SCIENCE CHINA
Information Sciences

• RESEARCH PAPER •

September 2013, Vol. 56 092116:1–092116:12
doi: 10.1007/s11432-011-4429-y

A fast calculation strategy of density function in ISAF reconstruction algorithm

WANG GongMing[1,2], ZHANG Fa[1*], CHU Qi[1,2], FAN LiYa[3], SUN Fei[4] & LIU ZhiYong[1*]

[1]*Institute of Computing Technology, Chinese Academy of Sciences, Beijing* 100190*, China;*
[2]*Graduate University of Chinese Academy of Sciences, Beijing* 100190*, China;*
[3]*IBM China Research Lab, Beijing* 100094*, China;*
[4]*Institute of Biophysics, Chinese Academy of Sciences, Beijing* 100101*, China*

Received September 7, 2012; accepted October 27, 2012; published online November 30, 2012

Abstract The ISAF reconstruction algorithm is a new method for reconstructing icosahedral molecules from their projections. This algorithm works in spherical coordinate system and can achieve higher resolution than the traditional Fourier-Bessel algorithm in cylindrical coordinate system; however this method needs huge computations, which limits its application in reality. The main bottleneck lies in the calculation of density function as it occupies 90% running time of the whole algorithm. A fast calculation strategy of density function is proposed to solve this problem. This strategy is composed of three components: the fast calculation method of density function of mesh point in spherical coordinate system, the transformation method of density function of mesh point from spherical coordinate system to Cartesian coordinate system and the fast two-phase mapping method. The time complexity of calculating density function is decreased from $O[(L_M)^8]$ to $O[(L_M)^7]$ in our strategy. The experimental results on Psv-F simulated data indicate that the speed of calculating density function is increased almost two orders of magnitude and the speedup of the whole algorithm could reach 30 times. In addition, the speedup could go up with the increase in the number of images and the requirement of accuracy.

Keywords ISAF, 3D reconstruction, density function, spherical coordinate system, quaternion interpolation

Citation Wang G M, Zhang F, Chu Q, et al. A fast calculation strategy of density function in ISAF reconstruction algorithm. Sci China Inf Sci, 2013, 56: 092116(12), doi: 10.1007/s11432-011-4429-y

1 Introduction

In the late 1960s, DeRosier and Klug [1] proposed the principle of reconstruction of a three-dimensional (3D) object from a set of EM (electron microscopy) images, which is thought to be of epoch-making significance in the history of structural biology. After many years' development, EM reconstruction methods have developed into three main categories: electron crystallography [1], electron tomography [2] and single particle technology [3]. Among these methods, the single particle technology has no requirement for crystallization or limitation of molecular size [4,5], which makes it play an important role in the study of macromolecular structure.

*Corresponding authors (email: zf@ncic.ac.cn, zyliu@ict.ac.cn)

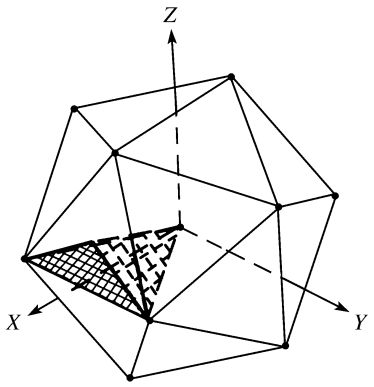

Figure 1 The asymmetric unit in digonal axis coordinate system.

Considering that the Fourier-Bessel algorithm [6], the traditional 3D reconstruction method in single particle technology, could not meet the demand for high-resolution reconstruction, Crowther [7] first proposed the idea of using spheric harmonics function to reconstruct 3D structure in spherical coordinate system. This idea can fully use the structural information of original data and enhance the reconstruction accuracy consequently. In 2003, Navaza [8] developed the ISAF (icosahedral symmetry-adapted functions) method based on the spheric harmonics function, which reveals finer structural details of molecule. In 2007, Liu [9] further improved ISAF reconstruction algorithm and the effective resolution was about 0.5 angstrom higher than that of the Fourier-Bessel algorithm.

However, the density function of tens of millions points must be calculated in ISAF reconstruction algorithm, and moreover, the operation for each point involves the computing of spherical Bessel function, intergration, double accumulation and ISAF. The statistical data shows that the step of calculating density function takes more than 90% of running time, which becomes the bottlenneck of ISAF reconstruction algorithm.

To solve this problem, a fast calculation strategy of density function is proposed, and the time complexity of calculating density function is decreased about one power law in theory. This strategy includes three items. The first item is the fast calculation method of density function in shperica coordinate system, which reduces a large number of repeated and irrelevant calculation of mesh point. The second item is the transformation method of density function from spherical corrdinate to Cartesian coordinate, which is based on quaternio interpolation and linear interolation. The third item is the fast two-phase mapping method, which can reduce about 3/4 computational load of mapping operation.

To validate this strategy, we compared it with the original ISAF reconstruction algorithm, recISAFs algorithm [9] and Fourier-Bessel algorithm, using the simulated data of Psv-F (penicillium stoloniferum virus F). The results show that the speed of calculating density function is increased by almost two orders of magnitude and the speedup of the whole reconstruction algorithm could reach 30 times without losing accuracy. In addition, the speedup could go up with the increasing number of images and the requirement of accuracy.

2 The ISAF reconstruction algorithm

2.1 The description of ISAF reconstruction algorithm

The ISAF reconstruction algorithm is composed of four stages [10]: mapping molecular images, constructing and solving fitting equation set, calculating density function, and storing into MRC file.

(1) Mapping molecular images

The main operation in this stage is to obtain structural information (positions, structural factors) in molecular images and map them onto the asymmetric unit, which is the basic symmetrical unit in ISAF reconstruction algorithm [11]. The asymmetric unit is shown in shadow area in Figure 1 and its region

is $\{(r,\theta,\varphi)|r\in[0,r_m],\theta\in[69.09,90.0],\varphi\in[-31.72,31.72]\}$. In this formula, (r,θ,φ) is the spherical coordinates and r_m is the molecular radius.

