33rd Annual International Conference of the IEEE EMBS
Boston, Massachusetts USA, August 30 - September 3, 2011

A Fast Mapping Method in the ISAF Reconstruction Algorithm

Gongming Wang, Fa Zhang, Qi Chu, Fei Sun, Zhiyong Liu

Abstract—**ISAF (icosahedral symmetry-adapted functions) algorithm is the new high-resolution algorithm of icosahedral molecules. But its running speed is very slow because of the time-consuming operations of mapping sampling points into 3D space. In this paper, a fast mapping method is proposed to increase the running speed of this stage. First of all, the angle corresponding to one pixel arc in the maximum Fourier ring was taken as the sampling angle and the same angle sampling was applied in every rings. After that, the sampling points in ring R=1 were mapped into 3D space. Finally, the 3D spatial positions of radial sampling points in other rings were deduced according to the rotate angle invariability of radial sampling points. The simulated data of PSV-F (Penicillium stoloniferum virus F) and experimental Cryo-EM data of CPV (cytoplasmic polyhedrosis virus) were used for validating this method. The results show that the whole speedup reaches to an order of magnitude at the premise of assuring accuracy. In addition, the speedup is increasing with the increase of the maximum Fourier radius and the number of projections.**

I. Introduction

Infectious diseases have emerged and threatened human existence, such as HBV, HIV, SRAS etc, most of which are caused by viruses. Thus, to wipe out these diseases, the first important step is to determine the structures of corresponding viruses, which is the major content of the structural biology. Among all methods to generate the molecular structures, ISAF algorithm [1,2] based on Cryo-EM images can reach the highest resolution. The atomic models of viruses have been built with this method [3,4]. But the running speed of this algorithm is very slow. Liu has increased the speed of constructing fitting coefficient equation set and calculating density function with the asymmetric unit [2]. Wang has speeded up the calculation of density functions [5]. However, its running speed remains slow. Thus, it is very important to increase its running speed.

The ISAF algorithm contains four stages: mapping molecular images, constructing and solving fitting coefficient equation set, calculating density function and writing MRC file. In the first stage, the uniform arc sampling policy is applied to EM images at first [6]. After that, these sampling points are mapped onto 3D space. The mapping operation is the matrix-vector multiplication. The number of sampling points reaches about tens of millions. Thus, this operation is time-consuming. Thus, reducing the number of mapping operation can improve the performance of ISAF algorithm.

To solve this problem, the uniform angle sampling policy is applied. The sampling angle intervals in different rings are all the same and the one in the maximum Fourier radius is one pixel. Because the polar angles of the mapped points derived from radial sampling points are all the same, our method only calculates the polar angles from one ring and the one in other rings can be deduced. Therefore, the speedup in this stage is $S=(R_m+1)/2$ and the whole speed of ISAF algorithm also increases to some extent.

To verify our method, the experiments were carried out with simulated data of PSV-F and experimental Cryo-EM data of CPV. The experimental results show that the whole speedup of ISAF algorithm is generally over 10 times and up to 15 times for PSV-F and is over 5 times and up to 7 times for CPV. Furthermore, the whole speedup is increasing with the increase of the maximum Fourier radius and the number of images.

II. Problem Analysis And Solution

A. Problem description

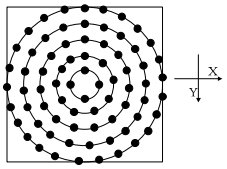


Fig. 1. The uniform arc sampling in the conventional ISAF algorithm

For ISAF algorithm, the first stage is mapping molecular images. One Cryo-EM image is divided into several rings and distance between two adjacent rings are one pixel. In every ring, uniform arc sampling is applied and the sampling intervals are one pixel according to Shannon sampling theorem [6]. It is shown in Fig. 1.

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} \cos\varphi\cdot\cos\theta\cos\omega - \sin\varphi\cdot\sin\omega & -\cos\varphi\cdot\cos\theta\sin\omega - \cos\omega\sin\varphi & \cos\varphi\cdot\sin\theta \\ \sin\varphi\cdot\cos\theta\cos\omega + \cos\varphi\cdot\sin\omega & -\sin\varphi\cdot\cos\theta\sin\omega + \cos\omega\cos\varphi & \sin\varphi\cdot\sin\theta \\ -\sin\theta\cos\omega & \sin\theta\sin\omega & \cos\theta \end{bmatrix} \cdot \begin{bmatrix} x \\ y \\ 0 \end{bmatrix} \quad (1)$$

Every EM image is placed XOY coordinate plane. 3D coordinate of point (x,y) in EM image is $(x,y,0)$ and must be mapped onto (X,Y,Z) with Eq. 1.

