AreTomo3 User Manual

Shawn Zheng

Chan Zuckerberg Imaging Institute

shawn.zheng@czii.org

AreTomo3 is a multi-GPU accelerated software package that enables real-time fully automated reconstruction of cryoET tomograms in parallel with cryoET data collection. Integrating MotionCor3, AreTomo2, and GCtfFind in a single application, AreTomo3 has established an autonomous preprocessing pipeline that, whenever a new tilt series is collected, can activate and repeat itself from correction of beam induced motion and assembling tilt series to CTF estimation and correction, tomographic alignment, and 3D reconstruction throughout entire session of data collection without human intervention. The end results include not only tomograms but also a rich set of alignment parameters to bootstrap subtomogram averaging. Our test shows that AreTomo3 can catch up to 9-target PACE Tomo data collection with 4 NVidia RTX A6000 GPUs when it was configured to perform 2D local motion correction on movies and 3D local motion correction on tilt series. The offline testing shows that AreTomo3 runs faster than the data collection. As a result, GPU resources can be shared with other tasks to expand the preprocessing capacity. Tomogram denoising and particle picking now can run concurrently with AreTomo3 to maximize the preprocessing workflow.

Version **2.0.6**

Revised on Jan 01, 2025

Contents

1. System Requirement and Installation

AreTomo3 is a multi-GPU accelerated Linux application written in C++/CUDA. It is recommended to run AreTomo3 on a Linux platform equipped with four or more state-of-the-art NVidia GPUs. Ideally each GPU has at least 20 GB memory. Systems with GPUs having smaller memory may also work but are not tested. It is recommended to have at least 512 GB CPU memory in total or 100 GB CPU memory per GPU on average. When AreTomo3 is run on fewer GPUs, less CPU memory is needed at the cost of reduced performance.

AreTomo3 is a single-program application. Once unzipped, no further installation is needed. However, AreTomo3 depends on **CUDA driver** and **toolkit** as well as **libtiff** library. Ideally they are installed by system administrators and thus shared across different user accounts. If not, users can install them using anaconda or miniconda in their local accounts. Once installed, their installation paths need to be added to **LD_LIBRARY_PATH**, a Linux environment variable in **.bashrc** or **.cshrc** file.

1.1 Install from source

When the source code is downloaded, the binary can be generated using two included **makefile** and **makefile11**. Open a Linux terminal and execute at the command prompt **make exe** if CUDA 10.x is installed or **make exe -f makefile11** when CUDA 11.x or 12.x is installed. If you do not know what CUDA version is installed, it can be found by running **nvcc -- version** at the command prompt.

Both makefile and makefile11 define **CUDAHOME**, a variable that holds the full path to CUDA toolkit installation point. Use CUDAHOME in the provided makefile and makefile11 as an example to help locate where CUDA toolkit is installed on your system. Then revise CUDAHOME accordingly.

When NVIDIA releases the next generation of GPU cards, the new compute capability may not be included in the existing makefiles. The compiled version of AreTomo3 will cease to function with an error message stating “invalid device ordinal”. In this case, user would need to recompile AreTomo3 from the source. The first step is to identify the compute capability of the new GPU card, which can be found by searching NVIDIA website. The next step is to add a line to the relevant section of the makefile, as shown below. Once that is done, run “make clean -f makefile11” followed by “make exe -f makefile11”.

```
CUFLAG = -Xptxas -dlcm=ca -O2 \
    -gencode arch=compute_90,code=sm_90 \
    -gencode arch=compute_89,code=sm_89 \
    -gencode arch=compute_86,code=sm_86 \
    -gencode arch=compute_80,code=sm_80 \
    -gencode arch=compute_75,code=sm_75 \
    -gencode arch=compute_70,code=sm_70 \
    -gencode arch=compute_61,code=sm_61
```

1.2 Compile the library files

There are two precompiled library files, **libutil.a** and **libmrcfile.a** in **LibSrc/Lib**. If you run into any problems in linking these two files, try recompiling them using the provided makefiles in their respective source directories. Compile libutil.a first followed by libmrcfile.a. Run **make clean** and then **make all** in LibSrc/Util directory and this will generate libutil.a, which is automatically moved into /LibSrc/Lib at the end of compiling. Repeat the same steps in LibSrc/Mrcfile.

1.3 AreTomo3 Configuration

Users need to set up environment variable LD_LIBRARY_PATH to include the full path of CUDA library. You use check the value of it by running "echo $LD_LIBRARY_PATH". In the following screen shot, the path of CUDA library is underscored with the yellow line.

```
(base) [shawn.zheng@czii-gpu-a-1 ~]$ echo $LD_LIBRARY_PATH
/hpc/apps/x86_64/cuda/12.0.0_525.60.13/lib64:/hpc/apps/x86_64/cuda/12.0.0_525.60.13/lib:/
home/shawn.zheng/miniconda3/lib:/hpc/slurm/installed/current/lib
```

When you see an error message like the below, it is likely that LD_LIBRARY_PATH is not set up properly.

```
/home/shawn.zheng/szheng/CuProjs/AreTomo2/AreTomo2: error while loading shared libraries:
 libcufft.so.11: cannot open shared object file: No such file or directory
```

2. MDOC or MRC(s) files

Starting with version 1.0.23, tilt series **MDOC** files are required only when running AreTomo3 in live processing mode. Users can process their **MRC**(s) tilt series in batch mode without MDOC files. AreTomo3 determines the correct starting based on user input to **-InSuffix**. Since AreTomo3 is no longer tied to MDOC files, we have replaced **-InMdoc** used in previous versions with **-InPrefix**. Currently, AreTomo3 can only perform offline processing of MRC(s) files.

