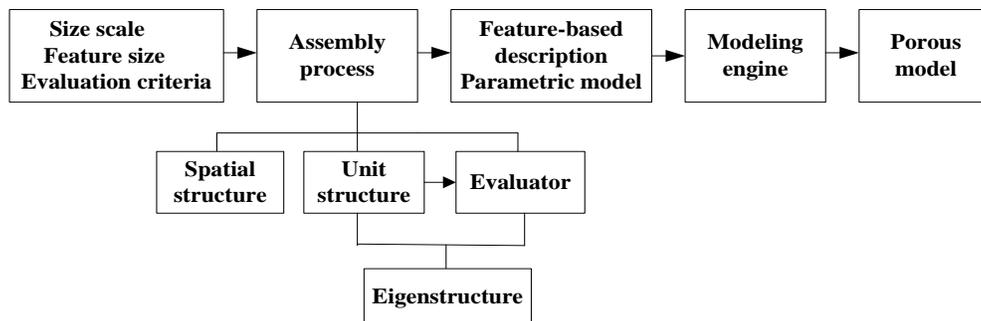


Fig. 2: Overall framework of algorithm for grid assembly

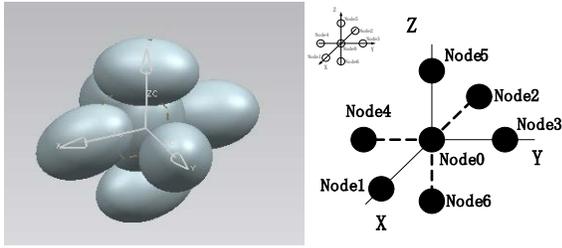


Fig. 3: The geometric construction and space index construction in the cell pore structure

showed, can be abstracted as mathematic index mode  $l\{$  Node 0(x, y, z), Node 1(x+1,y,z), Node 2(x-1,y,z), Node 3(x,y+1,z), Node 4(x,y-1,z), Node 5(x,y,z+1), Node6 (x,y,z-1) $\}$ .

KD tree (short for k-dimensional tree) is a kind of binary index tree for k-dimension space point. The internal node is strictly corresponding to the value in the data library. Also in k-dimension space each node is divided into two parts with a super plane, that's to say, the root node divide the space into two parts and the sub node further divide the space into smaller parts and the division of sub node don't cross the parental division (Tim and Jeremy, 2005).

Adopting this data structure for space data, the index speed can be improved significantly. Each node represents a data point in the KD tree and each record is notated by the attribute of the node. According to the calculated distance, the node is arranged from the smallest to the biggest. Here, the six neighbor nodes of one certain nodes are required and treated as a cell pore model.

Recently, scholars in the machine learning field have become more willing to adopt the technique of probability statistics to depict the uncertainty problems in the environment for obtaining a mathematical model. In the machine learning field, RDF includes several evaluators of decision tree and the output category of them is determined by the majority of one certain tree. RDF is the collection of evaluator trees  $\{DTree (X, \theta_k) \mid k = 1, 2, \dots\}$ , where cell classifiers  $DTree (X, \theta_k)$  is the classification regression tree without pruning branch constructed through CART algorithm.  $\theta_k$  is the independent, identical distribution random vector and determines the growing process of one certain tree. The technique of majority voting for the classifying and averaging is adopted for the regression to obtain the RDF for the output of final predicted value (Leo, 2001; Lee *et al.*, 2010; Gislason *et al.*, 2006). The feature of cell model can be depicted using the direction vector in the long direction  $l = \{l_0, l_1, l_2, l_3, l_4, l_5, l_6\}$ . Then one of the sample attribute set of the RDF can be expressed as  $l = \{l_0, l_1, l_2, l_3, l_4, l_5, l_6, D\}$ , where D is the sign feature of the sample:

D = 1 Good sample  
0 Bad samples

where,

1 = The good sample and shows that, in the cell pore model, the connectivity are achieved