Thousands of images are required in reconstruction.

This work was supported by knowledge innovation project of CAS (KGCX1-YW-13), and National Science Fundation of China (90912001).

Gongming Wang, Fa Zhang, Qi Chu and Zhiyong Liu are with Institute of Computing Technology, CAS(CO:100190, Beijing, China). Fei Sun is with Institute of Biophysics, CAS(CO:100101, Beijing, China). Besides, Gongming Wang and Qi Chu are also with Graduate University of CAS(CO:100190, Beijing, China).

The emails are wanggongming@ict.ac.cn, zf@ncic.ac.cn, chichisheep@gmail.com, feisun@ibp.ac.cn and zyliu@ict.ac.cn.

The corresponding authors are Fa Zhang (phone:(8610)62601012, fax:(8610)62601356) and Zhiyong Liu(phone:(8610)62600653, fax:(8610)62601356).

Thus, the number of sampling points reaches about tens of millions. Eq. 1 is a matrix-vector multiplication and applied to every sampling point. So, the mapping operation is time-consuming.

B. Rotated angle invariability of radial sampling points

Theorem: For one EM image, $(\varphi_0,\theta_0,\omega_0)$ is orientation, O is center, and R_1, R_2 are two radial sampling points. In 3D space, the polar radiuses of R_1 and R_2 are $|R_1|$ and $|R_2|$, and the polar angles are $(0.5\pi,\varphi)$. The two points are rotated with Z axis, Y axis, and Z axis orderly according to orientation $(\varphi_0,\theta_0,\omega_0)$. After that, they are mapped into points R_1' and R_2' in 3D space. Their positions are $(|R_1'|,\theta_1,\varphi_1)$ and $(|R_2'|,\theta_2,\varphi_2)$. Thus, we can conclude that $|R_1'|=|R_1|$、$|R_2'|=|R_2|$、$\theta_1=\theta_2$、$\varphi_1=\varphi_2$.

Proof: The line OR_1 is taken as a rigid body. The polar radius of R_1' is the distance between origin O and point R_1'. It is constant in the course of rotating according to the shape invariance of moving rigid body. So, $R_1'=R_1$ is true. Similarly, $R_2'=R_2$ is also true.

Let (x_1,y_1,z_1) and (x_2,y_2,z_2) are the Cartesian coordinates of R_1 and R_2. Let (x_1',y_1',z_1') and (x_2',y_2',z_2') are the Cartesian coordinates of R_1' and R_2'. We can obtain the flowing Eq. 2 and 3.

$$R_1\begin{cases}x_1=|R_1|\cos\varphi\\ y_1=|R_1|\sin\varphi\\ z_1=0\end{cases}\quad R_2\begin{cases}x_2=|R_2|\cos\varphi\\ y_2=|R_2|\sin\varphi\\ z_2=0\end{cases}\tag{2}$$

$$R_1'\begin{cases}x_1'=|R_1'|\sin\theta_1\cos\varphi_1\\ y_1'=|R_1'|\sin\theta_1\sin\varphi_1\\ z_1'=|R_1'|\cos\theta_1\end{cases}\quad R_2'\begin{cases}x_2'=|R_2'|\sin\theta_2\cos\varphi_2\\ y_2'=|R_2'|\sin\theta_2\sin\varphi_2\\ z_2'=|R_2'|\cos\theta_2\end{cases}\tag{3}$$

The $Rot(\varphi_0,\theta_0,\omega_0)$ is rotation matrix that maps 2D sampling points onto 3D space. Thus:

$$\begin{aligned}(x_1',y_1',z_1')^T&=Rot(\varphi_0,\theta_0,\omega_0)\cdot(x_1,y_1,z_1)^T\\ (x_2',y_2',z_2')^T&=Rot(\varphi_0,\theta_0,\omega_0)\cdot(x_2,y_2,z_2)^T\end{aligned}\tag{4}$$

