FISH QUANT v0b

Installation

Requirements

Functions were programmed and tested for Matlab 7.11. We use several function of *the Piotr's Image & Video Toolbox for Matlab* [1]. The following toolboxes are needed

- *Optimization toolbox*
- *Statistics toolbox*
- *Image Processing Toolbox*
- (Optional) *Parallel Computing Toolbox*

Matlab code

1. **Unzip archive** in user folder of Matlab. Under windows the Matlab work folder is usually called something like `C:\Users\muellerf\Documents\MATLAB`, where `muellerf` is the user name. In Matlab the user path can be found with the command `userpath`. After extracting there should now be a folder `C:\Users\muellerf\Documents\MATLAB\FISH_QUANT_v0a`.

2. **Update Matlab path definition**. This can be done with a few simple steps in Matlab

   a. In the Matlab menu select *File > Set Path*

   b. This will open a dialog box. In this box select *Add with subfolders …*

   c. This will open another dialog. Here select the folder of FISH_Quant from step 1, e.g. C:\Users\muellerf\Documents\MATLAB\FISH_Quant. Click *OK*.

   d. To save this settings press on *Save*. Depending on the settings of the installation of Matlab this might results in a warning saying that the changes to path cannot be saved. Matlab proposes to save the path-definition file *pathdef.m* to another location. Click *Yes*. Select a directory of choice, e.g. the Matlab work directory of the user.

IMPORTANT: if an older version of FISH-QUANT is already installed on the computer, either delete this folder, move it to a different location, or remove it from the path-definition of Matlab.

MIJ – Interface between Matlab and Imagej

FISH QUANT uses the software package MIJ to visualize results in ImageJ[2]. The package will be automatically installed during the installation process. The package also requires a proper installation of ImageJ[3]. More detailed description can be found on the website http://bigwww.epfl.ch/sage/soft/mij/.

Start FISH-QUANT

At the start of FISH-QUANT MIJ is installed. To guarantee that this is successful FISH-QUANT has to be called correctly. Several options are possible. Open Matlab and type in the command window and depending on installation type one of the following commands

- **`FISH_QUANT('Ver','v0b')`**
  If installation was performed as outlined above. `'Ver'` is a keyword indicating that a version number will be specified, `'v0b'` is the version number of the current release.

- **FISH_QUANT**

  Can be called without option if the Java-files of MIJ were saved in the java sub-directory of the Matlab root directory. This option can be used if the MIJ files are installed according to the instructions on the MIJ web site.

- **FISH_QUANT('Dir','path')**

  If the Java-files of MIJ were saved in a different directory. `'Dir'` is a keyword indicating that a directory will be specified, `'path'` indicates this directory. For example, if these files were saved in directory C:\java, then the function call will be called by `FISH_QUANT('Dir',' C:\java')`

Overview

Matlab package to quantify mRNA FISH data. There are two main parts

- Single mRNA detection in the cell. Allows estimating the number of transcripts in a cell.

- Quantification of number of transcripts at the transcription site. (not implemented)

These functions can either be performed cell-by-cell or automatically for a larger population of cells.

The main program is FISH_QUANT. This allows accessing and controlling all other functions.

A typical workflow involves

- Finding the optimized conditions for mRNA detection in the cell by analyzing a small population.

- Running the automatic quantification

- Quality check by manually looking at random samples.

- Analyzing the transcription site.

Single mRNA detection

Example data-set

In the main folder of FISH_QUANT an example data-set can be found (FISH_test_data.tif). There is also an already defined outline for the cell and the transcriptions site (FISH_test_data__outline.txt) and the results of a successful spot detection (FISH_test_data__spots).Plots below are from the analysis of this data-set.

Basic workflow

This is the actual workflow which is then also implemented in the automated analysis of the data. The GUI allows optimizing the various parameters. Then this optimized set of parameters can be used in the automated detection.

1. Define experimental parameters

In this panel the parameters of the experiment can be specified. The user can choose from a number of presets that can be modified. Values needed are

- Pixel-size in XY in nanometer
- Pixel-size in Z in nanometer
- Refractive index of medium
- Excitation wavelength of fluorophore
- Emission wavelength of fluorophore
- Type of microscope. Currently 'confocal', 'nipkow' and widefield are supported.

These values are used to calculate the size of the theoretical PSF in XY and Z^4 (shown in the panel). These values are used for the filtering step and also as initial starting points for the fit with the 3D Gaussian. See Appendix 1 for more details.

