MotionTracking Statistic Manual

MotionTracking by
Yannis Kalaidzidis
Alexander Kalaidzidis, Giovanni Marsico, Mikhail Chernykh
1997-2013

Manual by
Yury Bodrov (bodrov@mpi-cbg.de)
2010-2013

Zerial Lab, MPI-CBG

Last Update: November 29, 2013

Contents

1 Calculate Statistics with MotionTraching **2**

2 Statistic Calculation Tools **2**
2.1 Statistic Menu . . . 2
2.1.1 Distributions and Dependencies . . . 2
2.1.2 Time Course Analysis . . . 4
2.1.3 Analysis of individual Tracks . . . 5
2.2 Batch Statistic Window . . . 6
2.2.1 Automatic Enumeration . . . 6
2.2.2 Image Set Analysis . . . 8
2.2.3 Process Data . . . 10
2.2.4 Plotting Data from Batch Statistic Window . . . 11
2.2.5 Handling multiple projects or result files . . . 11
2.3 Graph Window . . . 12

3 Statistic Characteristics and Parameters **15**
3.1 Common Statistical Parameters . . . 15
3.2 Object Characteristics . . . 15
3.3 Track Characteristics . . . 16
3.4 Quality Control Parameters . . . 16
3.5 Shape Analysis . . . 16
3.6 Other Statistical Parameters . . . 16
3.7 Example Statistic List . . . 16

1 Calculate Statistics with MotionTraching

This Chapter is dedicated to the most insightful part of MotionTracking: the statistic calculations.

While object and track search are important to produce data it is useless without proper analysis. This is where the massive statistical tools come into play. There is a great variety of them and this guide will rather show how to access them than explain their mathematical conception.

MotionTracking offers a variety of tools that can be used to do this analysis. These are described in Section 2.

Objects (or Tracks) that were calculated in the previous stages have a very large number of possible properties which can be of interest for different experiments. These can be roughly grouped into

- Basic properties of the data, for example *Number of Objects* in a image (see Section 3.1).
- Fundamental properties ob Objects (Section 3.2) and Tracks (3.3), for example *Mean Object Intensity* or *Total Track Length*
- Quality Control Parameters used to evaluate the quality of the experiment, for example *In Focus Score* (Section 3.4)
- Specialized Parameters used in specific applications (Sections 3.5 and 3.6)

2 Statistic Calculation Tools

The fastest and simplest way to calculate statistics for a given set of images (one Motion-Trakcing project) is to use the "Statistics" (2.1) menu found in the MotionTracking Main Window. It is best used to quickly calculate a small number of interesting metrics, for example to estimate if the mean intensity is high enough over the whole project to confirm proper staining. In a most of the cases though the number of statistics of interest is high and it becomes impractical to calculate them one buy one and save each result seperately. Therefore the "Batch Statistic Window" (2.2) is used frequently to calculate and store multiple statistics simultaneously. This "Batch Statistic Window" is the fundamental *quantitative multi-parametric image analysis (QMPIA)* tool.

2.1 Statistic Menu

2.1.1 Distributions and Dependencies

Object Characteristic Distributions and Dependencies are the simplest statistics that can be calculated and are the main use for the "Statistic Menu". The results will be plotted immediately in a graph window (See section 2.3). Detailed descriptions of the most common characteristics can be found in sections 3.2 and 3.3.

The most important items of this menu are the Object (or Track) Characteristic Distribution (Figure 2) and Dependency (Fig 3). Both open a separate window which

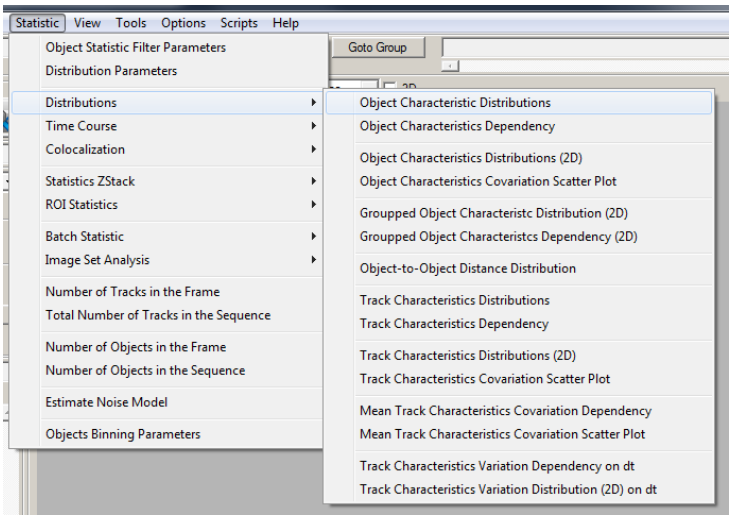


Figure 1: Statistic Menu with Distributions dialog highlighted.

allows you to select many different Object Characteristics like intensity or size. The Track Characteristics work exactly like the object ones except offering you parameters like track duration or speed instead of object size and so on.

The characteristics itself are listed on the left side and are used to change the characteristics of the selected item in the list on the right.
In both dialogs you can add elements and operators with the right click menu as seen in the right part of Figure 3. To add multiple object characteristics use the button found in the "Add" submenu, as clicking on the left will just change the current one. When adding a new item you will be asked which characteristic you want to add directly.
As you might imagine, this tool can be used to construct any mathematical formula containing the Object Characteristics and numbers.