A tilt series MDOC file contains multiple **ZValue** sections, one for each tilt angle. The following is an example of a ZValue section.

```
[ZValue = 0]
TiltAngle = -0.03
StagePosition = 54.95 56.35
StageZ = 0.82
Magnification = 81000
Intensity = 0.00
ExposureDose = 3.76
PixelSpacing = 1.54
SpotSize = 6
Defocus = -3.97
ImageShift = 3.24 0.33
RotationAngle = 83.94
ExposureTime = 1.00
Binning = 1
MagIndex = 29
CountsPerElectron = 2121.78
MinMaxMean = 502.00 4407.00 2154.42
TargetDefocus = -2.00
PriorRecordDose = 0.00
SubFramePath = 20240216_002_Krios1_RP_5proteinsMixB1s_grid1\Position_5_10_001_-0.01_20240216_111455_EER.eer
NumSubFrames = 1
FrameDosesAndNumber = 3.76 1
DateTime = 16-Feb-2024  11:14:58
FilterSlitAndLoss = 10.00 0.00
ChannelName =
CameraLength = NaN
```

Fig. 1 Excerpt of tilt series mdoc file showing the first ZValue section. Each tilt image has a ZValue section. AreTomo3 parses each ZValue section to extract the tilt angle and the file name of movie. In some cases, data collection software may also produce per-tilt mdoc file. Make sure to use tilt series mdoc file.

AreTomo3 uses **ZValue**, **TiltAngle**, **ExposureDose**, and **SubFramePath** to assemble motion-corrected tilt images into tilt series and produce the ordered list file needed by Relion subtomogram averaging. Since movie files can be transferred for one location to another, the directory path provided by SubFramePath is ignored in AreTomo3. Instead, AreTomo3 requires mdoc files to be with their movie files in the same and flat directory.

Note that some data collection software may also generate per-tilt mdoc files. **AreTomo3** needs **tilt-series mdoc** files that contain multiple ZValue sections.

3. Supported Movie Files

AreTomo3 supports both **EER** and **TIFF** movies. For EER movies, it is strongly recommended to use **frame integration file** to generate rendered movies for measuring and correcting beam induced motion, since processing raw movies is extremely time consuming and demanding on computing resources. Frame integration is a software implementation of variable frame rates within each exposure to account for the rapid early motion. An example of the frame integration file is given below.

```
50  10  0.006
100 20  0.006
300 30  0.006
```

Fig. 2 An example of frame integration file, a 3-column text file that specifies how raw frames are summed to generate rendered frames. The measurement and correction of beam induced motion are performed

on rendered frames, not raw frames. In this example, **Column 1** shows that raw movie frames are divided into three non-overlapping groups of sizes 50, 100, and 300, respectively. If the sum of the values in column 1 exceeds the total number of frames, AreTomo3 will automatically adjust by reducing the values. **Column 2** determines how many raw frames are summed to create each rendered frame within each group. **Optional Column 3** displays the raw frame dose, which is used to calculate the dose of the rendered frames. This column is relevant only when the ExposureDose field in the mdoc file is not filled in. The values in column 3 must be **constant** if provided. If they are not consistent, frame integration will be disabled, which may lead to inaccurate correction of beam induced motion and excessive use of system resources since AreTomo3 will work raw frames instead of integrated frames.

Use -**FmIntFile** followed by the path of the frame integration file to enable this function.

4. Use Cases

AreTomo3 can be invoked from Linux command line. The version number and built date can be found by running **AreTomo3 --version**. If this command fails, this usually means that there are some issues with CUDA toolkit installation. One problem could be the LD_LIBRARY_PATH is not configured to include the path to the proper CUDA library. Users can get quick help information about the command line parameters by running command **AreTomo3 --help.**

The major operations in **AreTomo3** include correction of beam induced motion, assembling full, odd, and even tomographic tilt series, CTF estimation, tomographic alignment, local CTF correction of tilt series, and tomographic reconstruction of local-CTF corrected full, odd, and even tomograms.

AreTomo3 has multiple entry points that allow users to repeat a subset of operations. These entry points are described below. The different entry points are controlled by **-Cmd**.