(2) Constructing and solving fitting equation set

The main operation in this stage is to construct fitting equation set according to the structural information and solve fitting coefficients. The fitting coefficients will be used for fitting structural factor of every point in 3D space with linear combination of corresponding ISAF basis functions.

(3) Calculating density function

The main operation in this stage is to calculate density functions $\rho(x,y,z)$ of mesh points inside the asymmetric unit and deduce density functions in other 59 symmetrical areas according to the sixty-symmetrical property of icosahedrons. Detailed analysis of the stage is described in subsection 2.2.

(4) Storing into MRC file

The main operation in this stage is storing density functions of mesh points in Cartesian coordinate system into MRC file.

2.2 Time complexity analysis of calculating density function

In ISAF reconstruction algorithm, the density function $\rho(r,\theta,\varphi)$ of point (r,θ,φ) is calculated as follows:

$$\begin{aligned}\rho(r,\theta,\varphi)&=\int F(R,\Theta,\Phi)\exp(-2\pi\mathrm{i}\boldsymbol{R}\cdot\boldsymbol{r})R^2\sin\Theta\mathrm{d}R\mathrm{d}\Theta\mathrm{d}\Phi\\&=4\pi\Bigg[\sum_{l=0\bmod 2}^{L_M}(-\mathrm{i})^l\sum_{u=1}^{u(l)}\bigg(\int_0^\infty f_{l,u}(R)j_l(2\pi Rr)R^2\mathrm{d}R\bigg)\\&=S_{l,u}(\theta,\varphi)+\sum_{l=1\bmod 2}^{L_M}(-\mathrm{i})^l\sum_{u=1}^{u(l)}\bigg(\int_0^\infty f_{l,u}(R)j_l(2\pi Rr)R^2\mathrm{d}R\bigg)S_{l,u}(\theta,\varphi)\Bigg],\end{aligned}\tag{1}$$

where $S_{l,u}(\theta,\varphi)$ is the basis function, $l(0\leqslant l\leqslant L_M)$ is the index of $S_{l,u}(\theta,\varphi)$, $u(1\leqslant u\leqslant ul)$ is the multiplicity of $S_{l,u}(\theta,\varphi)$, $f_{l,u}(R)$ is the fitting coefficient, $j_l(2\pi Rr)$ is the spherical Bessel function and R is the Fourier radius in reciprocal space. It can be seen that the spherical Bessel function $j_l(2\pi Rr)$ and the ISAF basis function $S_{l,u}(\theta,\varphi)$ are two basic elements in formula (1). A detailed description of them is as follows:

(1) Spherical Bessel function $j_l(2\pi Rr)$

The definition of spherical Bessel function $j_l(2\pi Rr)$ is given by [12]

$$j_l(z)=\frac{z^l}{1\cdot3\cdot5\cdots(2l+1)}\left\{1-\frac{\frac{1}{2}z^2}{1!(2l+3)}+\frac{(\frac{1}{2}z^2)^2}{2!(2l+3)(2l+5)}-\cdots\right\}=(-1)^lz^l\left(\frac{\mathrm{d}}{z\mathrm{d}z}\right)^l\frac{\sin z}{z},\tag{2}$$

(2) can be calculated by the Miller method [12], and its time complexity is $O(l^2)$.

(2) ISAF basis function $S_{l,u}(\theta,\varphi)$

$S_{l,u}(\theta,\varphi)$ is a linear combination of spheric harmonics functions [13]. The $S_{l,u}(\theta,\varphi)$ of angle (θ,φ) in index l, multiplicity u is given by

$$S_{l,u}(\theta,\varphi)=\sum_{m=-l}^{l}A_{l,m}^u\cdot Y_{l,m}(\theta,\varphi),\tag{3}$$

where $m\in[-l,l]$ and $A_{l,m}^u$ is the combinational coefficient generated by recursive algorithm [14]. $Y_{l,m}(\theta,\varphi)$ is the spheric harmonics function and its definition is expressed as [15]

$$Y_{l,m}(\theta,\varphi)=\sqrt{\frac{(2l+1)}{4\pi}\cdot\frac{(l-|m|)!}{(l+|m|)!}}P_l^m(\cos\theta)\cdot\mathrm{e}^{\mathrm{i}m\varphi},\tag{4}$$

where $P_l^m(\cos\theta)$ is the associated Legendre polynomial; it can be calculated by dynamic programming method. Thus, the time complexity of $S_{l,u}(\theta,\varphi)$ is $O(l^3)$ according to formule (3) and (4).

By formula (2), $j_l(2\pi Rr)$ has a threshold R_M, and $j_l(2\pi Rr)$ will become zero if $R > R_M$. Thus, the infinite integral in formula (1) can be converted into the definite integral, meaning that $\int_0^{\infty} f_{l,u}(R) j_l(2\pi Rr) R^2 \mathrm{d}R = \int_0^{R_M} f_{l,u}(R) j_l(2\pi Rr) R^2 \mathrm{d}R$. So, the equivalent description of formula (1) is

$$\begin{aligned}\rho(r,\theta,\varphi) = 4\pi\Bigg[&\sum_{l=0\mathrm{mod}2}^{L_M} (-i)^l \sum_{u=1}^{u(l)} \left(\sum_{R=0}^{R_M} f_{l,u}(R) j_l(2\pi Rr) R^2 \right) S_{l,u}(\theta,\varphi) \\ &+ \sum_{l=1\mathrm{mod}2}^{L_M} (-\mathrm{i})^l \sum_{u=1}^{u(l)} \left(\sum_{R=0}^{R_M} f_{l,u}(R) j_l(2\pi Rr) R^2 \right) S_{l,u}(\theta,\varphi) \Bigg]. \end{aligned} \tag{5}$$

The number of basic operations in formula (5) is

$$u(0)(R_M \cdot 1^2 + 1^3) + u(1)(R_M \cdot 1^2 + 1^3) + u(2)(R_M \cdot 2^2 + 2^3) + \cdots + u(L_M)(R_M \cdot (L_M)^2 + (L_M)^3).$$

