Wood Image Classification

Initial Results

# Research questions

The main problem is: given annotated (or labelled) microscopy images of section(s) from different, mostly CITES listed, wood species, can a new unlabeled microscopy image be automatically classified to which species it belongs? To tackle the main problem we defined a sub-problem to be solved first. Given a set of annotated microscopy images of wood species, can we define and automatically compute a (dis)similarity score between every pair of images? The magnitude of the score should reflect correctly the degree of similarity within the pair - large for images of the same species and small if the images are from different species.

# Data

A very small annotated dataset was available – 27 images in TIF format. Of these images only 19 could be used, because were of species with more than 1 image. Those 19 images were converted to PNG (to make usage of existing software easier). They corresponded to 9 species, each specie was represented by 2 images and only 1 - by 3.

## Species list (unused)

Only single images were available for some of the species: Anodendron rubescens, Carini decr, Crater letest, Dregia volubilis, Gymnema tingens, Napol vog, Ocinotis gracilis, and Periploca laevigata. These images were not used in the experiments.

## Species list (used)

The list of species, represented by more than 1 image, sometimes of different microscopy resolution is given below along with the number of images | and resolutions given in brackets:

1. Argania spinosa (3| 200 m)
2. Brazzeia congo (2| 500m)
3. Brazzeia soyaux (2| 500m)
4. Chrys afr (2| 200m)
5. Citronella sylvatica (2| 1 x 500m & 1 x 200m)
6. Desmostachys vogelii (2| 1 x 500m & 1 x 200m)
7. Gluema ivor (2| 200m)
8. Rhaptop beguei (2| 500m)
9. Stemonurus celebicus (2|500m)

On Figure 1 the first image per specie are illustrated.

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
| 7. | 8. | 9. |

Figure 1 Example images of each of annotated set of 9 wood species

# Approaches

The main idea is that the morphology of the wood cells and their spatial arrangement are the distinguishing characteristics between the species. A description of what the experts consider classifying microscopic features is given at the Wood anatomy of Central European species [website](http://www.woodanatomy.ch/micro.html). Inspired by these descriptions and having a [salient regions detector software](http://software.esciencecenter.nl/software/salient-region-detectors/) in house as part of NLeSc technology platform [eStep](http://software.esciencecenter.nl/), we applied the following approach:

1. Automatic detection of microscopic cells using the salient regions detection algorithm.
2. Computing features of all or some pre-filtered subsets of detected regions.
3. Defining a similarity metric between the computed features (or of their histograms) and computing a similarity matrix between all images

We have tried several features and techniques in step 2, thus resulting in a number of different approaches.

## Automatic salient regions detection

The wood cells from the microscopy images can be seen as salient regions of type “islands” [1]. We have tested both a classical salient region detector, the Maximally Stable Extremal Regions (MSER) [2], as well as our recently-proposed detector, the Data-driven Morphology Salient Regions (DMSR) [3]. Both detectors work with binarizations of gray-scale images, hence as pre-processing, the color images were converted to gray-scale. We have used both the original and the MATLAB Image Processing Toolbox implementation of MSER and also the MATLAB version of the DMSR software. Figure 2 illustrates some results of the wood cells detection using both detectors[[1]](#footnote-1). The conclusion of the comparison was that DMSR performed better than MSER- it consistently gave no redundant or overlapping regions and with much better fitting to the exact cell boundaries. A big advantage was also the option of selecting a specific type of salient regions (“islands”, i.e., lighter regions on darker background) in DMSR, unlike in MSER where all salient regions are detected.

|  |  |  |
| --- | --- | --- |
|  |  |  |
|  |  |  |
|  |  |  |

Figure 2 Salient regions detection on the same wood images. The regions are shown by their equivalent ellipse representation. Left: MSER original code (only every 3rd region is shown); middle: MATLAB MSER implementation; right: DMSR

We have opted in using only DMSR regions for the rest of the experiments. An example of the exact shaped regions (no their elliptic approximations) detected by DMSER is given on Figure 3.

