GraphScan 2.1 Manual

Table of Contents

1 Introduction

The GraphScan scans graphs in bitmap format to (x,y) data. The output can be used in spreadsheet programs or other scientific programs. The GraphScan is a freeware program but the source code is not. You may distribute this program freely but any modification, disassembling or using some parts of it without an agreement with the author is not allowed. I'll hope you enjoy using this program and find it useful for your scientific or other purposes.

GraphScan 2.1 is the third major release of this program. The features listed below are introduced in this version:

Save/open scan file   The scan data and most of the program options can be saved in a scan file (XML format). The scanning parameters are importable to the program from a scan file.

Tabbed sheets   The Image, Options, Scan, Data and Info windows are on different sheets which can be easily switched with the keyboard or mouse.

Origin reference   The origin reference mark is now cross styled.

New threshold options   It is now possible to use filter outside, inverse, minimum thresholds for the scan.

RGB thresholds   Red, green and blue channels can be filtered with their own threshold values.

Missing data points   The missing point mode enables to interpolate (linear) or fill constant values for missing points.

Missing data graph   Missing data can now be plotted separately on the graph sheet and saved in the data file as a new column (Calculated).

Multiple graphs   It is possible to show real, sdev, sdev/y and bitmap graphs on the same chart sheet. A logarithmic scale option is also available.

Filter preview   This program has two windows, where black and white, gray and RGB filtering previews can be generated.

Copy selected data   It is now possible to copy a selection of the scan data from the Data sheet.

1.1 System Requirements

The GraphScan 2.1 (ver. 2.1.1) is compiled with the Pentium instruction set option. Therefore at least a Pentium processor is needed. This version copies the whole image to a large 2 dimensional data array in order to enhance calculation speed. As a result of this, more memory will be used by the program.

Operating system requirements:

Windows XP (the program has not been tested on other versions of Windows)

1.2 Program Installation

Two installation options exist for this program: Microsoft Installer package (GraphScan 2.1.msi) and without an installer (graphscan2.zip). In order to run the packed version of the program, extract the files from the archive to a directory on your computer and run the graphscan2.exe. Both packages are downloadable freely from the following Web address:

http://www.chemistry.hut.fi/graphscan

2 How to Use this Program

2.1 Before You Start to Scan

Before you start to use this program, you must have a graph in a bitmap format. The minimum size of the image is 2x2. The color depth can be anything, but it is recommended to use gray scale images in favor of B&W images. You can also use color images which are interpreted as grayscale images by the program. The image can be opened from a file or copied from the clipboard.

2.2 Setting the Scanning Parameters

Selecting the Reference Screen Coordinates

The reference coordinates can be chosen with mouse when the image is loaded. The first step is to choose the reference origin. The next step is to choose reference X and Y coordinates. The last step is to mark the scanning area. All of these steps can be done with the mouse when the image is visible or by entering the coordinates manually to the *Options* sheet. The following table shows how to select the reference coordinates with the mouse:

Description	Coordinate(s)	Toolbar button or procedure
Clear the scan area	-	
Lock or unlock the scan area	-	
Scan start	SX_0 and SY_0	Mouse button 1 pressed down on the picture
Scan end	SX_1 and SY_1	Moving the cursor to the desired end position and releasing the mouse button 1
Reference origin	X_0 and Y_0	
Reference X	X_1	
Reference Y	Y_1	

Entering the Real Reference Coordinates

All other scanning parameters except the screen coordinates must be entered or chosen from the *Options* sheet. The real reference origin values are initially set to (0,0) which can be changed to any real value. The real reference X_1 and Y_1 fields are initially set to (1,1) which can be changed to any other real values than the real origin values.

Options Sheet

The scanning parameters can be entered or chosen from the *Options* sheet. This sheet is divided in four areas: Coordinates, Program Options, Parameters and a small Preview window. From the first area one can choose or adjust the coordinates. The second area defines how the program behaves on different tasks, which preview windows will be generated and gives the user options to perform fundamental tasks. The Parameters are separated in 6 different sub sheets where to user may choose all scanning and filtering parameters except coordinates. The preview window indicates how filtering will affect on the scanning area.