Its upper bound is

$$\begin{aligned} &u(L_M)[R_M \cdot (L_M)^2 + (L_M)^3] + \cdots + u(L_M)[R_M \cdot (L_M)^2 + (L_M)^3] \\ &\quad = (L_M + 1) \cdot u(L_M)[R_M \cdot (L_M)^2 + (L_M)^3], \end{aligned}$$

$u(L_M) = \left[\frac{L_M}{30}\right] + u(L_M \bmod 30) = \left[\frac{L_M}{30}\right] + \{0,1\}$ and $R_M = L_M \cdot \frac{D}{2\pi r_m}$ (D is the size of molecular image) [9]. Therefore, the time complexity of $\rho(r,\theta,\varphi)$ is

$$\begin{aligned} &O\left\{(L_M + 1) \cdot u(L_M)\left[R_M \cdot (L_M)^2 + (L_M)^3\right]\right\} \\ &\quad = O\left\{ L_M \cdot \left\{ \left[\frac{L_M}{30}\right] + \{0,1\} \right\} \cdot \left[L_M \cdot \frac{D}{2\pi r_m} \cdot (L_M)^2 + (L_M)^3 \right] \right\} \\ &\quad = O\left[L_M \cdot L_M \cdot (L_M)^3\right] = O\left[(L_M)^5\right]. \end{aligned}$$

Since the number of mesh points inside the asymmetric unit is $N_1 = \frac{1}{60} \cdot \frac{4}{3}\pi \cdot r_m^3 = \frac{\pi}{45} \cdot r_m^3$, the time complexity of density function is $O\left[(L_M)^5 \cdot \frac{\pi}{45} \cdot r_m^3\right] = O\left[(L_M)^5 \cdot r_m^3\right]$. It can be simplified into $O\left[(L_M)^5 \cdot (r_m)^3\right] = O\left[(L_M)^5 \cdot (L_M \cdot \frac{S}{2\pi R_M})^3\right] = O\left[(L_M)^8\right]$ because $r_m = L_M \cdot \frac{S}{2\pi R_M} S \geqslant 2R_M$.

3 Fast calculation strategy of density function

Here we propose a fast calculation strategy of density function to reduce the time complexity of calculating density function from $O\left[(L_M)^8\right]$ to $O\left[(L_M)^7\right]$. This strategy is composed of three components: the fast calculation method of density function in spherical coordinate system, the transformation method of density function of mesh point from spherical coordinate system to Cartesian coordinate system, and the fast two-phase mapping method.

3.1 Fast calculation method of density function in spherical coordinate system

This method utilizes the following three properties:

Property 1. The polar angles (θ, φ) of radial mesh points in different shells are equivalent.

In asymmetric unit, the mesh points in different shells can be generated by dividing every shell with the constant angle interval $\Delta\theta$ and $\Delta\varphi$, as shown in Figure 2(a). Suppose a ray that starts from the origin goes through the mesh point (θ, φ) in shell defined by Figure 2(a), as shown in Figure 2(b). Since the polar angles of all points in a ray are equal, the intersections of the given ray and other shells must be segmented by mesh points with polar angles (θ, φ). Thus the value of $S_{l,u}(\theta, \varphi)$ in different shells is the same, and we only need to calculate it in one shell.

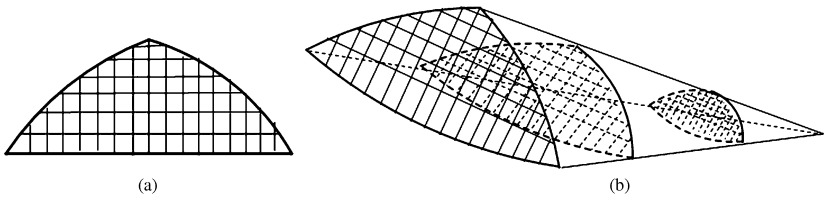

Figure 2 The mesh segmentation in asymmetric unit. (a) The mesh segmentation in a given shell; (b) the mesh segmentation in different shells.

Property 2. $S_{l,u}(\theta,\varphi)$ has the parity symmetry on the y-axis according to the parity of l.

$$S_{l,u}(\theta,\varphi)=\begin{cases} S_{l,u}(\theta,-\varphi), & l \text{ is even},\\ -S_{l,u}(\theta,-\varphi), & l \text{ is odd}. \end{cases} \tag{6}$$

Proof. For $S_{l,u}(\theta,\varphi)$, the combinational coefficient $A^u_{l,m}$ and the spheric harmonics function $Y_{l,m}(\theta,\varphi)$ have the following properties:

$$A^u_{l,-m}=(-1)^{m-l}A^u_{l,m}, \tag{7}$$

$$Y_{l,-m}(\theta,\varphi)=(-1)^{-m}Y^*_{l,m}(\theta,\varphi), \tag{8}$$

$A^u_{l,0}$ is zero when l is odd according to formula (7), i.e.

$$\begin{aligned} S_{l,u}(\theta,\varphi)&=A^u_{l,0}Y_{l,0}(\theta,\varphi)+\sum_{m=5}^{l}\left[A^u_{l,m}Y_{l,m}(\theta,\varphi)+A^u_{l,-m}Y_{l,-m}(\theta,\varphi)\right]\\ &=\begin{cases} A^u_{l,0}\sqrt{\frac{2l+1}{4\pi}}P^0_l(\cos\theta)+\sum\limits_{m=5}^{l}2A^u_{l,m}P^m_l(\cos\theta)\cos m\varphi, & l \text{ is even},\\ \sum\limits_{m=5}^{l}2A^u_{l,m}P^m_l(\cos\theta)\sin m\varphi, & l \text{ is odd}. \end{cases} \end{aligned}$$

Thus, only half of $S_{l,u}(\theta,\varphi),\varphi\geqslant 0$ need to be computed.