Entry Point	Operations
-Cmd 0	Full processing from movies to tomograms. Default setting
-Cmd 1	Start from tomographic alignment.
-Cmd 2	Repeat only tomographic reconstruction
-Cmd 3	Repeat only CTF estimation
-Cmd 4	Rotate the tit axis by 180^{o} and reconstruct tomograms

4.1 Live processing pipeline (-Cmd 0)

Live processing pipeline starts with the correction of beam induced motion and ends with the generation of 3D tomograms. Here is the command line that we used to run AreTomo3.

```
AreTomo3 -InPrefix ./Position -InSuffix .mdoc -InSkips _override -OutDir ./run001/
-gain ./20240209_114818_EER_GainReference.gain -FmIntFile ~/FmIntFile.txt
-Group 2 4 -PixSize 1.54 -McPatch 5 5 -AtPatch 4 4 -AtBin 4 4 6 -Wbp 1
-FlipVol 1 -TiltCor 1 -AlignZ 600 -VolZ 1200
-Gpu 0 1 -Cmd 0 -Serial 1000 -TmpDir ver23/temp 2>/dev/null
```

Fig. 3 A command line example for running live or offline full processing starting from the correction of bead induced motion. **-Cmd 0** is the default setting for processing from the correction of beam induced motion on movies to reconstruction of tomograms. **-Serial 1000** means that AreTomo3 will wait 1,000 seconds for a new tilt series saved into the current directory before exiting.

-InPrefix ./Position and -InSuffix .mdoc together instruct AreTomo3 to look for in the current directory all the mdoc files whose names start with Position.

-InSkips _override informs AreTomo3 to skip all the mdoc files with -override in their names. We encountered in several cases where two mdoc files in the same directory, for example, Position_1_3.mdoc and Position_1_3_override.mdoc, refer to the same EER movie files. **-InSkips _override** was used in the command line to avoid processing the same movies twice.

-OutDir specifies where the output directory is. In this example, the generated tilt series, tomograms, CTF estimation results, and alignment parameters are saved in ./run001/ directory. Note that the ending back slash is optional.

-Gain specifies where to load the gain reference file. **AreTomo3** can load gain reference saved in either a .gain or a .mrc file.

-InFmIntFile specifies where to load the frame integration file that converts raw frames into rendered or integrated frames, non-overlapping sums of raw frames. This is a software implementation of variable frame rates. Beam induced motion is measured and corrected on the integrated frames. It is recommended to integrate less in early frames and more in later ones to take into account of fast early beam induced sample motion. **-InFmIntFile** and **-Group** are usually used together for robust motion correction.

-Group 2 4 in this example informs AreTomo3 how to form non-overlapping groups of the integrated frames. Beam induced motion is measured on the group sums. The measurements are then interpolated and extrapolated to each individual integrated frames. For the measurement of the global motion, each group contains 2 integrated frames. For local motion measurement, each group contains 4 integrated frames.

-McPatch 5 5 informs AreTomo3 to run 5x5 patch-based motion correction. If not specified, only global motion is corrected.

-McBin 1 informs AreTomo3 not to perform Fourier cropping on the motion corrected images.

-PixSize specifies the pixel size of the input movies. The pixel size of motion corrected images is the multiplication of this pixel size and the **-McBin** value.

-AtBin 4 4 6 will generate **three** tomograms reconstructed with 4x, 4x, and 6x binnings, respectively. Binning in AreTomo3 is performed by Fourier cropping the aligned tilt series before reconstruction. The pixel size of the generated tomogram is the multiplication of the **-PixSize**, **-McBin**, and **-AtBin** values. The reconstruction of 2nd and 3rd tomograms can be disabled by setting the corresponding binning factors zero. The first binning factor is required.

-AtPatch 4 4 invokes 4x4 patch-based local alignment after global tilt series alignment.

-FlipVol 1 flips the generated tomogram by rotating it around x axis. When the tomogram is saved into a MRC file, each section is a x-y slice of the tomogram. By default, tomograms are not flipped. In this case, a MRC section is a x-z slice of tomograms.

-OutImod 1 generates needed Imod files to bootstrapping Relion subtomogram averaging. These Imod files are saved in a created sub-folder in the output directory.

-Wbp 1 specifies weighted back projection for tomogram reconstruction. Another option is SART reconstruction.

-Gpu is followed by GPU IDs that are involved in the computation.

-Serial 1000 means that AreTomo3 will wait maximum 1000 seconds for the next tilt series to be collected. If still no available, AreTomo3 will quit after finishing all the ongoing processing. This is specific for the live processing. When a new tilt series becomes available, AreTomo3 will reset the timer and wait another 90000 seconds before quitting.

4.2 Interrupting and Resuming

AreTomo3 can be interrupted by **ctrl c** command by pressing **ctrl** and **c** keys at the same time. This mechanism can terminate running AreTomo3 either at the end of data collection or when users need to adjust AreTomo3 command line parameters.

AreTomo3 can also be resumed after it is interrupted without reprocessing tilt series that have been reconstructed. This is implemented by tracking finished tilt series in a text file (**MdocDone.txt**) saved in the output directory. **-Resume 1** works only with **-Cmd 0** to resume the processing when AreTomo3 is terminated prematurely. This function instructs **AreTomo3** to skip tilt series that are already reconstructed and proceeds with newly corrected tilt series.

4.3 Starting from tomographic alignment with mrc(s) files (-Cmd 1)

Should users want to realign tilt series and reconstruct the tomograms, this is the command for this purpose. This command can be executed only after the tilt series and the associated tilt angle files have already been generated. If you have run AreTomo3 with -Cmd 0, these files should be saved in the output directory.

