pd Documentation
August 19, 2008
Lynn Epstein

Introduction

pd (for picture decompose) is an image analysis program in which the user identifies standards for each category of interest. For example, the user categorizes the image in Fig. 1A and 1E as comprised of background, fungal pustules and leaf. In an image, the color of each pixel can be represented by the percentage of the maximum value of red, green and blue (%RGB). First, the user selects a representative set of %RGB values for each category of interest; these are the standards. Then, for each standard, the algorithm computes a vector in %RGB space that includes 0,0,0 and the selected point. For each pixel in an image, the algorithm finds the vector in the set of standards whose direction from 0,0,0 is closest to the direction of the vector for that pixel in the image and assigns the pixel to the category of interest of that closest vector. pd can analyze batches of any number of ppm files. pd produces text output that has counts of the number of pixels in each category for each of the ppm files; and, for each ppm file, a -pd.ppm file that shows the assigned categories of interest with the assigned colors (Fig. 1C and 1G). After visually verifying that the assigned categories faithfully represent the image, the user can readily compute the percentage area of each category of interest.

pd was written by Susan Bassein in collaboration with Lynn Epstein.

Computer system

This distribution of pd is executable under the cygwin system (www.cygwin.com) on a Windows XP system running on an I86 PC.

How to use pd

1) Download required software, all for free.

A) pd. Download: http://plantpathology.ucdavis.edu/faculty/file/epstein.zip

B) Cygwin. Download: http://www.cygwin.com/.

C) pd requires that images be saved as ppm files. Similarly, pd produces a ppm file with the assigned categories; these should be viewed for quality control. Also, pd requires a collection of %red, %green, and %blue values of user-selected standards. The GNU Image Manipulation Program (GIMP) readily performs all three functions, including producing ppm files from a large number of types of graphical files (e.g., jpeg, jpg, jpe, tif, and tiff). Download free from: http://www.gimp.org/windows/

2) Record digital images. Although the following are not required for pd, we recommend the following regarding recording an image. If you have a background, use a color that is not in your image, e.g., for a leaf, a solid blue. Uniform lighting is best. For scanned images, we've had good experience with an Epson Perfection 2450 PHOTO scanner with leaves mounted perpendicular to the scanner light bulbs and images saved at 2400 dpi as tif files.

3) Convert digital images to ppm files and name them using pd conventions.

A) Using The GIMP, convert tiff files to ppm by saving as ppm with the "raw" option (not ASCI).

B) The names of the ppm files must conform to the following format. Each name must be one or more words, with multiple words separated by "-" (not "_"), followed by ".ppm". There must not be any spaces in the file names. For ease of processing pd output files through statistical analysis software, we recommend ppm files be named using strict conventions with treatments and replicates separated by dashes, e.g., treat1-rep1.ppm

4) Prepare a a file of user-selected color standards.

A) Overview. This process requires viewing the files with The GIMP (or other software) and using a "color picker" (the eyedropper tool) to select % red, %green & % blue values for a set of standards. You will record these values on a spreadsheet (Excel is fine) and graphically view them before saving the file.

B) Select standards, record %red, %green, %blue (%RGB) values and prepare the file for pd. This can be done with ppm (or other file types) and the eyedropper tool on The GIMP (or other software such as Photoshop). Note that the eyedropper tool provides both raw pixel values and percentage values; record % values. The standards file can be assembled in Excel and then saved as a tab separated text file; the standards used to analyze a batch of files that included the leaves shown in Fig. 1A and 1E is shown in Table 1.

Advice: For each category, pick standards that represent the full range of colors. This is best done at higher magnification. For example, for the leaf images in Fig. 1A and 1E, for the plant category, select at least one standard from the following: green leaf, chlorotic leaf, necrotic leaf, guard cell, and trichome.

C) File format guidelines for the tab-delimited text file with %RGB standards.

i) There are no headers in the file

ii) The first column has the name of the category and must be no longer than 63 characters; this is the name assigned to a category of interest.

iii)The second through 4th column have the % values of the assigned red, green and blue color, respectively, that correspond to the color that will be displayed for that category of interest. For example, we assigned red (100, 0, 0) to designate pustules, and green (0, 100, 0) to designate leaf tissue. Any color can be selected; for example (59, 49, 27) is brown.

iv) The fifth through seventh columns have the % red, green and blue values, respectively, for each selected standard point. *Use the eyedropper tool in The GIMP to select the range of colors observed for each category*. pd allows a maximum of 256 standard values (=lines of text). That is, within the limits of a total of 256 standards, the user can have any number of categories and standards per category; the number of standards can vary for different categories.