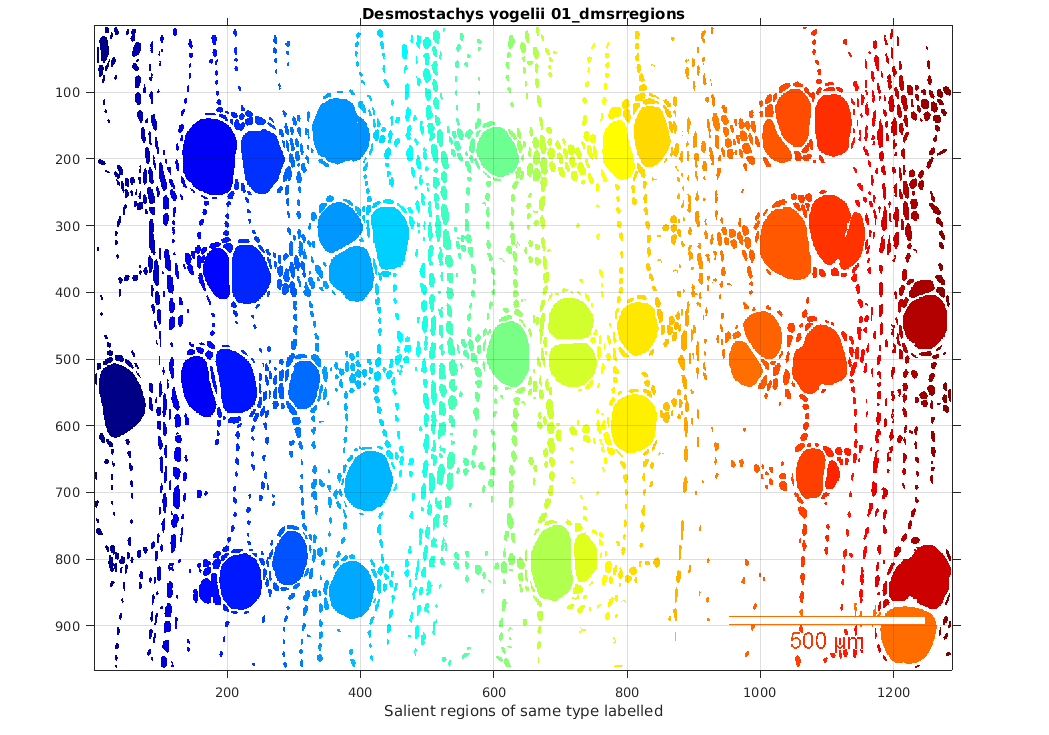


Figure 3 Exact shape of the detected DMSR regions of type "islands" from a wood microscopy image.

## Region features

After the detection steps, we have computed various features of the individual salient regions.

### Generic features for all regions

#### Basic region properties

The following properties were computed for all individual DMSR regions using the MATLAB Image Processing Toolbox command *regionprops*: Area, Convex Area, Eccentricity, and Equivalent diameter, Minor Axis Length, Major Axis Length and Orientation. For the exact definition of these properties, please visit the [*regionprops* web documentation](http://nl.mathworks.com/help/images/ref/regionprops.html?requestedDomain=www.mathworks.com). Histograms of the distribution of these properties for all regions per image and for all images have been computed (see Figure 4).