Assuming we have run AreTomo3 with -Cmd 0 and saved all the generated files in ./run001/ directory. The following command line is how we reprocess the tilt series to generate a new set of tomograms in batch mode and save them in ./run002/.

```
AreTomo3 -InPrefix ./ver206/Position_ -InSuffix .mrc
-InSkips _ODD,_EVN,_Vol,_CTF
-OutDir ./run002/
-AtBin 6 -AtPatch 4 4 -VolZ 1800 -OutImod 1 -Wbp 1 -FlipVol 1 -TiltCor 1
-Cmd 1 -Gpu 0 1 -Serial 1
```

Here, **-Serial 1** combined with -InSkips informs AreTomo3 to perform batch processing of tilt series whose names start with Position_ and end with .mrc excluding those files with _ODD, _EVN, _Vol, and _CTF in their names. We use this example to show you how to exclude files from further processing. If **-Serial 1** were not given, AreTomo3 would look for a single tilt series named Position_. Since it would not be able to find it, AreTomo3 would exit immediately.

4.4 Tomographic reconstruction only (-Cmd 2)

-Cmd 2 goes straight to reconstruction of tomograms. This command is intended for users to reconstruct tomograms differently using the same alignment parameters. Full processing (-Cmd 0) needs to be done first before running this command. Users can change how volume is reconstructed by changing the settings of **-VolZ**, **-FlipVol**, **-Wbp**, and **-Sart**.

```
AreTomo3 -InPrefix ./run001/Position_ -InSuffix .mrc
-OutDir ./run002/ -Cmd 2 -Serial 1
-Wbp 1 -FlipVol 1 -VolZ 1800 -AtBin 6
-Gpu 0 1
```

Note that run001 directory must contain _TLT.txt and .aln files. The reconstructed volumes are saved in run002 directory. If the output directory is the same as the input, the original volume files will be overwritten.

4.5 CTF estimation only (-Cmd 3)

-Cmd 3 lets users repeat only CTF estimation. Full processing (-Cmd 0) needs to be done before running this command. The **-InMdoc** and **-OutDir** setup in the command line must be the same as that in the previous full processing.

```
AreTomo3 –InPrefix ./run001/Position_ –InSuffix .mrc
–OutDir ./run002/ –Cmd 3 –Serial 1
–PixSize 1.54 –Kv 300 –Cs 2.4
–Gpu 0 1
```

If the output directory is the same as the input, the newly generated files will overwrite the old ones.

4.6 Rotate tilt axis by 180° (-Cmd 4)

-Cmd 4 lets users rotate tilt axis by 180°. The estimate of the tilt axis angle may be off by 180° in the early versions of AreTomo3. This command corrects this error without realign the tilt series of an entire data set. It corrects not only the .aln files but also the Imod related files in the _Imod subdirectory. A new set of tomograms are also generated. Full processing (-Cmd 0) needs to be done before running this command.

```
AreTomo3 –InPrefix ver204/Position_ –InSuffix .mrc
–OutDir run002/ –Cmd 4 –Gpu 0 1 –Serial 1
–SplitSum 0 –InSkips _CTF,_Vol –VolZ 1800
–Wbp 1 –AtBin 8 –FlipVol 1
```

If the output directory is the same as the input, the newly generated files will overwrite the old ones.

4.7 CTF estimation and local CTF correction

CTF estimation is enabled when the pixel size is passed into AreTomo3 using -PixSize. Otherwise, no CTF estimation will be performed, nor will be the CTF correction. The default values for high tension and spherical aberration are 300 and 2.7, respectively. Users can use **-kV** and **-Cs** to pass different values.

-CorrCTF 1 enables local CTF correction on original tilt series when CTF estimation is performed. The CTF correction is tile based with each tile having its own CTF calculated based on its distance from the tilt axis. CTF correction is done on odd, even, and full tilt series before they are reconstructed into tomograms.

-CorrCTF 0 disables CTF correction.

4.8 Generating multi-res tomograms

Since version 1.0.18, **-AtBin** can accept optional second and third float values. When they are given, AreTomo3 will reconstruct two more tomogram using the second and third binning factor. The corresponding file names will be embedded with _2ND_Vol and _3RD_Vol, respectively. The second tomogram will be reconstructed using weighted back-projection without CTF correction. The third one is reconstructed using SART.

4.9 Offline batch processing of multiple MRC tilt series

AreTomo3 2.0 provides an entry point for batch-processing MRC tilt series without MDOC files. File extensions of the tilt series can be .mrc, .mrcs, or .st. The following is an example of the command line.

```
AreTomo3 -InPrefix ./Position -InSuffix .mrc -InSkips override -OutDir ./ver23/
-PixSize 1.54 -AtPatch 4 4 -AtBin 4 4 6 -Wbp 1
-FlipVol 1 -TiltCor 1 -AlignZ 600 -VolZ 1200
-Gpu 0 1 -Serial 1 -TmpDir ver23/temp
```

In this case, AreTomo3 will process any tilt series whose name starts with **Position** and ends with **.mrc** in the current directory. To enable batch processing, **-Serial 1** must be present in the command line. Without it, AreTomo3 would run in single mode. Please see section 4.10 for more details.

