Black Spot Leaf Area Calculator, Version 1.0 Beta
User Manual

Varun Varma & Anand M Osuri

National Centre for Biological Sciences (NCBS),

Tata Institute of Fundamental Research (TIFR),

Bangalore – 560 065,

India.

License information

Black Spot Leaf Area Calculator version 1.0 beta. Copyright (c) 2013, Varun Varma and Anand M Osuri.

What is Black Spot?

Black Spot is a method and stand alone software for rapid and accurate estimation of leaf area from leaf images scanned using a flat-bed scanner. This easy to use software has been specifically designed keeping in mind field ecologists and the large volumes of data needed to be processed for ecosystem level studies.
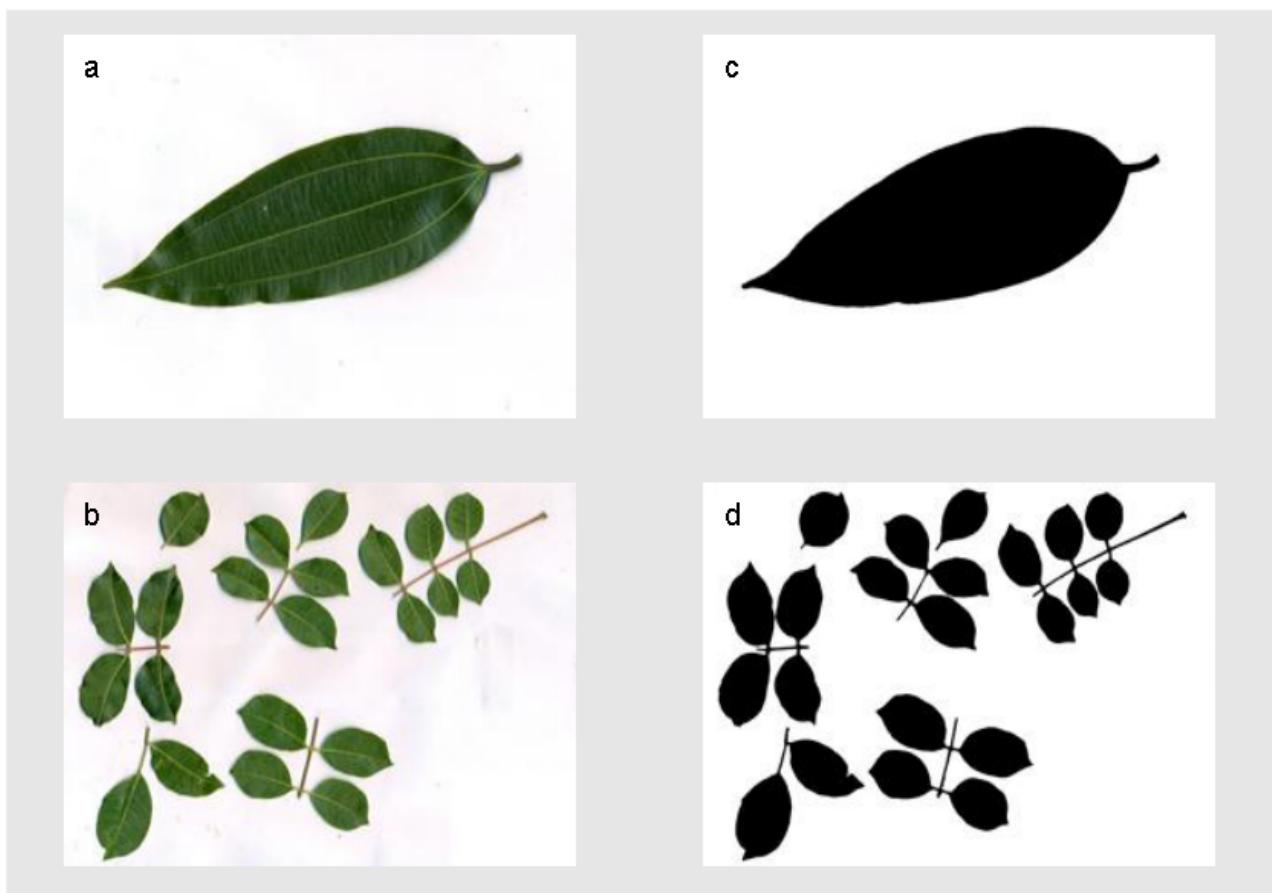

Figure 1. Example of scanned leaves of (a) *Cinnamomum malabatrum* and (b) *Sclrerocarya birrea* and leaf area detected by Black Spot Leaf Area Calculator (c) and (d).