2. Read-in image.

Opens a user dialog to specify the files which will be analyzed. Currently only single color stacks are supported. Alternatively a user-defined outline can be read-in (see next section).

3. Outline of cell and transcription site

Allows specifying the outline of the cell and the transcription-site(s). Definition is in XY and extends throughout the entire stack. mRNA will only be detected inside the cell and the transcription site will be ignored.

Alternatively a user-defined outline can be read-in from a file with 'Load Outline'. This option also automatically loads the associated image file. See next section "Saving and loading data' for more details.

4. Filtering

This step comes before the pre-detection. Filtering with a Gaussian Kernel will smooth the image. Two filtering steps are performed in 3D. First, the image is convolved with a larger Kernel to obtain a background image. This background image is then subtracted from the original image. Second, the background-subtracted image is smoothed with a smaller Kernel, which enhances the single particles.

The size of these two Kernels can be determined by the two multiplicative factors in the filtering panel.

5. Pre-detection

Pre-detection is performed on the filtered image with a non-maximal suppression. The threshold specified in the pre-detection panel (default 30) is the **minimum intensity** a spot has to have in order to be considered. This value has to be adjusted to the actual image data.

The **detection region** describes the size of a sub-region around each identified spot which will be considered in the following analysis steps. For each candidate spot a region +/- the specified number of pixel values with respect to the identified spot location will be extracted. This region is then used for the calculation of the quality scores and also for the fitting of the 3D Gaussian.

For each spot a **quality score** is calculated. Currently two scores are implemented

- *Standard deviation*. Standard deviation of all pixels in the sub-region. Works robust also for dimmer spots
- *Curvature*. 3D curvature [5]. Works best for bright spots.

When pre-detection is finished a **histogram of the quality score** is shown and the user has to specify a cut-off. In the title of the histogram the total number of spots and the currently consider number of spots are shown. The degree of thresholding depends on the quality of the data on how well the above parameters are adjusted. Good spots populate the right part of the histogram, whereas bad spots can be found on the right. If the threshold for the minimum intensity has been set optimally and the image data is not noise only one distribution can be seen (Fig A and B). Here the cut-off should be set that the entire distribution is considered.
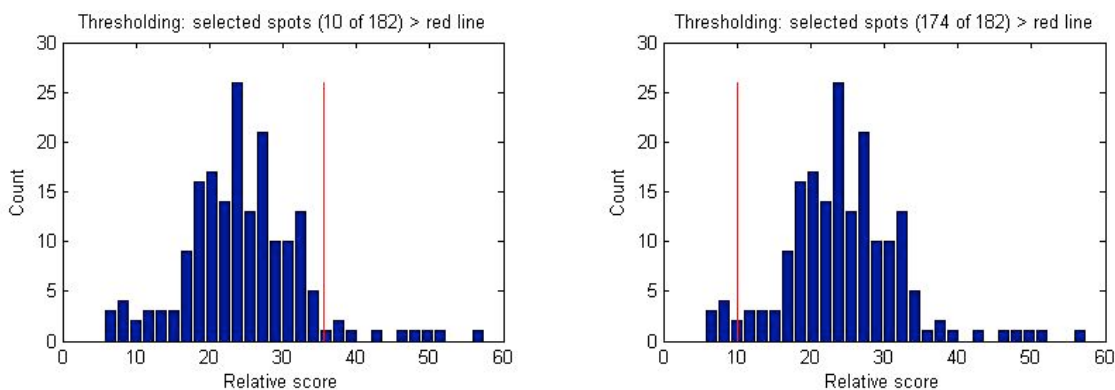
Fig A (left). Initial thresholding parameter for intensity is 30. Thresholding parameter for quality score is set to high and too many spots are rejected. Fig B (right). As in figure A but thresholding parameter for quality score is set correctly.