Property 3. The integral $\int_0^\infty f_{l,u}(R)j_l(2\pi Rr)R^2\mathrm{d}R$ in formula (1) is independent of polar angel (θ,φ) and can be calculated in advance.

According to the above properties, a fast computation method of the density function in spherical coordinate is proposed and a large number of repeated and irrelevant computations are reduced consequently. The method goes as follows:

Step 1: All integrals $\int_0^\infty f_{l,u}(R)j_l(2\pi Rr)R^2\mathrm{d}R$, $l\in[0,L_M]$, $u\in[1,u(l)]$, $r\in[1,r_m]$ are calculated with the accumulation [7] and then stored into the 2D matrix with r the first index and (l,u) the second index.

Step 2: Let $\Delta\theta=\Delta\varphi=\frac{1}{r_m}$. The mesh points in different shells of the asymmetric unit can be generated by dividing every shell according into angle intervals $\Delta\theta$ and $\Delta\varphi$.

Step 3: The value of $S_{l,u}(\theta,\varphi)$ in one shell is computed where $l\in[0,l_{\max}]$ and $u\in[1,u(l)]$.

Step 4: Density functions $\rho(r,\theta,\varphi)$ of mesh points in all shell are calculated finally.

3.2 Transformation method of density function from spherical coordinate system to Cartesian coordinate system

The density functions of mesh points in spherical coordinate system should be converted into that of integer mesh points in Cartesian coordinate system in order to meet the demand of MRC format, the uniform layout for storing density information.

This convertion problem can be described as in Figures 3 and 4.

In Figure 3, A_1, B_1, C_1 and D_1 are four mesh points in a certain shell, A_2, B_2, C_2 and D_2 are four corresponding mesh points in an adjacent shell and E is a point inside the surface hexahedral volume

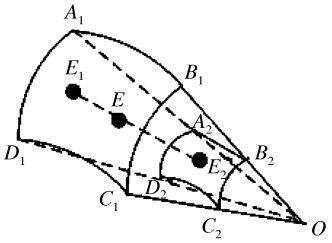

Figure 3 The conversion of density function in mesh points from spherical coordinate to Cartesian coordinate.

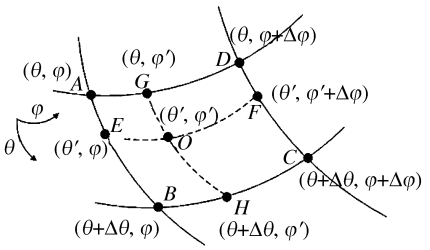

Figure 4 The quaternion interpolation in a certain shell.

$A_1B_1C_1D_1 - A_2B_2C_2D_2$. When the density functions of A_1, B_1, C_1, D_1, A_2, B_2, C_2, D_2 are given, how can we obtain the density function of E?

As we know, the quaternion interpolation is suitable for the interpolation in spheric surface. For example, A, B, C and D are four vertexes of spherical quadrilateral $ABCD$, as shown in Figure 4. Their cooridnaes are (r, θ, φ), $(r, \theta + \Delta\theta, \varphi)$, $(r, \theta + \Delta\theta, \varphi + \Delta\varphi)$ and $(r, \theta, \varphi + \Delta\varphi)$ respectively. O' is a point inside the spherical quadrilateral $ABCD$ and its coordiante is (r, θ', φ').

The density function of O' can be calculated by the quaternion interpolation, as shown by

$$\begin{aligned}\rho(r,\theta',\varphi') = {} & \frac{\sin(1-t_\theta)\omega}{\sin\omega}\cdot\frac{\sin(1-t_\varphi)\omega}{\sin\omega}\rho(r,\theta,\varphi) + \frac{\sin t_\theta\omega}{\sin\omega}\cdot\frac{\sin(1-t_\varphi)\omega}{\sin\omega}\rho(r,\theta+\Delta\theta,\varphi)\\ & + \frac{\sin(1-t_\theta)\omega}{\sin\omega}\cdot\frac{\sin t_\varphi\omega}{\sin\omega}\rho(r,\theta,\varphi+\Delta\varphi) + \frac{\sin t_\theta\omega}{\sin\omega}\cdot\frac{\sin t_\varphi\omega}{\rho}(r,\theta+\Delta\theta,\varphi+\Delta\varphi),\end{aligned} \tag{9}$$

where $t_\theta = \frac{\theta'-\theta}{\Delta\theta}$, $t_\varphi = \frac{\varphi'-\varphi}{\Delta\varphi}$, $\omega = \frac{\pi}{180}$.

It can be seen that the problem can be solved via combination of quaternion interpolation and linear interpolation. Suppose that a ray, starting from the origin and through E, intersects the two adjacent shells at E_1 and E_2, as shown in Figure 3. Then this method proceeds as follows:

Step 1: In the spherical quadrilateral $A_1B_1C_1D_1$, the density function of E_1 is generated by using formula (9).

Step 2: Similarly, the density function of E_2 in the quadrilateral $A_2B_2C_2D_2$ can be generated.

Step 3: The density function of E can be calculated by the linear interpolation

Thus we can calculate the density functions of every integer mesh point inside the asymmetric unit in Cartesian coordinate system.

3.3 Fast two-phase mapping method

To calculate the density function of mesh point (x, y, z) outside the asymmetric unit in Cartesian coordinate system, we first need to map this point onto the asymmetric unit using the sixty symmetrical properties of icosahedron, and then calculate the density function of the mapping point from the density functions of known points.

Since the mesh points outside the first (or fourth) quadrant can be mapped onto the first (or fourth) quadrant by only changing the signs of their coordinates, we present a fast two-phase mapping method, and reduce about 3/4 computational load of mapping operation.

The first stage: The mesh points in the first (or fourth) quadrant are mapped onto the asymmetric unit according to the sixty symmetrical property of icosahedron. The density functions of these points are calculated through interpolation.

The second stage: The mesh points outside the first (or fourth) quadrant are mapped onto the first (or fourth) quadrant by changing the signs of their coordinates. Then, the density functions of these points can be gained by assigning directly.