User-selected standards are shown in yellow in Table 1.

D) Recommendations for the Excel file used to prepare the text file

i) A header row

ii) The seven columns indicated above; (these columns will be copied without the header row(s), pasted into a new worksheet, and then saved as a tab-delimited text file)]

iii) An eighth and ninth column in which the data in the fifth through seventh column are used to compute the following: % R/(%R+%G+%B) and the %G/(%R+%G+%B). These are graphically displayed as the x and y coordinates.

(1) The graphical display of % R/(%R+%G+%B) and %G/(%R+%G+%B) of your selected standards as x and y coordinates is a two-dimensional projection of the standard values. An example is shown in Fig. 2.

(2) Trouble-shooting your standards. Use the graphical display to examine your selection of standards. For example, overlapping points in multiple categories are problematic.

iv) Delete the header row before saving as a tab-separated text file.

5) Run pd.

A) If you are not familiar with either UNIX or cygwin, you will need assistance from your local computer person to put pd in a location in which cygwin can find it. You may also need instructions to first organize your files (make a folder, most simply with only the txt file with standards and your ppm files) and to start cygwin. (To start cygwin, click on the icon).

i) Assuming that your folder with the .txt file with standards is on the d drive, and the folder is named pustules7-15-08, the command line is

cd "/cygdrive/d/pustules7-15-08"

B) To run pd, as an example, we will assume that the standards file is named "rgb.txt", the ppm files are named "treat1-rep1.ppm", "treat1-rep2.ppm", "treat2-rep1.ppm", and "treat2-rep2.ppm", and the file with the numbers of pixels in each category of interest will be "treatOUT.txt".

i) The command line (on cygwin) is then:

pd rgb.txt treat1-rep1.ppm treat1-rep2.ppm treat2-rep1.ppm treat2-rep2.ppm > treatOUT.txt

ii) You may use Unix command line facilities to simplify the command line. For example, if you want to analyze all ppm files in your folder, the command line can be

pd rgb.txt *.ppm >treatOUT.txt

iii) The program will report the progress of execution by announcing the beginning of processing of each ppm file. It will report the completion of execution by announcing "Done". Wait until you receive this confirmation of termination.

iv) The program will also produce files "treat1-rep1-pd.ppm", "treat1-rep2-pd.ppm", "treat2-rep1-pd.ppm", and "treat2-rep2-pd.ppm"

6) Visual verification of assigned categories. Using GIMP or other software, open the -pd.ppm files and compare to the original ppm files *at increased magnification*. Note that GIMP readily allows comparison of the original ppm and –pd.ppm file; use either the values on the axes or the coordinates of a pixel of interest, shown on the lower left corner of the GIMP window to compare specific portions on the image. If portions of the image were assigned inappropriate values, modify the standards; use the eyedropper tool to find the %RGB values of pixels that were incorrectly assigned. The graphical display of the two-dimensional projection of your standards in the Excel spreadsheet can be very helpful in deciding what changes in the standards file will be either useful or pointless.

7) Import the output txt file (treat-OUT.txt in this example) with count data into Excel, compute desired percentages, and prepare for export into your statistical analysis software.