You will notice several places to select weighting functions. It will open an almost identical dialog to that of Figure 2 where you can select the parameter you want the original characteristic to be weighted by. This allows you to take certain objects as more important than others.

For example a size distribution weighted by mean intensity will take brighter object into account stronger than dimmer ones and thus reduce the impact of the noise.
The weighting can, of course, be done in different ways: Mean, Mode, Median and Sum. Sometimes it makes sense to calculate a important parameter with all available weighting and without any weighting at all to compare the results. Mean is the most default setting, though.

The results will be plotted in an appropriate graph (see section 2.3)

Both windows (2 and 3) allow to save (and load) the setup so that it can be used on a different project. The buttons are in the menu bar at the top.

Important Note: To select a different channel you have to do it in the box shown in Figure **??**.

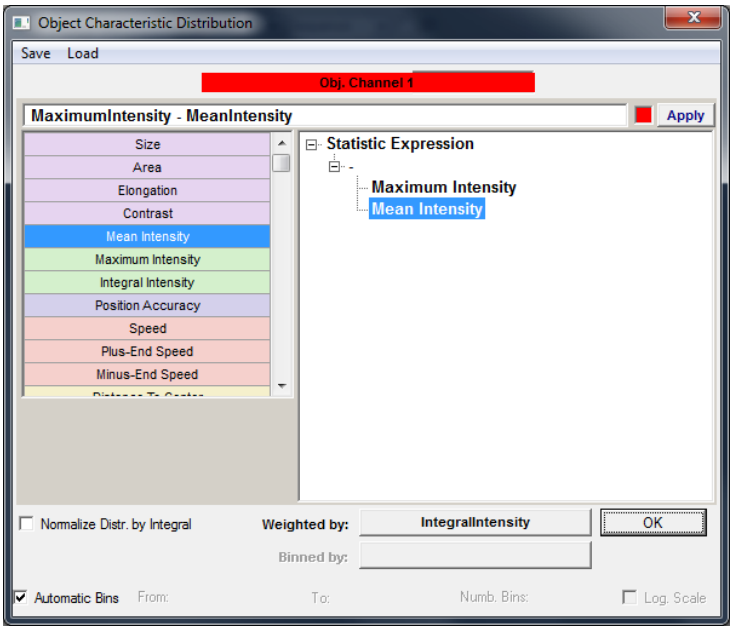


Figure 2: Object Characteristic Distribution Dialog. In this (not particulary meaniningful) example the "Characteristic" to be calculated is the difference between maximum and mean intensity which is weighted by the integral intensity. This will be calculated for all objects on Channel 1 (red) in the project (.mtj file). Note that it is possible to use custom binning for the resulting distribution in the bottom of this window.

2.1.2 Time Course Analysis

MotionTracking can measure how certain statistics change over the course of a movie. For example, how does the number of objects change with time? Does the number increase or decrease? Or again, how does the Mean Intensity of my objects change with time? These types of analysis are accessed through the "Statistics → Time Course" submenu shown below:

- Number of Objects Course: reports the number of objects found in each frame by Object Search versus the time of the movie.
- Number of Tracks Course: reports the number of tracks present in each frame as found by Track Search versus the time of the movie.
- Integral of Object Characteristic Course: reports the sum of the Object Characteristic Integral for all objects in the frame versus the time of the movie. The Characteristic you are interested in can be chosen from a dialog similar to figure 2.
- Mean Object Characteristic Course: reports the sum of the Mean Object Characteristic for all objects in the frame versus the time of the movie. Characteristic choice as above.
- Track Characteristic Course: reports the sum of the Object Characteristic Integral for all objects in the frame versus the time of the movie. Characteristic choice as above.

