

Manual for SHERPA

SHERPA, the tool for "shape recognition, processing and analysis", offers an image processing workflow focused on identification and measurement of object outlines (see Fig. 1). As the name implies, SHERPA is a helpful assistant in surmounting huge piles of image data. Though it was developed focusing on analyzing diatom valves, SHERPA can also handle other classes of objects, so feel free to replace the term "valve" with your own object of interest whenever it is mentioned within this manual.

Starting point are micrographs obtained by optical microscopy or similar images. For each depicted object, the respective outline is detected and compared to a set of templates which characterize representative shapes, and a ranking is applied according to the chance of representing a relevant object. Selected results can be exported along with a series of descriptors for further morphometric scrutiny.

Please note that the function for measuring costae distance is experimental, works only on few species, and is not recommended for routine use.

SHERPA's user interface consists of three main areas:

- Loading template data
- Managing image data and adjusting the settings for image segmentation
- Detecting, matching and ranking the detected objects

These main areas are completed by functions for managing settings profiles, reviewing results and exporting data into CSV format.

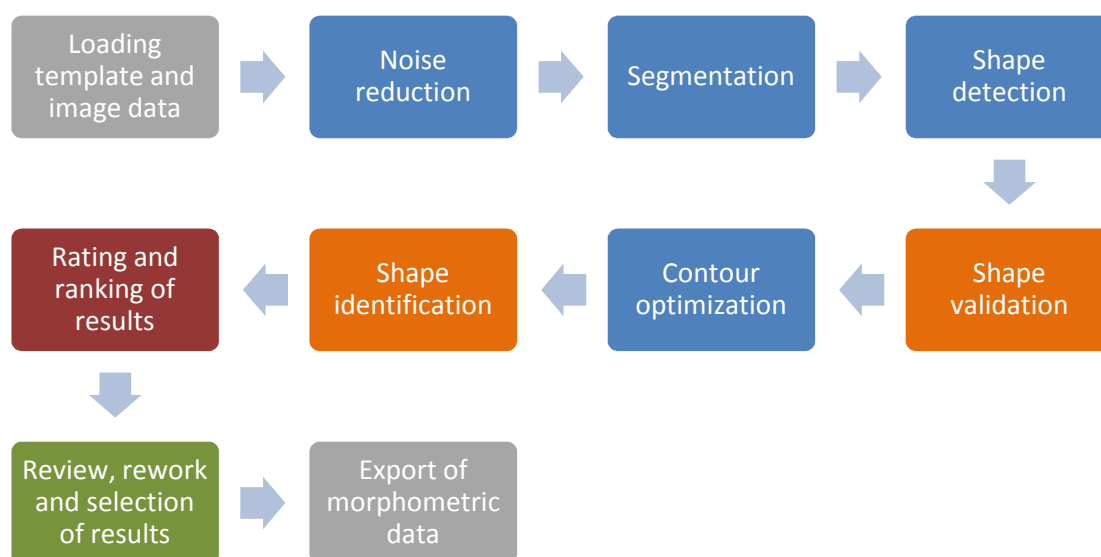


Fig. 1 Structure of SHERPA's image processing pipeline / workflow

For a fast overview of SHERPA's workflow, please read the "Quickstart Guide" linked within the "Help" menu. The "Straightforward Guide to Suitable Settings" (see the appendix at the end of this manual and also the "Help" menu) will explain the process of adapting the contour analysis parameters to your needs. For detailed explanations on the analysis process please consult the "Technical Information" document (also linked within the "Help" menu).

Content

1	Installation.....	4
2	Templates.....	4
3	Images	6
3.1	Segmentation procedures.....	7
3.2	Noise Reduction	7
3.3	Segmentation	8
3.3.1	Otsu's Thresholding.....	9
3.3.2	Histogram Equalization plus Otsu's Thresholding.....	9
3.3.3	RATS.....	10
3.3.4	Canny Edge Detector	11
3.3.5	Adaptive Thresholding.....	11
4	Objects.....	12
4.1	Analyzing data	12
4.1.1	Validation parameters for contours.....	12
4.1.2	Contour optimization	13
4.1.3	Shape identification.....	14
4.1.4	Matching, quality and convexity defect indicators used for ranking	14
4.1.5	A straightforward way to suitable settings	17
4.2	Working with results	17
4.2.1	Displaying shapes	17
4.2.2	Reworking contours manually.....	18
4.2.3	Exporting results.....	19
4.3	Other settings.....	20
4.3.1	Essential parameters: Micrometer factor	20
4.3.2	Essential parameters: Number of harmonies for elliptic Fourier analysis	20
4.3.3	Segmentation and optimization settings	20
4.3.4	Settings for displaying / exporting results.....	20
4.4	Settings profiles.....	21
5	Appendix: A straightforward way to suitable settings	22

1 Installation

SHERPA was developed for Windows7 64 Bit. In some cases “Microsoft .NET Framework 4” and the “Microsoft Visual C++ 2010 SP1 Redistributable Package (x64)” have to be installed prior to running SHERPA. Please consult the file “README.txt” for information on this subject and read the disclaimer.

2 Templates

Objects are identified based on shape information. Prototypes of relevant shapes are stored as image files of the formats TIFF, JPG, PNG or BMP. These contain outlines of each suitable object type (only one template per file!) as a black silhouette on a white background (see Fig. 2). A broad collection of templates depicting diatom valves is provided along with SHERPA, but for good results you need to create a set of templates depicting the morphological variability of the objects you are interested in. Usually several templates are needed to cover the range of shapes corresponding to one type (see Fig. 2 e-g). New templates can easily be created with a standard drawing program or by using the appropriate function within SHERPA (see Chap. 4.2.1).

Please note: Since templates are matched to object shapes by using elliptic Fourier analysis (see Chap. 4.1.3), the identification process is insensitive to size, rotation and position. But it is not invariant to mirroring, so for objects which do not have symmetry with respect to an axis, two templates need to be used (see Fig. 2 b-c).

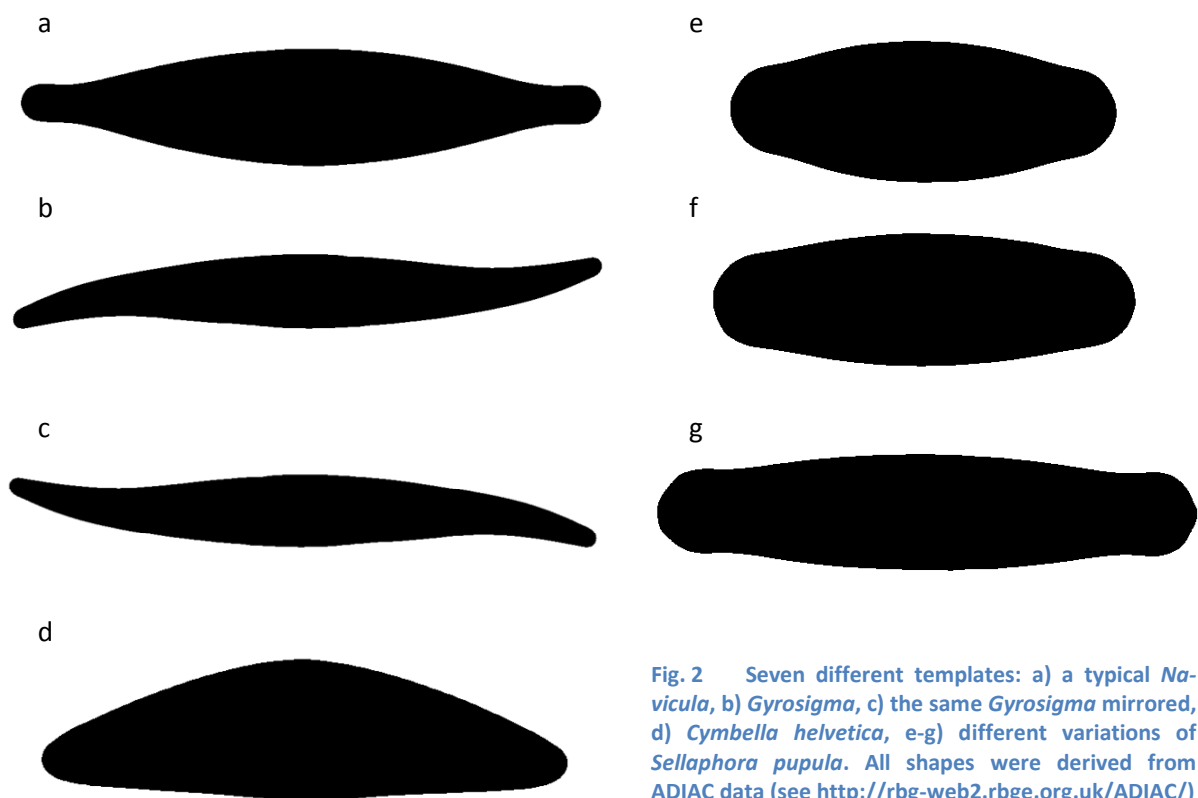


Fig. 2 Seven different templates: a) a typical *Na-vicula*, b) *Gyrosigma*, c) the same *Gyrosigma* mirrored, d) *Cymbella helvetica*, e-g) different variations of *Sellaphora pupula*. All shapes were derived from ADIAC data (see <http://rbg-web2.rbge.org.uk/ADIAC/>)

Using the main menu, templates can be loaded from selected files within one folder, a complete folder, or a complete folder plus its subfolders. This way, a structured library of object types can be organized in a simple and effective way.