By taking Eq. 2 and 3 into Eq. 4, and eliminating common items $|R_1|$ and $|R_2|$, we can obtain:

$$(\sin\theta_1\cos\varphi_1,\sin\theta_1\sin\varphi,\cos\theta_1)^T=(\sin\theta_2\cos\varphi_2,\sin\theta_2\sin\varphi_2,\cos\theta_2)^T$$

The corresponding items are equivalent, that is:

$$\begin{aligned}\sin\theta_1\cos\varphi_1&=\sin\theta_2\cos\varphi_2\\ \sin\theta_1\sin\varphi_1&=\sin\theta_2\sin\varphi_2\\ \cos\theta_1&=\cos\theta_2\end{aligned}\tag{5}$$

In Eq. 5, $\cos\theta_1=\cos\theta_2$ and $0\le\theta_1,\theta_2\le\pi$, $\cos\theta$ is monotonic decreasing in range $\theta\in[0,\pi]$. Thereby, $\theta_1=\theta_2$.

In Eq. 5, using $\theta_1=\theta_2$ produces: $\cos\varphi_1=\cos\varphi_2$. $\sin\varphi_1=\sin\varphi_2$

If $\varphi_1\ne\varphi_2$, produces: $0\le\varphi_1,\varphi_2\le2\pi$ $\varphi_1+\varphi_2=2\pi$

So, $\cos\varphi_1=\cos\varphi_2$ if and only if $\varphi_1=\varphi_2=\pi$, not for all $0\le\varphi_1,\varphi_2\le2\pi$. Thus, the hypothesis $\varphi_1\ne\varphi_2$ is false.

Then, we can conclude that $\varphi_1=\varphi_2$.

In light of this, we must have $|R_1'|=|R_1|$、$|R_2'|=|R_2|$、$\theta_1=\theta_2$、$\varphi_1=\varphi_2$.

C. Fast mapping method

The polar angles of mapped points from radial sampling points are all the same. Thus, we only map sampling points in one ring, not in all rings. The fast mapping method is proposed and contains three stages as shown in Fig 2.

Step 1: Unifrom angle sampling

Sampling is carried out as shown in Fig 2.a, the sampling

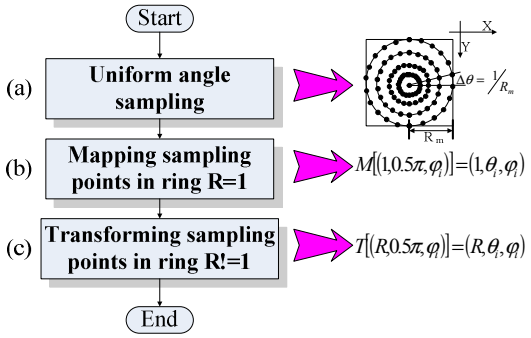


Fig. 2. The flow of the fast mapping method

angle interval in different rings are all $1/R_m$ radian (R_m is the maximum of Fourier radius).

Step 2: Mapping sampling points in ring R=1

The sampling points $(1,0.5\pi,\varphi_i)$ $i=0\ldots N-1$ whose radiuses are all one are mapped into 3D space. The mapped points are $(1,\theta_i,\varphi_i)$ $i=0\ldots N-1$. $N=[2\pi R_m]$ is the number of sampling points in this rings. $M[(1,0.5\pi,\varphi_i)]=(1,\theta_i,\varphi_i)$ is short for this stage.

Step 3: Transfroming sampling points in ring R!=1

In other rings $R_P\ne1$, the mapped points derived from sampling points $(R_P,0.5\pi,\varphi_i)\,i=0\ldots N-1$ are (R_P,θ_i,φ_i) $i=0\ldots N-1$. $T[(R_p,0.5\pi,\varphi_i)]=(R_p,\theta_i,\varphi_i)$ is short for this stage.