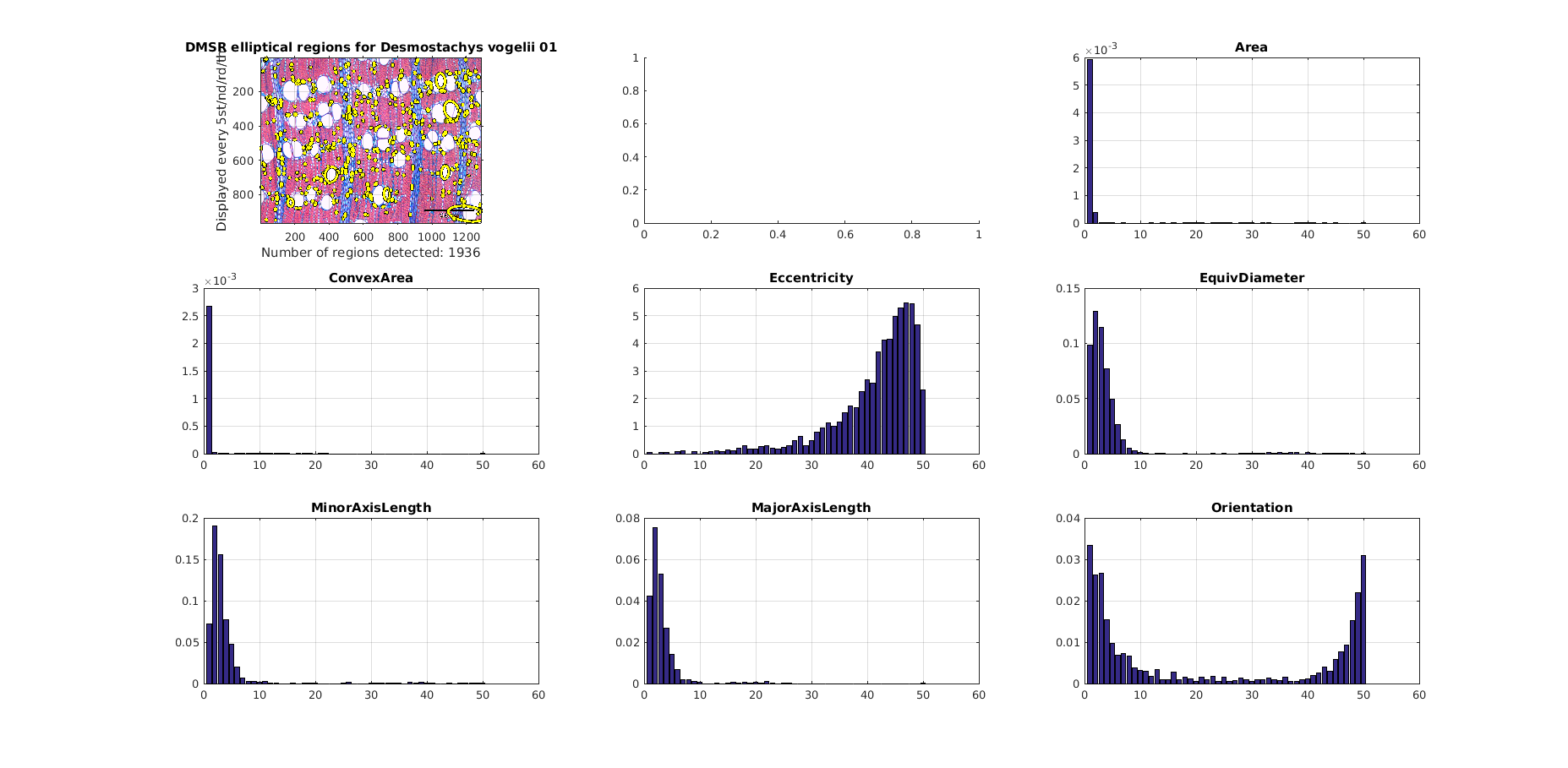


Figure 4 Histograms of 7 properties of the detected DMSR regions from a wood microscopy image.

By inspecting these histograms, we concluded that not all of these basic properties are discriminative enough across species and we opted for some derived and other more discriminative region properties.

#### Derived region properties

The selected 5 derived (or more discriminative basic) properties were:

1. Relative Area – the regions area divided by the image area and taking into account the microscopy resolution (manually).
2. Eccentricity (as above) - the eccentricity of the ellipse that has the same second-moments as the region. The eccentricity is the ratio of the distance between the foci of the ellipse and its major axis length. (The value is between 0 and 1. An ellipse whose eccentricity is 0 is actually a circle, while an ellipse whose eccentricity is 1 is a line segment.)
3. Orientation (as above) - the angle between the x-axis and the major axis of the ellipse that has the same second-moments as the region. (The value is ranging from -90 to 90 degrees.)
4. RatioAxesLength- the ration of the Minor and Major Axis lengths (see above) of the ellipse that has the same second-moments as the region.
5. Solidity - the proportion of the pixels in the convex hull that are also in the region. (Computed as Area/ConvexArea. For details see [*regionprops* web documentation](http://nl.mathworks.com/help/images/ref/regionprops.html?requestedDomain=www.mathworks.com)).