0 = The bad sample

The working process of RDF is shown in the following Fig. 4, after inputting the original samples, firstly randomly selected several ones are used as the new training set X ( $L_1, L_2 \dots L_n$ ) and train a serial decision trees  $\{DTree (X, \theta_1), DTree (X, \theta_2), \dots, DTree (X, \theta_m)\}$ . When training each tree, the best attribute is selected from the randomly selected attributes to best split the internal node. This can make each decision tree greatly independent, thus improving the generalization ability and classifying ability of RDF. Then with each trained decision tree, the set T is tested to decide and the output can be independently obtained. Finally, using decision model series  $\{DTree (X, \theta_1), DTree (X, \theta_2), \dots, DTree (X, \theta_m)\}$ , one multiple decision model system can be obtained. The simple majority vote technique has been adopted to output the decision result for this system. The final decision result is:

$$DT(x) = \text{Majorityvote} \{Dtree (X, \theta_m) = D\}^{m = N_{tree}}_{m = 1}$$

where

DT (x) = United decision model

DTree (X,  $\theta_m$ ) = Single decision tree classify model

D = Output variable ( target variable)

In this equation, the final classify is determined using majority vote technique. The procedures of RDF algorithm based on Open CV (Open Source Computer Vision Library) are expressed as follows:

**Algorithm 1:** Random Decision Forests Input:

1. Train sample  $L = \{L_i(l_0, l_1 \dots l_6, D) \mid i=1, 2, \dots, n\}$

2. Test sample  $T = \{T_j(l_0, l_1 \dots l_6, D) \mid j=1, 2, \dots, m\}$

For  $k = 1, 2, \dots, N_{tree}$

- (1) Obtain the training set S using the Bootstrap sampling method;
- (2) Construct one decision tree with training set
  - a. Randomly select m features from 21 ones
  - b. Determine the best segmentation method in each node and split the current node into right and left ones
- (3) Halt condition
  - a. The maximum splitting depth has been achieved
  - b. All samples corresponding to root node are the same category
  - c. The best split precision has been achieved and can meet the requirement

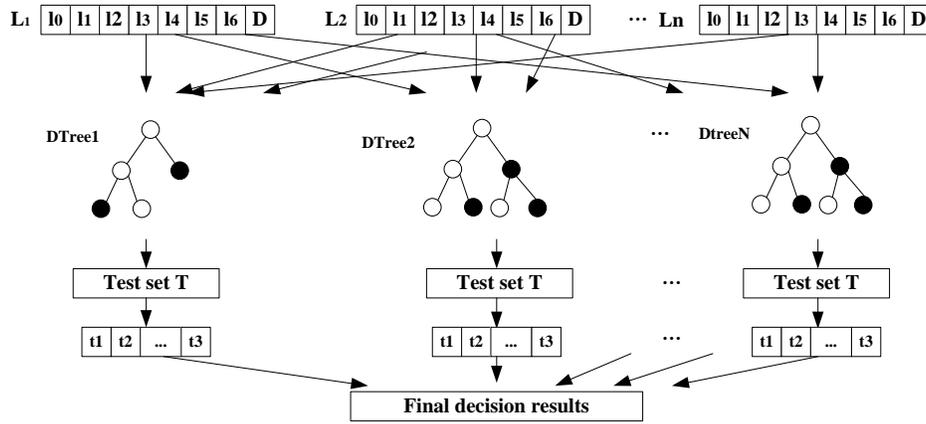


Fig. 4: Schematic diagram of RDF's working process

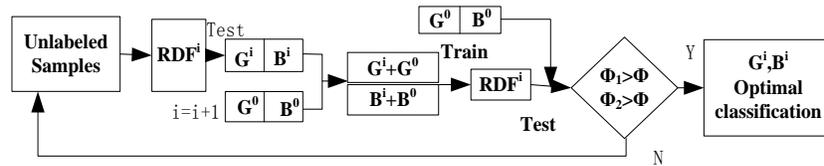


Fig. 5: Flow chart for the test (RDF<sup>0</sup> is obtained through training with the signed sample G<sup>0</sup> and B<sup>0</sup>, where G denotes good sample and B denotes bad sample, and  $\Phi_1, \Phi_2$  denote the recognition rate)

End

- Output: 1. Output the RDF structure
- 2. Output the decision output of RDF

When training the large, high dimension data in RDF, the excessive fitting phenomenon can be avoided in most cases, so it cost less time. The data set is good. The performance of date is evaluated by repeatedly training and testing the data in the following experiment. The principle of experiment is that the learning method of decision tree is mainly used for classifying and deciding and it is an inductive learning method based on cases and the decision trees are reasoned using a group of unordered and rule less cases and this can express the forming classifying rules. The common procedures in the test are shown as follow in Fig. 5, the classification label of unsigned sample were predicted by RDF, then add the signed sample G<sup>0</sup> and B<sup>0</sup> to iterate training RDF, through repeated training and testing of RDF to observe its reliability.