A set of descriptors is calculated for each template (see Tab. 1 & Tab. 2). These descriptors (see Fig. 3) are used for validation, identification and ranking of the objects which were detected during image processing. Values which are outside the ranges set for contour validation are highlighted in

dark red (please refer to Chap. 4.1.1 for detailed information on the validation process). This is an indication that the validation settings in the “Objects” area might not be correct, at least by not corresponding to the selected templates, because objects exactly matching this template would not pass validation and thus produce any result at all, even though they might be relevant. Values highlighted in light red do not coincide with the quality settings for contours (see Chap. 4.1.4) and would result in a ranking downgrade.

File	Folder	Area / μm^2	Perimeter / μm	Width / μm	Height / μm	Width/Height Ratio	Smooth- ness	CDF	PCAF	CHMDF	Template is convex	Show
Achnanthes delicatula.895.tif	C:\Users\mkloster\DiatoMorphoTo\Templates-Invertiert-geglättet\ADIAC\Achnanthes delicatula.895.tif	146.58	50.19	19.90	10.32	1.909	0.996	1.033	1.02	2.48	<input checked="" type="checkbox"/>	Show
Achnanthes lanceolata.126.tif	C:\Users\mkloster\DiatoMorphoTo\Templates-Invertiert-geglättet\ADIAC\Achnanthes lanceolata.126.tif	153.72	52.76	21.23	9.53	2.229	0.993	1.034	1.02	0.00	<input checked="" type="checkbox"/>	Show
Achnanthes lanceolata.132.tif	C:\Users\mkloster\DiatoMorphoTo\Templates-Invertiert-geglättet\ADIAC\Achnanthes lanceolata.132.tif	185.55	62.18	26.24	9.14	2.871	0.999	1.030	0.89	0.00	<input checked="" type="checkbox"/>	Show
Achnanthes linkiei.842.tif	C:\Users\mkloster\DiatoMorphoTo\Templates-Invertiert-geglättet\ADIAC\Achnanthes linkiei.842.tif	478.40	99.41	41.43	18.48	2.242	0.997	1.061	5.84	576.71	<input type="checkbox"/>	Show
Achnanthes pseudogroenlandica.945.tif	C:\Users\mkloster\DiatoMorphoTo\Templates-Invertiert-geglättet\ADIAC\Achnanthes pseudogroenlandica.945.tif	235.13	85.05	38.27	7.25	5.278	0.996	1.029	2.21	15.81	<input type="checkbox"/>	Show
Achnanthes saxonica.335.tif	C:\Users\mkloster\DiatoMorphoTo\Templates-Invertiert-geglättet\ADIAC\Achnanthes saxonica.335.tif	142.00	51.09	20.72	8.29	2.500	0.997	1.027	0.95	0.00	<input checked="" type="checkbox"/>	Show
Achnanthes saxonica.338.tif	C:\Users\mkloster\DiatoMorphoTo\Templates-Invertiert-geglättet\ADIAC\Achnanthes saxonica.338.tif	114.61	42.92	16.34	8.60	1.900	0.999	1.028	0.84	0.00	<input checked="" type="checkbox"/>	Show
Achnanthes saxonica.339.tif	C:\Users\mkloster\DiatoMorphoTo\Templates-Invertiert-geglättet\ADIAC\Achnanthes saxonica.339.tif	244.59	72.59	30.92	9.60	3.219	0.995	1.026	0.95	0.00	<input checked="" type="checkbox"/>	Show

455 template file(s) loaded

Fig. 3 Area „Templates“, showing features of loaded templates

Clicking the “Show” button or double clicking a row of data will open a window displaying the template image (see Fig. 4). A click with the right mouse button onto the image will open a context menu with functions for zooming and saving the displayed image into a new file. Zooming can also be accomplished by turning the mouse wheel or by using the “+” and “-” keys. Panning can be done by using the scrollbars or the cursor keys or by turning the mouse wheel whilst the “SHIFT” or “ALT” key is pressed. These functions are usually available whenever an image is displayed in SHERPA.

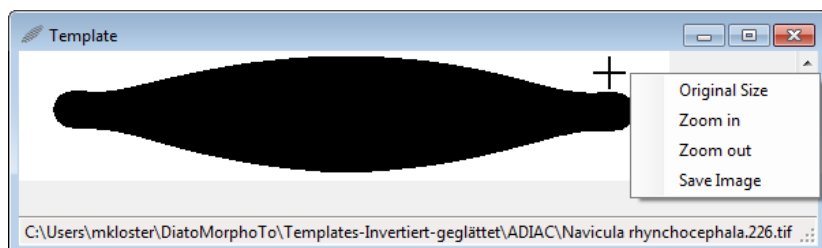


Fig. 4 Window showing a template image and its context menu

3 Images

SHERPA works for images depicting objects as dark structures on bright background (like obtained e.g. using brightfield microscopy) or as bright structures on dark background (like obtained e.g. using darkfield microscopy). The type of image data has to be configured using “Images → Configure Image Type” from the main menu. Objects imaged using phase or differential interference contrast (PhaCo resp. DIC) might not be measured correctly due to halo or relief artifacts. For proper results, object outlines should be focused as completely as possible. Minor blurring will affect accuracy of outline detection, while extensive fuzziness might impede usable results. The sample density should be sparse with objects isolated and clearly separable for a good identification yield.

Images to be analyzed can be organized and loaded the same way template files are (see Fig. 5). When loading complete folders, certain files can be selected using filename wildcards (e.g. using the wildcard “*01.tif” would load only files of type “TIF” whose names end on “01”).

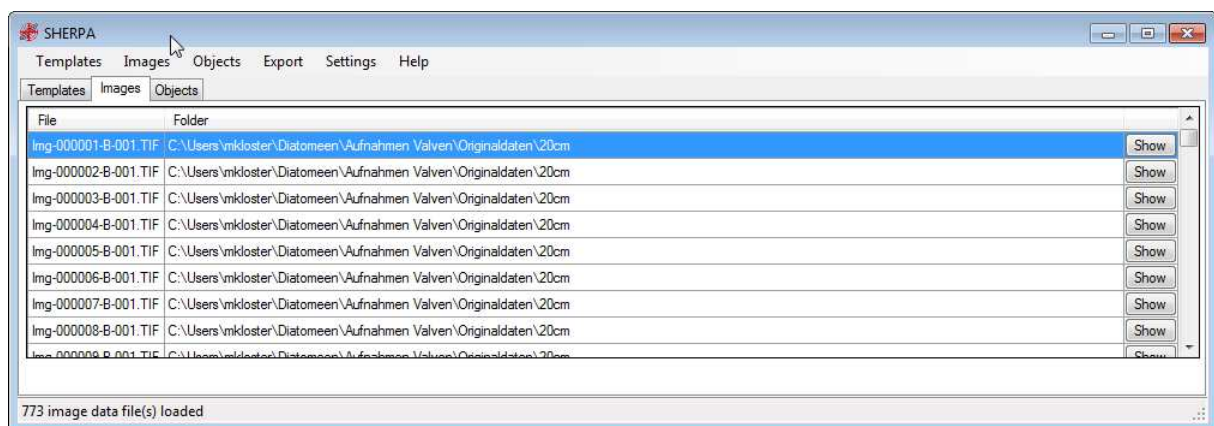


Fig. 5 Area “Images” showing the list of image files to be analyzed

Just like in the “Templates” area, clicking the “Show” button or double clicking a row of data will display the corresponding image (see Fig. 6). The buttons at the window’s bottom will open dialogs for adjusting and testing the segmentation procedures, which are used to separate objects from image background.



Fig. 6 Image window showing a brightfield micrograph of a valve and buttons for testing segmentation procedures

3.1 Segmentation procedures

Up to five different procedures employing four common segmentation algorithms can be used simultaneously. This allows for an increased yield of detected objects, since each of these approaches presents its own advantages and disadvantages depending on data quality and thus may successfully segment data where the other procedures fail. Subsequently several morphological operators can be applied for repairing minor segmentation faults (see Chap. 4.1.1). Only the one segmentation result which produces the best ranking for an object (i.e. seems to provide the best result, see Chap. 4.1.4) is taken into consideration for further analysis. For results having identical ranking values, the contour which matches better to a template is taken into account. So, despite of using multiple procedures, always only one single shape will result for each detected object. Two shapes are considered as belonging to the same object if the centroid of one shape lies within the area of the other. Segmentation and optimization procedures to be applied during image analysis can be set within the “Objects” area or via the main menu (see Chap. 4.3.3).

Settings for each procedure should be adapted so that object outlines are segmented as well as possible without gaps, indentations or bulges, and separated from other objects. This customization can be done using the test buttons provided when any of the images to be analyzed is displayed (see above and Fig. 6) by clicking on its “Show” button. Especially poor image contrast, objects lying partly out of focus, or debris and other objects superimposing them, might cause segmentation problems, but in many cases at least one procedure succeeds. Shapes that are considerably distorted by objects which are overlapping or merged due to close proximity cannot be identified correctly, whilst minor indentations and bulges of the outline can be repaired using contour optimization (see Chap. 4.1.1). The often abundant small debris particles will be rejected by contour validation (see Chap. 4.1.1).

