EMBO Practical Course on Image Processing for Cryo EM
1 –11 September 2015

Introduction and general info on how to use Imagic

Introduction

The information we wish to extract from EM images, the signal, is the projected density of the structure of interest. The recorded images contain, in addition to the signal, fluctuations in intensity caused by noise from many different sources. Sources of noise include background variations in ice or stain, damage to the molecule from preparation procedures or radiation, and detector noise. Many views of the particle are recorded in different orientations, but each individual image has a low SNR (signal to noise ratio). SNR = P_{signal}/ P_{noise} , where P is the integral of the power spectrum).

The main task in extracting the 3D structural information is to determine the relative positions and orientations of these particle images so that they can be precisely superimposed. Alignment is done by finding shifts and rotations that bring each image into the same orientation as a reference image. Assuming that the noise is not correlated to the structure, it can be suppressed by averaging images of the particles, thereby enhancing the structural information. For averaging, it is essential that the particle images are brought into orientation with the reference so that similar features superimpose in their average. In general, the alignment is an iterative process beginning with coarse features of the data set, for example, center of mass of each particle image, followed by grouping and averaging of individual images. Cross correlation is the main tool for measuring similarity of images, but it is not very reliable at low SNR. In practice, alignments are iterated so that successive averages contain finer details, which in turn improve the reference image, for subsequent rounds of refinement (See review Orlova EV, Saibil HR, 2011; and van Heel et al., 2000)

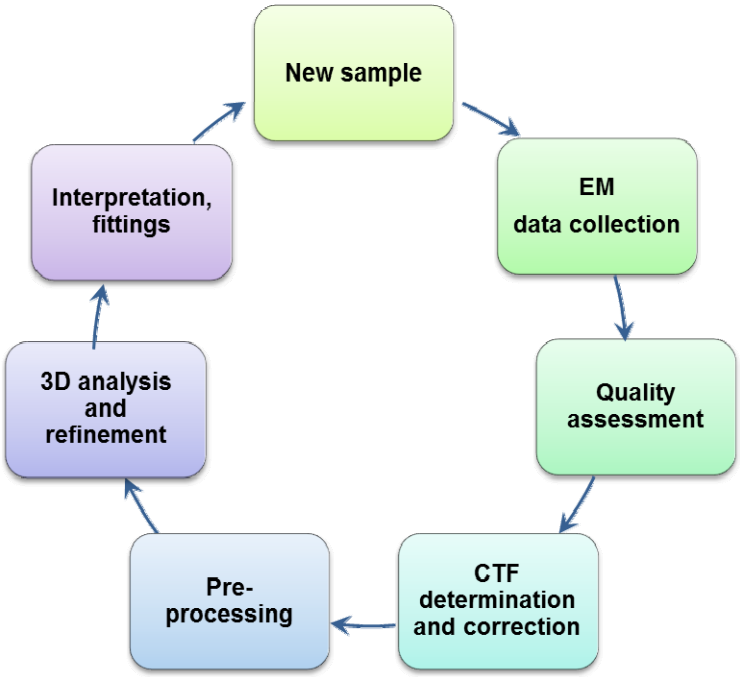
Averaging improves the SNR by a factor of $\sqrt{N}$, where N is the number of averaged images. This in turn facilitates the determination of relative orientations of the different group averages ("characteristic views"). Analysis of images followed by classification into different groups (clusters) according to their features is the basis of the statistical approach (see review van Heel et al., 2000).

To calculate the 3D map from a set of projection views, the relative orientations of the 2D projections must be determined. There are two general approaches to this problem. An experimentally based approach involves the collection of images of the same particles at different tilt angles. This method is particularly applicable for particles that adopt a preferred orientation on the support grid. The other approach is computationally based, in which untilted images are collected. For the second approach, it is essential to collect a range of views distributed over different orientations. The biggest challenge in orientation determination is to get the first set of assignments for a data set corresponding to an unknown 3D structure, especially if it is asymmetric. Once an initial model (starting model) is available, the orientations can be refined using a variety of approaches.

Angles (orientations) can be defined using common lines (Orlova & Saibil, Van Heel et al., 2000). It is based on the theorem: for any set of 2D projections of a given 3D structure each pair of 2D projections has at least one 1D (line) projection in common. The common line between two projections in Fourier space is the line of intersection of the corresponding two planes in Fourier space. Determination of common lines from individual raw images is difficult, but the presence of symmetry provides many more constraints and results in multiple common lines, both from the same image (self-common lines) and between image pairs (cross common lines). A common lines approach in real space for arbitrary symmetry was developed by van Heel and colleagues and implemented in IMAGIC (Van Heel et al., 1996). Once a consistent initial 3D map has been

obtained, the structure can be refined by further cycles of alignment, classification, and common line searching.

The procedure of projection matching is much easier to understand in principle, but it needs an initial model. Once a 3D structure is available, even at very low resolution, it can be used to generate reprojections at all possible orientations. In projection matching, for each image in turn, the Euler angles of the reference image that gives the best cross correlation are assigned to the raw image or class average. For each comparison, all possible in-plane alignments must be tested, so that this is a very lengthy calculation. Once the Euler angles are assigned, a new 3D map can be calculated and the procedure iterated with the new set of reprojections. A diagram showing the steps in image processing can be seen in the figure on the left.