The test data comes from 100 signed good data and 100 signed bad data. 80 pieces of data are selected from 100 good data and 100 bad data, respectively. The other 20% of data are used as test set. Through repeatedly training and testing, the classifying ability shown by RDF is obtained. Through experiment, the training and testing reliability of RDF in Fig. 6 showed that RDF can recognize 85% of data group and the fluctuation is

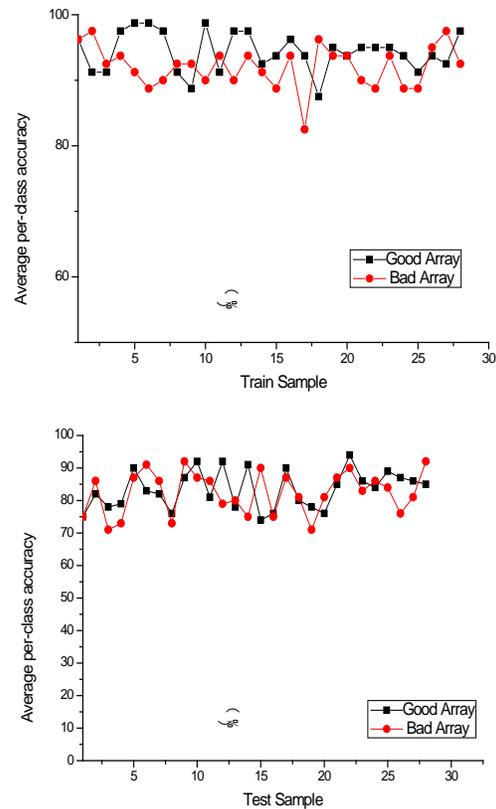


Fig. 6: Graph of training and testing reliability of RDF

small the RDF generalization ability can meet requirements to ensure the unit of ellipsoids can be filled in the grid.

**DESCRIPTION OF ALGORITHM FOR GRIS ASSEMBLY**

To ensure the connectivity among the cell pore models, process of grid assembly consists of 4 steps, as shown in Fig. 7: generating the spatial grid, building spatial index mechanism of KD tree, initially filling grid with the selected appropriate step length and completely filling entire grid with designed filling mechanism. Finally RDF was used as evaluator and KD tree as the most recently seeking area in the high dimensional space to optimize the designed model and complete the entire assembly process.

At the generalizing spatial grid stag, the requirement size of designed model is  $N(X * Y * Z)$  and the actual generalizing process needs to extend outward to add a grid, that is  $N+1\{(X+2) *(Y+2)*(Z+2)\}$ . The boundary can't be served as the centre of cell model and so they are not considered as accessing objects in the later filling and optimization process. The center of each regular hexahedron is numbered and served as index value for filling the ellipsoid to build the index mechanism for KD tree for subsequent filling and optimization.

At initially filling process. Firstly select appropriate step length. Then select M centers of grid from the generalized spatial grids as the center of cell model and simultaneously select M good cell models from the good sample library to fill the grid and initially fill the M cell models.  $M < N1$  can ensure initialized n cell models are independent to each other and the M cell models are uniformly dispersed among the entire grid, effectively utilizing the good data in the sample database.

To completely filling the grid and initially finishing assembling. Firstly search all the elements in the grid, use one grid as the centre to search for the adjacent 6 ellipsoids. If the cell model is empty, then select one cell model from the good sample library and assemble the ellipsoids to the grid with certain rules. The algorithm procedures are described as followed.