When a settings dialog is closed by clicking “OK”, its parameters are stored for the image processing pipeline’s segmentation step. Cancelling the dialog will restore the former settings.

3.2 Noise Reduction

Each segmentation procedure implements a Gaussian and a median filter for noise reduction. Gaussian filtering blurs the image data (and thus the object outlines), whilst median filtering preserves object edges better but tends to homogenize fine structures. The filter diameter should not be too large to prevent distortion of the outlines; the current settings’ effect can be inspected by clicking the “Noise Reduction” button (see Fig. 10). Hints on reasonable values are given via a “tooltip” which will show up when hovering with the mouse cursor above the input field; this applies to each configurable parameter anywhere in SHERPA.

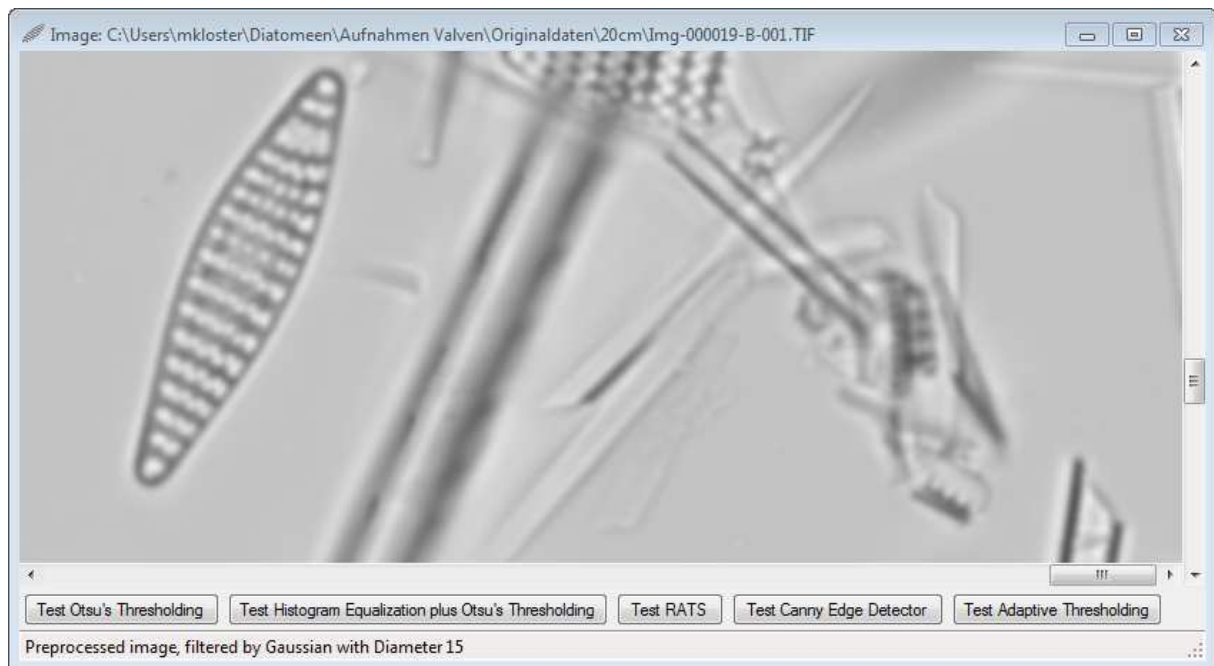


Fig. 7 Effect of denoising using a Gaussian filter (compare to Fig. 6). The filter diameter was exaggerated to illustrate the blurring effect.

3.3 Segmentation

The segmentation process is an essential part of the SHERPA image processing pipeline. It separates objects from image background by applying a threshold to each pixel. As a result, objects are pictured in white on a black background (see Fig. 8). Only the outline of each white object is inspected during contour analysis, so an incomplete segmentation of an object's interior is negligible as long as the outline is correct. The segmented result can be viewed by clicking the "Segmentation" button (see Fig. 10).

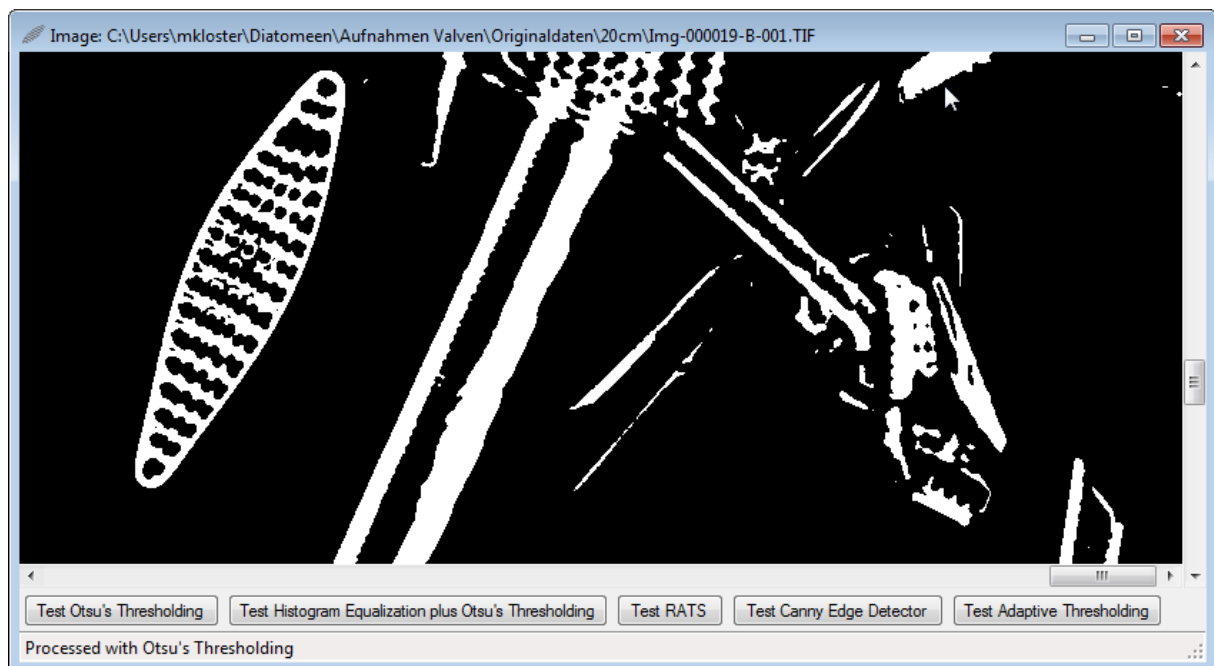


Fig. 8 Typical segmentation result (compare to Fig. 6). Detected objects are colored white on black background, a successfully segmented diatom valve can be seen on the left side.

Outlines derived from a segmentation result can be viewed by clicking the "Contour Detection" button (see Fig. 10), which will show each contour in a different color (see Fig. 9).



Fig. 9 Typical result of contour detection. The diverse detected objects are highlighted in different colors.

3.3.1 Otsu's Thresholding

Otsu's thresholding identifies objects by analyzing the global intensity histogram and calculating a threshold which separates them from image background. This often gives good results, even though typical intensity histograms found in brightfield images of diatom valves do not show the bimodal distribution usually mentioned as a precondition for applying Otsu's thresholding. Since no additional parameters need to be configured, this algorithm is a good choice for unattended segmentation (see Fig. 10).

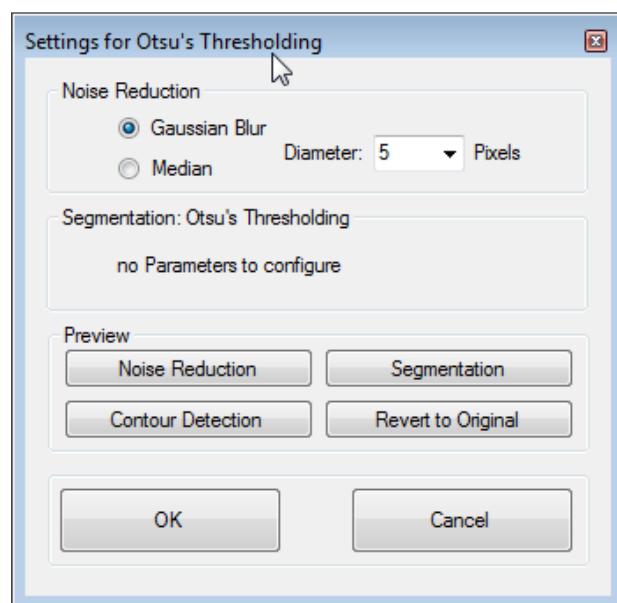


Fig. 10 Settings dialog for Otsu's thresholding

3.3.2 Histogram Equalization plus Otsu's Thresholding

For this procedure Otsu's thresholding is preceded by a step of histogram equalization, which tries to achieve a balanced distribution of image intensities. This enhances contrast and enables for segmentation of data with inadequate exposure. The settings dialog looks similar to that of Otsu's thresholding (see Fig. 10).