As before, histograms of the distribution of these properties for all regions per image and for all images have been computed (see Figure 5).

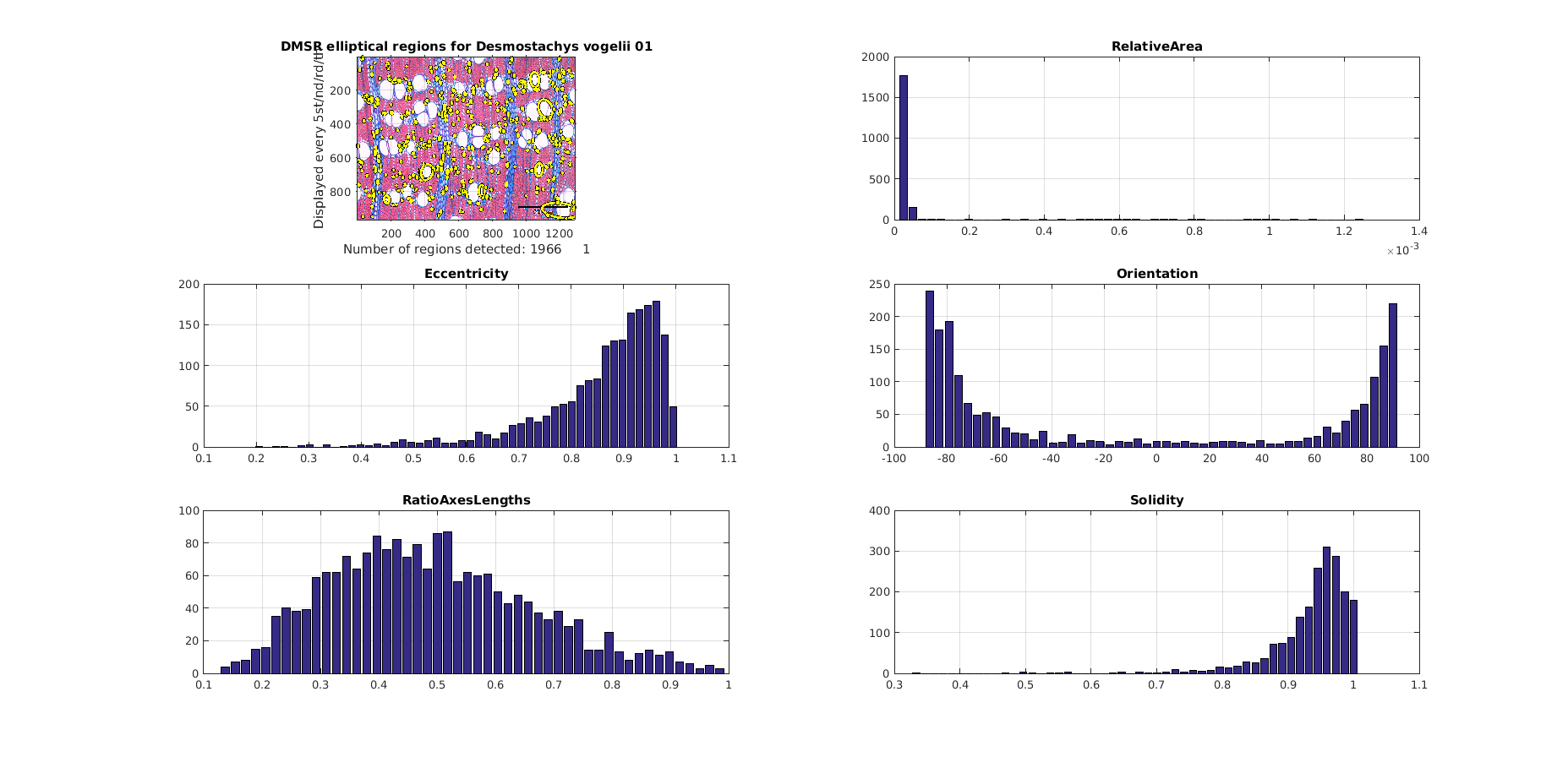


Figure 5 Histograms of 5 basic or derived properties of the detected DMSR regions from a wood microscopy image.