**Algorithm 2:** Completely the entire assembly

Input: sample library L

For k = 1,2...Node

1 search around the grid Node (  $X_k, Y_k, Z_k$  )

2 Node served as the centre of cell pore model A0, search the other 6 adjacent ellipsoids to form a cell unit { Node(  $X_k, Y_k, Z_k$  ) , Node(  $X_{k+1}, Y_k, Z_k$  ) , Node(  $X_k,$

$Y_k, Z_k$  ) , Node(  $X_k, Y_{k+1}, Z_k$  ) , Node(  $X_k, Y_{k-1}, Z_k$  ) , Node(  $X_k, Y_k, Z_{k+1}$  ) , Node(  $X_k, Y_k, Z_{k-1}$  ) }

3 If ( Unit is empty)

Scheme one

a. Randomly select a sample Temp from the sample library

b. Fill the empty node in the Unit with the eigenvalue corresponding to node Temp

Scheme two

a. Mandatorily set the empty node from the Unit with  $\sqrt{3}/3$

b. Put Unit into KD tree to search the most similar sample from the library

c. Fill the empty node in the Unit with the eigenvalue corresponding to node Temp

End

Scheme one is simple for filling process and can reduce the algorithm complexity in the basic assembly process. Scheme two increases the algorithm complexity. However it can reduce optimization nodes in each iteration process for the whole grid assembling.

Optimize the initially assembled grid to guarantee the porosity and connectivity among a scope of intersecting area for the whole ellipsoid. In this process, two tools are adopted. The first one is evaluator and the RDF was selected as the evaluator.RDF in this algorithm is designed for a framework that can extend and the data it can process have reached million-scale. So the RDF is packaged as the dynamic link library to improve the processing speed. The second one is using one kind of data structure called KD tree separating data points in the K dimension space. The procedures for the assembly algorithm are detailed as follows:

**Algorithm 3:** Assembly algorithm for the grid

Input: 1. Sample library L; 2. RDF dynamic link library; 3.Tree structure of RDF; 4. Iteration number D and cell model number optimized in every round

Begin

Construct KD tree with sample library

For k = 1, 2...Node

(1) Search all the grid elements in Node (  $X_k, Y_k, Z_k$  )

(2) Node serving as the centre of cell pore model A0, search the other 6 adjacent ellipsoids to form a cell unit, { Node( $X_k, Y_k, Z_k$  ) , Node(  $X_{k+1}, Y_k, Z_k$  ) ,Node(  $X_{k-1}, Y_k, Z_k$  ) , Node(  $X_k, Y_{k+1}, Z_k$  ) , Node(  $X_k, Y_{k-1}, Z_k$  ) , Node(  $X_k, Y_k, Z_{k+1}$  ) , Node(  $X_k, Y_k, Z_{k-1}$  ) }

(3) Send Unit to the RDF for predicting and the result is denoted as Result;

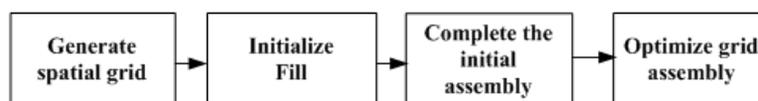


Fig. 7: Basic process of grid assembly

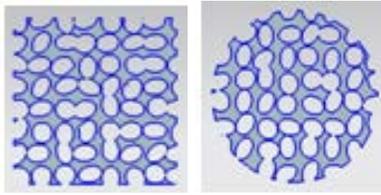


Fig. 8: Axial section and horizontal section in the scheme one

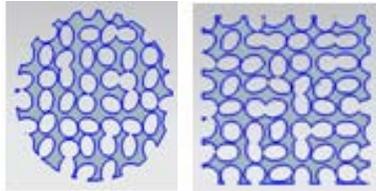


Fig. 9: Axial section and horizontal section in the scheme two

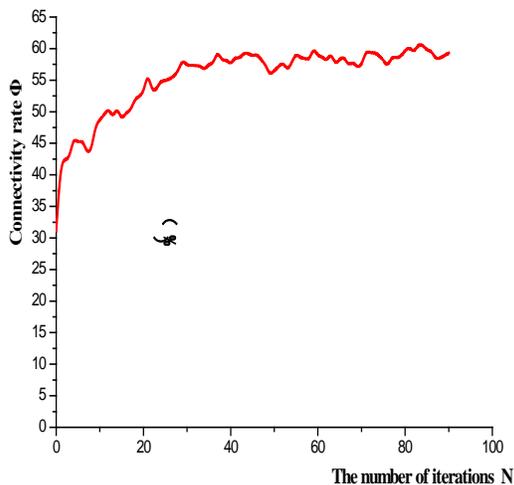


Fig. 10: Relation figure between iteration number and connectivity rate

- (4) If Result is bad sample
  - a. Send Unit to KD tree and search the most adjacent good sample Temp
  - b. Don't handle the optimized node temporarily and the none optimized node is replaced by  $(\alpha * \text{Unit}[i] + \beta * \text{Temp}[i])$  and scaled as the unit one
  - c. Mark down the cell model number n and the iteration number d in the whole iteration process
- (5) Halt condition: a.  $n < N$ ; b. Iteration number  $d < D$ ; End

