

TOMOMAN Manual

William Wan
Max Planck Institute for Biochemistry
Department of Molecular Structural Biology
82152 Martinsried, Germany

Last update: 23-04-2018

General

Tomoman is a set of MATLAB wrapper scripts for pre-processing tomography data. The general aim of the pipeline is to start from unsorted raw data and arrive at a point where the stacks are ready for tomogram alignment. Metadata is accumulated during processing, and is stored in a matlab struct array.

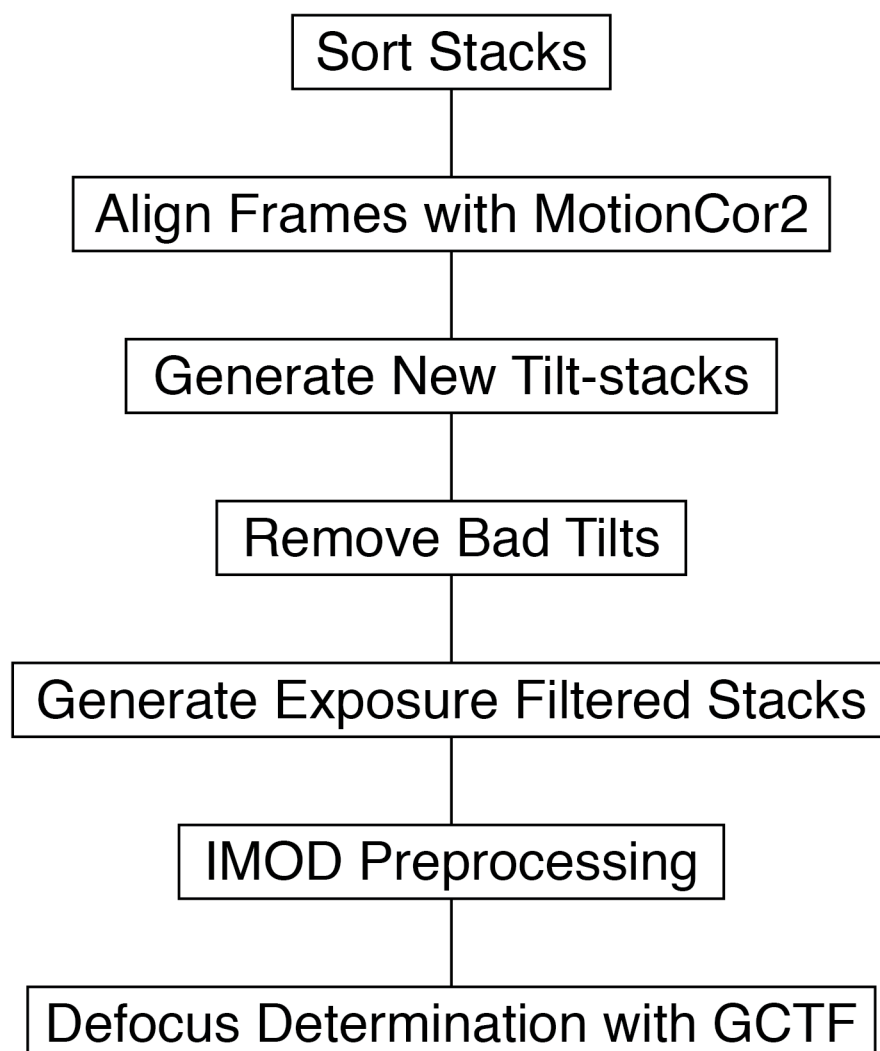


Fig 1. Flowchart of preprocessing workflow.

Tomoman can be run during data collection, and re-run as more data comes in. Metadata, including progress through the various pre-processing steps is kept in a MATLAB struct array called the “tomolist”.

Tomoman is run using either the `tomoman_run.m` or `tomoman_run_full.m`, with the former providing a smaller list most used parameters, and the latter providing all parameters. Each step is broken into a small block of options in the input area; each processing step can be enabled or disabled using the first option. If the tomolist shows that a processing step has already been performed, it is skipped for that processing step, but will be processing for steps that have not been performed. Enabling the “force” parameters will force reprocessing.