General Scan Parameters

The *scan width* specifies the width of the scanning rectangle. The minimum value is 1, which scans each vertical lines of the scanning area. A larger width decreases the amount of output data points and reduces the effect of random noise.

Another parameter that is used to minimize the effect of random noise is *threshold*. All points which colors are below threshold are ignored by the scan procedure. The initial value for the minimum threshold is 0 and for the maximum threshold 128 which equals 50% gray. The minimum threshold value 0 stands for a white pixel and the value 255 (maximum threshold) stands for a black pixel.

Other filtering options are described in the Scan Thresholds and Filtering section.

Weighted X points option is unchecked as default. It means that X points are middle points of the scanning rectangles. If the option is checked, the X points are weighted means of the scanning rectangles.

Weighted Y points option is checked as default. This means that weighted coefficients are used when calculating Y coordinates for the points. This option reduces the effect of background noise when scanning gray scale or color images. If the option is unchecked, all non filtered pixels are assumed as equally important.

RGB Scan Parameters

The red, green and blue filter can be used in order to extract a desired colored graph from the image. The color channel value of the pixel (red, green or blue) gives the weighted coefficient for the scan in this mode. A threshold value of 0 means black and 255 means 100 % red, green or blue.

Clipboard and Text File Options

Copy the output to clipboard checkbox is checked as default. If this option is checked, the scan output will be copied to the clipboard automatically after a scan.

Include scan parameters checkbox is by default unchecked. When this option is checked, a list of scan parameters will be copied or saved with the scan data.

Scanning Parameters checkbox is by default unchecked. It specifies whether the scan parameters will be included to the output or not.

Include column names checkbox is by default unchecked. When this option is checked, the column names will be copied or saved with the scan data.

The *Delimiter* field specifies which string will be used as a delimiter in the output. It is by default set to the semicolon character. The length of this field can be between 0 and 8 characters.

Calculation Selections

These options enable or disable calculation of real coordinates, sdev, sdev/y or Bitmap coordinates. From the *Primary Calculation* it is possible to choose which calculation is the primary graph (enabled) on the *Scan* sheet.

Real Coordinates checkbox is by default checked. It specifies whether the real coordinates are included to the output.

SDEV checkbox is by default unchecked. It specifies whether the SDEV (standard deviation) of Y field is included to the output.

SDEV/Y checkbox is by default unchecked. It specifies whether the SDEV/Y field is included to the output. The SDEV/Y value will not be calculated if Y equals 0.

Bitmap Coordinates checkbox is by default unchecked. It specifies whether the bitmap coordinates are included to the output.

Data Column Labels

The default data column names can be changed from these fields.

Missing Data Points

The missing point mode enables to interpolate (linear) or fill constant values for missing points.

Image Info

On this information sheet are listed name of the image, width, height and color information.

2.3 Scanning the Graph

After the scanning parameters are successfully defined, the *Scan* command is executable. The program starts to scan the graph in the scanning area after the *Scan* command. When the scan is complete, the program finalizes some calculations and generates the output. When all is done, a "Scan finished" text appears on the status bar.

2.4 Viewing the Result of the Scan

The GraphScan has two scan result sheets: *Image* and *Data*. The *Scan* sheet displays the scan result as a (x,y) scatter graph. The *Data* sheet shows the scan result as numerical data on a spreadsheet table from where it can be copied to the clipboard. The *Image* and *Data* sheets can be used as visual tools to check the goodness of the scan.

2.5 Using the Scan Data

When the data is copied to the clipboard, it is easy to paste it to a spreadsheet program. The data is represented in columns, which has to be split to multiple columns. The fields are though separated with a delimiter, so it is easy to split them with the spreadsheet program. The output data can also be saved to a text file.