Functioning macromolecular complexes often coexist in multiple states in solution and these are captured in the vitrified sample. Structural heterogeneity can be caused by dynamic structural states,the presence of reaction intermediates, variable ligand binding, multiple oligomeric states, or changes in environmental conditions such as temperature and solutes. Heterogeneity imposes significant limitations on the achievable resolution, since information from different molecular states in a heterogeneous ensemble will be combined into one reconstruction.

There are three main approaches currently used to identify and sort molecular heterogeneity (Orlova, Saibil, 2010). All of them are based on statistical methods. The first approach is based on statistical analysis of raw images (in 2D)—*a priori* analysis to detect the heterogeneity of the sample in its images. Here, the initial sorting is done on 2D images only, prior to any three-dimensional (3D) reconstruction. In the second approach, an initial 3D map is required in order to separate the images into subsets of images presenting a molecular complex in similar orientations. Analysis of structural heterogeneity is then done in 2D for each orientation subset. It has been applied to the reconstruction of heterogeneous ribosome complexes (Fu et al., 2007; Klaholz et al., 2004). The third approach is based on *a posteriori* analysis of 3D reconstructions by considering a population (as many as possible) of 3D reconstructions in order to examine variations in 3D maps (Ohi et al., 2004; Penczek et al., 2006a,b; Scheres et al., 2007). It should be noted that all these methods in their present form require involvement of the researcher in the separation procedure, to decide which parameters must be taken into account as significant during the analysis.

During these two practicals you will familiarise yourself with the alignment procedure, the principal component statistical analysis implemented in IMAGIC (Van Heel et al., 1996), how to separate images according to their features (analysis of heterogeneity), how to use angular reconstitution based on the “anchor” set and how to calculate three dimensional reconstructions.

These materials provide general information on:

1. Main commands in IMAGIC
2. Major steps in pre-processing images in analysis: normalisation and filtering of images.
3. Alignment of images.

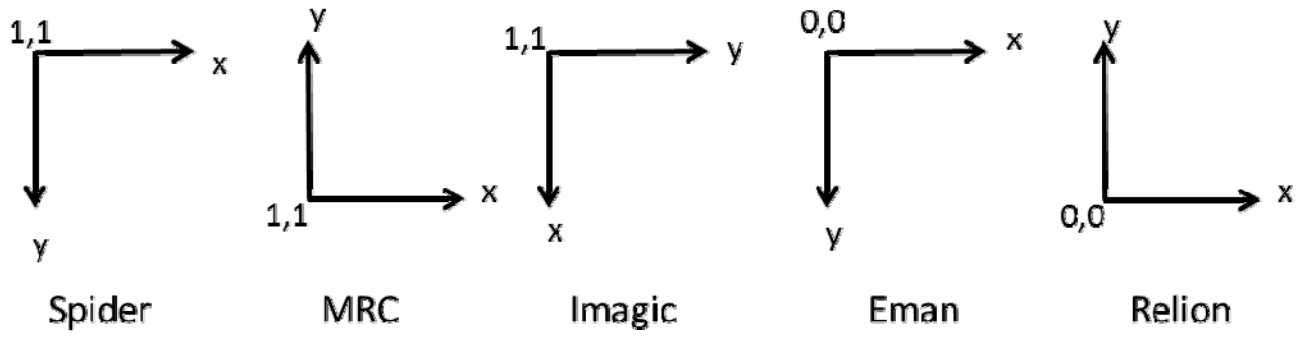
Further reading:

1. Orlova EV, Saibil HR. Structural analysis of macromolecular assemblies by electron microscopy. Chem Rev. 2011 111(12):7710-48 (Free article)
2. van Heel M, Gowen B, Matadeen R, Orlova EV, Finn R, Pape T, Cohen D, Stark H, Schmidt R, Schatz M, Patwardhan A. Single-particle electron cryo-microscopy: towards atomic resolution. Q Rev Biophys. 2000 Nov;33(4):307-69.
3. van Heel M, Harauz G, Orlova EV, Schmidt R, Schatz M. A new generation of the IMAGIC image processing system. J Struct Biol. 1996 Jan-Feb;116(1):17-24
4. Orlova EV, Saibil HR. (2010) Methods for Three-Dimensional Reconstruction of Heterogeneous Assemblies, Methods in Enzymology, ed G. Jensen,Volume 482, 321-341
5. Elad, N., Clare, D., Saibil, H. R., and Orlova, E. V. (2008). Detection and separation of heterogeneity in molecular complexes by statistical analysis of their two-dimensional projections. J. Struct. Biol. 162, 108–120.
6. Fu, J., Gao, H., and Frank, J. (2007). Unsupervised classification of single particles by cluster tracking in multi-dimensional space. J. Struct. Biol. 157, 226–239.
7. Klaholz, B. P., Myasnikov, A. G., and Van Heel, M. (2004). Visualization of release factor 3 on the ribosome during termination of protein synthesis. Nature 427, 862–865.
8. Ohi, M., Li, Y., Cheng, Y., andWalz, T. (2004). Negative staining and image classification - Powerful tools in modern electron microscopy. Biol. Proc. Online 6, 23–34.
9. Penczek, P. A., Frank, J., and Spahn, C. M. T. (2006a). A method of focused classification, based on the bootstrap 3D variance analysis, and its application to EF-G-dependent translocation. J. Struct. Biol. 154, 184–194.
10. Scheres, S. H., Gao, H., Valle, M., Herman, G. T., Eggermont, P. P., Frank, J., and Carazo, J. M. (2007). Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization. Nat. Methods 4, 27–29.