In the meantime, AreTomo3 also looks for the corresponding tilt angle file that has the same file name except a different file extension. The extension of tilt angle files must be either **.rawtlt** or **_TLT.txt**. Tilt series files and the associated tilt angle files must be in the same flat directory. In a tilt angle file, **column 1** in the angle file is mandatory, which contains the tilt angles in the same order as in the tilt series. **Column 2** is optional and the order of the acquisition. **Column 3** is the image dose in $e^{-}/\text{Å}^2$.

4.10 Offline processing of a single tilt series

In case users want to process a single tilt series at a time, they can use the single processing mode of AreTomo3 as shown below. This is equivalent to setting **-Serial 0**, which is the default setting. Similarly, AreTomo3 assumes tilt angles are saved in a text file that shares the same file name but ended with either **.rawtilt** of **_TLT.txt**. The tilt series file and the associated tilt angle file must be in the same directory.

In single processing mode, AreTomo3 reads the full name of the tilt series including the path provided after **-InPrefix** and ignores -InSuffix. **Do not split** the file name into prefix and suffix as is done in batch processing mode.

```
AreTomo3 -InPrefix ./Position.mrc -OutDir ./run001/
-PixSize 1.54 -AtPatch 4 4 -AtBin 4 4 6 -Wbp 1
-FlipVol 1 -TiltCor 1 -AlignZ 600 -VolZ 1200
-Gpu 0 -TmpDir ver23/temp
```

4.11 Runtime estimation of sample thickness

When **-AlignZ** is not given or set to zero, AreTomo3 will estimate sample thickness and use this value for AlignZ during the alignment.

When **-VolZ** is not given or set to -1, AreTomo3 will use the estimated sample thickness plus an extra space above and below the sample for VolZ. The value for extra space can be specified by users using **-ExtZ**, which is default to 300.

4.12 Defocus handedness

AreTomo3 uses a right-hand coordinate system that has the positive defocus handedness. The *z*-axis of this coordinate system points to the electron source. In this coordinate system, moving the sample closer to the electron source makes it less underfocused.

Note that the tilt axis angle is with respect to the *y*-axis of this coordinate system.

4.13 Local CTF correction

When pixel size, Cs, and high tension is given, AreTomo3 will perform local CTF correction for each tilt image before tomographic reconstruction. This is the default setting for the first tomogram corresponding to the first binning value following **-AtBin**. However,

Users can disable the local CTF correction using **-CorrCTF 0**.

4.14 Suppress generation of odd and even tilt series and tomograms.

-SplitSum 0 disables generation of the odd and even tilt series and tomograms.

5. Generated Files

5.1 Tilt series

After each movie is motion corrected, sums of odd frames, even frames, and all frames are generated. After all tilt movies are corrected, the odd, even, and full sums are assembled into odd, even, and full tilt series in the ascending order of tilt angles. A tilt series is named after the corresponding mdoc file. The odd and even tilt series are appended with _ODD and _EVN in their names, respectively. For example, **Position_5_10.mrc**, **Position_5_10_ODD.mrc**, and **Position_5_10_EVN.mrc** are generated and associated with **Position_5_10.mdoc**.

5.2 Tomograms

Three tomograms are reconstructed for each mdoc file. They are odd, even, and full tomograms reconstructed from odd, even, and full tilt series. They are named after the corresponding tilt series with

_Vol appended in the file names such as **Position_5_10_Vol.mrc**, **Position_5_10_ODD_Vol.mrc**, and **Position_5_10_EVN_Vol.mrc**. The odd and even tomograms are used to train in real time the model of CZII denoiser (**DenoisET**). Upon completion of model training, DenoisET then starts denoise full tomogram in real time.

5.3 Tilt angles and acquisition sequence

A tilt angle file, **Position_5_10_TLT.txt** for example, also named after the corresponding mdoc file, contains two columns. The first column is tilt angles of the corresponding odd, even, and full tilt series. The second column shows the order of tilt images acquired during the data collection and is 1-based indices.

```
-21.01    15
-18.01    12
-15.01    11
-12.02     8
 -9.02     7
 -6.02     4
 -3.02     3
 -0.03     1
  2.98     2
  5.98     5
  8.98     6
 11.98     9
 14.98    10
 17.98    13
 20.98    14
```

Fig. 4 An excerpt of the content in a TLT.txt file. The first column lists the tilt angles of images in a tilt series. The second column lists the indices that show the order of images acquired during data collection.

5.4 CTF estimation files

Two text files are generated per tilt series, one compatible with CTFFind4 format and one compatible with Imod. Again these two files are named after the corresponding mdoc file. The CTFFind4 compatible one is appended with _CTF in its name. The Imod compatible one is further appended with _CTF_Imod in its file name.

5.5 Alignment file

Tomographic alignment parameters are saved in the **aln** file that saves both global and local alignment information. When the local alignment is enabled with -AtPatch, the aln file has two sections, one for global and one for local alignment. Otherwise, only global section is present. An aln file starts with the header section where each line starts with the number sign # as shown in the following example.

```
# AreTomo Alignment / Priims bprmMn
# RawSize = 4096 4096 31
# NumPatches = 16
# DarkFrame =    28     0    29.98
# DarkFrame =    29     0    32.98
# DarkFrame =    30     0    35.98
# AlphaOffset =    -9.00
# BetaOffset =     0.00
```

Fig.5 Header section of the aln file. DarkFrame entries exist only when dark images are detected.

