Software Engineering Project “Bugfind”
User Manual

Project group
Perttu Haimi, Petteri Hintsanen, Philipp Masche, Toni Ruokolainen and Osku Salerma

Instructor
Peter Lamberg

Helsinki 7th May 2001
User Manual
UNIVERSITY OF HELSINKI
Department of Computer Science

Version history

- v0.1, 25.3.2001 (first draft)
- v0.2, 4.4.2001 (first version)
- v1.0, 7.5.2001 (final version)

Contents

1 Preface

The program Bugfind is intended to find particles of a specific macromolecular complex (virus, ribosome, protein complex, etc.) in cryoelectron microscopy images. It makes 2D projections from a 3D model of a macromolecular complex and compares these with the image. The projection and comparison is a very time consuming task. Bugfind is implemented to be used in a Mandrake Linux Beowulf cluster, where it distributes the calculation to different nodes for parallel computing.

2 Basic usage of Bugfind

The user runs the program from the commandline in the cluster. The program name is `bugfind`. Bugfind has only one command line parameter, the filename of the configuration file.
Bugfind has to be started with `mpirun`, which needs some additional parameters to specify how many nodes the program should use.
For example

```
mpirun -np 3 bugfind myconfig.cfg
```

runs Bugfind on three nodes with the configuration file `myconfig.cfg`.

3 The configuration file

All parameters for Bugfind have to be specified in a configuration file. A configuration file is a simple text file, which contains in each line one parameter name and its value.

The following assignments are possible:

`IMAGEFILE=`*filename*	*filename* is the name of a micrograph file. The file format is MRC.
`MODELFILE=`*filename*	*filename* is the name of a 3D model file. The file format is MRC.
`OUTPUTFILE=`*filename*	*filename* is the name of the file where the output data will be stored.
`DIAMETER=`n	n is a number which describes the diameter for the gaussian-filter.
`DENSITY=`n	n is a number, which describes the projections-raster-density.
`SEARCHRADIUS=`n	n is the diameter of the circular area searched around the candidate points. Value of 0 means just search the candidate points.
`THRESHOLD=`n	n is a threshold value used by the matching algorithm. It's between -1.0 - 1.0.

```
IMAGEFILE=rod_noise.mrc
MODELFILE=rod.mrc
OUTPUTFILE=rodoutput.txt
DIAMETER=64
DENSITY=30
SEARCHRADIUS=0
THRESHOLD=0.6
```

A sample configuration file

4 The output file

All results of bugfind are placed in the output file. This file is a normal text file whose name is specified in the configuration file.
For each particle found one record of several lines will be printed to the output file. The records are ordered by the correlation score, with the best hits coming first. The records are divided by the separator '%'.

The output has the following format:

SCORE=f	f is the quality of the correlation, a value between 0.0 and 1.0
POSX=n	n is the x-coordinate of the particle in the micrograph
POSY=n	n is the y-coordinate of the particle in the micrograph
ANGLEALPHA=x	x is the alpha-orientation of the particle
ANGLEBETA=x	x is the beta-orientation of the particle
ANGLEGAMMA=x	x is the gamma-orientation of the particle
%	record separator

5 Hints how to use Bugfind

Bugfind is a relatively simple to use and robust tool, but there are some small things the user has to take care on his own:

- The image and the model should be in `unsigned char` MRC file format and the files should be readable for the user who runs Bugfind.
- The destination of the output file needs to be write accessable by the user who runs Bugfind.
- The numeric parameters in the configuration file should fit to the image and model. Bugfind just takes them as they are. If it finds nothing because of bad chosen parameters, it is the user's problem.
- The image and the model should have the same contrast. That means, that if the particle in the model is white on black background but the micrograph has black particles on white background, the micrograph has to be inverted.

The following things should be considered for each parameter:

- The diameter should have the pixel size of typical particles in the image. Nevertheless, it can be helpful to vary this a little.
- A high value for density can increase the runtime of Bugfind polynomially (x^3). But for asymmetric particles it can be necessary to search with a high density to pick all different shapes of particles.
- Search radiuses other than 0 should not really be used. The expected gain is very little and the running time of the program can balloon out of all practicality too easily.
- In micrograph images with a high noise-level, the best corrrelation results may not rise above 0.2. Some particles have a worse correlation value than some false matches. So it can be very difficult to find the right threshold.

The main difficulty in using Bugfind is to find out the right parameters for each particle picking. There are two important things to consider before running Bugfind:

- For large images it can easily happen that the runtime rises up to hours, even if a low density and a small search radius is chosen. Especially in these cases it is good to know in advance if the results will be satisfying. It is recommended to run Bugfind with a small piece of the image to find out the optimal parameters and run Bugfind with the whole image with these parameters.
- Since only results whose correlation values are greater than or equal to the the treshold are placed in the output file, it may be better to choose a small treshold and maybe get some bad results. These can easily be cut off from the result file afterwards.