Introduction to IMAGIC

Please do not forget about the system of coordinates for IMAGIC on the screen.

Image coordinates



We will be using the latest version of IMAGIC.

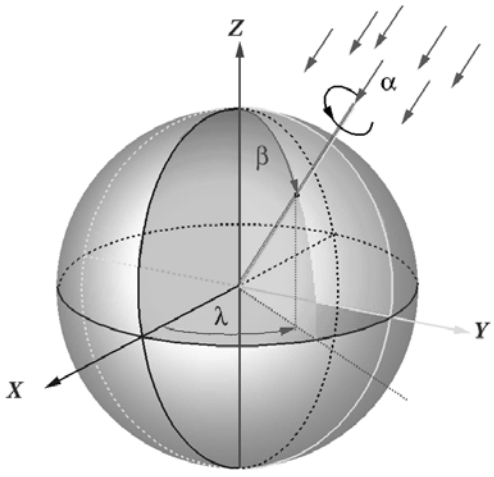
To start the program, type **imagic**.
A list of some commands that will be used in the practicals:

MENU	COPY-IMAGE	MASK-IMAGE
DISPLAY	SUM-IM	MASK-AUTOMATICALLY
SURVEY	ALIGN-MAGNIFICATION	EULER
SUBFILE	ALI-PAR	CONVERT-IMAGIC-TO-FREALIGN
EXC-COPY	ALI-DIR	TRUE-THREED-RECONSTRUCTION
COPY-IMAGE	MULTI-REFERENCE-ALIGNMENT	THREED-FORWARD-PROJ
RENAME-IMAGE-FILE	PROJECTION-MATCHING	THREED-FOURIER-TRANSF
NORM-VARIANCE	EXTRACT-ALIGNED-IMAGES	THREED-FILTER
PRETREAT-IMAGES	EQUIVALENT-ROTATION	THREED-ZOOM
APPEND-IMAGES	MSA-RUN	THREED-SUBFILE
ARITHMETIC-WITH-IMAGE	MSA-SUM	THREED-SUM-VOL
SHIFT-IMAGE	MSA-EXT	THREED-AUTOMATIC-MASK
MASK-IMAGE	EXTRACT-IMAGES	THREED-BINARIZE-MASK
MASS-CENTER	IMPOSE-SYMMETRY	THREED-IMPOSE-SYMMET
CTFFIND3-RUN	FOURIER-TRANSFORM-IMAGE	THREED-SURFACE-V
AVERAGE-NFOLD-ROTATIONAL	FOURIER-OPERATIONS	FOURIER-SHELL-COR
ROTATE-RANDOMLY	BAND-PASS-FILTER	MOVIE
TEST-ROT-SYMM	CONVERT-FORMAT-IMAGE	EM2EM

To see a listing of all commands, type **menu**.
To get help when using the program or within a particular command type 'help' or 'help' followed by the command name, e.g.: 'help EULER '↵

Angles (orientation) in Imagic.

Conventions in IMAGIC:
$\lambda = \gamma$ – rotation around the Z-axis of the object
β – rotation around the X-axis of the object
α – rotation in the plane of the image (around the new Z-axis)



Euler Angles

If you need some explanation to what should be the answer for the parameter, type "**?**"-> help information will be printed,
More detailed information can be found on the site:
http://www.imagescience.de/manuals/

Some versions of Imagic do not accept file names with dots in the middle of the name, therefore try to avoid it. You can have numbers in file names. Commands can be typed with either small or capital letters, but file names are case sensitive

Command **EM2EM** is used to convert images to and from various data formats that are used within the "3-D Electron Microcopy Community".
If you want to convert between different IMAGIC formats (REAL, INTG, PACK etc.) please use command CONVERT-IMAGE. Same command as **IMPORT-EXPORT-IMAGIC-FORMAT**

```
IMAGIC-COMMAND : EM2EM

Convert 2D image(s) or a single 3D volume:  2D  or 3D

Data format of the input to be converted:
  BROOKHAVEN_STEM       CCP4
  DATA_ONLY             DIGITAL_MICROGRAPH_2
  EM                    FEI_RAW_IMAGE
  FABOSA                FORMATTED
  IMAGIC                JPEG
  KONTRON               MDPP
  MEDIPIX               MRC
  OFFSET                PIF
  PGM                   RAW_IMAGE
  RAWIV                 RGB_TIFF
  SHF                   SPIDER
  SUPRIM                TIA/EMI
  TIFF                  TVIPS
  VOLUMETRIC
Export to which data format:
  CCP4                  DATA_ONLY
  DIGITAL_MICROGRAPH_2  EM
  FORMATTED             FEI_RAW_IMAGE
  IMAGIC                JPEG_GREYSCALE
  KONTRON               MDPP
  MRC                   OFFSET
  PIF                   PGM
  POSTSCRIPT            RAW_IMAGE
  RAWIV                 RGB_TIFF
  SHF                   SPIDER
  SUPRIM                TIFF_GREYSCALE
  TVIPS
```