Often, the histogram looks bi-modal, and then the cut-off should be chosen such that the two populations are well separated. If there is only one distribution, the pre-detection worked either very well or the initial cut-off was already too stringent.
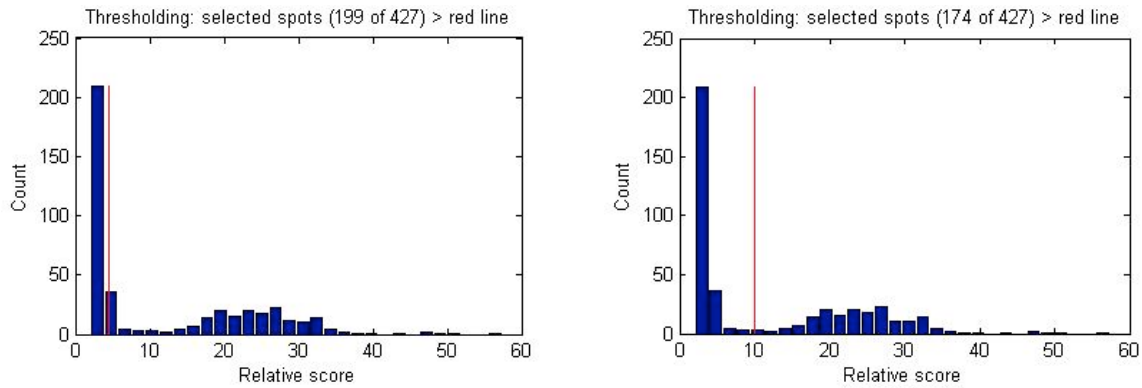
Fig C (left). Initial thresholding parameter for intensity is 15. Therefore more spots are considered. Ad in Fig A. Thresholding parameter for quality score is set to low and too many spots are considered. The spots identified in Figure A and B are on the right side of the histogram. The false-positives are on the left side (with smaller quality scores) . Fig D (right). As in figure C but thresholding parameter for quality score is set correctly.

Note: for this steps a relatively modest thresholding strategy can be applied since a second round of thresholding will be performed on the fitted spots.

Note: this step supports parallel computing. This option can be enable by checking the respective option in the lower part of the GUI.

6. Fit spots

All pre-detected spots will be fit with a 3D Gaussian considering the same sub-region used in the pre-detection. In the first round of fitting all fitting parameters are free, i.e. for each spot the

- width of the Gaussian in XY,
- width of the Gaussian in Z,
- position in X,Y,Z,
- amplitude, and
- local background

are calculated. As soon as the fitting process is finished various plots are shown. See next section 'thresholding' and the chapter on visualization for more details. The averaged values for the estimated parameters are shown in the panel 'Experimental PSF'.

The option '**Fixed width'** allows fitting all spots with fixed values for the width of the Gaussian in XY and Z. For this fit the values shown in the experimental PSF are used. Here different sets of parameters can be chosen, see 'Thresholding' and 'Averaging of spots' for more details.

The results of the spot detection can either be inspected in the GUI directly or by using the ImageJ visualization functionality (see 'Visualization' for more detail).

Note: this step supports parallel computing. This option can be enable by checking the respective option in the lower part of the GUI.

7. Thresholding

In this panel the detected spots can be thresholded with the different fitting parameters. The parameter can be selected in the pop-up menu. This changes the histogram to this parameter. Minimum and maximum values for this parameter can be specified. These values are indicated in the histograms by a

red and a green line. To activate this threshold the button 'Apply threshold has to be pressed'. The histogram on the right shows always the entire histogram of the selected parameter, the histogram on the left after thresholding.

The option '**Lock values'** allows saving these thresholding parameters. Then the next parameter can be inspected. The option 'Unlock all' removes all applied thresholds.

The results of the spot detection can either be inspected in the GUI directly or by using the ImageJ visualization functionality (see 'Visualization' for more detail).

Detection in batch model

… comes next …

Advanced detection features (not implemented)

… comes next …

Multi-Gaussian Kernel detection to resolve spots which are spatially close to each other [5] .

Visualization

Double click on all plots will open them in a separate figure.

Histogram of estimated parameters

Two histograms are in the thresholding panel. They show the histogram of the selected fitting parameter. The right plot shows the histogram of all parameters, the left after thresholding.

Location of detected spots

In the main plot a maximum projection in XY of the image is displayed. After the spot detection the identified spots are shown. After thresholding the thresholded spots among the detected spots are shown. Next to the plot window is a panel which allows selecting several plot options

- **Image**: raw image or filtered image

- **Spots**: all detected spots, detected spots and thresholded spots, no spots

- **Outline of cell and transcriptions site**: on / off

Quality of detected spots

There are three smaller plots situated under the main window. The first two panels show a maximum projection of all detected spots in XY and XZ is shown. The last panel shows the maximum projection of the residuals in XY.

Visualization in ImageJ

In this panel the results of the spot detection can be shown in ImageJ. Again different combinations of the image shown and the spots shown can be selected. The resulting images can either be displayed in a 3D orthogonal projection or as independent z-stacks.

Saving and loading data

FISH QUANT allows loading and saving of different parameters. Can be accessed by the Load/Save menu.