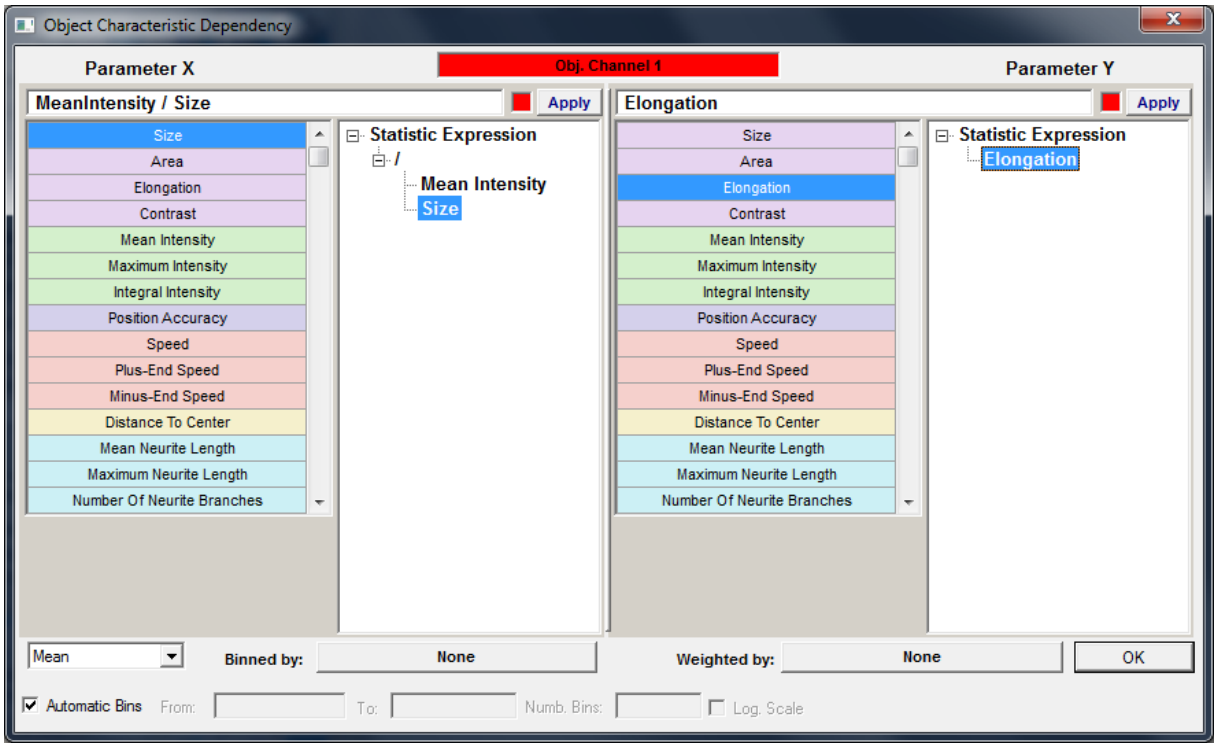


Figure 3: Object Characteristic Dependency Dialog showing an example. In this case the unweighted Dependency between *the ratio of mean intensity to size* and the *elongation* of objects is set up to be analysed for all objects on channel 1 (red). The mean of the distribution will be displayed as the numerical result as selected in the box on the bottom left of the window.

- Total Intensity Course: reports the sum of the intensity values of every pixel versus the time of the movie. Note: these values will be very large numbers (imagine an average intensity of 250 a.u. on a pixel for a 1000 × 1000 pix image) for larger images. This statistic is also useful in characterizing photo-bleaching.
- Background Intensity Course: reports the sum of the background fluorescence in each frame versus the time of the movie. Note: the sum of the background fluorescence will be a very large number for larger images. This statistic is useful in characterizing the photo-bleaching of the background fluorescence.
- Mask Area Course: If a mask has been applied to the images, this feature calculates the relative area of the mask with respect to the area of the image as a whole. The relative area is plotted versus the time of the movie.
- Object Characteristic Distribution Course: reports the change of a Object Characteristic Distribution over time.

Additionally, there are 5 more Time Courses based on dependencies between different Object Characteristics or Intensity in particular, which can be very handy for certain tasks but are too specialised to be explained here in detail.

2.1.3 Analysis of individual Tracks

MotionTracking allows you to examine the properties of individual tracks as well. Position the mouse cursor over the track of interest. Right-click to bring up a context-sensitive

menu. Select the appropriate statistic.

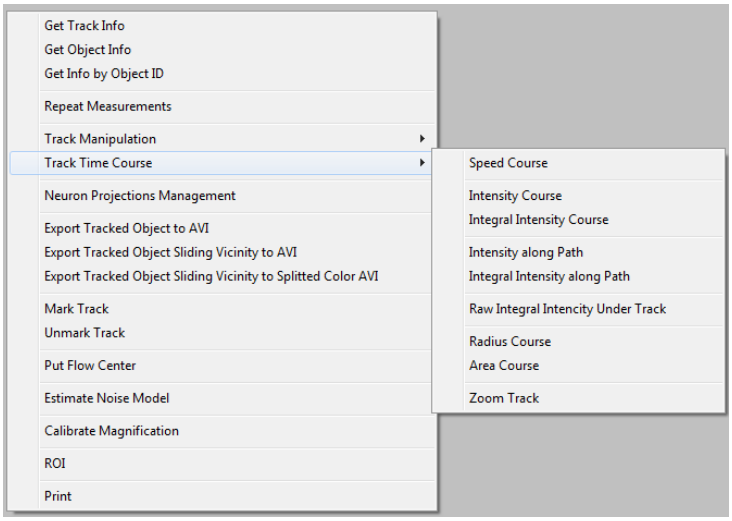


Figure 4: Overview of Statistics for Individual Tracks.

The various options in the Track Time Course menu will produce the value (e.g., the Intensity) of the track versus the time of the track. In the case of object properties such as Intensity, Radius, and Area, the program measures the objects associated with the track in each frame.

2.2 Batch Statistic Window

From "Statistics → Batch Statistics → Open Batch Statistic Window" you can access the "Image Set Analysis" window where a row in the table will be created for each frame in the picture (if you have a project loaded). The file-name column will be filled with the path and filename of each picture. From here on you can save the table as an .srl file which works independently from the project files once created. It will copy the statistic filters and requires masks and objects (and tracks if applicable) to be already calculated and present.

The first option, "Open Batch Statistic Window" will do just that: open an "Image Set Analysis" window where a row in the table will be created for each frame in the picture. The file-name column will have the corresponding full file name of the image data for each frame in the order they are in the project. From there on you have to add additional columns depending on what you want to calculate. This Image Set Analysis has three core menu items: Image Set Analysis (Figure 8), Process Data (9) and Automatic Enumeration (6).

2.2.1 Automatic Enumeration

Enumeration is essential for MotionTracking to handle the data set properly. This means that the "Sequence Value" column will be filled out according to some rules. Although any statistics can be calculated without it, no processing can be done and no data can be plotted unless all rows of the table are enumerated.