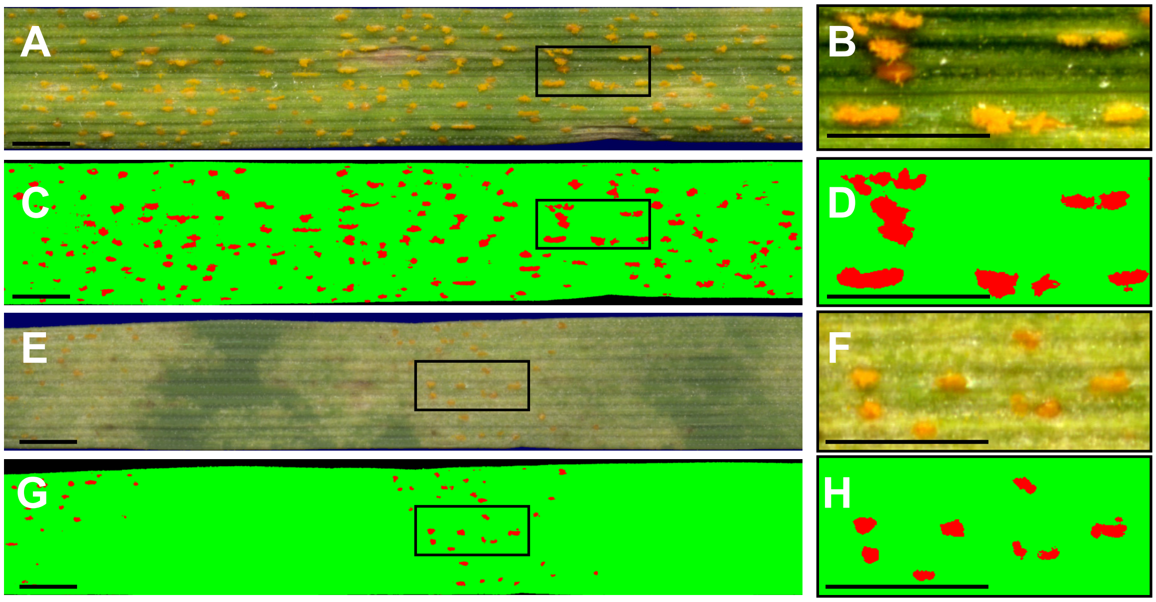


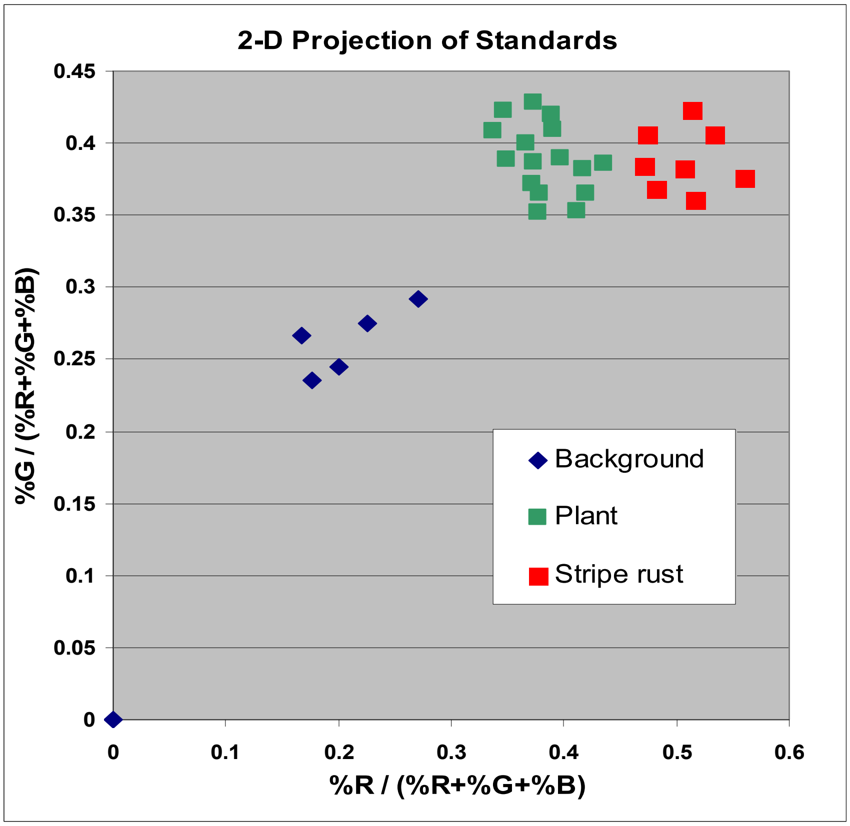
Fig. 1. Visual display of digital quantification of the percentage of leaf area covered with *Puccinia striiformis* pustules. A,E) Scanned images of five-cm segments of wheat leaves. C,G) Each pixel in A and E was categorized by the program "pd" as either leaf (green), *P. striiformis* (red) or background (black). Images in the right column are an enlargement of the image in the rectangle in the left column. A-D) Susceptible RSL 11-19 with pustules covering 6.9% of the surface area in the segment shown in A. E-H) Resistant RSL3-28 with pustules only on 0.9% of the segment shown in E. Contrast and brightness were manipulated in B and F to better show pustules. The white areas in E are the hypersensitive response. Bar = 5 mm. The pd program is freely available at http://plantpathology.ucdavis.edu/faculty/epstein/.