The main features of Black Spot is as follows:

1. **No scale calibration required –** Black Spot does not require the user to manually provide any scale calibration to estimate leaf area. The software, automatically extracts pixel resolution information from the scanned leaf image meta data and uses it to output leaf area in usable units (cm^2).

2. **No spectral calibration required –** The image processing rules for Black Spot have been designed by analysing the spectral properties of over 1500 leaf images spanning multiple species, to arrive at general, yet accurate spectral rules to identify leaf from non-leaf. As a result the user is not required to carry out spectral calibration (e.g. selecting colours,

manually specifying a sample of the leaf area, etc.) to estimate leaf area.

Currently, in its beta stage, Black Spot's spectral rules have been tested on Canon and Hewlett Packard scanners. We hope to add more scanner brands to this list and are expecting feedback from users to increase the range of scanner brands that can be used with Black Spot. As a user, you can contribute to the development of Black Spot by writing to us at the address (blackspot@ncbs.res.in).

3. **Process multiple species at the same time** – As Black Spot's spectral rules have been calibrated using a large number of samples across species, the software is capable of processing multi-species image data at the same time.

4. **Batch processing** – With its in-built spectral rules and automated scale calibration, Black Spot's leaf area estimation procedure allows users to batch process a large number of leaf images with minimal user intervention. The user is only required to capture leaf images bearing a few basic rules in mind, organise images into a single folder on their computers, then provide Black Spot with the location of these images, the desired output location and 3 other simple image processing parameters. The software then sequentially process all images in the input folder and stores results in the user specified output folder.

5. **Fast** – Each leaf, when scanned at a resolution of 300 dpi for an A4 size scan area can be processed by Black Spot in approximately 15 to 20 seconds, resulting in considerable savings in user time.

6. **Ready to use output** – Estimates of leaf area are stored in a comma separated value (csv) file which can be imported into any statistical environment for analysis.

7. **Error checking** – Black Spot's output also includes a copy of the original leaf image and a classification mask (a binary or black and white) image of the area detected as leaf. These are stored in the output folder which can be accessed by the user at any time to visually check the accuracy of the software's image processing. Images where users detect error can be corrected with simple image editing and reprocessed in Black Spot. A detailed description of commonly found errors and methods to correct these are provided in the 'troubleshooting' section of this manual.

8. **Portable installation** *(currently for Windows and Linux only)* – The software does not require installation and even though written in python, does not require python installed on your computer. Provided as a zip folder, Black Spot can be run once uncompressed off of your computers hard drive or any portable storage device.

9. **Free software** – Black Spot is licensed under GNU – GPL version 3. Users are free to distribute and modify Black Spot under the terms of the GNU General Public License as published by the Free Software Foundation. Please refer to the Licence section of this manual for additional details.

Using Black Spot

Downloading and installation

Black Spot can be downloaded from [Source forge]. When uncompressed, the downloaded folder will contain a user manual, a readme text file, a copy of the python program for Black Spot and a sub-folder 'Black_Spot_v1_0_beta' which contains the software.

The software has been designed as a portable install and on Windows machines can be run after decompressing the downloaded folder. Users need to navigate to the 'Black_Spot_v1_0_beta' folder and run the application 'Black_Spot_v1_0_beta'

For computers running Linux and Mac OSX, a windows emulator will be required. For example, Linux users can use Wine (http://www.winehq.org), while OSX users can use VirtualBox (http://www.virtualbox.org) (VirtualBox requires a valid Windows version installed on your computer).

Alternatively, users can also run Black Spot using the python code available within the downloaded Black Spot file. For this you will need to install python 2.6 or 2.7 (http://www.python.org) as well as the following python modules associated with the python version you install and your OS.