MotionTracking has different pre-defined fields to handle certain time of information that

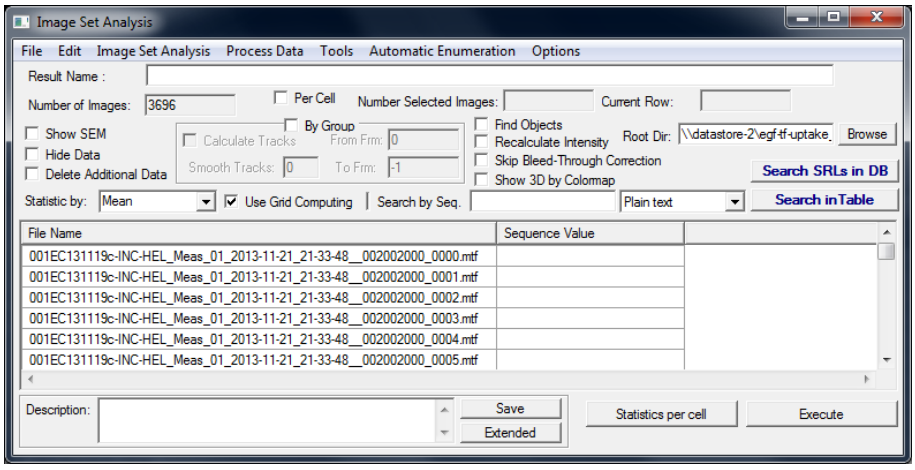


Figure 5: Overview of an empty Batch Statistic Window. Each line represents one frame of the project. The first column shows the path to the file on the hard drive, the second column is added by default and will contain some sort of description of the conditions later on.

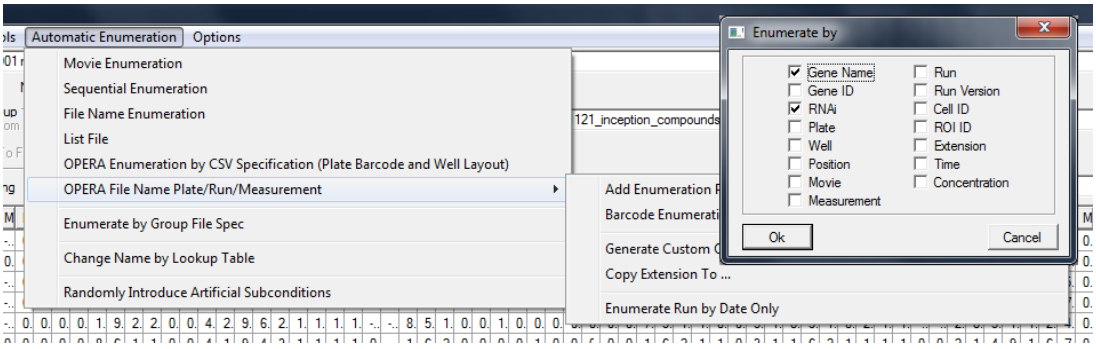


Figure 6: Overview of the automatic enumeration menu. It is used to assign meaningful identifiers like the experimental conditions to each image in an (semi-)automatic fashion.

is used to describe the conditions of each of the images in the table (Gene, siRNA, concentration etc.).

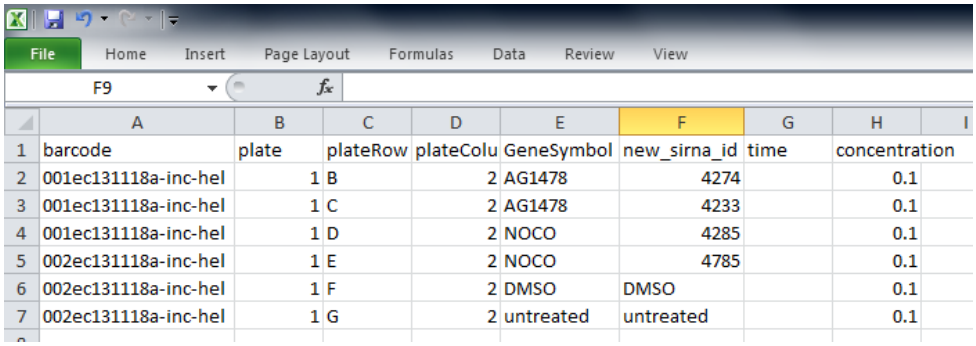
There are multiple ways to enumerate the data:

- **Manual Enumeration** is the easiest solution for small numbers of lines. Just double-click on each well in the Sequence Value column and fill in the condition of the data. To avoid confusion it is best to keep the description short to one character sequence.
  Although it is technically possible to manually enter a sequence that will math the MotionTracking format for the Gene Name, siRNA etc. fields it can be quite confusing.

- **Random or Sequential Enumeration** will introduce arbitrary values. Both can come in handy if you just want to plot the data quickly, but can't be used to process data properly.

- **Automatic Enumeration for OPERA** can be used if the barcode of the images can be parsed into their exact location. An enumeration file that maps position and

condition can then be applied to identify all images of the table at once. See Figure 7 for the necessary layout of such enumeration files.
In addition it's possible to extract certain information like plate number, date or run number directly from the barcode via "Opera File Name Run/Plate/Measurement".