Table 1. The spread sheet that was produced in preparation for analysis of a batch of files that included the images shown in Fig. 1A and 1E.

	Visual Display[a]			Selected Standards[b]			2-D Projection of Selected Standards[c]	
	R%	G%	B%	**R%**	**G%**	**B%**	%R/(%R+%G+%B)	%G/(%R+%G+%B)
Bkgd	0	0	0	0	0	39	0	0
Bkgd	0	0	0	6	8	20	0.176470588	0.235294118
Bkgd	0	0	0	9	11	25	0.2	0.244444444
Bkgd	0	0	0	5	8	17	0.166666667	0.266666667
Bkgd	0	0	0	9	11	20	0.225	0.275
Bkgd	0	0	0	13	14	21	0.270833333	0.291666667
Plant	0	100	0	61	57	44	0.37654321	0.351851852
Plant	0	100	0	35	30	20	0.411764706	0.352941176
Plant	0	100	0	56	54	38	0.378378378	0.364864865
Plant	0	100	0	39	34	20	0.419354839	0.365591398
Plant	0	100	0	64	64	44	0.372093023	0.372093023
Plant	0	100	0	48	44	23	0.417391304	0.382608696
Plant	0	100	0	44	39	18	0.435643564	0.386138614
Plant	0	100	0	53	55	34	0.373239437	0.387323944
Plant	0	100	0	44	49	33	0.349206349	0.388888889
Plant	0	100	0	56	55	30	0.397163121	0.390070922
Plant	0	100	0	55	60	35	0.366666667	0.4
Plant	0	100	0	33	40	25	0.336734694	0.408163265
Plant	0	100	0	60	63	31	0.38961039	0.409090909
Plant	0	100	0	51	55	25	0.389312977	0.419847328
Plant	0	100	0	36	44	24	0.346153846	0.423076923
Plant	0	100	0	47	54	25	0.373015873	0.428571429
StripeRust	100	0	0	63	44	15	0.516393443	0.360655738
StripeRust	100	0	0	64	49	20	0.481203008	0.368421053
StripeRust	100	0	0	79	53	9	0.560283688	0.375886525
StripeRust	100	0	0	69	52	15	0.507352941	0.382352941
StripeRust	100	0	0	65	53	20	0.471014493	0.384057971
StripeRust	100	0	0	79	60	9	0.533783784	0.405405405
StripeRust	100	0	0	63	54	16	0.473684211	0.406015038
StripeRust	100	0	0	73	60	9	0.514084507	0.422535211

[a] User selects values for a color that represents that category in the –pd.ppm output file.
[b] Using a “color picker,” user records the %R, %G, %B values for the range of colors in that category. There are no pre-set numbers of standards that are necessary for a category.
[c] The 2-D projection shown in Fig. 2 is useful in deciding whether additional standards will be helpful or not.

Fig. 2. The two dimensional projection of standards shown in Table 1[a].



[a] Note that standards in each category form a cluster and that none of the clusters overlap.

Frequently asked questions

How many standards should I select for each category? This depends upon the range of %RGB values within each category and how similar the values are in two different categories. For example, addition of more points within any of the clusters above would be neither helpful nor harmful (except for increasing computer processing time). Indeed, nothing is gained by having some of the points (particularly the internal points) in all three clusters above. However, when selecting standards from a range a colors in a collection of images, it may be easier to simply pick the entire range of perceived colors without focusing on the number of standards.

I'm not satisfied with pd's recognition of categories. Would addition of more standards help? The key to answering your question is to compare the original image and the –pd.ppm file at high magnification. Find locations on the original image in which pd assigned an inappropriate category. Add the %RGB standard(s) to your Excel file and note where the new points are in the 2-D projection. If standards from multiple categories overlap, then pd is not appropriate for your analysis. If you need to simply expand a cluster or to better define the margin of a cluster, then addition of more standards will help.

How does pd compare with other image analysis software? "Assess," a program sold by the American Phytopathological Society for quantification of plant disease is easier to use than pd. However, Assess uses a simpler algorithm and was unable to reliably differentiate between stripe rust pustules and the (whitish) necrotic tissue shown in Fig. 1E. Neither program was able to reliably differentiate between necrotic and non-necrotic plant tissue.