Python Imaging Library (PIL) - http://www.pythonware.com/products/pil/

Numpy – http://www.numpy.org

Scipy – http://www.scipy.org

Capturing leaf images

Collection and storage of leaf samples: In the likely event of a time lag between sample collection and scanning, care must be taken to minimize water losses and the breakdown of leaf structure. In dry, hot conditions, these losses can be mitigated by storing and transporting the samples in portable ice boxes, which maintain cool, but not freezing temperatures. The collected samples should preferably also include some length of the stem to which the leaves are attached. Prior to scanning, users are advised to rehydrate the leaves by immersing the stem in a tub of water for at least 6-8 hours.

Scanning leaf samples: At the time of leaf scanning, detach the leaf from its stem and wipe dry. Place the leaf on the scanning surface. Refer to manual provided by the scanner manufacturer for the procedure to create and save a .jpg image of the leaf. For most accurate results and efficient operation, users are advised to:

1. Paste a white paper on the inside of the scanner lid so that it forms a background to the scanned leaf. This will prevent permanent marks forming on the scanner lid which can affect data quality. Replace the background paper periodically, as and when stains accumulate on it.

2. Give some thought to which side of the leaf to scan. Since the dorsal surfaces of leaves are usually smoother and more reflective than ventral surfaces, scans of dorsal surfaces of certain leaves might contain bright reflections which induce errors in Black Spot image

classification. On the other hand, the leaf venation, which is usually more prominent on the ventral surface, might result in errors under certain circumstances, particularly as the colour of the veins tends towards yellow. A few practice runs of scanning leaves and running Black Spot would help users to take better decisions on which side of leaves to scan.

3. Cut leaves into smaller pieces and and distribute over the scan area, and if necessary, across multiple scans. Leaf material that is closer to the margins of the scanner bed are likely to raise the scanner lid enough to allow light leakage, which is a potential source of error in Black Spot image classification. This problem is exacerbated by particularly thick structures, such as petioles. Further, large leaves are prone to getting folded when the scanner lid is closed, which often results in discolouration along the folds in the image and errors in Black Spot image classification. Both these problems can be reduced by cutting leaves into smaller pieces and avoiding to place them close to the edge of the scanner bed. It is often more efficient to create a larger number of uncluttered images than fewer, crowded ones. When distributing parts of individual samples across multiple images, use appropriate tags in the image file name so that data can be easily consolidated.

4. When working with compound leaves, it is advised to detach leaflets from the rachis and distribute them over the scan area, and if required over multiple scans. This reduces errors of leaflets overlapping and an underestimate of leaf area.

5. While scanning samples the user should keep a record of scanner settings used, especially pixel resolution and scanner brand. Black Spot utilises pixel resolution of the image to provide scale to the leaf image. It then utilises this data to report leaf area in units of cm^2. There are instances where this meta data can be stripped off an image and therefore, scale reference information is not available to Black Spot. In such a scenario, Black Spot is not capable of reporting leaf area in cm^2 and will only return the number of pixels classified as leaf. The user can then use the manually recorded pixel resolution in a simple formula (see output section of this manual) to calculate leaf area.

Preparing for Black Spot: Create a folder on your computer within which images for the Black Spot run are to be stored. Transfer all the images for analysis to this folder. Ensure that all the images are in .jpg format. When running, Black Spot will ignore all other file types and sub-folders that might be present within this folder.

Note to Linux users using a Windows emulator (Wine): the folder where images should be stored before processing with Black Spot is within the 'C Drive' that WINE creates. This is located in '/home/<username>/.wine/drive_c'. The '.wine' folder is a hidden folder. To view it press CTRL+H in your file browser. Therefore the folder where you store your images will have a path something like '/home/<username>/.wine/drive_c/leaf_images'.

Running Black Spot: Once your images are organised into a single folder, Black Spot is launched by running the 'Black_Spot_v1_0_beta.exe' file within the downloaded and decompressed Black Spot files. After, displaying the initial license information and instructions, the user is prompted with five questions:

`Enter path to folder with images >`
Here the user needs to enter the full path of the folder where the images to be processed have been stored. On Windows, the folder path can be viewed by browsing to an image within the input folder and right clicking on one of the images in that folder. The path required will be listed under

'location'. If an invalid folder path is provided (folder does not exist or path is specified incorrectly), Black Spot will display an error and prompt the user for this information again. On providing a valid response, Black Spot will list out all .jpg images in the folder and provide a count of files to be processed.