Outline of cell

The outline of the cell and the transcription site(s) can be stored in a simple text file (see Appendix 1). This files does, however, not only define the outlines but contains the file-name and information about the experimental details as well.

Results of spot detection

Results of the spots detection can be stored in a simple text file. This file contains all the estimated parameters for the detected spots of one cell. In addition it repeats the information contained in the outline-definition file.

Detection settings

… comes next …

Other Features

There are a number of other features implemented which simply the workflow.

Double click opens in new window

Double click on any of the plots in the GUI will open them in a separate figure.

Close all windows.

The button 'Close all windows' will close all windows except the GUI.

Parallel computing

This option enables parallel computing for the spot detection and spot fitting. The Parallel Computing Toolbox of Matlab is required.

Advanced features

Averaging of spots

FISH-QUANT allows averaging the identified spots. This is useful to obtain a PSF with improved signal-to-noise ratio. Some of the functionality is implemented in the GUI itself. More functions are implemented in the menu.

Visualization

The results of the various steps involved in the averaging process can be visualized in ImageJ. This plot functions can be found in the menu 'averaged spot'

Averaging

The user can average the spots in the GUI. The respective controls are in the panel 'Average spots'.

The user can define the **size of the resulting averaged spot** in pixels.

During the averaging process the spots are aligned based on the estimated center of the PSF. Since the center is known with sub-pixel accuracy, the alignment can be done with sub-pixel accuracy as well. This is implemented in FISH-QUANT as **oversampling (OS)** by sub-dividing the pixels in smaller sub-pixels. The corresponding factors are specified in the fields 'Factor OS'. The default settings are 3 for xy and z. This means that each pixel is sub-divided in 3*3*3 = 27 sub-pixels. Each sub-pixel has the same intensity than the larger pixel but now spots can be aligned with higher precision.

When starting the averaging process the user is asked if the **local background** of each spots should be subtracted before the averaging process. The local background for each spots is obtained from the fit with the 3D Gaussian. The results are shown in a separate Matlab figure.

Fitting

The resulting spot can be fit with the 3D Gaussian function. The size of the fitting region can be changed in the panel 'Predetection of spots'. The estimated numbers are shown in the panel 'Experimental PSF'. The width in xy and z can now also be used as fixed parameters to fit the individual mRNAs again.

Radial averaging

To reduce the noise levels the resulting image can further be radial averaged (Menu > Averaged spot > Radial average). The function determines the center of mass in XY and calculates for each z-plane the radial intensity distribution with respect to this center. This is done on the over-sampled data.

Constructing an image from radial averaged data

The radially averaged data can then be used to construct and xyz image. The constructed image will be on normal sampling size, e.g. with the original pixel-size. Since the construction is based on the over-sampled data the constructed image can have off-set positions of the center with respect to xy. When using this function (Menu > Averaged spot > Construct from radial average) the user will be asked for this offset in pixels.

Fit of constructed image

The constructed image can be fit with the 3D Gaussian function (Menu > Averaged spot > Fit constructed image). The size of the fitting region can be changed in the panel 'Predetection of spots'.

Appendix

Appendix 1 – definition of files defining outlines

These files are defined for each recorded stack. They define the parameters of the experiment and for each cell in the field of view the outline of this cell and the location of the transcription site(s). An example for such a file can be found below (grey box).

```
ROW   FILE
 1    OUTLINE DEFINITION, 05-Mar-2011
 2    FILE   FISH_test_data.tif   v1
 3    Pix-XY Pix-Z RI      Ex      Em      NA      Type
 4    160    300   1.458   568     568     1.25    widefield
 5    CELL   Cell1
 6    X_POS  60     5       40      238     230     100     END
 7    Y_POS  7      90      197     195     104     11      END
 8    TxSite 1
 9    X_POS  43     45      75      72      END
10    Y_POS  73     103     101     72      END
```

- **Line 1** Header row, arbitrary text.
- **Line 2** starts with the key word **FILE** followed by the filename, followed by the version number of the software.
- **Line 3** contains a description of the parameters in line 3.