IMAGIC works with two files at the same time: one file contains information about the image and the other contains the image itself. The file with the information has the extension '**hed',** the second has the extension **'img'. EM2EM** will create both files automatically from any other format during converting into Imagic format**.**

Viewing images

To display an image in IMAGIC we have to convert the image into **IMAGIC** format. The command **EM2EM** (see above) converts other formats into IMAGIC format so that they can be processed.
To display the image use the command **'DISP'.** IMAGIC provides information from the header about the data file.

```
IMAGIC-COMMAND: disp

Input image file                          :Image_test

Size of the display window     800,800    -> the window to display images, its
                                             size is given in pixels. It can
                                             be altered later. The first
```

number is the vertical size, the second it the horizontal size

```
Type of cursor:
  CROSS  SQUARE  CIRCLE
Please specify option
```
CROSS, *usually this is used as default.*

```
Then you will see the table

      Current DISPLAY settings:
Input image FILE name        : Image_test
LOCATION numbers             :        1       to:          1
Output DEVICE                : X_WINDOWS
DEVICE window size           :      800        ,         800
SCALE factor                 :        1.0000
MINX, MAXX                   :        1       to:        512
MINY, MAXY                   :        1       to:        512
GREYVALUES                   : Local
ERASE screen                 : No
STARTING point    (top left) :        1        ,          1
Display of NAME and location : Name and location
Video lookup table (VLT)     : Linear Black/White
---------------------------------------------------------------------
Image name, parameter & story:
TIFF input image = /d/ipcourse/prac5/image_test.tif
SIZE:  512,  512  LOC:    1  TYPE: PACK

Parameters to be changed:
NO_CHANGES(=DISPLAY), SETTINGS, OPTIONS [NO]       :

You can use some other options

NO_CHANGES(=DISPLAY), SETTINGS, OPTIONS [NO]       : options

Display options available:
  SCAN        COORD       PROFILE     FILM_MOVIE  WATCHDOG
  SELECT_LOC  PLOT        CLSPLOT     ZOOM        WRITE
  SAVE        SAVE        WAIT        BACK

Please specify option [scale]                       :
```

If you type **scale** then you can change the pixel size at which the image will be shown, you can manipulate its size, and grey values. The following parameters can be changed.

Options for the "ordinary" display:

FILE	: Allows you to change the input image file
DEVICE	: Allows you to change the display window (if it is possible on your computer)
LOCATION	: Change the sequence of images to be displayed (location numbers of the input image file)
SCALE	: Scale factor
MINX,MAXX	: Line range to be displayed
MINY,MAXY	: Pixel range to be displayed
GREYVALUES	: Change grey values of the input image (brightness and contrast)
ERASE	: Erase screen first before displaying
START	: Specify starting point of the image on the monitor (with cursor or explicitly by giving the coordinates)
NAME	: Display filename and location numbers
LOCATIONS	: Display location number only
NAME/NOTHING	: No filename and no location number displayed
VLT	: Change the standard video lookup table

(to get another brightness, contrast and/or colors)
NO , or "enter" : No further changes desired ... go ahead and display

NOTE1 : *, EXIT, QUIT or BYE allows you to leave the program and any of the IMAGIC commands. Please do not leave by typing ctrl+C.

How to change greyvalues:

```
NO_CHANGES(=DISPLAY), SETTINGS, OPTIONS [NO]              grey
 Please give option to get the grey values
(SURVEY,HEADER,INTERACTIVE,WINDOW)  [I]                : i
 Give black, white levels (0,0 = automatic) [-0.25,0.25]          : 50,200 ( as
an                                                                  example)
Overwrite old DISPLAY window [YES]                                : yes
```

Some other useful options in **display:**

How to do measurements and contouring of images:

```
NO_CHANGES(=DISPLAY), SETTINGS, OPTIONS [NO]             coos
Should the values be stored in a (plt) file? [YES]       yes
Output coordinate (plt) file (NO ext.) [Image_test.plt]  press "ENTER"
NO_CHANGES(=DISPLAY), SETTINGS, OPTIONS [NO]             press "ENTER"
```

Information written on the screen:

```
Active display option: COOS    : On
 Coordinates STORED in a file : Yes
 Output PLT file                          : Image_test.plt
Parameters to be changed
NO_CHANGES(=DISPLAY), SETTINGS, OPTIONS [NO]           : press "ENTER"
```

Now the program is ready to write the coordinates and measurements. You can pick particles or do contouring using the option **"COOS".** Contouring is used to generate a specific mask around your image.

Coordinates are written in the file **Image_test.plt**

NOTE: You can select as many coordinates as you want. If you want to stop selecting coordinates, please click on the (last) selected point once again. Select point of interest with mouse and press the middle mouse button.

X-coordinate = 23 Y-coordinate = 311

Store these values (YES or NO) or STOP [YES] : "enter" if **yes**

Select point of interest with mouse and press the middle mouse button. Please note, that if you hold down the middle mouse button and drag the mouse away from the selected point, a circle will appear on the image, with its centre located in the selected point.
Then you will see:

X-coordinate = 34 Y-coordinate = 185 Radius = 27

This option is useful to do size measurements on images.
Store these values (YES or NO) or STOP [YES] : **"enter"** *(if yes)*

Type **STOP** to end the process of reading of coordinates.