Enter output folder name (this will be created) >
Here the user is asked to provide a name for the output. This name will be used to create a sub-folder within the input folder specified above where all output files will be stored. Additionally, the output .csv file with leaf area estimates will also have the same name as specified here and will be stored in the output folder. If the user specifies an output name which already exists (i.e. exists within the input folder), Black Spot will return an error and prompt you for another response. The user can either provide a new output name or delete the existing folder with the output name specified and retry. Black Spot will not over-write pre-existing folders.

What % of the image along the edge would you
like to impose this border to (integer between 0 and 50) >
The edge of scanner lids of most scanners is often not white, resulting in the formation of a dark margin to the images, which is usually broader in the case of thick leaves. This can cause errors in Black Spots classification. To avoid the need for manual correction of this problem, the user can specify the size of an area around the edges of the image (in percentage units) they wish to exclude from the analysis. Black Spot only accepts integer values between and including 0 and 50 for this question.

Define filter size (3, 5, 7 or 9) >
The user is asked to specify the size of a despeckling filter (either a 3 x 3, 5 x 5, 7 x 7 or 9 x 9 pixel filter). As the name suggests, this reduces the amount of speckling in the image classification which can add up to a substantial amount of error. As a rule of thumb, if your images contain thin linear leaves (e.g. grasses), use a 3 x 3 filter. For broader leaves use larger filter sizes. In invalid response will generate an error message and the user will be prompted again. Valid responses are either 3, 5, 7 or 9.

Would you like to resize images (y/n)?
Here the user specifies if they want images to be resized before image processing. Resizing is more of a time saving step. Default scanner settings usually scan at a resolution of 300 dpi for an A4 size scan area. This resolution is often far greater than required. By indicating yes to this answer, Black Spot will resize each image to 1/4 of their original dimensions before processing. As a guideline, images which have been scanned at a resolution of 300 dpi or more may be resized without noticeable loss in accuracy. Valid responses are either 'y' or 'n'.

On completion of all five questions, Black Spot can be left to run. You will see a series of messages in the console window that indicates Black Spots progress.

The output: Before Black Spot starts processing images, it creates an output folder with a name as specified by the user as a sub-folder within the input folder. Within this output folder Black Spot stores processing images (which can be used for error checking) and creates a new comma separated value (.csv) file which stores estimates of leaf area from the supplied images.

1. Output images: As part of the output, Black Spot saves a copy of the original image and

classification image (leaf mask; figure 1c and 1d) in the output folder. The copy of the leaf image will have '_leaf_image.jpg' appended at the end of the original image name. The leaf mask will have '_leaf_mask.jpg' appended at the end of the original file name. The leaf mask is a pure black and white image (binary image) where black pixels are what the software has classified and counted as leaf. Once Black Spot has finished processing all images, the user can open the output images in any image viewer and scroll through the leaf image and its leaf mask to visually check for errors in classification. Refer to the trouble shooting section of this manual if you find large errors that need to be corrected.

2. Output .csv file: Black Spot stores leaf area estimates in a .csv file within the output folder. Names and descriptions of the column headers are as follows:

Image_name: name of the input image

exif_found: Flags if image contains exif data. If exif data is found and complete, this field will read 'Exif info found'. If exif data is found but does not contain either scanner brand name or pixel resolution information, the field will read 'Exif info found but incomplete'

Scanner_brand: Reports scanner brand name used for image capture. If no scanner brand name is found, field will read 'Not detected'

px_resolution: Returns the pixel resolution at which image was scanned. If not detected, will return a value of 0.

Leaf_pixels: A count of the number of leaf pixels detected by Black Spot.

Leaf_area_sq_cm: Conversion of number of leaf pixels detected to units of square centimetres (cm^2). If exif data is not found for the image, or even if found but lacks pixel resolution information (px_resolution field = 0), leaf pixel count will be reported here. If the user is aware of the pixel resolution at which the image was scanned (in dpi), he/she can calculate leaf area by dividing the value in this field by $(dpi/2.54)^2$, i.e. leaf area = number of leaf pixels/$((dpi/2.54)^2$.