### Features from groups of similar regions

While the distributions of the generic features of all regions reflect the individual cell properties, one can observe that the grouping of regions with certain properties into constellations is also characteristic of some species. For instance, we can observe that the large round-shaped cells in Stemonurus celeb. are evenly distributed across the tissue, while such regions come in clusters of 2 or 3 with Chrys afr. (see 9. and 4. on Figure 1). Other examples are the elongated cells grouped in vertical “bands” for many species, e.g. # 5, 6 and 9, while small round-shaped cells are grouped in vertical “ribbons” in Gluema ivor (7. on Figure 1). These constellations (groups) are harder to automatically detect and describe in comparison with the individual cell properties. We have opted for a 2 stage approach: firstly we filter similar regions based on their individual properties (like Area, Eccentricity and Solidity) and next, we try to describe their constellations.

### Filtering of similar regions

We have defined 4 groups of regions, we would like to detect and have (manually) defined a filtering criteria:

|  |  |
| --- | --- |
| Regions description | Filtering condition |
| Large | Big Relative Area ([0.2, 1]) AND big Solidity ([0.85, 1]) |
| Small | Small Relative Area ([0, 0.199]) AND big Solidity ([0.85, 1]) |
| Small round-shaped horizontally oriented | Small Eccentricity ([0, 0.85]) AND big Solidity ([0.85, 1]) AND small Relative Area ([0, 0.199]) |
| Small elongated vertically oriented | Vertical Orientation ([-90🌣,-55🌣] OR [55🌣, 90🌣]) AND big Eccentricity ([0.75,1]) AND small Relative Area ([0,0.2]) |

We have filtered the above 4 types of regions for every image using the binary masks of all salient regions obtained by the DMSR detector. An illustration of the filtering of the large round-shaped regions can be seen on Figure 6 and of the small elongated vertically oriented ones - on Figure 7.