	A	B	C	D	E	F	G	H
1	barcode	plate	plateRow	plateColu	GeneSymbol	new_sirna_id	time	concentration
2	001ec131118a-inc-hel	1	B	2	AG1478	4274		0.1
3	001ec131118a-inc-hel	1	C	2	AG1478	4233		0.1
4	001ec131118a-inc-hel	1	D	2	NOCO	4285		0.1
5	002ec131118a-inc-hel	1	E	2	NOCO	4785		0.1
6	002ec131118a-inc-hel	1	F	2	DMSO	DMSO		0.1
7	002ec131118a-inc-hel	1	G	2	untreated	untreated		0.1

Figure 7: Layout of an enumeration file. It can be crated with Excel or similar, but must be saved as a .csv.

2.2.2 Image Set Analysis

The first one, "Image Set Analysis" is used to add parameters which will then be calculated when the "Execute" button is used. Which parameters to use depends on your experiment in particular, but almost certainly you will be interested in the object characteristics of "size" and "mean -", "total -" and "integral intensity". Typical track characteristics are "speed" and "duration". Of course MT has many more to offer and you can create any combination or ratio between the characteristics as described in via the dependencies menu. See section 3 for a list of available statistics.

If you have calculated a part of the table and click execute you will be asked if you want to recalculate all data (which will erase all previous data) or not (in which case only the lines with *at least one* empty cell will be calculated). You can select a specific number of lines before clicking the Execute button, in which case you will be asked if you want to narrow your calculation on the selected rows. Additionally there is a way to (re-)calculate single columns, rows or wells out of the list as well: "right click → calculate → calculate column/row/well". To clear all data from the rows use "right click → clear rows".

The Object and Track Characteristic (Dependency) menu items work exactly as described in section 2.1. Additionally to the possibilities that they give you there are some quite useful features mostly based on Distances. Remember that all Cell or Nucleous related options require nuclei and cells to be calculated during Object Search.
Additionally to all statistics that you want to use it is advised to include the following quality control parameters to filter out empty, overfilled and out-of-focus images:

- Number of Objects on the Nuclei-Channel
- Masked Area
- any of the parameters specified under "Image Quality Controls"

To filter the results, sort the statistic table by one of the quality control parameters (left click on the column header and select "Sort"). Then select all rows falling into your category as you would normally select multiple files in windows, and "right click → delete" them. This delete will only affect the current .srl file and will never touch any project or image files.

Typical thresholds for bad pictures are the following numbers:

- any with less than 5 (too empty) or more than 65 (too full) Nuclei
- less than 5% masked area (too empty again)
- about 10% of the images with the least contrast (out-of-focus score), although this strongly depends on the quality of your imaging. You should take a look on the low-on-contrast pictures before deleting them by selecting them and using "Right Click → Go To Image/ Activate Image Set"

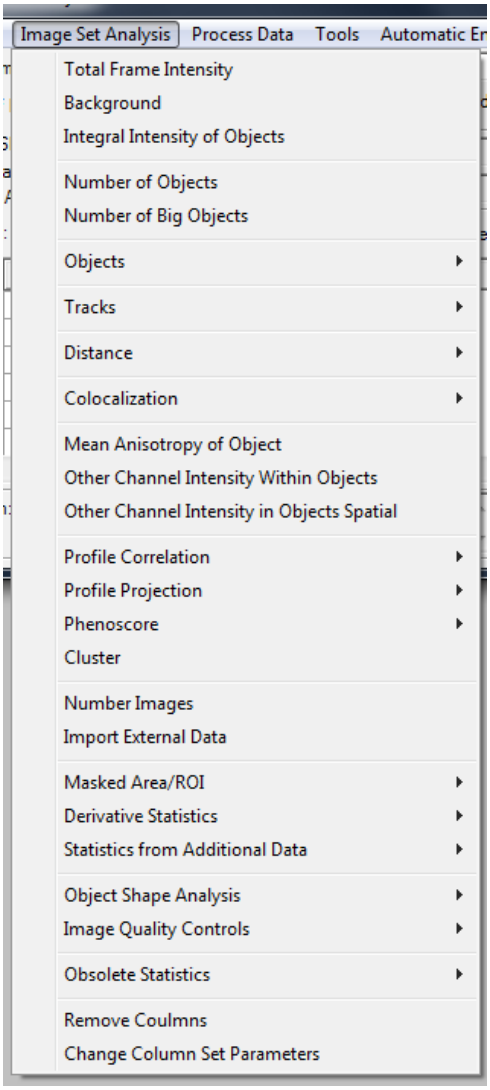


Figure 8: Overview of the image set analysis. It is used to add statistics to be calculated.

After you have selected the parameters you need you can get an overview of the whole parameter set by using “Change Column Set Parameters” from Figure 8.

This list of parameters can be saved as a “Statistic Template” and loaded (applied) to other .srl files in the future via “File → Statistic List Template → Load/Save Statistic List”.

As you might have noticed, the table can grow very big if you want to know many things about the pictures. Besides, every cell of this statistic contains the whole distribution of this parameter with all necessary information labelled as “additional data”. Thus, the size of a .srl table can easily grow to 100 Mbyte or more. If you don’t need the whole distributions and are satisfied with the Mean, which is displayed as a number in the cell, select “No Additional Data” and the file size will be considerably smaller. You can view the “Additional Data” at any time by selecting it from the context menu of any cell. To get the additional data back you have to recalculate the respective cells, rows or columns.