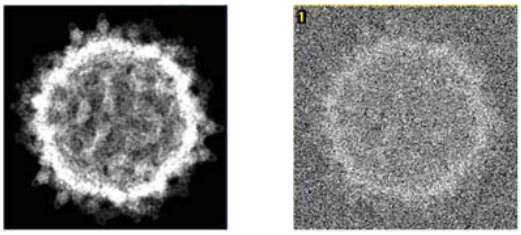
With conventional method, $\sum_{R=1}^{R_m}[2\pi R]$ points be mapped. But $[2\pi R_m]$ points are mapped with our method. So, its speedup is $S=\sum_{R=1}^{R_m}[2\pi R]\Big/[2\pi R_m]=(R_m+1)/2$.

III. Experiments And Discussion

To validate fast mapping method, the simulated data of PSV-F (Penicillium stoloniferum virus F) and experimental Cryo-EM data of CPV (cytoplasmic polyhedrosis virus) were applied. The following experiments were all implemented on one PC with CPU AMD Dual-Core 2.6G

Hz, and DDR 2G.Byte memory.

A. The reconstruction experiment with simulated data



(a) (b)

Fig. 3. Simulated images of PSV-F. (a) Projection without noise. (b) The same projection with added noise, with SNR of 0.1

The simulated data were projections of map generated from PSV-F structure (PDB ID 3ES5), which was solved by X-ray crystallography to 3.3 Å [7] and has icosahedral symmetrical structure. Firstly, this model was generated at a resolution of 3.8 Å by using pdb2mrc [8]. The key parameters were apix=1.3 and res=3.8. The size of model was $304\times304\times304$ pixels. Secondly, 2000 projections with random orientations were generated by using prjectIcosFile [9]. Finally, the random noise with SNR of 0.1 was added to these projections. The simulated images with and without noises are shown in Fig 3. The conventional mapping method and our fast mapping method were both tested. The maximum Fourier radius was 1/3.9Å^{-1}. For the purpose of comparison, the original maps at resolution of 3.9Å were generated with filtering.

The reconstruction results are shown in Fig 4. It can be seen that the maps with two methods are all the same.

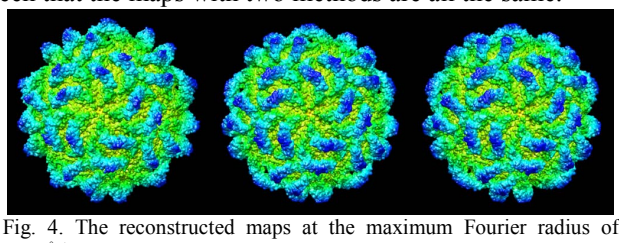


Fig. 4. The reconstructed maps at the maximum Fourier radius of 1/3.9Å^{-1}. From left to right, the first column is the original map. The second column is the reconstructed map with the conventional method. The third column is the reconstructed map with the fast method.

To validate its accuracy quantitatively, the FSC (Fourier Shell Correlation) curves [7] of the reconstructed maps with two methods at the maximum Fourier radius of 1/3.9Å^{-1} were calculated. They are shown in Fig 5. The FSC curves with two methods are identical and they drop to 0.5 at 1/3.9Å^{-1}. It indicates that the effective resolutions with two methods are both 3.9Å. So, the fast method could not influence the accuracy.

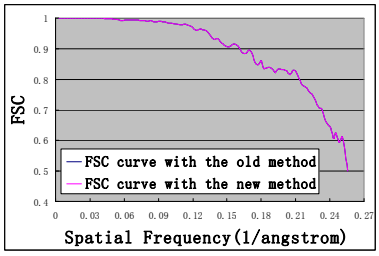


Fig. 5. The FSC curves of the reconstructed maps with two methods at the maximum Fourier radius of 1/3.9Å^{-1}

The Eq. 6 was used for calculating its speedup.

$$A=\frac{Ori\_t\_prj+t\_nonprj}{t\_prj+t\_nonprj}=\frac{t\_prj\cdot S+t\_nonprj}{t\_prj+t\_nonprj} \quad (6)$$

Where, Ori_t_prj, t_prj are mapping times of conventional method and fast method. t_nonprj is other running time. $S=(R_m+1)/2$ is local speedup.