 # RawSize shows the image width and height in pixels and number of images in the tilt series.

 # NumPatch indicates number of patches used in local alignment. **# DarkFrame** lists the dark tilt image that is excluded in the alignment. There are three entries of # DarkFrame rejected in this example. The first value is the zero-based image index in the raw tilt series, i.e. the one saved onto disk. The second value is reserved for future use. The third one is its tilt angle.

 # AlphaOffset gives the detected tilt angle offset that is the actual tilt angle of sample at nominal zero degree. By default (**-TiltCor 0**) this angular offset is automatically measured unless users specify **-TiltCor -1**. The measured tilt angle offset is added to the nominal tilt angles only when **-TiltCor 1** is used. Please be cautious with this option when subtomogram average is the next processing. Adding the tilt angle offset computationally rotates the sample in the tomograms and thus changes the z coordinates of particles. In this case, the corrected tilt angles should be used for subtomogram averaging, not the nominal tilt angles.

 # BetaOffset is zero and reserved for future implementation.

The next section is the **global alignment parameters** organized in 5 columns in a 10-column table. **SEC** column shows the zero-based indices of tilt images in the corresponding mrc file. **ROT** column lists the angle of tilt axis in degree relative to the y (vertical) axis of the unaligned tilt images. **TX** and **TY** are the translational shifts in pixel in unaligned image plane. **TILT** lists the tilt angles of images. The values of tilt angle are the sums of nominal tilt angles (microscope readings) and **# AlphaOffset**. Users can retrieve the nominal values by subtracting **# AlphaOffset** from the **TILT** column. Note that there are no entries for rejected dark images.

Local alignment data is presented right below **# Local Alignment**. This section contains a 7-column table as shown in the following example. Column 1 shows zero-based indices of image indices in the corresponding mrc file. Column 2 lists indices of patches in each tilt image. Columns 3 and 4 present the x and y coordinates of patch centers, respectively. Columns 5 and 6 show the measured translations in pixel in x and y directions, respectively. Note that the x and y coordinates are relative the image center. Column 7 indicates whether the measurement is reliable or not with 1 denoting reliable. The unreliable measurements are excluded in the local correction of sample motions. Importantly, local alignment data presented here represents the residual alignment errors after the global alignment. Therefore, the coordinates and translations are relative to the global translation aligned tilt images without rotation.

# SEC	ROT	GMAG	TX	TY	SMEAN	SFIT	SCALE	BASE	TILT
0	84.0033	1.00000	-26.113	-137.927	1.00	1.00	1.00	0.00	-45.02
1	84.0033	1.00000	-20.276	-79.669	1.00	1.00	1.00	0.00	-42.01
2	84.0033	1.00000	-11.406	-112.696	1.00	1.00	1.00	0.00	-39.02
3	84.0033	1.00000	-20.730	-79.387	1.00	1.00	1.00	0.00	-36.01
4	84.0033	1.00000	-16.668	-114.443	1.00	1.00	1.00	0.00	-33.01
5	84.0033	1.00000	-14.818	-68.274	1.00	1.00	1.00	0.00	-30.02
6	84.0033	1.00000	-22.440	-92.762	1.00	1.00	1.00	0.00	-27.01
7	84.0033	1.00000	-8.281	-52.114	1.00	1.00	1.00	0.00	-24.02
8	84.0033	1.00000	-23.853	-74.965	1.00	1.00	1.00	0.00	-21.01
9	84.0033	1.00000	9.043	-36.378	1.00	1.00	1.00	0.00	-18.01
10	84.0033	1.00000	-31.795	-65.251	1.00	1.00	1.00	0.00	-15.01
11	84.0033	1.00000	2.055	-22.090	1.00	1.00	1.00	0.00	-12.02
12	84.0033	1.00000	-16.796	-49.033	1.00	1.00	1.00	0.00	-9.02
13	84.0033	1.00000	-3.999	-21.075	1.00	1.00	1.00	0.00	-6.02
14	84.0033	1.00000	-18.390	-93.674	1.00	1.00	1.00	0.00	-3.02
15	84.0033	1.00000	-5.265	-36.435	1.00	1.00	1.00	0.00	-0.03
16	84.0033	1.00000	-0.940	46.147	1.00	1.00	1.00	0.00	2.98
17	84.0033	1.00000	5.306	-17.315	1.00	1.00	1.00	0.00	5.98
18	84.0033	1.00000	0.000	0.000	1.00	1.00	1.00	0.00	8.98
19	84.0033	1.00000	3.946	-11.816	1.00	1.00	1.00	0.00	11.98
20	84.0033	1.00000	-6.927	11.334	1.00	1.00	1.00	0.00	14.98
21	84.0033	1.00000	-0.986	-6.024	1.00	1.00	1.00	0.00	17.98
22	84.0033	1.00000	-12.766	7.682	1.00	1.00	1.00	0.00	20.98
23	84.0033	1.00000	-1.761	4.175	1.00	1.00	1.00	0.00	23.98
24	84.0033	1.00000	-9.999	19.771	1.00	1.00	1.00	0.00	26.98
25	84.0033	1.00000	-6.673	7.913	1.00	1.00	1.00	0.00	29.98
26	84.0033	1.00000	-10.817	10.732	1.00	1.00	1.00	0.00	32.98
27	84.0033	1.00000	-6.994	41.365	1.00	1.00	1.00	0.00	35.98

Fig. 6 An example of the global alignment parameters in an aln file. AreTomo3 uses only column SEC, ROT, TX, TY, and TILT. There are no entries for dark images since they are excluded from alignment process. TX and TY are the translational shift in pixel with respect to tilt images, not movies since tilt images are likely binned after motion correction.