Error_Message: Summarises any errors during processing of an image. If there are no errors (exif data found and complete) then will return a value of 'OK', otherwise will report 'Exif data incomplete or not found'.

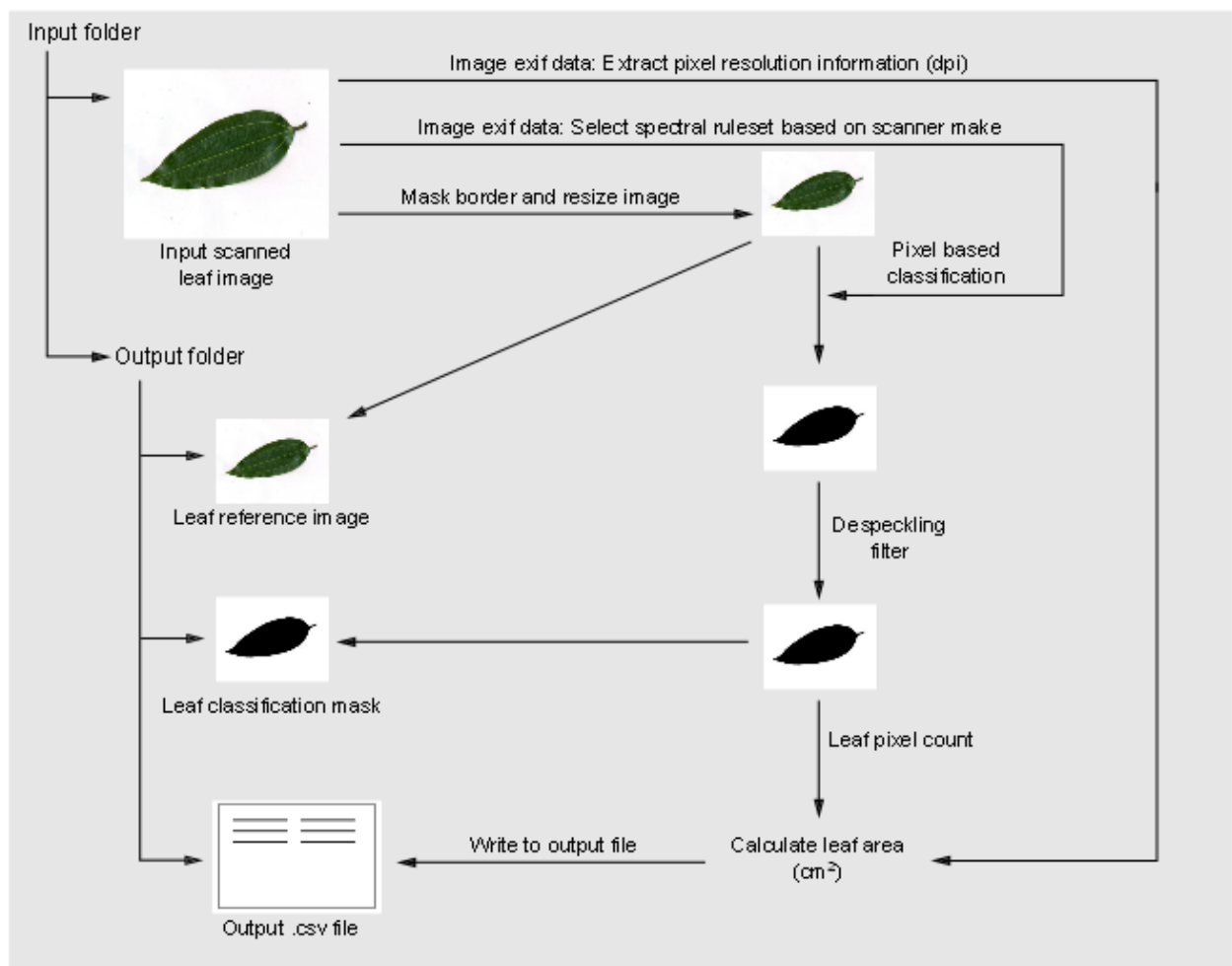# How does it work? - Identifying leaf from non-leaf