TABLE 1. THE SPEEDUPS IN DIFFERENT NUMBERS OF PROJECTIONS AT THE MAXIMUM FOURIER RADIUS OF 1/3.9Å^{-1}

	t_prj	S	t_nonpuj	A
250 projections	7.44	51	216.72	2.66
500 projections	16.98	51	227.81	4.47
1000 projections	28.02	51	243.268	6.16
2000 projections	72.3	51	257.232	11.19

TABLE 2. THE SPEEDUPS AT THE DIFFERENT MAXIMUM FOURIER RADIUS WITH 2000 PROJECTIONS

	t_prj	S	t_nonpuj	A
1/32.825Å^{-1}	9.36	6.5	17.985	2.88
1/13.13Å^{-1}	18.78	15.5	64.562	4.27
1/6.565Å^{-1}	33.66	30.5	35.361	15.39
1/3.9Å^{-1}	72.3	51	257.232	11.19

The speedups were shown in table 1 and table 2. The units in two tables are second. It can be seen that the whole speedup using our method is increasing with the increase of the target resolution and the number of projections. Under the circumstance of high resolution (less than 10Å), the speedup is often over 10 times and up to 15 times.

B. The reconstruction experiment with experimental Cryo-EM data

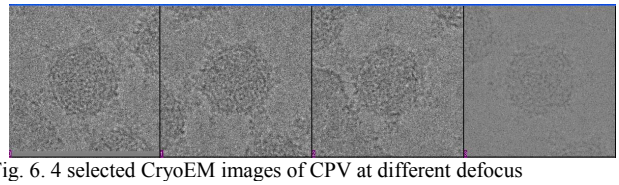


Fig. 6. 4 selected CryoEM images of CPV at different defocus

The CPV has icosahedral structure. Currently, a full atomic model of the CPV capsid was built [4]. The selected 2460 images were used for validating our method.

Fig 6 shows 4 images at different defocus. The imaging parameters are as follows: image size is 1024×1024 pixels, sampling step is 1.3Å/pixel, electron microscopy is FEI 300-kV Titan Krios, speed voltage is 300kV, spherical-aberration coefficient is 2.7 mm, and defocus [10] is between 0.68μm and 3.26μm.

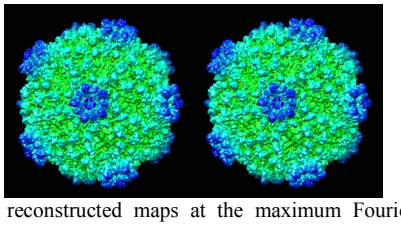


Fig. 7. The reconstructed maps at the maximum Fourier radius of 1/7.61Å^{-1}. The left maps are generated with conventional method. The right maps are generated with fast method.

The conventional method and the fast method were used for reconstruction experiments. The maximum Fourier

radius was 1/7.61Å^{-1}. The reconstruction results are shown in Fig 7 and indicate that the reconstruction results with two methods are identical.

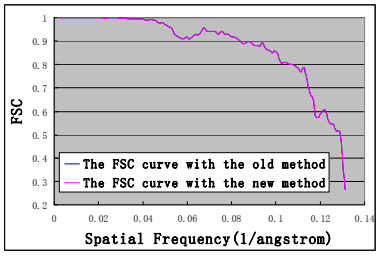


Fig. 8. The FSC curves of the reconstructed maps with two methods at the maximum Fourier radius of 1/7.61Å^{-1}

The FSC curves of the reconstructed maps with two methods at the maximum Fourier radius of 1/7.61Å^{-1} were calculated. They are shown in Fig 8. They are identical and

TABLE 3. THE SPEEDUPS IN DIFFERENT NUMBERS OF IMAGES AT THE MAXIMUM FOURIER RADIUS OF 1/7.61Å^{-1}

	t _ prj	*S*	*t _ nonpuj*	*A*
308 images	15.06	50	706.077	2.02
615 images	29.1	50	717.905	2.91
1230 images	63.78	50	741.187	4.88
2460 images	137.52	50	882.536	7.61

TABLE 4. THE SPEEDUPS AT THE DIFFERENT MAXIMUM FOURIER RADIUS WITH 2460 IMAGES

	t _ prj	*S*	*t _ nonpuj*	*A*
1/30.43Å^{-1}	26.34	12.9	212.733	2.31
1/15.22Å^{-1}	57	25.3	296.457	4.92
1/10.14Å^{-1}	91.92	37.6	423.578	7.53
1/7.61Å^{-1}	137.52	50	882.536	7.61

drop to 0.5 at 1/7.61Å^{-1}. It indicates the effective resolutions with two methods are both 7.61Å. Thus, the fast method could not influence the resolution.