Option PROFILE – it is useful for assessment of profile of the power spectrum

```
NO_CHANGES(=DISPLAY), SETTINGS, OPTIONS [NO]       :  profile
```

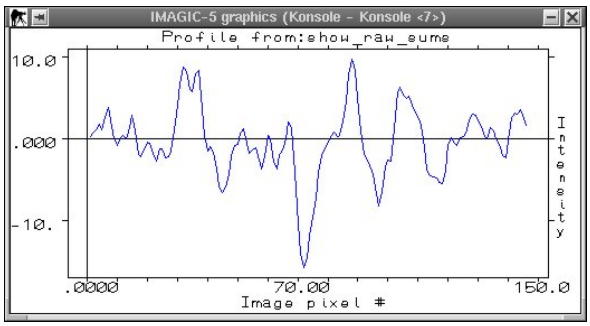
**ATTENTION: Profile can only be used for the LAST image displayed

Use cursor to position profile **YES**

Need starting point for profile: Select a point of interest within the window with image and click with the mouse (press the middle button) where you would like to **start** and then click at the **end** point for profile, where you would like to finish. Press **"Enter",** then you will see a request for the output device.

```
Output device (X_WINDOWS, PS) [X_WINDOWS]          :   press "Enter"
-----------------------------------------------
 Active option: PROFILE       : On
 PROFILE start (image coord.) :     3, 650
 PROFILE end   (image coord.) :     351, 370

The curve will appear in a separate window  →
```



The coordinates of the starting and end points for the profile can be given exactly with the following options:

```
NO_CHANGES(=DISPLAY), SETTINGS, OPTIONS [NO]          profile
 Use cursor to position profile                            no
Starting point (IMAGE coordinates X,Y)                     1,256
End point      (IMAGE coordinates X,Y)                     512,256
 Output device (X_WINDOWS, PS) [X_WINDOWS]                 "ENTER"
```

Now, if you would like to finish your work with DISPLAY, please type **'*', ''bye', or 'exit'**

NO_CHANGES(=DISPLAY), SETTINGS, OPTIONS [NO] **bye**

SURVEY - the command (**SURVEY-DENSITIES**) performs simple statistics of (sets of) images. It is used to determine **min, max, mean**, and **sigma** of densities in images and writes these values into headers.

```
IMAGIC-COMMAND: survey

Mode of survey:
  2D-LOCAL  3D_LOCAL GLOBAL  MAXPOS
Please specify option [GL]                                    ?
        Options available in this program
        LOCAL   -  calculate max and min, etc., of each image separately
        GLOBAL  -  calculate max and min of ALL images together
        MAXPOS  -  print position of maximum pixel in each image plane
Chose the default option = GLOBAL  by pressing  "ENTER"   ->   ↵
Mode of output:
  PRINT_ONLY, UPDATE_HEADER
Please specify option [PRINT_ONLY]                            PR
Input file, image loc#s []                                    Image_test
```

The program will print on the screen MAX, MIN, AVERAGE values of image densities

Preprocessing of images consists of several main steps

Extracting images of single particles from the image of the micrograph, changing size of images (Particle picking can be done using other packages).

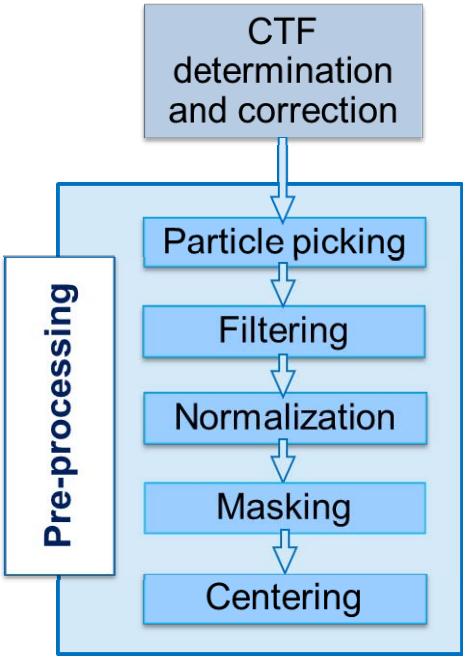
There is a command **SUBFILE (**or **CUT-IMAGES**) that allows you to extract a smaller part of the original image into a new file. It can be done from the center, or from any place within the original images. Option **APERIODIC** is used to extract small images of selected single molecules from the

original digitized micrograph using a **plt file** created by **Ximdisp**, **Boxer**, or some other programs, that have an output file of particle coordinates in **IMAGIC** format.

```
IMAGIC-COMMAND : SUBFILE

Input file, image loc#s              Image_test_nv
Output file, image loc#s             Image_test_big

Mode of operation:
  APERIODIC                CENTER
  GENERAL                  PAD_GENERAL
  LARGE_PAD                SMALL_PAD
  PERIODIC                 SQUARE
  CHECKERS                 FIXED_OVERLAP_CHECKERS
  BEST_CHECKERS
Please specify option [APERIODIC]          cent
Output image dimensions X,Y (integers)     768,768
```