3.3.3 RATS

The “Robust Automated Threshold Selector” (RATS) calculates intensity thresholds for each pixel by examining its local neighborhood and weighing this according to its gradient magnitudes to improve edge detection.

The RATS algorithm requires two parameters, the sigma for calculating the intensity gradient and an exponent which describes the influence of the gradient magnitudes on the threshold calculation (see Fig. 11).

RATS runs quite slowly compared to the other procedures, but it might provide a successful segmentation where the other algorithms fail. To enhance this ability, not only a single but a whole range of sigmas can be applied. The sigmas from start to stop value will be used for calculation in steps of one, so a setting of “1 to 10” will result in running RATS ten times using sigmas respectively. For calculating previews using the “Segmentation” or “Contour Detection” button, but only a single sigma value will be used (the one which was edited or clicked at most recently).

Due to its slow execution speed, it is advisable to test if additional usage of RATS really yields a significant increase in usable results. For samples with adequate contrast and sparse object density usually the other four segmentation procedures will suffice.

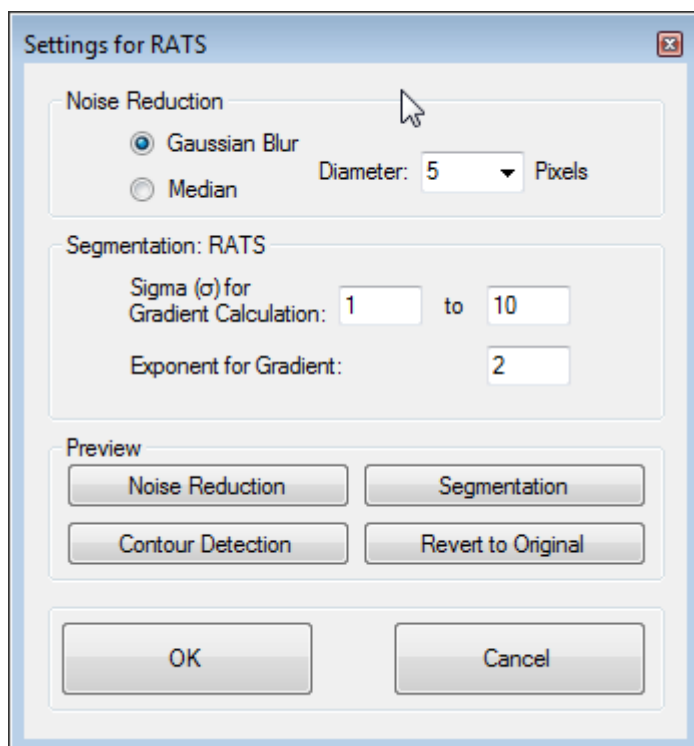
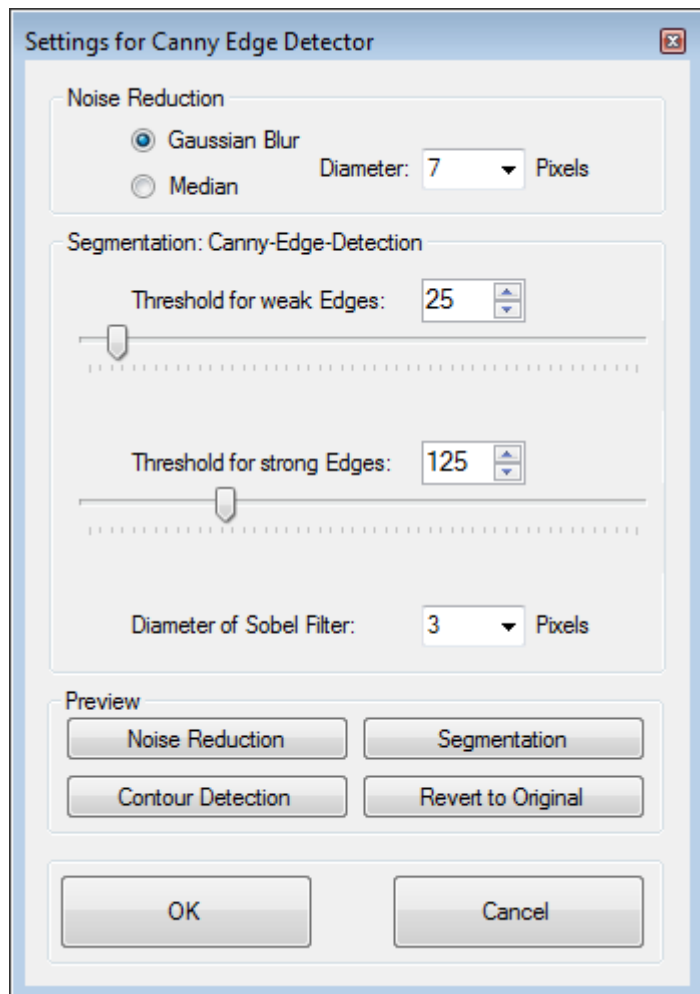


Fig. 11 Settings dialog for RATS

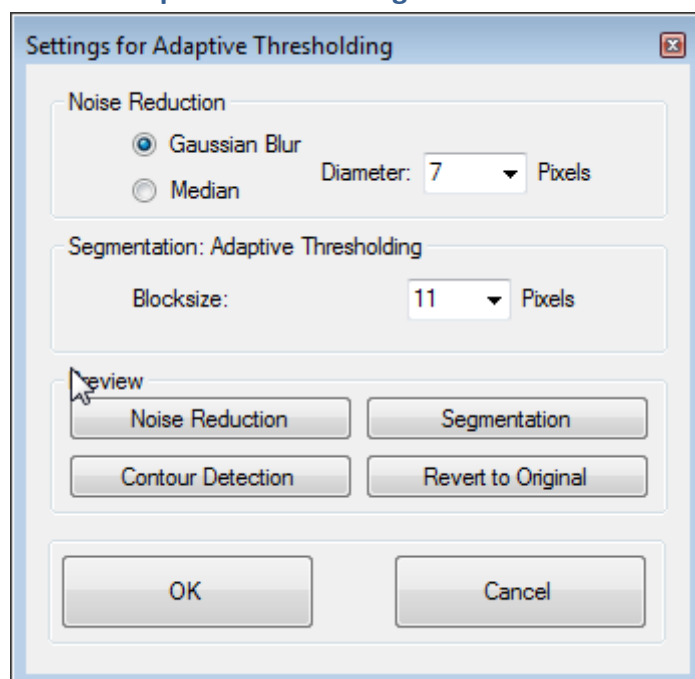
3.3.4 Canny Edge Detector

The Canny edge detector (see Fig. 12) tries to identify object edges by calculating intensity gradients using a sobel filter and generating one pixel wide lines along the highest gradient magnitudes (strong edges). These are subsequently elongated along lower gradient magnitudes (weak edges). The threshold for strong edges should be about three to four times the threshold for weak edges.

Fig. 12 Settings dialog for Canny edge detector



3.3.5 Adaptive Thresholding



Adaptive Thresholding calculates a local intensity threshold depending on the local neighborhood of each pixel, which is weighted using a Gaussian distribution. The parameter "Blocksize" represents the neighborhood diameter (see Fig. 13). Sometimes this procedure tends to segment not the real object boundary but a surrounding halo, so please check the results to decide if this procedure should be applied.

Fig. 13 Settings dialog for Adaptive Thresholding

4 Objects

The main control center for detection and identification of objects offers a variety of parameters to be configured (see Fig. 14). Don't be confused by their large number, the approach to obtain suitable settings is straightforward (see the appendix at the end of this manual), and the default values will usually work with only slight adaption.

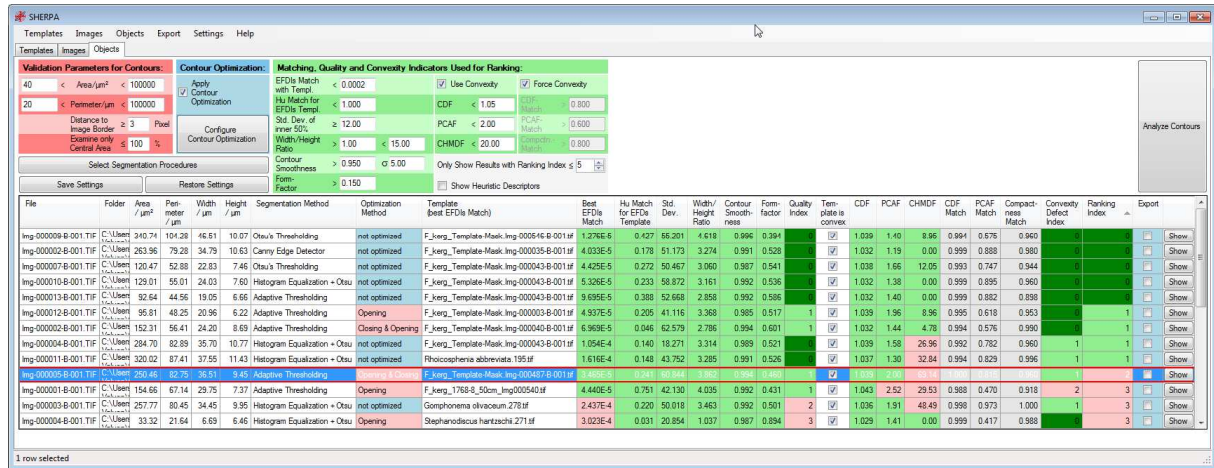


Fig. 14 The „ Objects“ area showing various settings and a list of analysis results.