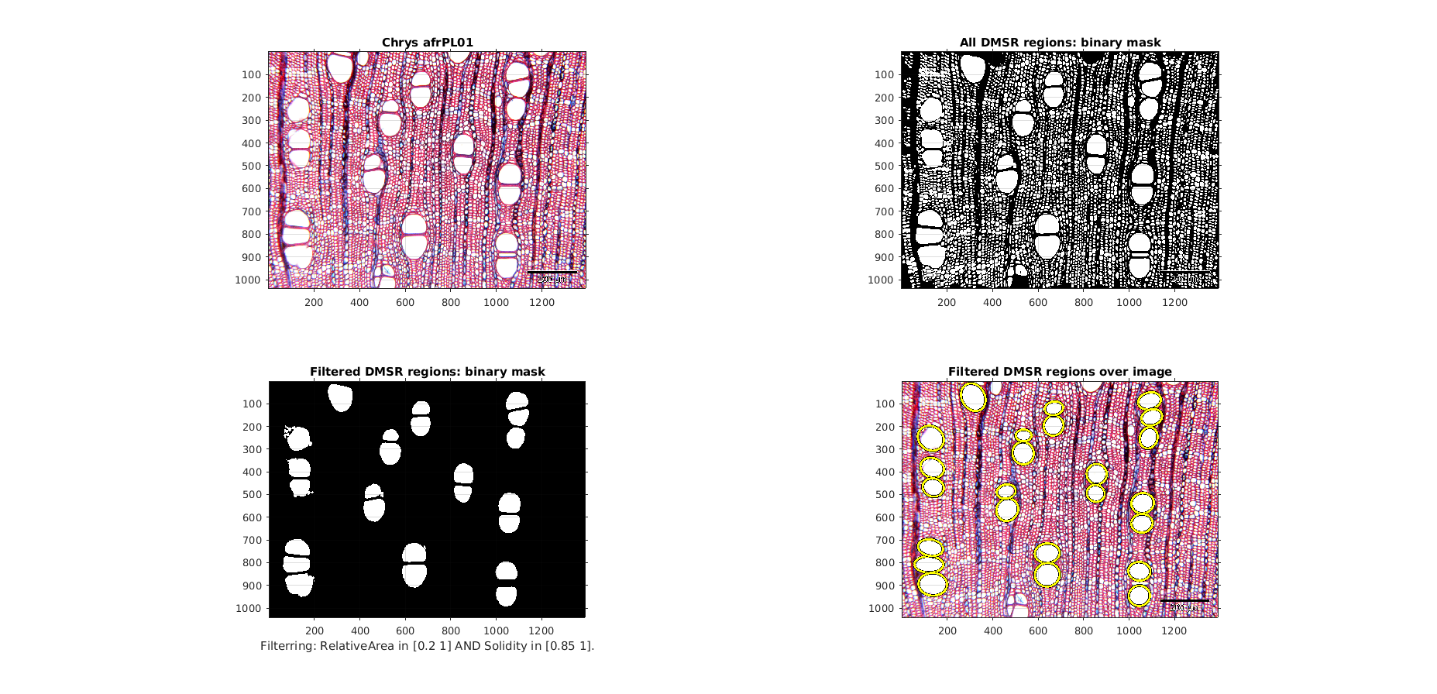


Figure 6 Large DMSR regions filtered from all regions from a wood microscopy image.

