

TJHSST Computer Systems Lab Senior
Research Project
Reverse Engineering Graphs: Obtaining Data
Points from Scatterplots
2008-2009

Maya Wei

October 31, 2008

Abstract

Various programs exist to take data points and use them to render a graph. However, once the data are put into visual form, there is a loss of numerical information if the original data cannot be obtained. This project seeks to take data from a graph; in essence, the purpose is to reverse engineer a given graph. This will provide for a set of data points which can be used for various other numerical purposes, not simply the graph form in which they are presented.

Keywords: image analysis, scatter plots, statistical graphs

1 Introduction - Elaboration on the problem statement, purpose, and project scope

1.1 Scope of Study

In trying to read a graph, various image analysis techniques will be used: edge analysis, and from that, shape recognition – or at the very least, shape differentiation. The intention is to develop a method which will be capable

of successfully reading a point's location to its axes regardless of point shape: various graph-creating programs (OpenOffice Calculate, Microsoft Excel) use different formatting with different colored backgrounds, different guidelines, and different shapes of points.

1.2 Expected results

The results expected are an accurate recreation of the points which were utilized to create the graph. With these results, it would be possible to represent the data in other graphs, make statistical calculations, etc.

1.3 Type of research

The program is created for personal use; it is applied research.

2 Background and review of current literature and research

The field of image analysis and computer vision is highly advanced at this point in time. There have been many papers written on shape identification, image recognition, and graphic rendering. While there is no project similar to what is being done here with graphs and data points, undoubtedly somebody has sought to do it before. I will simply utilize basic techniques of image analysis.

3 Procedures and Methodology

This project utilizes Java. As of first quarter, I have created a working GUI, worked with various image readers, and have created a very simplistic method to identify edges. From here, the goal is to utilize these found edges to split them up between guidelines, axes, numbers, and points.

The input data is to be found in the form of a graph. Various graph inputs – square points, triangle points, circle points – will be tested to as to determine the accuracy of the program.

Currently, the program has a very primitive form of edge detection. It takes the color of the point at $(0, 0)$ and makes the assumption that the points color is also the background color. It then looks at each individual point (x, y) and the points neighbors $(x+1, y)$, $(x-1, y)$, $(x, y+1)$, $(x, y-1)$. If any of the points neighbors are the same color as $(0, 0)$, then that means that (x, y) is touching an edge. This method will become more developed throughout the next three quarters.

When actual data is obtained, the error analysis will be comparing the original data set to the accuracy of the data set produced from the graph.

4 Expected Results

The desired result of this program is to find data points from a given graph. The returned result will be displayed in a chart.

While application of this field is slim, the various image analysis techniques could be utilized. One could also expand upon the concept to be able to read bar graphs, pie charts, line graphs, etc.