4.1 Analyzing data

Data analysis is the central function of SHERPA's image processing pipeline. After image segmentation was executed, an extensive number of shapes will be validated, identified and rated. For each of these steps you need to configure a number of parameters with settings appropriate to your particular data set and type of investigation. Analysis results are computed or updated by clicking the "Analyze Contours" button or the appropriate entry from the "Objects" main menu.

4.1.1 Validation parameters for contours

Each image segmentation can result in hundreds or even thousands of separate objects. Most of these do definitely not depict relevant objects, and should be sorted out by applying validation parameters (see Tab. 1 and portions of Fig. 15 highlighted in red). Only objects with an area and perimeter within a reasonable size are processed further to increase analysis speed; also objects which are close to the image boundary can be excluded from further analysis, because most likely they are truncated by the camera's field of view. Last, if images are known to be centered on objects of interest, further analysis can be restricted to objects lying in a central user-specified portion of the image.

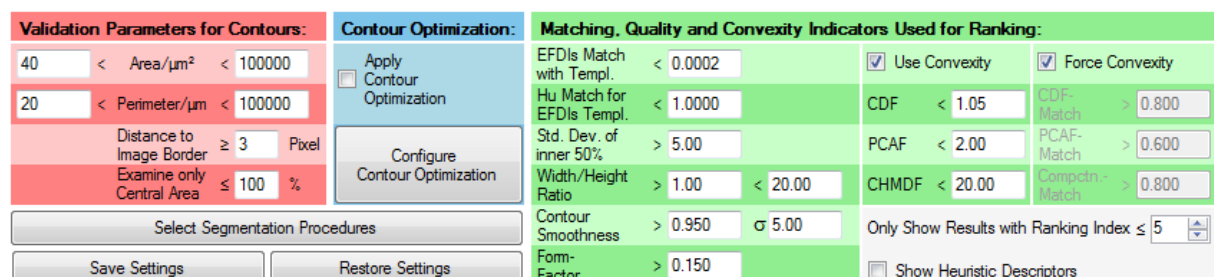


Fig. 15 Settings for contour validation, optimization, identification and ranking. Validation parameters are highlighted red, optimization settings blue and ranking indicators (see Chap. 4.1.4) green.

Objects being too large are rare and therefore don't reduce analysis speed very much, so the maximum size for area and perimeter can be set to a fairly high value. But usually a lot of objects are detected which are too small to be of interest, most of them being debris or dust speckles. The analysis

process can speed up drastically if adequate minimum sizes for area and perimeter are set. Please adjust these values so that the smallest objects relevant for your study are still detected. Setting minimum size values too low does only extend the time needed for contour analysis, it will not affect result quality. Setting the minimum value too high however might reject relevant objects from analysis, so be careful to choose an appropriate value. Also the templates are verified according to the size settings; values which do not match validation parameters will be highlighted dark red in the “Template” area to indicate this problem.

Sizes are defined in μm resp. μm^2 , the ratio between pixel size and μm can be configured via “Settings” in the main menu (see Chap. 4.3.1)

Tab. 1 Parameters used for contour validation

Area	An object’s area must be larger resp. smaller than a certain size.
Perimeter	The perimeter of an object must be larger resp. smaller than a certain size.
Distance to Image Border	The distance between object and image border must be equal or larger than this. Due to the algorithm used for contour detection, this distance measures at least one pixel.
Examine only Central Area	Same as “Distance to Image Border”, but not measured in absolute pixels but in percent of image width / height.

4.1.2 Contour optimization

Due to debris, overlapping structures, damages or segmentation faults, not all objects can be segmented successfully. To increase the yield of usable results, contours can be “repaired” by applying morphological operators “Opening”, “Closing” and combinations of these two. Small indentations and bulges can be removed this way and the yield of usable results can increase significantly. But this will also reduce the accuracy of the derived outlines and the reliability of convexity defect measures for identifying segmentation faults (see Chap. 4.1.4), so if shape accuracy is more important than identification rate, contour optimization should be switched off.

Clicking the button “Configure Optimization Settings” or the appropriate item in the “Objects” main menu, or double-clicking a result’s “Optimization Method” column will open a dialog (see Fig. 16) where the applicable operators can be chosen and configured. For all optimization methods the same parameter settings for the number of iterations and the diameter of the disk shaped structuring element are used. Increasing these values enlarges the size of defects which can be repaired, but will also apply a stronger smoothing effect on the contours, which might affect important shape features.

Depending on the type of contour defects, each of the four optimization methods will show different effects. So if you decided to apply contour optimization at all, selecting all four of them usually is a good choice, even though this will extend the time needed for contour analysis drastically. If you have a lot of small debris overlapping the objects, at least “Opening” should be selected. Just like with using multiple segmentation procedures, only the optimization result which produces the best ranking (see Chap. 4.1.4) will be used for contour analysis, the results of the other morphological operators will be ignored.

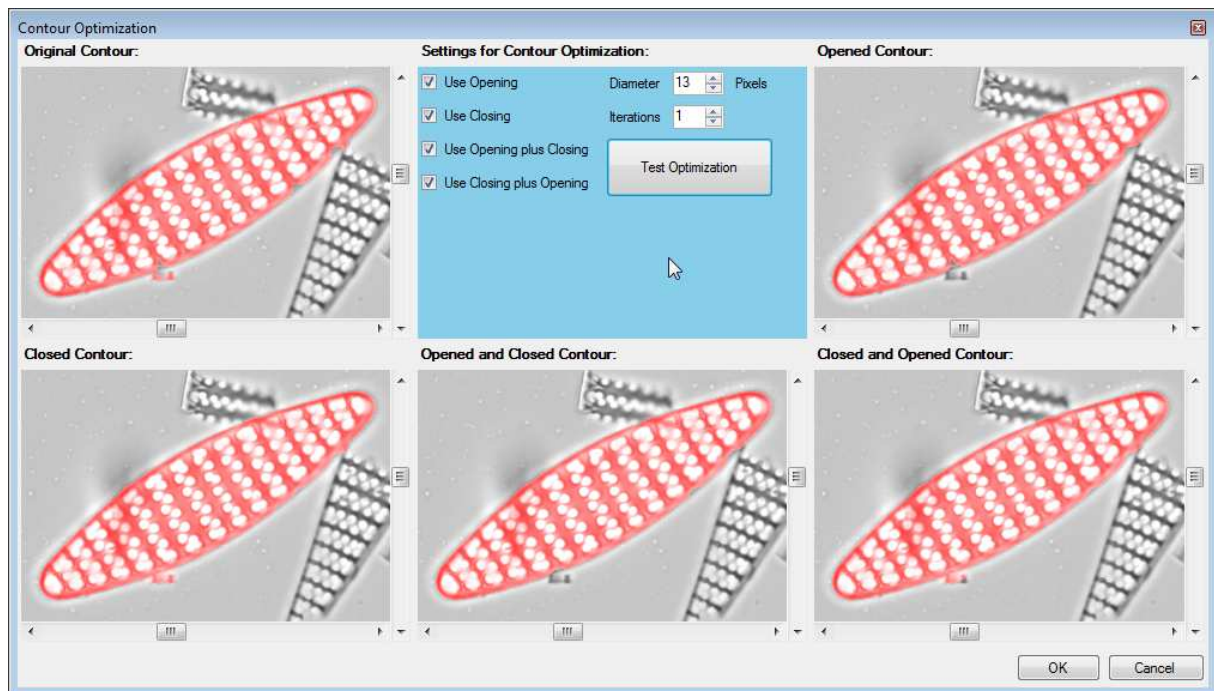


Fig. 16 Dialog for configuring contour optimization. Effects of the four optimization methods will be displayed simultaneously, zooming and panning of the contour images is synchronized.

4.1.3 Shape identification

Objects are identified by comparing their shape with templates, based on elliptic Fourier analysis. Invariants of elliptic Fourier descriptors are matched by summing up their squared differences; the number of harmonics employed can be configured using the “Settings” menu (see Chap.4.3.2). A comprehensive assortment of templates appropriate to your data and type of examination is essential for good performance of SHERPA. New templates can easily be created using the built in function (see Chap. 4.2.1).

4.1.4 Matching, quality and convexity defect indicators used for ranking

After validation, detected objects are rated using the parameters highlighted in green (see Fig. 15 and Tab. 2 - Tab. 4). Each feature which conforms to these settings serves as a positive indicator and improves the ranking. The better the ranking of an object (i.e. the smaller its ranking-value), the higher is the chance for it to depict a relevant shape.

A set of seven parameters is used to rate the matching between a contour and its best fitting template, to estimate the inhomogeneity of the object texture and to judge some geometrical features of the shape (see Tab. 2).