The speedups were shown in table 3 and table 4. The units in two tables are second. It can be seen that the whole speedup using the fast mapping method is increasing with the increase of the maximum Fourier radius and the number of projections. Under the circumstance of high resolution, the speedup is often over 5 times and up to 7 times.

From the above experiments, it can be seen that the fast mapping method can sharply decrease the running time of mapping molecular images without losing accuracy. Thus, it can increase the speed of whole reconstruction.

IV. CONCLUSION

The large number of mapping operations has influenced the running speed of ISAF algorithm greatly. To solve this problem, a fast mapping method is proposed. Its essence is uniform angle sampling, not conventional uniform arc sampling. The polar angles of mapped points from radial sampling points are all identical. Thus, mapping operations are only performed in one ring, not in all rings. Thus, it can decrease the number of mapping operations and the overall speed of ISAF algorithm can be increased.

The simulated data of PSV-F and experimental Cryo-EM data of CPV were used to validate this method. The results indicate that the overall speedup reaches an order of magnitude without losing accuracy. Furthermore, its speedup is increasing with the increase of the maximum Fourier radius and the number of images.

There are other-hedral symmetry molecules except for icosahedral molecule. The mapping strategies of them are similar to that of icosahedral molecule. Therefore, our future work lies on performing this method for 3D reconstruction of other-hedron symmetrical molecules.

REFERENCES

[1] Navaza.J, "On the three-dimensional reconstruction of icosahedral particles," J. Struct. Biol, vol. 144, pp. 13-23, 2003.

[2] Hongrong Liu, Lingpeng Cheng, Songjun Zeng, Canying Cai, Z. Hong Zhou, and Qibin Yang, "Symmetry-adapted spherical harmonics method for high-resolution 3D single-particle reconstructions," J. Struct. Biol, vol. 161, no. 1, pp. 64-73, 2008.

[3] Hongrong Liu, Lei Jin, Sok Boon S. Koh, Ivo Atanasov, Stan Schein, Lily Wu, et al. "Atomic Structure of Human Adenovirus by Cryo-EM Reveals Interactions Among Protein Networks." Science, vol. 329, no. 1038, pp. 1038-1043, 2010.

[4] Lingpeng Cheng, Jingchen Sun, Kai Zhang, Zongjun Mou, Xiaoxing Huang, Gang Ji, et al. "Atomic model of a cypovirus built from cryo-EM structure provides insight into the mechanism of mRNA capping," PNAS, vol. 108, no. 4, pp. 1373-1378, 2011.

[5] Gongming Wang, Fa Zhang, Qi Chu, Liya Fan, Fei Sun, and Zhiyong Liu, "A Fast Calculation Strategy of Density Function in ISAF Reconstruction Algorithm," SCIENCE CHINA. F Series, to be published.

[6] R.A.Crowther, D.J.DeRosier and A.Klug, "The Reconstruction of a Three-Dimensional Structure from Projections and its Application to Electron Microscopy," Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences, vol. 317, no. 1530, pp. 319-340, 1970.

[7] Junhua Pan, Liping Dong, Li Lin, Wendy F. Ochoa, Robert S. Sinkovits, Wendy M. Havens, et al. "Atomic structure reveals the unique capsid organization of a dsRNA virus," PNAS, vol. 106, no. 11, pp. 4225-4230, 2009.

[8] Ludtke. S.J, Baldwin. P.R, and Chiu. W, "EMAN: semi-automated software for high resolution single particle reconstructions," J. Struct. Biol, vol, 128, pp. 82–97, 1999.

[9] Liang. Y, Ke. E.Y, and Zhou. Z.H, "IMIRS: a high-resolution 3D reconstruction package integrated with a relational image database," J. Struct. Biol. vol. 137, pp. 292–304, 2002.

[10] Erickson,H.P. and A.Klug, "The Fourier transform of an electron micrograph: efects of defocussing and aberrations, and implications for the use of underlocus contrast enhancement," Berichte der Bunsengesellschaft für physikalische Chemie, vol. 74, no. 11, pp. 1129-1137, 1970.