3 Scan Thresholds and Filtering

In this program filtering means that the point will be rejected if its color is in a filtering region. Normal and outside filtering principles are shown in the following two pictures.
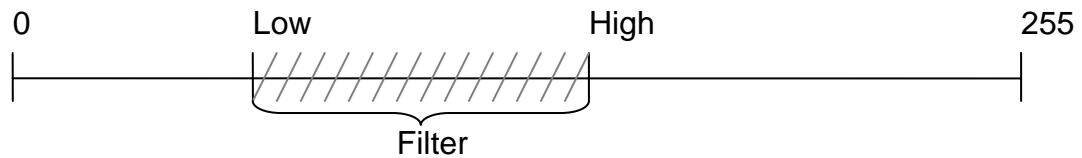
Normal filter (filtered points are inside threshold boundaries)
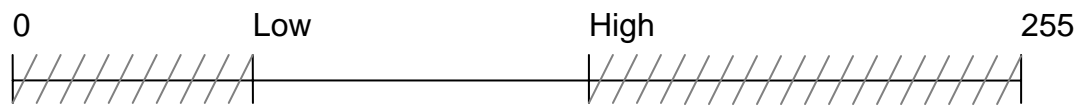
Outside filter (filtered points are outside threshold boundaries)

At the inverse mode the scale is inverted (value=255-value).

4 The Scanning Procedure and Statistics

The basic principle of this program is to select a scanning <u>area</u> from the image. The actual scan starts from the upper left corner of the scan area and proceeds column by column towards the bottom right corner of the scanning area. The scan width defines the width of the scanning rectangle (column). Only one data point or none is calculated inside the scanning rectangle.

Each point of the scanning rectangle is red by the procedure. Following steps describes the principles how to determine the position of the pixel inside the scanning rectangle.

1. Calculate how many scanning rectangles are needed (=maximum number of data points in the scan)
2. Scan all points inside the scanning rectangle:
   - Read a pixel and calculate the average of R, G and B values (=average grayscale)
   - Invert the value
   - If the value is less than threshold ignore the point and scan the next point
   - Use the value as weighting coefficient when calculating the weighted mean
   - Add the calculated value and the screen coordinates to their sums.
   - Scan the next point
3. If there was a point or points inside the scanning rectangle do following things:
   - Calculate the weighted mean to Y and convert it to a real coordinate
   - Calculate X as a middle of the scanning rectangle or as a weighted mean and covert it to a real coordinate
   - Calculate standard deviation to the Y point
4. Scan the next rectangle until all rectangles has been scanned

<u>The standard deviation</u>
In order to estimate the goodness of any scan, the program calculates weighted standard deviations: SDEV and SDEV/Y. The latter is a relative weighted standard deviation in respect to Y. The SDEV (weighted) is defined in this program as

$$SDEV_W = \sqrt{\frac{\sum_{i=1}^{N} \left(\frac{a_i}{\bar{a}} x_i - \bar{x}_W \right)^2}{N}} \, ,$$

where N is the number of points equal or above threshold,
   a_i is the weighted coefficient for the pixel,
   $\bar{a}$ is the mean of the weighted coefficient inside the scanning rectangle,
   x_i is pixels coordinate and
   $\bar{x}_W$ is the weighted mean coordinate for the pixel.

5 The Menu Items

5.1 File

File|Open Image…
This command opens a *File Open* dialog which is used to load a graph image in BMP format to the program. The size of the image must be at least 2x2 pixels.

File|Open Scan…
This command opens a *File Open* dialog which is used to load a scan file in XML format to the program.

File|Save Scan…
This command saves the scan data and program options to a file (XML format).

File|Save Data…
This command opens a *File Save As* dialog where the user can save the output of a scan to a text file.

File|Exit
This command terminates to program.

5.2 Edit

Edit|Copy Data
This command copies the scan output to the clipboard.

Edit|Paste Image
This command loads a graph image from the clipboard to the program. The image must be in BMP format with a minimum size of 2x2 pixels.

5.3 Scan

Scan|Options…
This command opens the *Options* sheet.

Scan|Preview
This command generates a filter preview image of the scanning area if the preview option is enabled on the *Options* sheet.

Scan|Scan
This command performs a scan if all parameters are successfully validated.

5.4 Help

Help|GraphScan Info
This command opens the *Info* sheet of the program.

Help|About
This command displays the *About* box of the GraphScan program.

6 Troubleshooting

If you have any question, find a bug in the program or you have any ideas that could improve this program, please send me an e-mail to the following address:

Kim.Miikki@tkk.fi