Settings for “Standard Deviation” should be adjusted according to image contrast, settings for “Width / Height Ratio” should cover the scope of relevant objects. However, inappropriate settings will only affect ranking, so you just might have to do a little bit more manual review when suboptimal settings are applied.

Tab. 2 Matching and quality indicators used for ranking

EFDIs Match with Template	Matching between elliptic Fourier descriptor invariants (EFDIs) of object and template shape. The lower the value, the better is the similarity between the two. The number of harmonics to be used for elliptic Fourier analysis can be defined via “Settings” in the main menu. Failure of this indicator causes a penalty twice as high as for the other indicators.
Optimization Method	Morphological Operator used to improve the object contour. If an optimization was applied to derive a shape, its ranking is degraded, because the resulting contour might be inaccurate.
Hu Match for EFDIs Template	Matching between the Hu invariants of the object and the template which matches best according to EFDIs.
Standard Deviation of inner 50%	Standard deviation of the intensity distribution within the object boundaries. Only the inner 50% of the area is analyzed. This way, diatom valves (normally containing striae / costae / areolae) can be distinguished from empty girdle bands (which can produce good outline matching, but have a homogenous interior). If you are not interested in the object’s texture, set this parameter to zero.
Width / Height Ratio	Ratio between object width and height. Usually objects of a certain type have a ratio within a certain range.
Contour Smoothness	Estimation of the object contour smoothness. Contours distorted due to segmentation inaccuracies or failures can be identified this way, because usually they are quite rough. This is accomplished by comparing the outline perimeter to that of the outline smoothed by a Gaussian filter of width σ .
Formfactor	Heuristic descriptor “formfactor”. $\text{Formfactor} = 4\pi \cdot \text{area} / \text{perimeter}^2$

One major decision that must be made is how to utilize convexity defect measures. This is especially important when you are investigating diatoms, depending on if you are looking for convex or concave valves or a mixture of them. Convexity defect measures are evaluated based on differences between a contour and its convex hull (i.e. the smallest area which encloses the contour without containing any concave parts). If only convex shapes are of interest, convexity defect measures (see Tab. 3) are excellent features to decide about segmentation quality. This works because in convex shapes even small indentations or bulges will cause noticeable concave parts within the outline, which significantly increase the convexity defect measures. But this approach will not work with objects whose shapes naturally contain concave parts.

If only convex objects are important, choose “Force Convexity” for contour analysis. This way, only convex templates are taken into consideration, segmentation faults are detected clearly, and segmentation quality can be judged quite precisely based on absolute values of the convexity defect measures (see Tab. 3).

If your data contains convex as well as concave objects, “Use Convexity” can be chosen. In this case, convexity defect measures are judged directly (like when using “Force Convexity”) only if the best matching template is convex. Otherwise, convexity defect measures plus the heuristic descriptor “compactness” will be compared between the object and its best matching template (see Tab. 4). Please note: this method can bias the ranking because for convex objects even small segmentation flaws will cause a significantly ranking downgrade, whilst for concave objects detection of segmentations flaws will be much less sensitive and thus cause a less significant ranking downgrade. To prevent this bias to hamper your results you will need to check the results at least up to ranking 4 and sort out any unwanted data. Also note: if no template is provided which matches the object sufficiently, a convex template might be the best (though not a really good) match for a not convex

shape. This will usually result in a really bad ranking because not only the matching quality but also the convexity defect measures will fail. So when using the mixed “Use Convexity” approach, make sure the morphological variability of the objects you are interested in is depicted by an adequate amount of templates. This usually is a good idea for any contour analysis, but especially in this case it is even more important.

If neither “Use Convexity” nor “Force Convexity” is selected, only comparison of convexity defect measures between object and template plus an evaluation of the formfactor takes place, regardless if the best matching template is convex or concave. This is usually a good choice if you do not know in advance if all relevant objects are convex or you don’t have an extensive library of templates yet, but notice that the detection of segmentation flaws is less sensitive this way.

If you are interested in convex objects only, choosing “Force Convexity” will provide a more precise ranking and might save you from a lot of manual reviewing.

Tab. 3 Absolute convexity defect measures used for ranking

CDF	“Convexity Defection Factor”, depicts the percentaged difference between area resp. perimeter of contour and convex hull
PCAF	The “Percent Concave Area Fraction” compares the areas of contour and convex hull.
CHMDF	For the “Convex Hull Maximum Distance Factor” each convexity defect’s maximum distance between contour and convex hull is calculated. For the distances larger than $\sqrt{2} \cdot \text{pixelwidth}$ the squares of the distances are summed up to the CHMDF.

Tab. 4 Relative convexity defect measures used for ranking

CDF-Match	Compares CDF of object and template
PCAF-Match	Compares PCAF of object and template
Compactness-Match	Comparison of the heuristic descriptor “compactness” between object and template. Compactness = $\sqrt{\frac{4 \cdot \text{area}}{\pi}} / \text{max. diameter}$

Please reconsider if you want to use contour optimization to remove small indentations and bulges from the contour. This can increase the yield of usable results significantly, but will also reduce the reliability of convexity defect measures for identifying segmentation faults due to a smoothing effect, so if shape accuracy is more important than identification rate, contour optimization should be switched off.

Results for all quality, convexity and matching indicators are displayed in the contour data table: fulfilled indicators are highlighted green, failed indicators red and unused indicators gray. All quality indicators sum up to the quality index, all convexity defect indicators to the convexity index, the ranking index is the sum of both. The lower the ranking index, the higher is the chance that a shape depicts a relevant object with good segmentation quality. An excellent ranking index of zero is highlighted in dark green, a good value of one in light green. Depending on the available templates and the utilization of convexity defect measures, results of ranking index zero usually are excellent, whilst results of ranking index one might need manual inspection (especially when contour optimization was applied), but most likely represent good results with only minor segmentation faults. Results having higher ranking indices demand manual verification. Failure of the indicator “EFDIs Match with Template” causes a penalty twice as high as for the other indicators. Accordingly, for objects which

do not have a closely matching template, adding an appropriate template and repeating contour analysis will help to improve their ranking significantly.

You can sort the results list by clicking on the appropriate column header, clicking the header again reverses the sorting direction. Usually you would sort by column “Ranking Index” in ascending order.

4.1.5 A straightforward way to suitable settings

Please refer to the appendix at the end of this manual for a procedure to optimize settings for the matching, quality and convexity defect indicators according to your own data and investigation.

4.2 Working with results

After contour analysis is completed, you usually need to check the results and export data for further inspection or analyses using other tools than SHERPA. Results can be narrowed down by skipping entries of insufficient quality (“Only Show Results with Ranking Index \leq ”) and by sorting the list according to the appropriate column, e.g. “Template” or “Area”.

Sorting by “Ranking Index” works best (i.e. sorts according to segmentation quality and chance of depicting a relevant object) if an appropriate set of templates was used. If you are interested in convex objects only (using the “Force Convexity” feature), sorting by “CHMDF” or “Convexity Defect Index” will show results with good segmentation quality first, even if appropriate templates are missing.

4.2.1 Displaying shapes

By double clicking onto different columns of the results list, the appropriate data is shown, i.e. the underlying image data, the best matching template or the coefficients / invariants of the elliptic Fourier analysis. Clicking checkboxes in the blue “Export” column marks / unmarks results for export, clicking with the right mouse button onto this column opens a context menu offering functions for marking / unmarking multiple results. Double clicking any other data column will open a window where the contour data is visualized, as does the “Show” button (see Fig. 17).

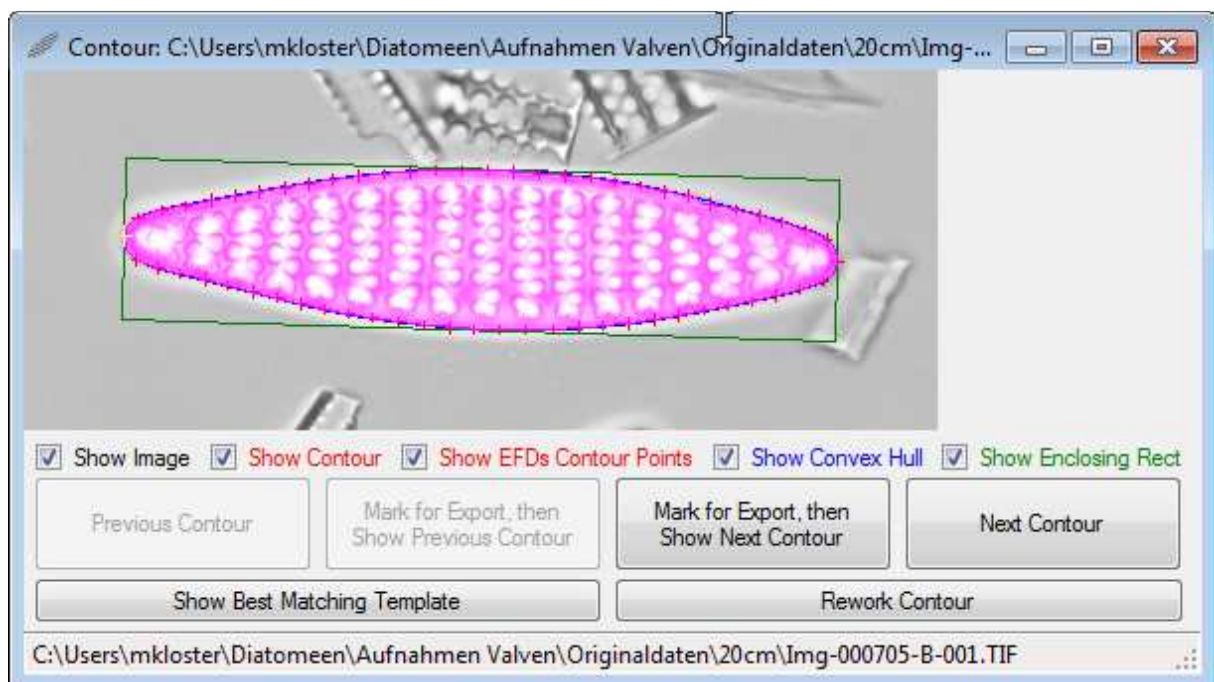


Fig. 17 Contour visualization. Diverse properties of the shape are displayed using different colors.