Sort stacks

Raw data is expected in to be placed in two folder: one containing the raw tilt-stacks and mdoc files generated by SeralEM, and a folder containing frames. Tilt-stacks and their associated mdoc files should consists of an optional prefix and a tomogram number, and an extension of either ‘.st’ or ‘.mrc’ ([prefix][tomo_num].[ext] and [prefix]][tomo_num].[ext].mdoc). The .mdoc file is key to stack sorting; the presence of a .mdoc in the raw data folder is what causes tomoman to recognize a new stack and generate a new tomolist entry.

Sorting will generate a tomogram folder named [prefix][tomo_num] for each tilt-stack, move the stacks and .mdocs into each folder. It will parse the frame names from the .mdoc file and place them into a ‘frames’ subfolder in each tomogram folder.

Align frames / new stacks

The first step is to align the frame images and place them into a new stack. Frame alignment is performed with MotionCor2 and the new stacks are generated in MATLAB.

In general, the number of frames and amount of dose/frame is relatively low, so all frames are used for alignment. I would generally not recommend performing dose filtering in MotionCor2, as the approach used is not well suited for tomography. Super resolution data should be binned by 2. Frame corrected stacks can also be written; these are particularly useful for performing per-frame dose-filtering at a later step.

Cleaning stacks

For efficient stack cleaning, tomoman opens each stack using IMOD’s 3dmod viewer. The numbers of bad tilts, as numbered in 3dmod, can be input into the MATLAB command window; this triggers the generation of a new stack (suing IMOD’s newstack command) with the bad tilts omitted, and updates the tomolist metadata. Supplying no bad tilts indicates there are no bad tilts to be removed. For completely bad tomograms, writing “skip” tells tomoman stop processing this tilt stack. NOTE: tomoman does not the control the 3dmod windows it opens, so remember to close them as you work.

Dose Filtering

Exposure filtering is performed by applying the exposure-dependent attenuation functions described in Grant and Grigorieff (doi: 10.7554/eLife.06980.001). Exposure filtering can be applied to images in the tilt stacks, which is essentially a type of dose-dependent low-pass filtering, or on aligned frame stacks generated by MotionCor2.

For frame stacks, tomoman's implementation is different from Unblur, where frequencies are reweighted after frame-filtering to produce a roughly flat frequency spectrum. This method is valid for 2D methods, as those frame-stacks contain information in all frequency ranges; the amplitude reweighting after filtering essentially produces a weighted sum at each frequency shell. However, in tomography the data is distributed across different views. The consequence of this is high-resolution information only exists in the first few collected frame-stacks, but not in the later ones. Therefore, reweighting later frame-stacks using the Grant and Grigorieff scheme re-amplifies noise, minimizing the tomogram contrast-enhancing properties of an exposure filtered tilt-stack. Instead, tomoman reweights the frequencies of each frame-stack so that the amplitudes are bounded by the exposure filter of the first frame.

Exposure filtering frame stacks requires more temporary storage space and more processing time than filtering tilt-stack images. However, filtering frame-stacks should preserve more high-resolution information in the earlier tilts; later tilts show little difference between the two approaches.

IMOD Preprocessing

A number of IMOD preprocessing steps can be automated using default parameters. Tomoman can perform .com script generation, X-ray/Outlier pixel correction, generate coarse aligned stacks, and run automated fiducial tracking by generating directive files and running IMOD.

In general, I find that automated fiducial tracking works poorly, so I only preprocess up to generating the coarse aligned stacks.

Defocus Determination

Defocus determination is performed using GCTF. Tomoman runs GCTF, concatenates the outputs and converts the measured defocii and astigmatisms to a file that can be read by IMOD's ctfphaseflip and NovaCTF.

In general, the signal is fairly low in tilt-stack images. I find that GCTF works better using smaller box sizes (i.e. 512 pixels), and using tilt-images rather than frames.