

# subTOM Tutorial

Dustin Morado

June 2019

## 1 Introduction

subTOM is a subtomogram alignment and averaging extension to the TOM toolbox, similar to the AV3 package, but in an easier-to-use pipeline.

The software aims to be a complete processing environment from the point after tilt-series have been collected, excluding manual tilt-series alignment which is already well serviced by IMOD's eTomo GUI.

subTOM is very similar to other subtomogram packages namely PEET, Pytom, I3, and Dynamo, and can be used to determine structures to resolutions near current method limits. However, it does not include any GUI interfaces, but attempts to still be approachable to beginner users through the means of ample documentation. It also makes some improvements over the AV3 package on which it is based, most noticeably in the ability to do almost all processing in a MATLAB independent manner through the use of compiled MATLAB code that runs on by way of the MATLAB Compiler Runtime (MCR).

I hope you find the software useful, and am always open to hear how the software could better serve users.

## 2 Installation

Installation of subTOM should be relatively straight forward, but the package does have several dependencies:

- IMOD (Version  $\geq 4.10.29$ )
- NovaCTF (Version  $\geq 25$ th commit)
- MATLAB Compiler Runtime (Version 2016b)
- TOM Toolbox [optional]
- CTFFIND4 [optional] (Version  $\geq 4.1.13$ )
- GCTF [optional] (Version  $\geq 1.06$ )

IMOD, CTFFIND4, GCTF, and the MCR can all be found online at their respective websites, while NovaCTF and the TOM Toolbox can be found on GitHub.

Once the dependencies you plan to use have been installed, you can get subTOM from my teraraid folder. Do this after changing to the directory in which you want to install subTOM, in this example we are doing it in the software directory in my home folder. From there we change into the newly created subTOM directory and run the installation script passing it the current installation directory and the location of the MCR libraries. At the MRC-LMB we have a copy of the 2016b MCR at the location shown below.

```
$ cd ~/software
$ git clone /net/dstore2/teraraid/dmorado/software/subTOM
$ cd subTOM
$ ./install.sh --help
  USAGE: install.sh <INSTALL_DIR> <MCR_DIR>
$ ./install.sh ${PWD} /public/matlab/jbriggs
```

This should create a new scripts directory in the subTOM directory that holds BASH scripts that you edit to perform the processing. The scripts are simply text files with only options, which in turn call the actual BASH scripts that run the pipeline itself.

I update subTOM regularly as a git repository, which makes updating very easy, and it is not a bad idea to update the software often:

```
$ cd ~/software/subTOM
$ git pull origin master
$ ./install.sh ${PWD} /public/matlab/jbriggs
```

### 3 Tutorial Setup

To walkthrough the basics of using subTOM we will be using a subset of EMPIAR-10164. This is the same subset that was used in the publication describing 3D-CTF correction of tomograms using NovaCTF, and went to a final resolution of 4 Angstroms.

Download the data (frames and mdocs) from EMPIAR-10164 for tilt-series TS\_01, TS\_03, TS\_43, TS\_45, and TS\_54. SubTOM works best when data is organized in a particular hierarchy, but the code does not enforce this hierarchy itself as in RELION or Scipion (although perhaps it should). After unzipping the data I have it organized in a folder for the tutorial:

```

$ cd /teraraid/dmorado/subTOM_tutorial/data
$ tree
.
|--- Frames
|   |--- TS_01_000_0.0.mrc
|   |--- TS_01_001_3.0.mrc
|   |--- ...
|--- TS_01.st.mdoc
|--- TS_03.st.mdoc
|--- ...

```

With your own data you may have more (hopefully), or less (hopefully not) files than this to start from directly from the microscope. At the minimum subTOM needs the tilt-image movies with the naming format:

```
<basename>_<image_idx>_<angle>.{mrc,tif}
```

Where `basename` is ideally "TS-", `image_idx` is a three digit number starting from 000 or 001, and `angle` is obvious. To save space many facilities often collect compressed unnormalized TIFF movies, and if this is the case you will also need the gain-reference file and the defects list that is output in this case by SerialEM.

However, in the more likely case you will also have a raw-summed tilt-series in the same data folder with the MDOC files and also perhaps a log file as well. The more files the better as much of the extended header information in the raw-summed tilt-series and moreover the information in the MDOC file makes later processing much smoother.

We now have our tutorial directory setup and can begin the preprocessing stage.