Adopt  $10 * 10 * 10$  grid model to conduct assembling experiment for the cell pore structure to obtain the pore model. Then with the aid of second development platform of UG for building 3D, the program is written with the Open Grip to transform from parametric model to entity model and perform Boole operation between solid model and pore model. Through conducting many

experiments, for the best optimization effect, the coefficient of association  $\alpha$  is set as 0.618,  $\beta$  as 0.382. In the completely filling stage in the assembling process, two methods are adopted. The following, as shown in Fig. 8 and 9, are the axial section and horizontal section of the two schemes two show the connectivity was improved. The porosity in the first scheme is 73.51% and the second is 75.73% by calculating.

On the whole, comparatively connectivity among internal pore of model can be seen from the section of model. The scheme two has better connectivity. As described above, the cell model reveals better connectivity when good samples are used. In the model used in this study, number of the good samplers NG can be reflected by the ratio  $\Phi$  defined by  $NG / (X * Y * Z)$ . When fully adopting scheme two, the optimization associated coefficient  $\alpha$  is set as 0.618,  $\beta$  as 0.382 for the connecting ratio in each iteration round. The results as shown in Fig. 10, when the iteration number is 0, the entire grid is randomly filled. When the proposed assembly algorithm is not adopted, the connecting ratio is about 31.14. When N changes from 1 to 50, the connecting ratio is increasing gradually, but when N reaches between 50 and 90, the connectivity rate is becoming stable and can reach the number of 60%. For in this study the proposed method is based on samples learning to build sample library and it belongs to probability statistics field and the abnormal value may appear in the optimization and iteration process. However, the entire connecting ratio has improved. So the proposed assemble algorithm for cell pore model can improve the dependence among cell models, thus improving the connectivity in the pore structure.

## CONCLUSION

Based on sample learning, the proposed assembly algorithm for grid in the porous structure modeling can significantly improve the dependence among cell pore models, thus improving the connectivity among porous structure and even ensure the randomness of the position and the size of the pore space. Generalization capability of the spatial index mechanism and evaluator is the key factor for the algorithm due to the high dimension data processed in this study. So increasing the generalization capability of evaluator, building object-orient spatial index mechanism for multiple dimension data and index mechanism for supporting simultaneous accessing data are important solutions to improve the proposed algorithm.

## ACKNOWLEDGMENT

The authors express deep gratitude to National Natural Science Foundation of China (No.51075253) and National Science Foundation for Distinguished

Young Scholars of China (No.51105239) for the financial support of this research project.

#### REFERENCES

- Cheng, M. and Y. Yuan, 2012. Learning based porous structure design approach. *CAD/CG'20*, 12: 343-346.
- Gislason, P.O., J.A. Benediktsson and J.R. Sveinsson, 2006. Random forests for land cover classification. *Pattern Recogn. Lett.*, 27(4): 294-300.
- Lee, S.L.A., A.Z. Kouzania and E.J. Hu, 2010. Random forest based lung nodule classification aided by clustering. *Comput. Med. Imaging Graph.*, 34(7): 535-542.
- Leo, B., 2001. Random forests. *Mach. Learn.*, 45(1): 5-32.
- Tim, F. and S. Jeremy, 2005. KD-tree acceleration structures for a GPU raytracer. *Proceeding of the ACM Siggraph/Eurographics Conference on Graphics Hardware (HWWS)*, pp: 15-22.
- Wang, B. and Z. Ning, 2012. Research on the reconstruction method of the micro-model of porous medium. *Reserv. Evaluat. Dev.*, 2(2): 45-53.
- Wang, L.D., Y. Yu, K. Zhou and B. Guo, 2011. Multiscale vector volumes. *ACM T. Graphic.*, 30(6): 167.
- Yanlong, W., Z. Ting, L. Jinhua and Z. Jin, 2009. The study of porous media reconstruction using a 2D micro-CT image and MPS. *Proceeding of the International Conference on Computational Intelligence and Software Engineering (CISE)*, pp: 1-5.