Figure 2. Schematic representation of Black Spot work flow: Each image in the input folder is processed sequentially. First, for each image, the exchangeable image file format (exif) information is extracted. The exif information contains the scanner brand used to capture the leaf image, as well as the pixel resolution during image capture. The image is resized, and a border mask applied as per user input. Based on the scanner brand identified, the software applies the appropriate spectral rule set to identify leaf from background. A first pass of the pixel based classification is followed by the application of a despeckling filter over the classified image, the size of which is specified by the user. The classification and filter processes result in a binary image (leaf and non-leaf) where the number of leaf pixels is determined. Leaf area is then calculated in square centimetres (cm^2) by utilising pixel resolution values extracted from the image exif data. As processing of each image is completed, an output comma separated value (csv) file is updated with one row of data containing the image name, a flag reflecting whether image exif data was found, scanner brand, pixel resolution, number of leaf pixels detected, estimated leaf area and an error flag indicating any discrepancies during processing. To allow for error checking, Black Spot also stores each leaf image (leaf reference image) and its classification image (leaf classification mask) in the output folder. The user can access these images at a later time to visually check for errors in leaf classification.

Calibration of scale - Leaf images taken using scanners are usually associated with image meta data known as the Exchangeable Image Format or exif data. Part of this exif data pertains to the pixel resolution used to capture the image. This resolution is expressed in units of dots per inch or dpi. Black Spot extracts this information from the image and converts the number of pixels classified as leaf to an area value in square centimetres (cm^2).

Detection of leaf pixels – Black Spot has in-built spectral rules to detect which pixels of the scanned image are that of leaf (pixel based classification). The basic principle of the spectral rule is to use band ratios instead of absolute pixel values. The background of the image (a white sheet) will have band ratios (i.e. red/green, green/blue, blue/red) tending towards one, while the leaf in the image will have band ratios which deviate from one. Along similar lines, shadows in the image, though of lower pixel intensities, also share similar properties of band ratios: i.e. shadows falling on the white paper will have band ratios close to one, while shadows on the leaf surface will have band ratios which deviates from one. These band ratio properties should in principle, hold true regardless of the scanner make and model that is used to capture leaf images. However, small variations (make and model specific) do exist, especially between scanner brands. To take this into account, Black Spot has slightly different rules depending on the scanner brand used. As mentioned in the discussion above on calibration of scale, the image exif data also stores the scanner make used to capture the leaf images. This information is used by Black Spot to apply the relevant spectral rules. Currently, Black Spot has been tested on Canon and Hewlett Packard scanners. Users of other scanner brands are urged to contact the developers if the software does not perform satisfactorily on their leaf images.

Trouble shooting

Using an image editing software to correct errors: Based on the leaf mask generated in the Black Spot output, users can identify any leaf images that have not been correctly classified. Users may then rectify these images to remove the source of error using any image editing software, and rerun on Black Spot. **An important requirement of the image editing software used is that it not modify the images' exif data file**, which contains information that is used by Black Spot. We recommend the use of GNU Image Manipulation Program (GIMP: http://www.gimp.org/), which is freely available on the Internet, for editing images that are to be processed with Black Spot.

Errors in image classification may arise either because of misclassification of leaf as background, or vice versa. The former may arise when the image of the leaf surface contains reflective highlights. Also, while Black Spot works effectively across a wide range of leaf colours, certain colours such as yellow may result in some classification error. In such cases, users may employ a variety of approaches to first isolate the leaf surface(s) and assign it a uniform green colour. In GIMP, foreground isolation can be achieved through any of the following three tools (including combinations of them):

1. Foreground Select

2. Select by Color

3. Intelligent Scissors

Once the foreground has been selected, users may Cut out the foreground colour and Fill in its place a uniform green colour (R: 50 G: 150 B: 1) using the Color Dialog.

In case the error is a result of discolouration to the background such as marks and stains, or from light leakage, users may Select the problem areas on the background and, Cut them out. Users may also use the Paintbrush Tool set to white colour (R: 255, G:255, B: 255) to paint over discoloured background areas. Alternately, users may use the foreground selection procedure above to select the leaf surface and Paste it into a new image. When doing this, users must ensure that the image resolution of the new image (dots per inch [DPI]) is the same as the image from which the leaf surface was extracted.

Reporting bugs, errors and getting help

For any queries and feedback please write to blackspot@ncbs.res.in