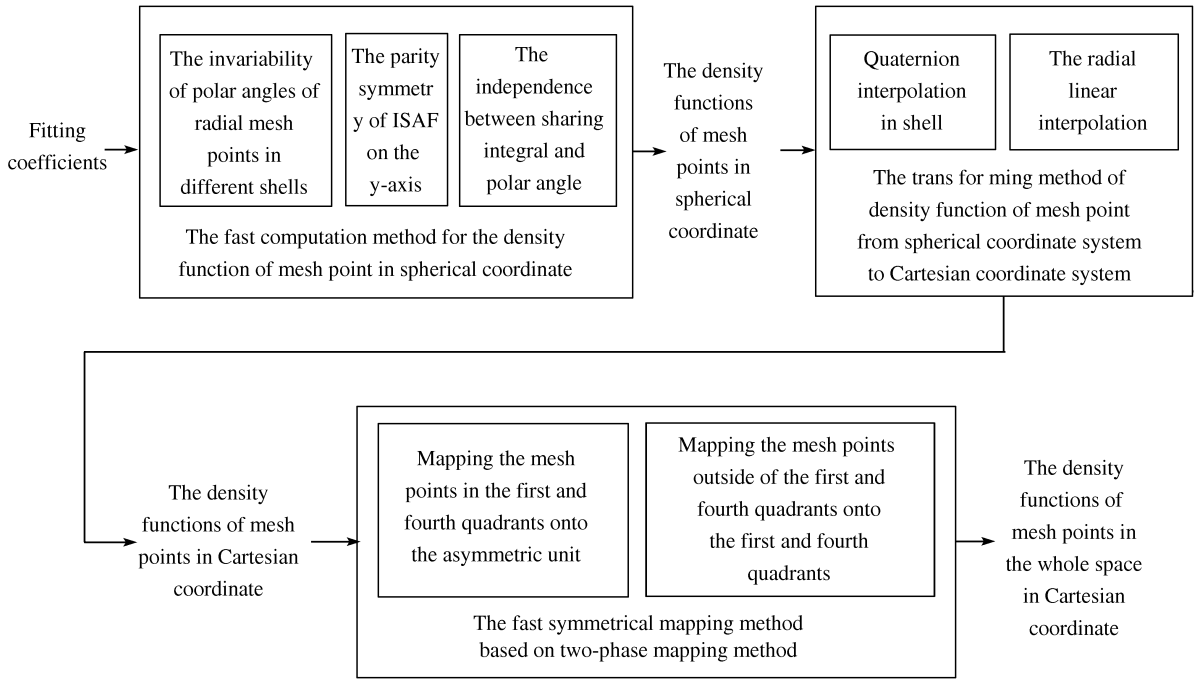

Figure 5 The fast calculation model of density function.

3.4 Fast calculation strategy of density function

The fast calculation strategy of density function is constructed by combining the above three methods. The details of three stages in the strategy is shown in Figure 5. The time complexity analyses of this strategy are as follows:

There are four main steps in the frist stage (see subsection 3.1).

In step 1, the upper bound of time complexity of integral $\int_0^\infty f_{l,u}(R)j_l(2\pi Rr)R^2\mathrm{d}R$ is $O\left[(L_M)^2R_M\right] = O\left[(L_M)^3\right]$, and the number of this integral is $r_m\sum_{l=0}^{L_M}u(l)$. Thus, the time complexity of this step is $O\big[(L_M)^3r_m\cdot\sum_{l=0}^{L_M}u(l)\big] = O\left[(L_M)^3r_m\cdot L_M\cdot u(L_M)\right] = O\left[(L_M)^3L_M\cdot L_M\cdot L_M\right] = O\left[(L_M)^6\right]$.

In step 2, the time complexity of calculating polar angle of one mesh point is $O(1)$. The number of mesh points is $\frac{\pi}{15}r_m^2$. Thus, the time complexity of this step is $O\left[r_m^2\right] = O\left[(L_m)^2\right]$.

In step 3, The number of mesh points is $\frac{\pi}{15}r_m^2$ and the upper bound of time complexity of $S_{l,u}(\theta,\varphi)$ is $O\left[(L_m)^3\right]$. The number of $S_{l,u}(\theta,\varphi)$ is $\sum_{l=0}^{L_M}u(l)$. Thus, the time complexity of this step is $O\big[(L_M)^3r_m^2\cdot\sum_{l=0}^{L_M}u(l)\big] = O\big[(L_M)^3r_m^2\cdot L_M\cdot u(L_M)\big] = O\big[(L_M)^3\cdot(L_M)^2\cdot L_M\cdot L_M\big] = O\big[(L_M)^7\big]$.

In step 4, The number of mesh points is $\frac{\pi}{15}r_m^3$ and the time complexity of calculating density function of one mesh point is $O\big[\sum_{l=0}^{L_M}\sum_{u=1}^{u(l)}1\big] = O\big[(L_M)^2\big]$. Thus, the time complexity of this step is $O\big[r_m^3\cdot(L_M)^2\big] = O\big[(L_M)^3\cdot(L_M)^2\big] = O\big[(L_M)^5\big]$.

Therefore, the time complexity of the first stage is $O\left[(L_M)^7\right]$.

In the second stage, the number of mesh points is $\frac{\pi}{45}\cdot r_m^3$ and the complexties of quaternion interpolation and linear interpolation are $O(1)$. Thus, the time complexity of this step is $O\left[(L_M)^3\right]$.

In the third stage, the main operations are matrix-vector multiplication and changing signs of coordinates. The number of matrix-vector multiplication is $\frac{1}{4}(2r_m)^3 = 2r_m^3$ and the number of changing signs of coordinates is $\frac{3}{4}(2r_m)^3 = 6r_m^3$. The two operations can be carried out within the constant time C. Thus, the time complexity of this step is $O\left[(L_M)^3\right]$.