The contour visualization window (see Fig. 17) offers a set of functions for displaying and managing results. The detected shape is highlighted in red, its convex hull (the smallest area which encloses the contour and does not have any concave parts) is highlighted in blue, the enclosing rectangle (which describes width and height of the contour along its major / minor axis) is drawn in green. The normalized contour refers to the set points for elliptic Fourier analysis and is displayed as a series of red crosses, which are about equally spaced (depending on the curvature of the shape). The starting point for Fourier analysis is the leftmost point with respect to the major axis and is marked yellow. Below the visualization, some checkboxes allow to configure which of these features will be displayed. Pressing the “SPACE” key will blank out all features but the original image data.

The right-click-menu of the image area allows for saving a shape as template file. It is recommendable to rework and smooth the contour prior to doing so (see Chap. 4.2.2), especially when it is of convex forming and “Enable / Force Convexity” might be used for analysis. New templates will only be taken into account after they are loaded within the “Templates” area and contour analysis is restarted.

Clicking on “Next / Previous Contour” or turning the mouse wheel within the lower part of the window (where the buttons are placed) scrolls through the results list. Using these buttons while pressing down the “SHIFT” key or using the “Mark for Export, then Show Next / Previous Contour” also marks a contour for export.

The button “Show Template” displays the best matching template, “Rework Contour” opens a window providing functions for manual contour correction.

4.2.2 Reworking contours manually

To increase the yield of usable results, shapes distorted by segmentation faults can be reworked manually (see Fig. 18). This is accomplished by “painting” within the object area (highlighted red) using the mouse. When entering the image, the mouse cursor changes into a blue painting cursor which acts as a brush / eraser. A click with the left mouse button will add, the right button will erase the area below the cursor to / from the shape. Whilst pressing the “CTRL” key, the underlying image will be blanked out, so you can see the painted contour more clearly. Pressing the “SPACE” key will blank out the red contour image, so you can see the object more clearly. Zooming is helpful for precision work and can be accomplished using the “+” and “-” keys or the mouse wheel. Panning can be performed using the scrollbars or the cursor keys, plus using the mouse wheel whilst the “SHIFT” or “ALT” key is pressed.

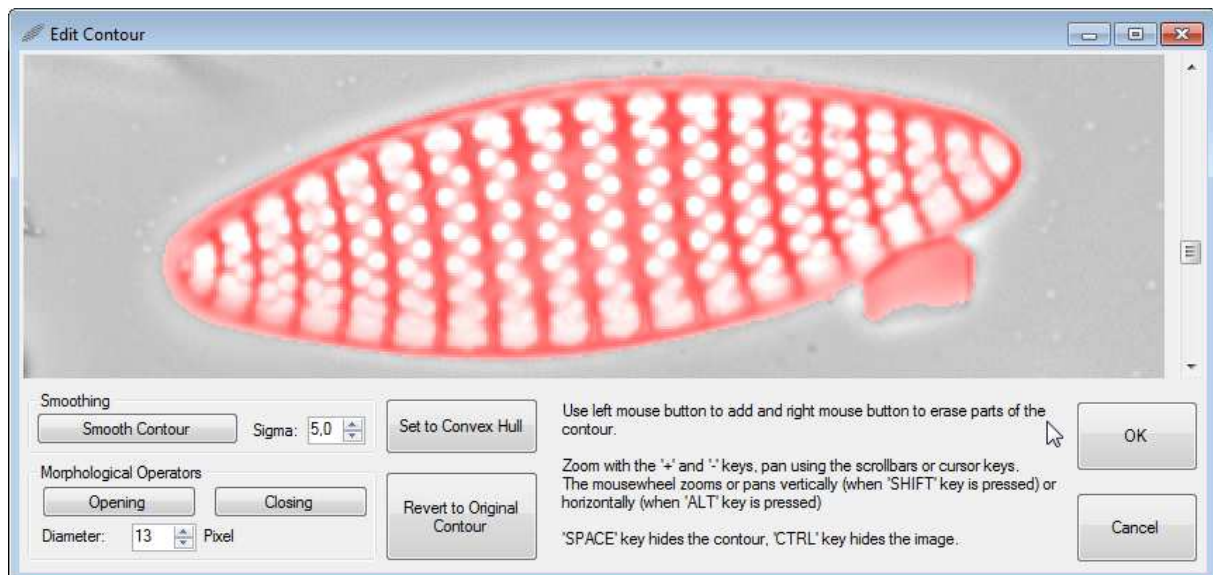


Fig. 18 Manual reworking a contour. The shape (colored red) can be repainted to correct segmentation faults.

“Smooth Contour” will apply Gaussian smoothing to the contour, using the selected sigma; to do so is recommendable especially if you are preparing a new template. “Set to Convex Hull” will expand a contour to its convex hull (guaranteeing it is completely convex), and “Revert to original contour” will restore the original shape. Also the morphological operators “opening” and “closing” can be applied with the selected settings. The former deletes bulges, the latter fills indentations, both of them apply a smoothing effect to the contour.

It is advisable to first erase bulges and fill indentations by painting or opening / closing and to finally apply a contour smoothing.

When closing the window using the “OK” button, the largest shape highlighted red is taken for calculating the new contour (since a shape might break down into several pieces when reworking it). The results list is updated accordingly, usually with an improved ranking.

4.2.3 Exporting results

Selected results can be exported into a series of CSV (readable e.g. by Excel or R) and image files for further morphometric analysis. Mark results to be exported by clicking the “Export” checkboxes, using the right-click-menu of these fields, or by marking them while scrolling through the results list using the contour visualization window (see Chap. 4.2.1).

Features displayed in the results list as well as the settings used for contour analysis will always be exported. Additional features to be exported can be selected using “Export → Select Features to export” from the main menu. These are: the coordinates of the contour (with respect to the original image), coordinates of points used for elliptic Fourier analysis, coefficients and invariants of the elliptic Fourier analysis and a set of images depicting the detected shapes as well as the appropriate cut-out of the image data. Use “Export → Export Contour Data” resp. “Export Template Data”, depending on which data you want to be exported, and assign a proper filename.

“Export Metafer VSAI File” saves shape outline and position of selected results in a format which allows to import these data as annotations into virtual slides produced by the Metafer Slide Scanning System.

4.3 Other settings

Several other parameters can be set within the “Objects” area or by using the main menu. Some of them are essential for calculating analysis results, while others only affect which results are displayed or exported.

Settings for contour validation and ranking calculation will be saved permanently only on request. This enables for trying out the effects of different values. Use “Save Settings” or “Restore Settings” to save the current values, resp. to restore the former values.

4.3.1 Essential parameters: Micrometer factor

The most essential parameter is the correct scale between image pixels and physical dimensions of the micrographs. The number of pixels which corresponds to a distance of 100 μm can be set by using “Adjust Micrometer Factor” from the “Settings” menu.

4.3.2 Essential parameters: Number of harmonics for elliptic Fourier analysis

Another important parameter concerns the number of harmonics to be used for elliptic Fourier analysis. A fixed number of harmonics is used to produce comparable results, and for analysis each contour is broken down to a corresponding set of points of about equal distance. The number of harmonics can be changed by using “Adjust Settings for Elliptic Fourier Analysis” from the “Settings” menu.

4.3.3 Segmentation and optimization settings

“Select Segmentation Procedures” opens a menu to select the segmentation procedures to be applied (see Chap. 3.3). Usually “Otsu’s Thresholding”, “Histogram Equalization + Otsu” and “Canny Edge Detector” are a good choice. Using “Adaptive Thresholding” might produce a halo surrounding the object, but this has been observed in rare cases only, so it is a good addition to the former three. RATS is extremely time consuming, so it is advisable to use it only for problematic image data.

“Configure Optimization Settings” opens a dialog to enable and adjust different methods for repairing minor segmentation faults (see Chap. 4.1.1).