2.2.3 Process Data

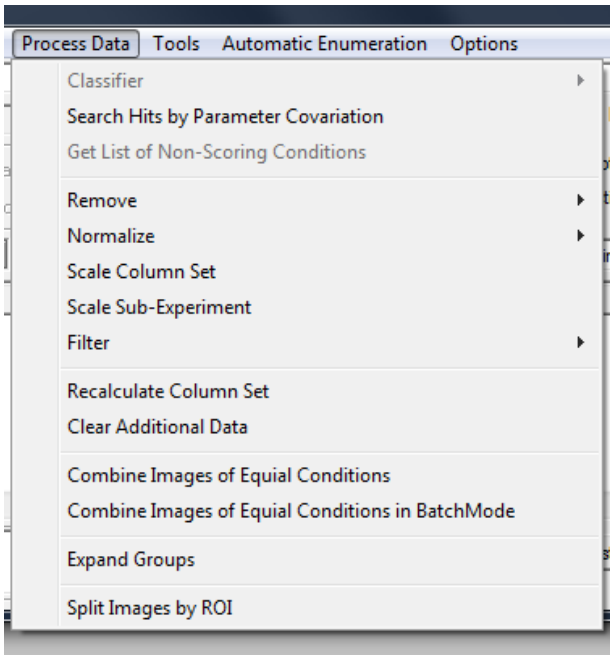


Figure 9: Overview of the processing menu. It is used to process the table as a whole after calculation.

This menu is used to manipulate the whole data set. Remember that most of them won’t work unless the data is properly enumerated (see section 2.2.1) It’s most important functions are:

- **Combine Images of Equal Condition:** combines the data from all lines which have the same values in the selected fields of the enumeration (for example combine all images with the same Gene and siRNA).
- **Normalize:** allows you to normalize the complete data set by either Total Set or a Control condition. In addition there are some options to shift the base value and scale the data sets.

- **Remove:** offers various functions to remove certain frames from the result file. For example outliers or those with bad values (Empty lines, NaN, etc.)

The common procedure is to enumerate the data, calculate everything and then process the data by combining all images of equal conditions and then normalize them by the control group (Mock, Untreated, etc/).

2.2.4 Plotting Data from Batch Statistic Window

There are multiple ways to plot a certain data set in MotionTracking. Note that the data set needs to be enumerated, so that none of the "Sequence Value" cells is empty, or the plots will not be created.

- **Profile (Curve/Bar Graph)**: Plots the profile(s) of the selected line(s) by displaying the value of each cell on the y-axis and the parameters on the x-axis. Note that it is helpful to normalize the data beforehand to get reasonable profiles as different parameters can be on different orders of magnitude to be compared in a non-normalized fashion.
- **Create Curve/Bar Graph**: Plots the values of the selected column for all lines. This is useful to compare a specific parameter across all conditions.
- **Get Additional Data**: Plots the distribution underlying the selected cell. Note that some parameters are just values (for example Number of Objects on Channel 1) but others have a distribution (Mean Intensity of Objects on Channel 1) and the mean value is the one that is displayed in the table.
- **(Cumulative) Parameter Distribution**: Plots the distribution of the mean value of the selected parameter of all rows (conditions).

2.2.5 Handling multiple projects or result files

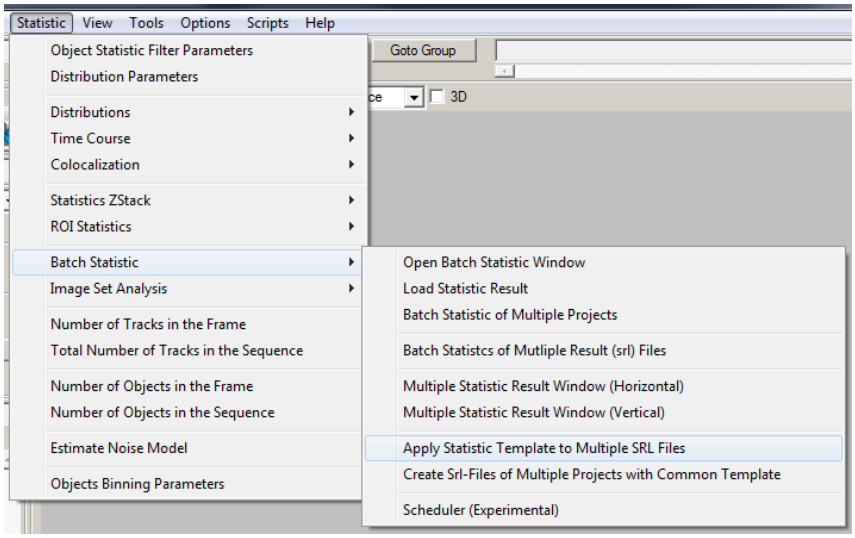


Figure 10: Overview of the available ways to handle multiple projects and result files.

Batch Statistic of Multiple Projects: After selecting the project files, an "Image Set Analysis" window will open with the lines representing whole projects instead of simple frames.
This is useful if you have grouped your images in projects in such a way, that you can compare complete projects instead of single frames (for example a movie of one condition vs. a second movie of a different condition).

Create .srl files of Multiple Projects with Common Template: creates a .srl file for each project in the list and uses the selected template for the other columns so that you don't have to do this manually.