# Local Alignment						
0	0	-1272.49	-1050.00	-3.27	-21.57	1.0
0	1	-455.00	-1020.88	-4.86	-20.73	1.0
0	2	570.77	-1005.25	-2.59	-2.58	1.0
0	3	1598.27	-974.22	682.84	-2083.76	0.0
0	4	-1555.11	-490.34	7.90	28.16	1.0
0	5	-503.87	-437.69	-3.67	-9.53	1.0
0	6	519.96	-311.99	1790.80	2056.45	0.0
0	7	1545.34	-331.74	8.26	3.15	1.0
0	8	-1597.80	52.57	0.58	1218.57	0.0
0	9	-541.09	377.55	-5.85	-40.03	1.0
0	10	498.38	579.78	2.25	20.35	1.0
0	11	1529.58	516.46	7.23	41.23	1.0
0	12	-1510.65	1345.49	3.71	15.67	1.0
0	13	-584.90	1325.02	2.51	14.29	1.0
0	14	471.12	1292.97	1.65	20.05	1.0
0	15	1524.38	1268.09	-0.90	16.34	1.0
1	0	-1279.45	-1112.42	-3.43	-26.24	1.0
1	1	-461.72	-1081.16	-3.97	-23.64	1.0
1	2	564.30	-1063.30	-2.09	-9.49	1.0
1	3	1591.97	-1030.66	586.73	-2132.66	0.0
1	4	-1556.77	-505.24	4.27	14.12	1.0

Fig. 7 An excerpt of the local alignment parameters in an aln file. This section appears only when users enable local alignment using -AtPatch. Columns 1 and 2 are the image and patch indices, respectively. Columns 3 and 4 are x and y coordinates of patch centers, respectively. Columns 5 and 6 are the translational shifts in x and y directions, respectively. Column 7 shows whether the measured shifts are reliable or not. The reliable measurements are labeled with 1.

6. Extra Command Line Options

6.1 Motion correction

-Group is a very useful option for measuring beam induced motion on challenging movies of low SNRs. This option decides how to divide movie frames or the rendered frames into non-overlapping groups. Beam induced motion is measured on the group sums and then interpolated/extrapolated to individual frames. This option takes two integers that are the group sizes of which the first one is for global measurement and the second for local measurement. The default values are 1 and 4. Users can selected different numbers but make sure the number of frames divided by the group size be larger than or equal to 3. The second number should be larger since individual patches have less SNRs than the full image and thus need more frames to increase SNRs.

-DefectFile is the same as that in MotionCor2/3. If there are fixed-pattern defects in micrographs, it is a good idea to compile a defect file and pass it to AreTomo3.

-Gain canload both mrc and gain files.

-EerSampling is the same as that in MotionCor2/3. It might be beneficial to combine -EerSampling 2 and -McBin 2 instread of -EerSampling 1 and -McBin 1 since working on the super-resolution frames may restore more information at lower tilts and in particular Tygress data collection.

-McPatch is the same as -Patch in MotionCor2/3. It enables local measurement and correction of beam induced motion.

6.2 Tomographic alignment

-AlignZ is the same as that in AreTomo/AreTomo2. The default value works well on thin and medium thick samples. Its value needs to be increased to be approximately the sample thickness in pixel.

-VolZ is the same as that in AreTomo/AreTomo2 and in pixel. It is recommended to reconstruct a slightly larger volume in z to cover the entire sample without clipping. This value is relative to the input tilt series. When -AtBin is used, the actual value z dimension is equal to the input value divided by the binning following -AtBin.

-AtPatch followed by two positive integers enables local tomographic alignment.

-AtBin enables floating-number binning by Fourier cropping to generate binned tomograms. Since version 1.0.18, users can supply a second parameter. It is optional. When it is provided, AreTomo3 will reconstruct another tomogram using weighted backprojection. The corresponding file name will be embedded with _2ND_Vol.

-FlipVol is the same as that in AreTomo/AreTomo2. -FlipVol 1 rotates the volume around its x axis. The reconstructed volume is saved in a mrc file as a stack of xz slices. When this option is enabled, the mrc file contains a stack of xy slices.

-OutImod is the same as that in AreTomo/AreTomo2.

-DarkTol is the same as that in AreTomo/AreTomo2. However, the implementation for detecting the dark images is different that considers the Tygress tilt series where the zero tilt image has more dose than the others.

-TiltCor is the same as that in AreTomo/AreTomo2. It is not recommended to enable this option (-TiltCor 1) when subtomogram average is part of your workflow since this option rotates the volume around its tilt axis. The z coordinates of particles picked from the rotated tomogram would be different from the coordinates in sample.