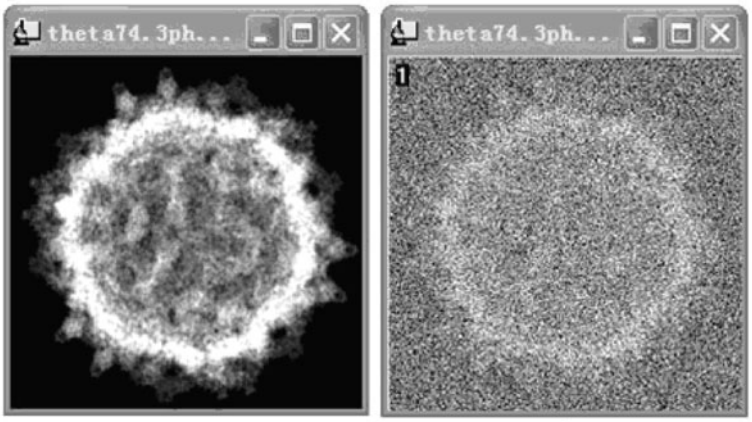

Figure 6 Simulated projection images of Psv-F. (a) Without noise; (b) with an SNR of 0.1.

From above analysis, we can find that the time complexity of calculating density function is reduced from $O\left[(L_M)^8\right]$ to $O\left[(L_M)^7\right]$.

For brevity in expression, the ISAF algorithm with the original calculation method of density function is called original ISAF algorithm and the one with the fast calculation model of density function is called accelerated ISAF algorithm.

4 Results and discussion

To validate the performance of our strategy, we first compared the accuracy and speed of calculating density function of our strategy to those of the original calculation method, using the simulated date of Psv-F (penicillium stoloniferum virus F). Furthermore, we compared the accelerated ISAF algorithm (the original ISAF algorithm with our strategy), the original ISAF algorithm, the recISAFs algorithm and the Fourier-Bessel algorithm from the whole 3D reconstruction procedure. All the experiments were all implemented on a PC with CPU AMD Dual-Core 2.6 G Hz and memory DDR 2 G byte.

4.1 Simulated data of Psv-F

The Psv-F structure (PDB ID 3ES5) was solved by X-ray crystallography to 3.3Å [16] and has icosahedral symmetrical property.

Firstly, the simulated 3D density map of Psv-F was generated at a resolution of 3.8Å by using program pdb2mrc [17] with parameters apix=1.3 and res=3.8. Its size is 304 × 304 × 304 pixels. Secondly, 2000 projection images with random orientations were generated by using program prjectIcosFile [18]. Finally, the random noise (an S/N of 0.1) is added to each of the projection images. The images with and without noises are shown in Figure 6.

4.2 Performance comparison in the stage of calculating density function

The original calculation method of density function and our strategy are both used for calculating the density functions at differernt L_M. The reconstructed maps with the two methods are shown in Figure 7. There is no significant difference between them in appearance.

Formula (10) is used for calculating the similarity between map A and map B at a given resolution in real space. In this formula, N is the number of mesh points inside the map and ρ_A^i and ρ_B^i are the density function of i-th mesh point inside map A and map B respectively [19]. The similarities at differernt resolutions can be gained by using formula (10) to maps at different resolutions. Formula (11) is used for calculating the similarities between map A and map B at different shells in reciprocal space,

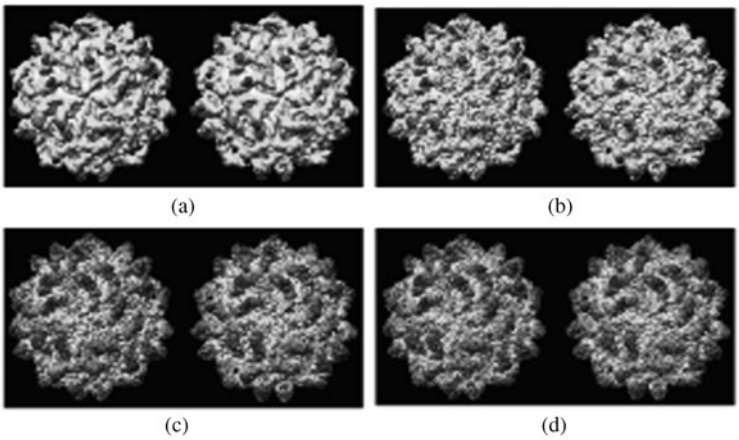

Figure 7 The reconstructed maps at different L_M. The left in each picture is the reconstructed map with the original method, and the right is the reconstructed map with the fast calculation model.

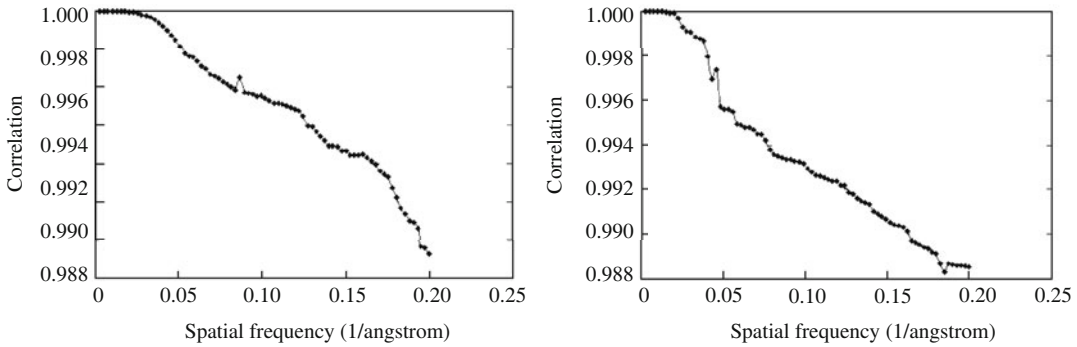

Figure 8 The similarities between reconstructed maps by the two methods at different resolutions in real space and reciprocal space. (a) Similarities in real space; (b) similarities in reciprocal space.