To display the image use the command **'DISP'** with the window size **800,800**

How to check how many particles you have picked? Please use the command **HEADERS-INFORMATION:**

```
IMAGIC-COMMAND : HEAD

Options available:
  READ/LOOK   WRITE/SET   HOW_MANY    PLT_OUT     TAKE_OVER
  SHRINK      SORT        COMPARE     HISTOGRAM   MEANINGS
Please specify option [HOW]                  :

Input (header) file, loc#s [file name]       : E1_data_or

Image name: model data
Size: 200, 200  Loc: 1  Type: REAL  Cre.Date: 17-Aug-2015  Time: 14:04:52
 Number of 2D images in file: 3000
```

Pre-processing of images:

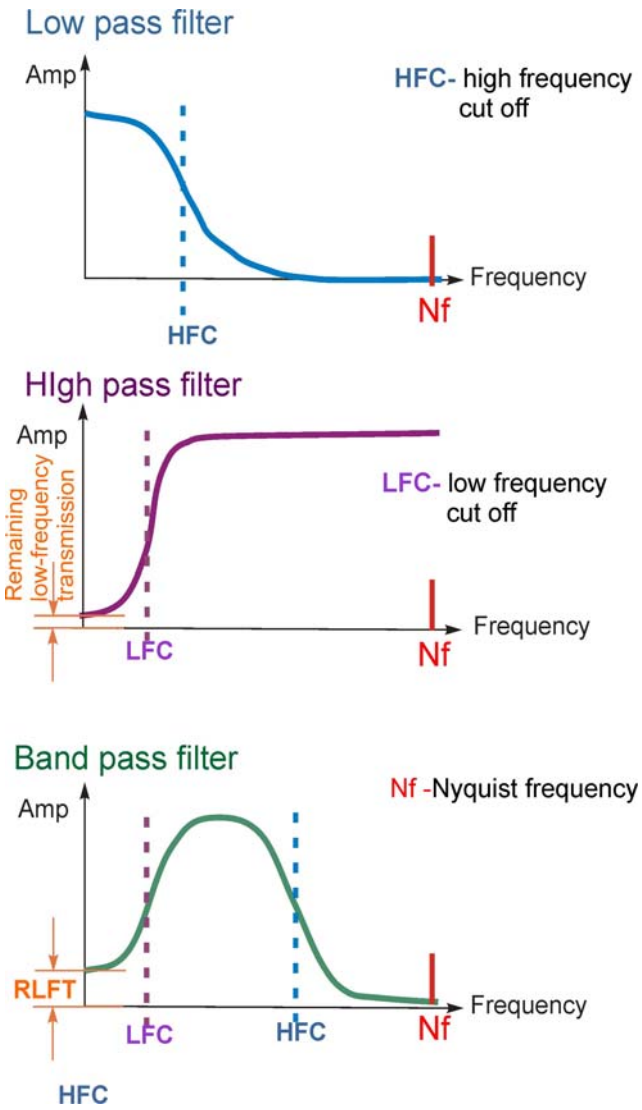
Using the command **PREPARE-IMAGES,** perform band-pass filtering of images. Apply a circular mask with $\mathbf{R_{mask} = 0.9}$, masking in real space (R_{mask} = 1 corresponds to a half of the box size). If R > 1.0 the program will understand it as pixels. Remember that the filter radius (R_{filt}) in transform pixels is given by formula

$$R_{filt} = 2 \times \text{pixelsize (Å)}/\text{(filter cut-off in Å)},$$

and that low frequencies correspond to long distances. That formula can be used to obtain low frequency cut –off (LFC) and high frequency cut (HFC) for the band pass filtering on the next step

Filtering of images

There are three main types of filters: low-pass filters that remove fine details of images (high spatial frequencies); high-pass filters remove or reduce large details (low spatial frequencies) and enhance small details (high spatial frequencies). Band-pass filters suppress both large and small details, allowing only details of certain size be enhanced – in the frequency space it would be a band of frequencies that passed through the filter.



Filters used in image processing

In order to recover all Fourier components of a periodic waveform, it is necessary to use a sampling rate (pixel size) at least twice the *highest* waveform frequency.

The **Nyquist frequency (Nf),** also called the Nyquist limit, is the highest frequency that can be coded at a given sampling rate in order to be able to fully reconstruct the signal:

$f_{Nyquist}$ = 1/(2pxl size)

Images can be normalized and filtered at the same time by using the operation **prepare-images**

```
IMAGIC-COMMAND : prep-images
Mode of operation                                  : PREPARE_IMAGES
Input file, image loc#s                            : ref_mix
Output file, image loc#s [ref_all_nv_bnd]          : ref_mix_jnk

  The image will be filtered. Please specify:

Low frequency cut off [0.1]                        : 0.1 ->LFC (See details
above)
Remaining low-freq. transmission [0.3]             :
High frequency cut off [0.65]                      : 0.7 -> HFC

  The image will be masked by a circle. Please specify
  the mask radius (pixels or fraction of inner radius).
  If you specify a drop-off it will be a soft mask.

Mask radius, drop-off (0: no mask) [0.65,10]       : 0.65,10
Desired new sigma [1.0]                            : 1.0
Remove (dust) outliers [NO]                        :
Invert the image densities [NO]                    :
Use MPI parallelisation [YES]                      :  YES
```

```
Number of processors to be used [3]                    :  5
```

It is recommended to check results of filtering using DISPLAY option how the filtering was done. If one will be not able to see the particles in the images it recommended to make the low frequency cut off smaller.