-OutImod is the same as that in AreTomo/AreTomo2.

-Wbp is the same as that in AreTomo/AreTomo2

-Sart is the same as that in AreTomo/AreTomo2.

6.3 CTF estimation and deconvolution

-PixSize is needed to start CTF estimation. It is the pixel size of movies, not tilt series. AreTomo3 calculates the pixel size of tilt series based on this input value and -McBin value specified at command line.

-Kv is default to 300 kV. Users should change it if the data collection is done at different voltage.

-Cs is default to 2.7 mm. Users can choose a different value using this option.

-AmpContrast is default to 0.07.

-ExtPhase should be used when data collection is done with a phase plate. The option takes two values, one is the estimated phase shift and the second is the search range. They are both in degree.

-CorrCTF 1 enables local CTF deconvolution of each tilt image. A tilt image is first divided into tiles. Each tile has its own CTF based on its location from the tilt axis. CTF deconvolution is done on each tile. Then CTF deconvolved tiles are put together to form the CTF deconvolved image.

7. Extra Tools

Since version 1.0.18, AreTomo3 provides a set of tools to help process their data. These tools are placed in tools subdirectory.

7.1 remap3D

remap3D is a python package that maps 3D targets picked in one set of tomograms to another set of tomograms that are reconstructed from the same set of tilt series but with different alignment parameters. In practice, we sometimes need to re-align the tilt series and reconstruct the tomograms. However we certainly do not want to repick the subtomograms. **remap3D** is designed to automatically map the 3D coordinates of picked particles to the newly reconstructed tomograms.

For each pair of tomograms reconstructed from the same tilt series, **remap3D** calculates the 3D coordinates based on forward- and back-projections defined in the corresponding pair of .aln files generated by **AreTomo3**.

remap3D requires an input star file of the 3D targets picked in one set of tomograms and the corresponding .aln files. The star file and the .aln files must be placed in the same directory. The output directory should contain the .aln files that are used to reconstruct the tomograms to which the 3D targets are mapped.

remap3D requires pandas, numpy, starfile. The following is an example of command line.

```
python ~/PyProjs/AreTomo3/Remap3D/remap3D.py \
       -ovs 4096 4096 1200 \
       -nvs 4096 4096 1200 \
       -ops 1.54 \
       -nps 1.54 \
       -os 24mar08a/stars/20240308_002_ribosome.star \
       -ns Temp/ribo_new.star   \
       -oa 24mar08a/run002/alns \
       -na 20240308_002_Krios1_RP_Lys6prtns/run006/alns/ \
       -oap Position_
```

-ovs or **--old_vol_size** should be followed by the *xyz* sizes of the original tomograms where the targets are picked. In practice, the tomograms are reconstructed at lower resolution. The coordinates of picked particles in the star file are often scaled back to the unbinned resolution. If this is true, we should use the unbinned volume size.

-nvs or **--new_vol_size** should be followed by the *xyz* sizes of the new tomograms to which the 3D targets are mapped. If **–old_vol_size** is unbinned, we should also use the unbinned size for the new tomograms here. Note that the unbinned *xy* sizes are most likely the same for the original and the new tomograms. However, their z-dimensions may have different sizes.

-ops or **--old_pix_size** is the pixel size in angstrom of the original tomograms where the particles are picked. If the coordinates of picked particles are scaled back to the unbinned resolution, we should use the unbinned pixel size.

-nps or **--new_pix_size** is the pixel size in angstrom of the new tomograms to which the particles are mapped. If the coordinates of picked particles are scaled back to the unbinned resolution, we should use the unbinned pixel size, which is the same as **-ps1**.

-os or **--old_star** is the old particle star file that contains the 3D coordinates of picked particles we need to map to the new tomograms.

-ns or **--new_star** is the new star file of mapped 3D particle coordinates.

-oa or **--old_aln** is the path to the directory that contains the .aln files used to reconstruct old tomograms.

-na or **--new_aln** is the path to the directory that contains the .aln files used to reconstruct new tomograms.

-oap or **--old_aln_prefix** is the prefix of the aln file names. Since the tomogram names listed in the old star file may have a different prefix from the corresponding aln files. In the above example, the tomograms

were named TS_xx_xx.mrc whereas the associated aln files were named Position_xx_xx.mrc. remap3D uses this user input to find the correct aln file for each tomogram.

Reference

Zheng S, Brilot A, Cheng Y, Agard D. Beam-Induced Motion Mechanism and Correction for Improved Cryo-Electron Microscopy and Cryo-Electron Tomography. Chapter 10, Cryo-Electron Tomography Structural Biology In Situ, Springer 2024.

Zheng S, Wolff G, Greenan G, Chen Z, Faas FGA, Bárcena M, Koster AJ, Cheng Y, Agard DA. AreTomo: An integrated software package for automated marker-free, motion-corrected cryo-electron tomographic alignment and reconstruction. J Struct Biol X. 2022 May 10;6:100068. doi: 10.1016/j.yjsbx.2022.100068. PMID: 35601683; PMCID: PMC9117686.

Technical Support

Shawn Zheng

Chan Zuckerberg Imaging Institute

shawn.zheng@czii.org