which is also known as FSC [20] (Fourier shell correlation). In this formula, $F_A(r)$ and $F_B(r)$ are the structural factors of two corresponding points inside map A and map B in Fourier shell $r = r_i$.

$$S_{A,B} = \frac{\sum\limits_{i=1}^{N} \rho_A^i \cdot \rho_B^i}{\sqrt{\sum\limits_{i=1}^{N} \left(\rho_A^i\right)^2 \sum\limits_{i=1}^{N} \left(\rho_B^i\right)^2}}, \tag{10}$$

$$\mathrm{FSC}_{A,B}(r_i) = \frac{\sum_{r\in r_i} F_A(r) \cdot F_B(r)^*}{\sqrt{\sum_{r\in r_i} F_A^2(r) \cdot \sum_{r\in r_i} F_B^2(r)}}. \tag{11}$$

The similarities are shown in Figure 8. It can be seen that the similarities in the real space are more than 99.3% and the similarities in the reciprocal space are more than 98.8%, suggesting that the fast calculation stratgey of density function does not lose the accuracy of reconstructed result.

Table 1 shows the running time of the two methods. It indicates that our fast calculation strategy can speed up the calculation of density function up to 170 times without losing accuracy. In addition, the speedup will grow up with the increase of the target resolutions.

4.3 Performance comparison in the stage of the whole 3D reconstruction

To verify the contribution of our fast calculation strategy to the whole 3D reconstruction, we compared the accelerated ISAF algorithm, original ISAF algorithm, recISAFs algorithm and Fourier-Bessel algorithm. In the experiment, four different target resolutions were set at 32.825Å, 13.13Å, 6.565Åand 3.9Å,

Table 1 The running time of calculating density function with the two methods at different L_M.

	Index 100	Index 150	Index 200	Index 250
Target resolution(angstrom)	12.309375	8.20625	6.1546875	4.92375
The running time with the original method(second)	188.533	547.144	1181.406	2153.379
The running time with the new method(second)	3.891	5.592	8.767	12.779
Speedup	48.45	97.84	134.76	168.51

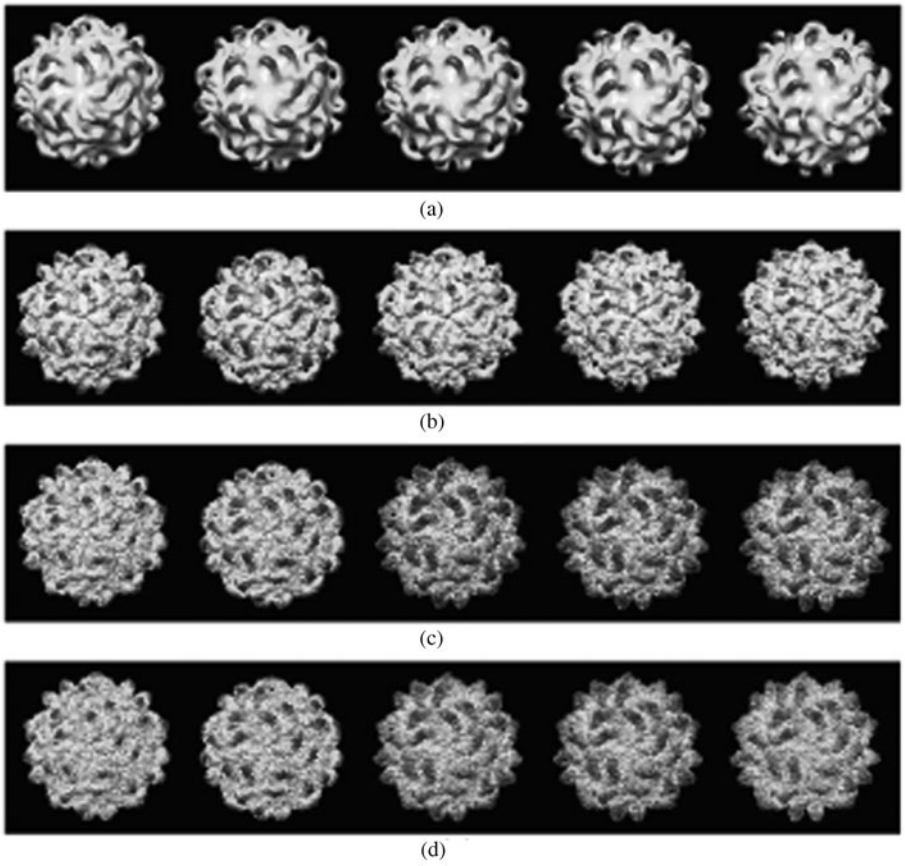

Figure 9 The reconstructed maps at different target resolutions. From left to right, the first column is the original map; the second column is the reconstructed map with the accelerated ISAF algorithm; the third column is the reconstructed map with the original ISAF algorithm; the fourth column is the reconstructed map with the recISAFs algorithm; the fifth column is the reconstructed map with the Fourier-Bessel algorithm.

respectively. For the purpose of comparison, the original maps were generated by filterring to the 3D density map in subsection 4.1 at the four corresponding resolutions. The reconstructed maps are shown in Figure 9.

To test the accuracy of this strategy quantitatively, the FSC curves of the four reconstructed maps at 3.9Å target resolution were calculated as shown in Figure 10. With the 0.5 FSC criterion [20], it can be seen that the effective resolution of accelerated ISAF algorithm is 3.9Å, the effective resolution of original ISAF algorithm is 3.94Å, the effective resolution of recISAFs algorithm is 3.98Å and the effective resolution of recISAFs algorithm is 4.19Å. That is to say, the accelerated ISAF algorithm has higher resolution than other methods.

To validate the acceleration effect of this strategy, the running times with different numbers of projection images at 3.9Å target resolution were calculated as shown in Table 2. And the running times with 2000 projection images at different target resolutions were calculated as shown in Table 3.