Normalization of images

Before you start any processing (e.g. calculating Fourier transforms, alignment, MSA analysis…) it is important to NORMALISE all images. This means that all images need to have an average density equal to ZERO (0), and the same standard deviation (σ sigma) of density distribution. This is performed using the command **NORM-VARIANCE.**

IMAGIC-COMMAND: **norm-var**

Please note: This command is used to normalize images to an average density equal to zero and with a specific standard deviation (sigma). During normalization you can use a circular, rectangular, or a contouring mask (using contours stored in a "plt" coordinate file).

```
Type of variance mask:
WHOLE_IMAGE  CIRCLE  SOFTCIRCLE  CONTOUR  MASK_FILE
Please specify option [WHOLE]                                  whole
Input file, image loc                                          data_set
**Use different input/output files if type PACK
Output file, image loc                                         data_set_nv
Please specify option                                          soft
Circle radius, drop off                                        0.9,7
Desired new sigma:                                             1
```

You may use any value for sigma, however, to get a reasonable range of densities in averaging and reconstructions the value “2” is recommended. **Remember** that, during the processing of your dataset, you **have to** use the sigma value you started with. Therefore, choose your favorite value for life and stick to it.

Centering of images.

Here you will find description of the procedure step by step. However in you practicals you will use different files and run all jobs from a batch file (see page 14). Here the file names are given as examples:

Calculate the total sum of all filtered images

```
IMAGIC-COMMAND : SUM-IM

Mode of summing:
  CONDITIONAL_SUM  SOME_SUM  TOTAL_SUM
Please specify option                          total_sum
Input file, NO loc#s                           set_sm_n
Output file, image loc#s                       sum_1
Variance images                                none
```

Rotational averaging of the image to prepare a reference to center the set of images

```
IMAGIC-COMMAND : aver-rot (AVERAGING-ROTATIONALLY)

Input file, image loc#s                        sum_1
Mode of output:
  IMAGE  PLT  BOTH
```

```
Please specify option                        IMAGE
Output file, image loc#s [sum_aver]          sum_aver_1
```

Next useful IMAGIC command: **DISPLAY**

Center the filtered images **"set_sm_n"** to the average-rotation of the total sum **"sum_aver_1"** with **ALI-DIR**

```
IMAGIC-COMMAND : ali-dir (ALIGNMENT-DIRECTIONALLY)
Alignment modes available:
  TRANSLATIONAL, ROTATIONAL, HORIZONTAL, VERTICAL, ALL
Parameters Option                        TRANS
Correlation functions                    CCF
Input file                               set_sm_n
Output file                              set_sm_cn
Reference file                           sum_aver_1
Give this reference a number             1
Options to filter the reference          NO_FILTER
Max shift (pixels/fraction of radius)    0.15
Full output of all alignment values [NO]            :NO
```

Now you will repeat this process (calculation of the total-sum, rotational average of it and translational alignment to it) several times. Repeating the cycle, using the new rotational average as reference each time and the original filtered set of images, will improve the alignment of the dataset. You can do this by running a batch job. For that we are using **MODE ACCUMULATIVE** option

See **page 14.** This procedure works well for the globular samples, if the molecule is elongated you should not do the rotational averaging but **"MASS-CENTER"**

Preparing the mask for statistical analysis

```
IMAGIC-COMMAND : test-im
File name:                          msa-mask
Image dimensions X,Y                200,200
Data                                REAL
Please specify option               DISC
Disc radius (fract /pixel)          0.8
```

You can look at the **MSA-MASK** file using the command **DISPLAY**

Statistical analysis of centered images

```
IMAGIC-COMMAND : msa-run (MULTI-STATISTICAL-ANALYSIS-RUN)
Use MPI parallelisation              NO
Please specify option                FRESH_MSA
MSA distances
Please specify option                MODULATION
Input   file (aligned "images")      set_bnd_c
Input MSA mask file                  msa-mask  (file create above)
Eigenimages output file              eig_1
Pixel coordinates output file        pixcoos
Eigenpixel vectors output file       eigenpix
Number of iterations (<65)           20 (number of iteration increases during
the                                      refinement process)
Number of eigenimages (< 70)         20 (number of iteration increases during
the                                      refinement process))
```

```
Overcorrection factor              0.8
Rootname for results file, NO ext.  msa_1
```

Over correction factor is ~ 0.8 (recommended default value). Higher values will give more accuracy but would take more time.

Display file **eig_1**. If when looking at the eigen images (eig-> your-own from German) the images appear black, change the grey values to automatic local scaling. You can show only the eigenimages you want by typing LOCATION in the display window and choosing the number of images you want to visualize.