4.3.4 Settings for displaying / exporting results

The quality of results to be displayed can be adjusted by changing “Only Show Results with Ranking Index \leq ”. This will shorten the results list to entries with a ranking better or equal to the entered value. Using an adequate assortment of templates and image data of good quality usually provides a ranking of zero for excellent, and a ranking of one for good results. When using contour optimization, at least results of ranking one and higher should be checked manually for correctness, otherwise at least results of ranking two or higher should be checked and, if they are correct, probably added as additional templates (see Chap. 4.2.1). Results with a ranking worse than four to five usually do not depict any relevant data. Contours which have a bad ranking due to segmentation faults can be reworked manually to increase the yield of good results (see Chap. 4.2.2). This is accomplished by “painting” the correct shape or by smoothing it

“Show heuristic Descriptors” controls, if a set of standard heuristic descriptors should be displayed. Since these descriptors are not used for ranking or validity checking, they don’t provide essential information for SHERPA’s contour analysis. Nevertheless, when exporting results they will be included for further morphometric examination.

4.4 Settings profiles

Settings can be managed by using different profiles. This way, settings tailored to a specific data set or investigation can be saved for later reuse. You will find the appropriate function via “Settings” in the main menu.

5 Appendix: A straightforward way to suitable settings

When it comes to finding settings suitable for your data and investigation, a simple procedure will help:

1. Load your templates. If you don't already have templates specific to your data, using those templates delivered with SHERPA is a good starting point.
2. Load your image data. In the beginning it is recommended to start with a decent amount of typical samples (some dozens to a couple of hundreds).
3. Select segmentation procedures "Otsu's Thresholding", "Histogram Equalization + Otsu", "Canny Edge Detector" and "Adaptive Thresholding".
4. Decide if you are interested in convex (valve) shapes only. If so, enable "Force Convexity", otherwise enable neither "Use Convexity" nor "Force Convexity".
5. Use default values for contour validation, rating and ranking. You might need to adjust the validation settings for area and perimeter (red area) according to the minimum object size you are expecting, otherwise objects being too small will be ignored.
6. Start contour analysis.
7. Sort the results by "Best EFDIs Match" in ascending order.
8. Display the shape visualization window (see Chap. 4.2.1) by clicking the "Show" button for the topmost (= lowest EFDIs matching) result. Here make sure "Show Contour" is checked and click onto "Show Template" to additionally display the best matching template.
9. Scroll through the results list by clicking onto "Next Contour", and for contours which are segmented well (i.e. red area really coincides with the object area), check if you feel comfortable with the assigned template. If their shapes don't match, you need to add the current contour to your template library (see Chap. 4.2.1, paragraph about "saving a shape as template file"). It is advisable to manually refine and smooth the shape using the "Rework Contour" functions before doing so (see Chap. 4.2.2).
10. When you are finished with checking, reload the templates (now also containing your new templates) and start over at step 6. Repeat this procedure several times until you are satisfied with the matching between objects and templates.
11. Now adjust settings for each rating / ranking indicator, starting with the indicator "Best EFDIs Match".
12. Sort the results list according to the indicator you selected at the previous step. Set the sorting order (by clicking once or twice onto the column header) depending on whether a low or a high value indicates a good measure, so that the best results for an indicator are listed topmost.
13. Display the shape visualization by clicking the "Show" button for the topmost data row.
14. Scroll through the results list by clicking "Next Contour" (or turning the mouse wheel within the window's lower button area) until you find the detected shape unsatisfying. A good value for a corresponding threshold of the indicator will be around the value shown for the result you find unsatisfying. Sometimes outliers might distort the choice, so please inspect a few more results until you are sure you reached an acceptable limit.
15. Set the value for the selected indicator to the new threshold.
16. Select another indicator and repeat from step 12. Indicators to be checked are:
 - Best EFDIs Match
 - Hu Match for EFDIs Template
 - Standard Deviation
 - Width / Height Ratio
 - Contour Smoothness

- Formfactor
 - CDF, PCAF and CHMDF (only for convex shapes)
 - CDF Match, PCAF Match and Compactness Match (if you are not using “Force Convexity”).
17. Restart “Analyze Contours”, sort by “Ranking Index” in ascending order, and recheck if you are satisfied with the results. Usually a ranking of zero indicates really excellent results and a ranking of one stands for good results. Results of ranking two to three contain good and bad results mixed, whilst worse rankings mostly contain undesired results. If you choose to “Use / Force Convexity”, ranking might be messed up by convex templates matched to concave shapes. In this case, you are either missing adequate templates or you should analyze your data without using or forcing convexity.
 18. Don’t forget to click “Save Settings” when you are finished with tuning values, otherwise your customized thresholds will be lost after restarting SHERPA! You can use multiple settings profiles customized to different studies (images coming from different sources, different sets of taxa, ...) see Chap. 4.4

The process of optimizing settings might look complicated and time-consuming at first glance, but it is pretty simple to accomplish and usually you only have to do so when you start analyzing a new type of data.

To check for redundant templates, you can run a shape analysis using the same files as templates and also as image data. In this case, not the best match will be shown (which would be the template matched to itself), but the second best, which is the template with the most similar shape. If the matching value is very close to zero, the template data might be redundant. For reliable results please disable contour optimization, use only Otsu’s thresholding without Gaussian or median filtering for segmentation and sort the results list by column “Best EFDIs Match”.

Ab hier einzeln drucken für "Settings Guide

SHERPA: A straightforward way to suitable settings

When it comes to finding settings suitable for your data and investigation, a simple procedure will help:

1. Load your templates. If you don't already have templates specific to your data, using those templates delivered with SHERPA is a good starting point.
2. Load your image data. In the beginning it is recommended to start with a decent amount of typical samples (some dozens to a couple of hundreds).
3. Select segmentation procedures "Otsu's Thresholding", "Histogram Equalization + Otsu", "Canny Edge Detector" and "Adaptive Thresholding".
4. Decide if you are interested in convex (valve) shapes only. If so, enable "Force Convexity", otherwise enable neither "Use Convexity" nor "Force Convexity".
5. Use default values for contour validation, rating and ranking. You might need to adjust the validation settings for area and perimeter (red area) according to the minimum object size you are expecting, otherwise objects being too small will be ignored.
6. Start contour analysis.
7. Sort the results by "Best EFDIs Match" in ascending order.
8. Display the shape visualization window (see manual, Chap. 4.2.1) by clicking the "Show" button for the topmost (= lowest EFDIs matching) result. Here make sure "Show Contour" is checked and click onto "Show Template" to additionally display the best matching template.
9. Scroll through the results list by clicking onto "Next Contour", and for contours which are segmented well (i.e. red area really coincides with the object area), check if you feel comfortable with the assigned template. If their shapes don't match, you need to add the current contour to your template library (see manual, Chap. 4.2.1, paragraph about "saving a shape as template file"). It is advisable to manually refine and smooth the shape using the "Rework Contour" functions before doing so (see manual, Chap. 4.2.2).
10. When you are finished with checking, reload the templates (now also containing your new templates) and start over at step 6. Repeat this procedure several times until you are satisfied with the matching between objects and templates.
11. Now adjust settings for each rating / ranking indicator, starting with the indicator "Best EFDIs Match".
12. Sort the results list according to the indicator you selected at the previous step. Set the sorting order (by clicking once or twice onto the column header) depending on whether a low or a high value indicates a good measure, so that the best results for an indicator are listed topmost.
13. Display the shape visualization by clicking the "Show" button for the topmost data row.

14. Scroll through the results list by clicking “Next Contour” (or turning the mouse wheel within the window’s lower button area) until you find the detected shape unsatisfying. A good value for a corresponding threshold of the indicator will be around the value shown for the result you find unsatisfying. Sometimes outliers might distort the choice, so please inspect a few more results until you are sure you reached an acceptable limit.
15. Set the value for the selected indicator to the new threshold.
16. Select another indicator and repeat from step 12. Indicators to be checked are:
 - Best EFDIs Match
 - Hu Match for EFDIs Template
 - Standard Deviation
 - Width / Height Ratio
 - Contour Smoothness
 - Formfactor
 - CDF, PCAF and CHMDF (only for convex shapes)
 - CDF Match, PCAF Match and Compactness Match (if you are not using “Force Convexity”).
17. Restart “Analyze Contours”, sort by “Ranking Index” in ascending order, and recheck if you are satisfied with the results. Usually a ranking of zero indicates really excellent results and a ranking of one stands for good results. Results of ranking two to three contain good and bad results mixed, whilst worse rankings mostly contain undesired results. If you choose to “Use / Force Convexity”, ranking might be messed up by convex templates matched to concave shapes. In this case, you are either missing adequate templates or you should analyze your data without using or forcing convexity.
18. Don’t forget to click “Save Settings” when you are finished with tuning values, otherwise your customized thresholds will be lost after restarting SHERPA! You can use multiple settings profiles customized to images coming from different sources or for different sets of taxa, see manual, Chap. 4.4.

The process of optimizing settings might look complicated and time-consuming at first glance, but it is pretty simple to accomplish and usually you only have to do so when you start analyzing a new type of data.

To check for redundant templates, you can run a shape analysis using the same files as templates and also as image data. In this case, not the best match will be shown (which would be the template matched to itself), but the second best, which is the template with the most similar shape. If the matching value is very close to zero, the template data might be redundant. For reliable results please disable contour optimization, use only Otsu’s thresholding without Gaussian or median filtering for segmentation and sort the results list by column “Best EFDIs Match”.