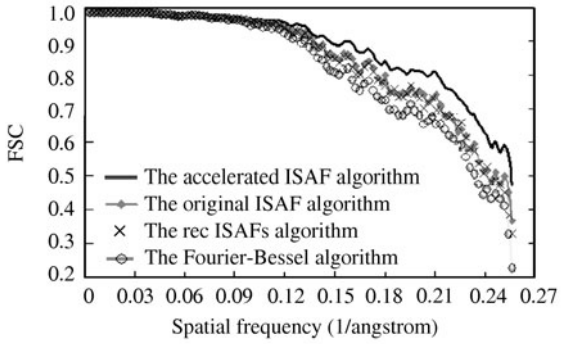

Figure 10 The FSC curves of the reconstructed maps with four methods at 3.9 Å resolution.

Table 2 The running times with different numbers of projections at 3.94 Å resolution(unit: second).

	Accelerated ISAF algorithm	Original ISAF algorithm	recISAFs algorithm	Fourier-Bessel algorithm
250 projections	241.486	7798.473	696.769	551.573
500 projections	246.871	7824.061	702.855	695.259
1000 projections	252.104	7863.921	718.489	956.190
2000 projections	259.673	7946.828	768.853	1583.489

Table 3 The running times with 2000 projections at different target resolutions(unit: second).

	Accelerated ISAF algorithm	Original ISAF algorithm	recISAFs algorithm	Fourier-Bessel algorithm
32.825 Å	15.454	81.442	151.821	36.386
13.13 Å	25.441	344.938	174.327	92.055
6.565 Å	52.352	1913.249	255.106	383.569
3.9 Å	259.673	7946.828	768.853	1583.489

From Table 2 and Table 3, it can be seen that the accelerated ISAF algorithm can speed up remarkably. Its speed is about 30 times as fast as that of the original ISAF algorithm. In addition, the speedup will grow up with the increase in the number of projections and target resolutions.

5 Conclusion

Calculation density function is the most time-consuming part in ISAF reconstrution algorithm. To tackle this problem, a fast calculation strategy of density function is proposed, and the time complexity of calculating density function is reduced from $O[(L_M)^8]$ to $O[(L_M)^7]$. The strategy includes three components: the fast calculation method of density function in spherical coordinate system, the transformation method of density function of mesh point from spherical coordinate system to Cartesian coordinate system, and the fast two-phase mapping method.

The simulated data of Psv-F (penicillium stoloniferum virus F) was used for validating the fast calculation strategy of density function. The results show that the speed of calculating density function was raised by about two orders of magnitude and the speed of whole 3D reconstrcution was about 30 times higher without losing accuracy. In addition, the speedup of this strategy will grow up with the increase in the number of projections and target resolutions.

Acknowledgements

This work was supported by the Key Knowledge Innovation Project of CAS(Grant No. KGCX1-YW-13), and the National Nature Science Foundation of China (Grant Nos. 61003164, 61070129).

References

1 Shannon D D J, Klug A. Reconstruction of three-dimensional structures from electron micrographs. Nature, 1968, 217: 130–134
2 Colsher J G. Iterative three-dimensional image reconstruction from tomographic projections. Comput Graph Image Process, 1977, 6: 513–537
3 Frank J, Goldfarb W, Eisenberg D, et al. Reconstruction of glutamine synthetase using computer averaging. Ultramicroscopy, 1978, 3: 283–290
4 Dubochet J, Adrian M, Chang J J, et al. Cryo-electron microscopy of vitrified specimens. Q Rev Biophys, 1988, 21: 129–228
5 Chiu W. Electron microscopy of frozen, hydrated biological specimens. Annu Rev Biophys Biophys Chem, 1986, 15: 237–257
6 Crowther R A. Procedures for three-dimensional reconstruction of spherical viruses by Fourier synthesis from Electron Micrographs. Philos Trans Roy Soc B, 1971, 261: 221–230
7 Crowther R A, DeRosier D J, Klug A. The reconstruction of a three-dimensional atructure from projections and its application to electron microscopy. P Roy Soc Lond A Mat, 1970, 317: 319–340
8 Navaza J. On the three-dimensional reconstruction of icosahedral particles. J Struct Biol, 2003, 144: 13–23
9 Liu H R, Cheng L P, Zeng S J, et al. Symmetry-adapted spherical harmonics method for high-resolution 3D single-particle reconstructions. J Struct Biol, 2008, 161: 64–73
10 Liu H R. The new method for high-resolution 3D reconstrcution of virus. PhD Thesis. Xiangtan: Xiangtan University. 2008. 2
11 Baker T S, Olson N H, Fuller S D. Adding the third dimension to virus life cycles: three dimensional reconstruction of icosahedral viruses from cryo-electron micrographs. Microbiol Mol Biol R, 1999, 63: 862–922
12 Abramowitz M, Stegun I A. Handbook of Mathematical Functions with Formulas, Graphs and Mathematical Tables. New York: Dover Press, 1965
13 Bradley C J, Cracknell A P. The Mathematical Theory of Symmetry in Solids. New York: Oxford University Press, 1972
14 Prandl W, Schiebel P, Wulf K. A recursive algorithm for the generation of symmetry-adapted functions: principles and applications to the icosahedral group. Acta Crystallogr A, 1996, 52: 171–175
15 Jackson J D. Classical Electrondynamics. 3rd ed. New York: Wiley Press, 1999
16 Pan J, Dong L, Lin L, et al. Atomic structure reveals the unique capsid organization of a dsRNA virus. P Natl Acad Sci USA, 2009, 106: 4225–4230
17 Ludtke S J, Baldwin P R, Chiu W. EMAN: semi-automated software for high-resolution single-particle reconstructions. J Struct Biol, 1999, 128: 82–97
18 Liang Y R, Ke E Y, Zhou Z H. IMIRS: a high-resolution 3D reconstruction package integrated with a relational image database. J Struct Biol, 2002, 137: 292–304
19 Willett P. Chemical similarity searching. J Chem Inf Comp Science, 1998, 38: 983–996
20 Harauz G, van Heel M. Exact filters for general geometry three dimensional reconstruction. Optik, 1986, 73: 146–156