The next command **MSA-CLASSIFICATION** sorts images with similar distribution of densities into groups. You have to define the number of such groups. If the number of groups will be big, each group will contain too few images, therefore the classes will be noisy. If the number of classes is too small, each group of images will contain too many images and different characteristic views could be mixed up. To understand this difference you can generate two sets of class-averages: one with 25 images per class (125 classes) and one with 100 images per class (30 classes).

```
IMAGIC-COMMAND : msa-class (MULTI-STATISTICAL-ANALYSIS-CLASSIFICATION)
Input to be classified
Please specify option                          IMAGES
Input (=output) file (treated by MSA)          set_bnd_c (it is an example name, you
                                               can use the file that is in your
                                               practiclas)
Percentage of images to be ignored             0
Active eigenimages for classification          30
Use default classification options             yes
What number of classes do you wish             30
Name of output results files                   classes_1_XXX  (typically it would be
                                               useful if XXX = # of classes that you
                                               want to have)
```

```
The results have been stored in the following files:
 ====================================================
 Text (listing) file containing classification results:    classes_1_XXX.lis
 File containing various plots of the variance behavior:  classes_1_XXX.plt
 File containing variance changes during classification:
    set_bnd_c_classify.plt
 File containing all classes and their members:
    classes_1_XXX.cls
 Image file containing coordinates of the classes:
    classes_1_XXX
```

The next command will calculate class averages – a total sum of images that compose a group (class) of images with similar features.

```
IMAGIC-COMMAND : msa-sum (MULTI-STATISTICAL-ANALYSIS-SUMMATION)

Input images to be summed               set_bnd_c
Rootname of input classification files  classes_XXX (recommended XXX=# of
classes)
Output class averages                   classums_XXX
Downweight small classes                NO
Mode of summing statistics:
  NONE  VARIANCE  S-IMAGE  I-IMAGE  FT
```

```
Please specify option                    NONE-> Try also the option S-image,
                                        and check on the screen what you obtain.
                                        White spots indicate the areas of the high
                                        variation of densities in images composing
                                        the class.

Fraction of worst class members to ignore 0
```

Small classes are groups that contain small number of images: from 1 to 3. If you look at such a class, you can see very often that the image has a strange distribution of densities caused by dust, a black line, or some spots. Visual inspection helps to identify them. So the option '**downweight small classes'** is not recommended; otherwise it will be difficult to interpret the results.
It is useful to use the option **S**-image, and check on the screen what you obtain. White spots indicate the areas of the high variation of densities in images composing the class. You can look at the images using the command **DISPLAY**

Writing a script for a batch job

Sometimes you need to repeat your calculation of certain steps during image analysis. For that one can use **MODE ACCUMULATIVE** option. MODE-ACCUMULATE is an important IMAGIC switch. After having switched over to MODE-ACCUMULATE operation, the commands are no longer executed directly but are rather accumulated on a file for later execution on the batch processor (MODE-SEND-TO-BATCH) or on the terminal (MODE-RUN). After having built a job in **MODE-ACCUMULATE** the **MODE-STOP** command will create a command file **NAME.b** which may be edited by the user.. See an example below:

```
IMAGIC-COMMAND : mo-ac

IMAGIC-COMMAND (ACC.) : exc-copy
What should be copied:
  2D_IMAGES/SECTIONS  3D_VOLUMES  1D_IMAGES/CURVES
Please specify option [2D_IMAGES/SECTIONS]         :

Exclusive copy operations:
  EXTRACT           EXCLUDE           REVERSE
  RANDOM_EXTRACT    SKIP_N_TAKE_M     TAKE_N_SKIP_M
  SORT              SELECT            SPLIT_FILE
  MERGE_FILES       MOVIE_DE_CHECKER
Please specify option [SORT]                       : extract
Input file [clsum_r2_no_den_300]                   : data_set_all_6000
Output file, image loc#s [clsum_r2_no_den_300_3er] : set_1500

Source of image loc#s:
  INTERACTIVE      PLT_FILE         NON_ZERO_IMAGES
  ACTIVE_IMAGES    INACTIVE_IMAGES  REF3D_UNCHANGED
  ORIGINAL_IMAGES
Please specify option [INTERACTIVE]                :
Location number(s) of image(s) wanted
[1-1000;3000-4000]                              : 1-1000;3000-4000
( it is important to put ";"
Numbers wanted:
  ALL  ODD  EVEN  INCREMENT
Please specify option [ALL]                        : inc
Increment wanted [2]                               : 4
```

Then you can type some other IMAGIC commands that are realted to the processing on the selected images such as normalization, filtering and any other. When you have finished your set of commands you have to use the command **MODE-STOP**

```
IMAGIC-COMMAND (ACC.) : mo-st

Filename for batch/script file [bigjob]            : extr-im-1500 (here you
have give a name for the batch job that would reflect what it does)

   Command (batch/script) file extr-im-1500.b
   is available now

   To run the job on the monitor please use
   UNIX command:   extr-im-1500.b
```

To run it as a hidden job please use the
UNIX command: job-name.b >& job-name.log &

To run it as a hidden job and display results on the monitor you can use the UNIX command: **tail -300 job-name.log** that will show the last 300 lines form the log file. You can look at the log file using an editor.

Applying alignment parameters to the images with high resolution information

```
IMAGIC-COMMAND : equ-rot (EQUIVALENT ROTATION)
                 The shift and rotation for two subsequent alignments
                 (centering and MRA) is done in one step in order to avoid
                 fuzziness caused by possible interpolation artifacts on high
                 resolution data.

Option used for current IMAGIC command             : EQUIV_ROTATION
Input file, image loc#s                            : set_bnd
Input header file with equiv. parameters           : set_sm_a1
Output file, image loc#s [set_100_bnd_5eqrt]       : set_bnd_a1
Max shift (pixels or fraction of inner radius)     : 0.2 (0.2 =0.15+0.05 this
                                                     comes for the previous
                                                     alignment parameters for
                                                     the smoothed images)
```