Batch Statistic of Multiple Result (.srl) Files: is a little bit different. Instead of using projects like the other two, it uses already existing result files. After choosing all the .srl files you want it goes through them and executes each one. In the end it works in the same way as any batch operation like "batch find objects" or "batch find tracks".

Apply Statistic Template to multiple SRL files: loads all of the selected result files and applied the selected template to them. Useful if you create a bunch of .srl files at some point and want to apply a new template to them (Warning: overwrites old template and deletes all calculated statistics).

2.3 Graph Window

The "Graph" window is used to both visualize data and do simple or complex manipulations of the resulting curve (for example averaging multiple curves). In MotionTracking *graphs* are the coordinate systems and *curves* are the plotted datasets. You can add as many graphs or curves as you want and move the curves freely between the graphs. In Addition to the main "Graph" Window there are "BarGraph", "3d Graph", "Time Course" and "Scatter Plot" Windows that have essentially the same functionality and user iterface, but handle different types of data.

Let's start with the simple stuff: you can find all scale-related options in the "Scale" menu in the menu bar. You may use "Find Scale" to let the program figure out the optimal scale by itself, but sometimes, for example if you want a logarithmic scale, you have to set it up manually. Select "Set Scale" in this case. You will see a window similar to Figure 12. Using x/y min and max will let you set up the minimal and maximal values of the view range. The Checkbox behind these values allows you to set up a logarithmic scale for the respective axis. Remember that once you select log scale you can't set the minimal (or maximal if working with negative values) to 0. (Because of $log(0) = -\infty$). The "Enable Sweeps" option allows you to display an split in the x-axis with the values specified below.

The "Graph" Submenu allows you to label your axis, create a legend and turn on a second y-axis on the right side of the graph. The "Marker" submenu is related and allows to modify shape and size of the curve points (markers) and thickness of the axis.
To add text, arrows or similar try the "Text" submenu. The text-fields you can add there behave just like normal text fields in any image editing program and can be moved around or manipulated. If you don't like the colors of the plot, go to "Color" and select the appropriate option.

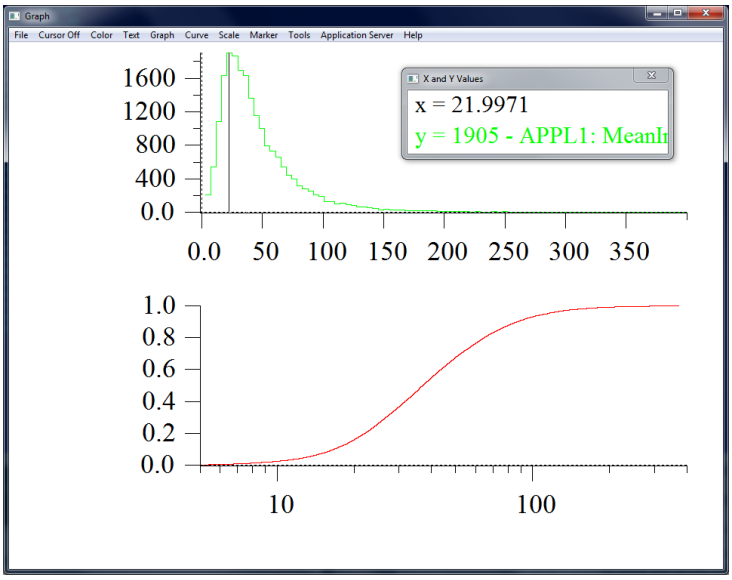


Figure 11: Plot of the distribution and the cumulative distribution of the *mean intensity* of the same objects. To switch to the calculation of cumulative distributions you need to select it in the "disribution build parameters" which can be accesses via the button right above the illumination correction interface entry.

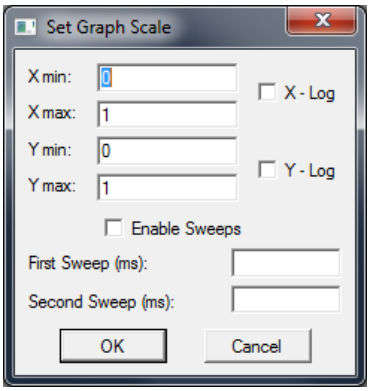


Figure 12: Scale setup dialog.

If you want to know the exact values, select "Cursor On" in the Submenu bar to get a line on where your cursor currently is located. A separate window with the exact coordinates where all curves cross this line will be displayed.

It is possible to save graphs and curves either together (*Save Window*) or just the curves (*Save Curve (Set)*) for later use through the file menu. This will use a MotionTracking-specific file format (.gwd) so the calculations can be continued properly when the graph is loaded again.

The whole plot can be exported as an image in various formats (*Save Image As*). The data sets (curves) can be exported as a .csv file for further analysis with Excel or similar programs (*Export Curve as CSV/Excel*).