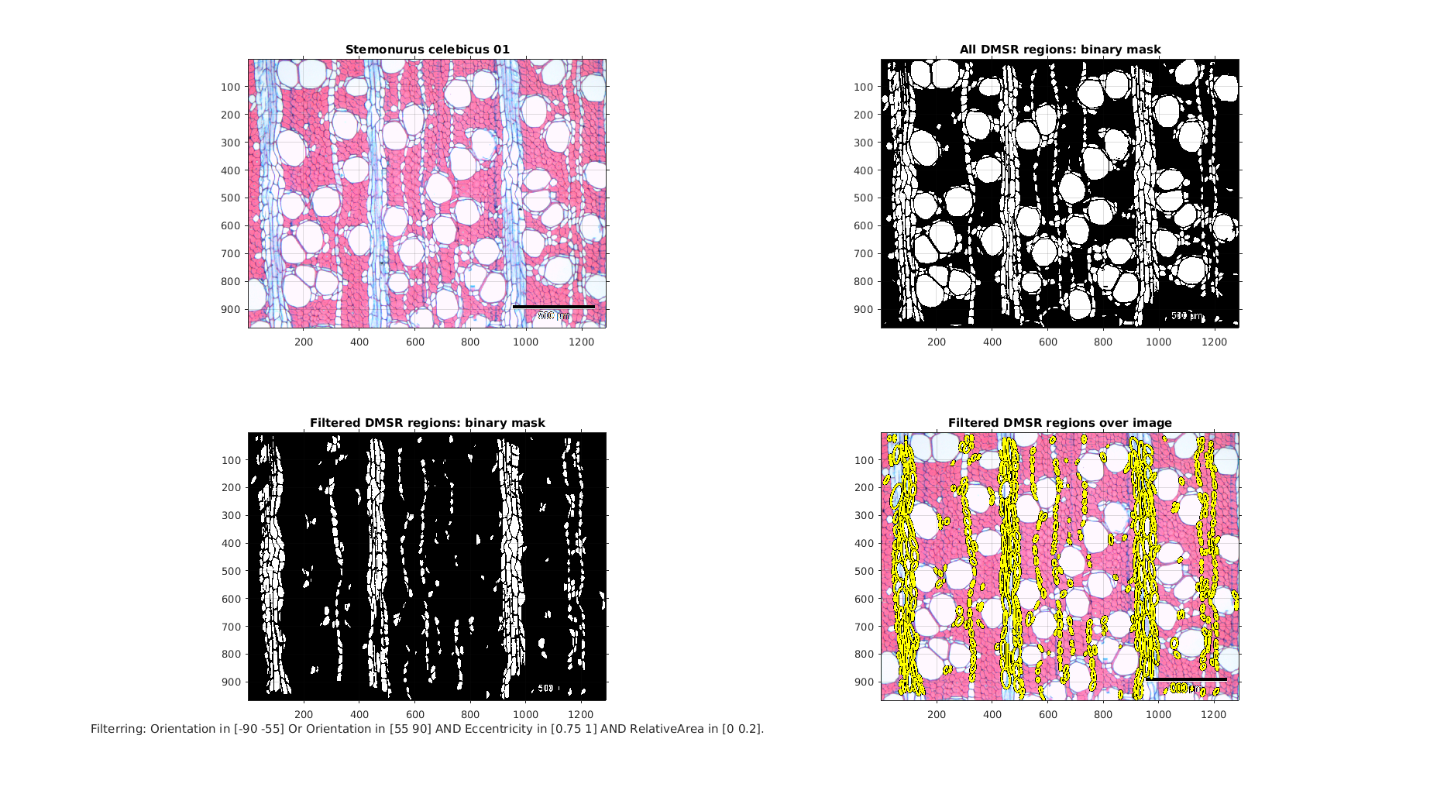


Figure 7 Small vertical elongated DMSR regions filtered from all regions from a wood microscopy image.

### Features for regions groups

After obtaining the 4 types of regions by filtering we have tried 2 approaches for constructing discriminating features of the constellation of the filtered regions.

#### DBSCAN clustering

The Density-based spatial clustering of applications with noise (DBSCAN) is one of the most common clustering algorithms and also most cited in scientific literature. Given a set of points in some space, it groups together points that are grouped together (points with many nearby neighbors), marking as noise (outliers) points that lie alone in low-density regions (whose nearest neighbors are too far away). In our problem we considered the centroids of each region as the set of points to be clustered. DBSCAN requires two parameters: ε - the distance between 2 points to be considered neighbors and minPts -the minimum number of points required to form a dense region. We have used the open-source MATLAB implementation of DBSCAN from [yarpiz](http://yarpiz.com/255/ypml110-dbscan-clustering) on the set of centroids of each filtered image using minPts = 2 and 3 different distances, where, *l* is the image diagonal length and ** is the microscopic resolution. The DBSCAN algorithm on the Chrys afr image from Figure 6 is shown on Figure 8.

|  |  |
| --- | --- |
|  |  |
|  | |

Figure 8 DBSCAN clusters of large salient regions’ centroids with increasing distance between neighboring points .

To describe the constellations of regions via these clusters we have constructed a vector of elements per cluster vector (EPCV) and have defined a number of quantitative features:

1. Total number of clusters
2. Number of elements in cluster ‘noise’
3. Maximum of the EPCV
4. Minimum of the EPCV
5. Mean of the EPCV
6. Standard deviation of the EPCV

Therefore, we obtain an 18 (3\*6)-dimensional feature vector per filtered version of every image (e.g. would sum up to 4 features vectors per wood image in total). The graphs on Figure 9 illustrate these features for the filtered large salient regions masks for 2 species (2 images per specie). We can see that these vectors are very similar for species that have more clusters of large regions (e.g. 8. Rhaptop), while are not so useful for species that don’t (e.g. 7.Gluema). Therefore we decided to abandon this approach.

|  |  |
| --- | --- |
|  |  |

Figure 9 Features of DBSCAN clusters of large salient regions' centroids for 2 species of wood.

#### Successive morphological dilation

# References

|  |  |
| --- | --- |
| [1] | E. Ranguelova and E. Pauwels, "Morphology-based Stable Salient Regions Detector," in *International conference on Image and Vision Computing New Zealand (IVCNZ’06)*, 2006. |
| [2] | J. Matas, O. Chum, M. Urban and T. Pajdla, "Robust Wide Baseline Stereo from Maximally Stable Extremal Regions," in *British Machine Vision Conference (BMVC)*, 2002. |
| [3] | E. Ranguelova, "A Data-driven Region Detector for Structured Image Scenes," in *International Conference on Image Processing (ICIP'16)*, submitted, under revision, 2016. |

1. All results for all images and methods are available as HTML pages with figures and can be shared on request. [↑](#footnote-ref-1)