Pix-XY	Pixel-size in XY in nanometer
Pix-Z	Pixel-size in Z in nanometer
RI	Refractive index of medium
Ex	Excitation wavelength of fluorophore
Em	Emission wavelength of fluorophore
Type	Type of microscope. Currently only widefield is supported.
BGD	Estimated background
RES	Squared sum of residuals

- **Line 4** specifies the experiment.
- **Line 5** specifies the first cell. Starts with keyword **CELL**, followed by a unique identifier of this cell.
- **Line 6** specifies the x-coordinates of the polygon describing the outline of the cell. Line is terminated with key-word **end**
- **Line 7** specifies the y-coordinates of the polygon describing the outline of the cell. Line is terminated with key-word **end**
- **Line 8** specifies a transcription site. Starts with keyword **TxSite**, followed by a counter.
- **Line 9** specifies the x-coordinates of the polygon describing the outline of the cell. Line is terminated with key-word **end**
- **Line 10** specifies the y-coordinates of the polygon describing the outline of the cell. Line is terminated with key-word **end**

Notes
- It is not necessary to defined transcription-sites. This block can be omitted.

- Additional transcription-sites can be defined by repeating block from line 7-9 and increasing the counter in line 7.

- Multiple cells can be defined by repeating the block from line 4 to line 9 with unique cell label.

- All values are separated by tabs.

- No empty rows in between definitions.

- Files are saved as .txt files.

Appendix 2 – definition of files summarizing results of spot detection

These files are defined for each cell in the recorded stack. Relevant information from outline definition files is repeated (experimental parameters, outline of cell and transcription site. An example for such a file can be found below (grey box).

ROW	FILE
1	RESULTS OF SPOT DETECTION PERFORMED ON 05-Mar-2011
2	FILE FISH_test_data.tif v1
3	Pix-XY Pix-Z RI Ex Em NA Type
4	160 300 1.458 568 568 1.25 widefield
5	CELL Cell1
6	X_POS 60 5 40 238 230 100 END
7	Y_POS 7 90 197 195 104 11 END
8	TxSite 1
9	X_POS 43 45 75 72 END
10	Y_POS 73 103 101 72 END
11	SPOTS

12	Pos_Y	Pos_X MuX	Pos_Z MuZ	AMP ITER	BGD SCORE	RES TH	SigmaX	SigmaY	SigmaZ	Cent_Y	Cent_X	Cent_Z
	MuY											
13	19618.4	20287.7 258.366	6732.74 287.676	129.545 1332.74	476.225 7	25197.6 20.2169	181.743 1	181.743	882.791	316.289	318.481	1203.91
14	20598.2	13092.7 278.157	7329.4 292.663	198.508 1329.4	463.098 7	33560.3 30.5706	178.049 0	178.049	836.684	316.336	318.219	1209.74

- **Line 1** Header row, arbitrary text.

- **Line 2-10** repeat information about file-name, experimental parameters, cell position and transcription site location as specified above.

- **Line 11** Key word SPOTS indicates that results of spot detection start

- **Line 12** Header for spot detection

Pos_Y	Estimated position of center in image (Y, in nanometer)
Pos_X	Estimated position of center in image (X, in nanometer)
Pos_Z	Estimated position of center in image (Z, in nanometer)
AMP	Estimated amplitude
BGD	Estimated background
RES	Squared sum of residuals
SigmaX	Estimated width = sigma of Gaussian (X, in nanometer)
SigmaY	Estimated width = sigma of Gaussian (Y, in nanometer)
SigmaZ	Estimated width = sigma of Gaussian (Z, in nanometer)
Cent_Y	Centroid of sub-image (Y, in nanometer)
Cent_X	Centroid of sub-image (X, in nanometer)
Cent_Z	Centroid of sub-image (Z, in nanometer)
MuY	Position in sub-image (Y, in nanometer)
MuX	Position in sub-image (Y, in nanometer)

MuZ	Position in sub-image (Y, in nanometer)
ITER	Iterations used for fit
SCORE	Score of pre-detection
TH	Considered after thresholding (1 = Yes, 0 = No)

- **Line 13- end** Results

References

1. Piotr's Matlab Toolbox. at <http://vision.ucsd.edu/~pdollar/toolbox/doc/>

2. Sage, D., Prodanov, D. & Ortiz, C. MIJ - A Java package for bi-directional communication and data exchange from Matlab to ImageJ. at <http://bigwww.epfl.ch/sage/soft/mij/>

3. ImageJ. at <http://rsb.info.nih.gov/ij/>

4. Zhang, B., Zerubia, J. & Olivo-Marin, J. Gaussian approximations of fluorescence microscope point-spread function models. *Appl. Opt.* **46**, 1819-1829 (2007).

5. Thomann, D., Rines, D.R., Sorger, P.K. & Danuser, G. Automatic fluorescent tag detection in 3D with super-resolution: application to the analysis of chromosome movement. *J Microsc* **208**, 49-64 (2002).