The last and most complex submenu is "Curves". It contains all the mathematical options available in the graph menu. Most of them have a lot of sub-options and parameters so a full explanation would be too long. At this point you should know what you

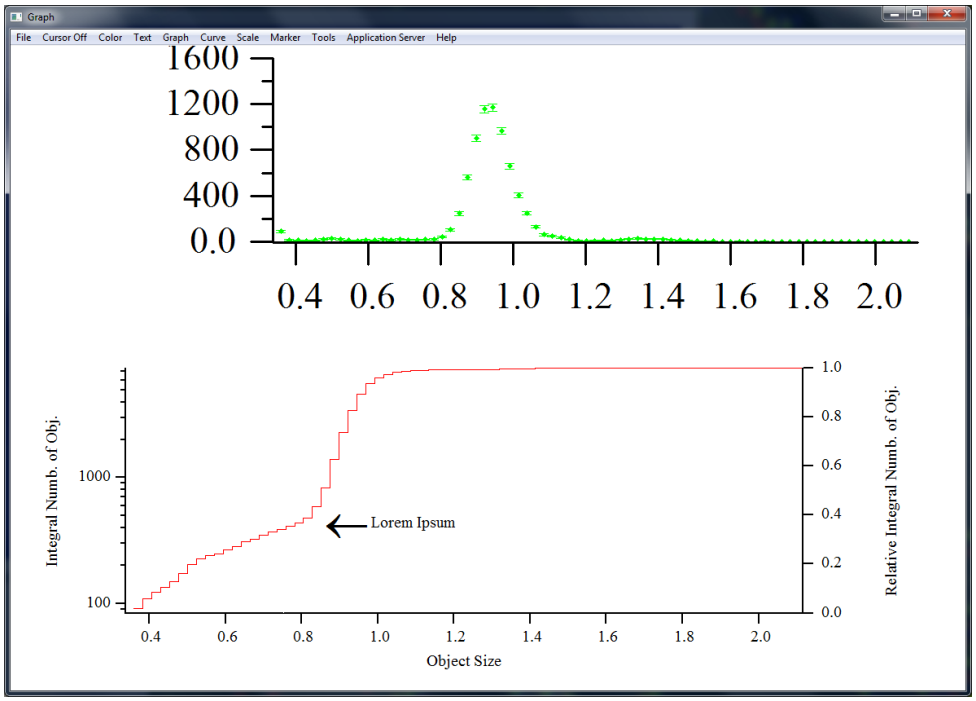


Figure 13: Plot of the distribution of the mean intensity of objects as well as an integrated (or summed up since it's discrete) version of the same plot. Note that the bottom graph has been labelled and resized for demonstation purposes.

want to calculate so that you can look for the functions you need. Usually the . Still, some important ones will be pointed out:

- **Error Bars On/Off** can be used to visualize the statistical errors of the underlying data. It is not possible to add manual error bars here.
- **Curve Manipulation and Conversion** contain all the options to move curves between graphs, duplicate them and convert them to other types of curve like histogram etc.
- **Curve Arithmetics** contain basic mathematical operations like addition and multiplication.
- **Analysis** contains all complex mathematical operations that MotionTracking can perform on curves on the fly, for exaple different types of deconvolution and fitting
- **Curve Set Operations** contain different methods to combine data from multiple curves like averaging.
- **Curve Transform** allows to integrate, differentiate, Fourier Transform etc. curves.
- **Curve Statistic** allows to estimate statistical properties of curves.
- **Distribution Analysis** used to calculate different parameters of given distributions.
- **Calculate** calculates and returns values of mean, SD, integral and others.

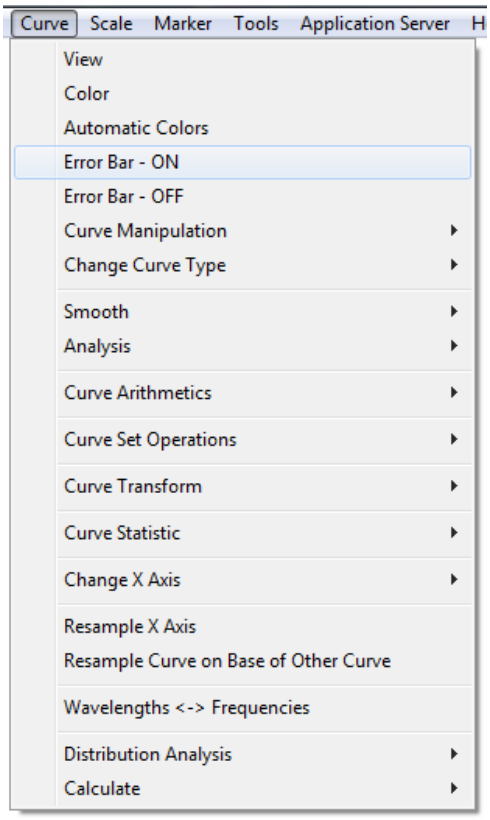


Figure 14: Curves submenu.

3 Statistic Characteristics and Parameters

The following sections will provide a brief overview over the available statistics for calculations. Note that some of them only exist in the Batch Statistic Window.

3.1 Common Statistical Parameters

-

3.2 Object Characteristics

- **Size**
- **Area**
- **Elongation**
- **Contrast**
- **Mean Intensity**
- **Maximum Intensity**
- **Integral Intensity**
- **Position Accuracy**

- **Speed** (Plus-End, Minus-End)
- **Distance to Center**
- **Volume** (of Nucleous)
- **Number of Neighbours**
- **Apical/Basal/Lateral Area**
- **Density**

3.3 Track Characteristics

-

3.4 Quality Control Parameters

nothing complicated

3.5 Shape Analysis

buggy

3.6 Other Statistical Parameters

rare

3.7 Example Statistic List