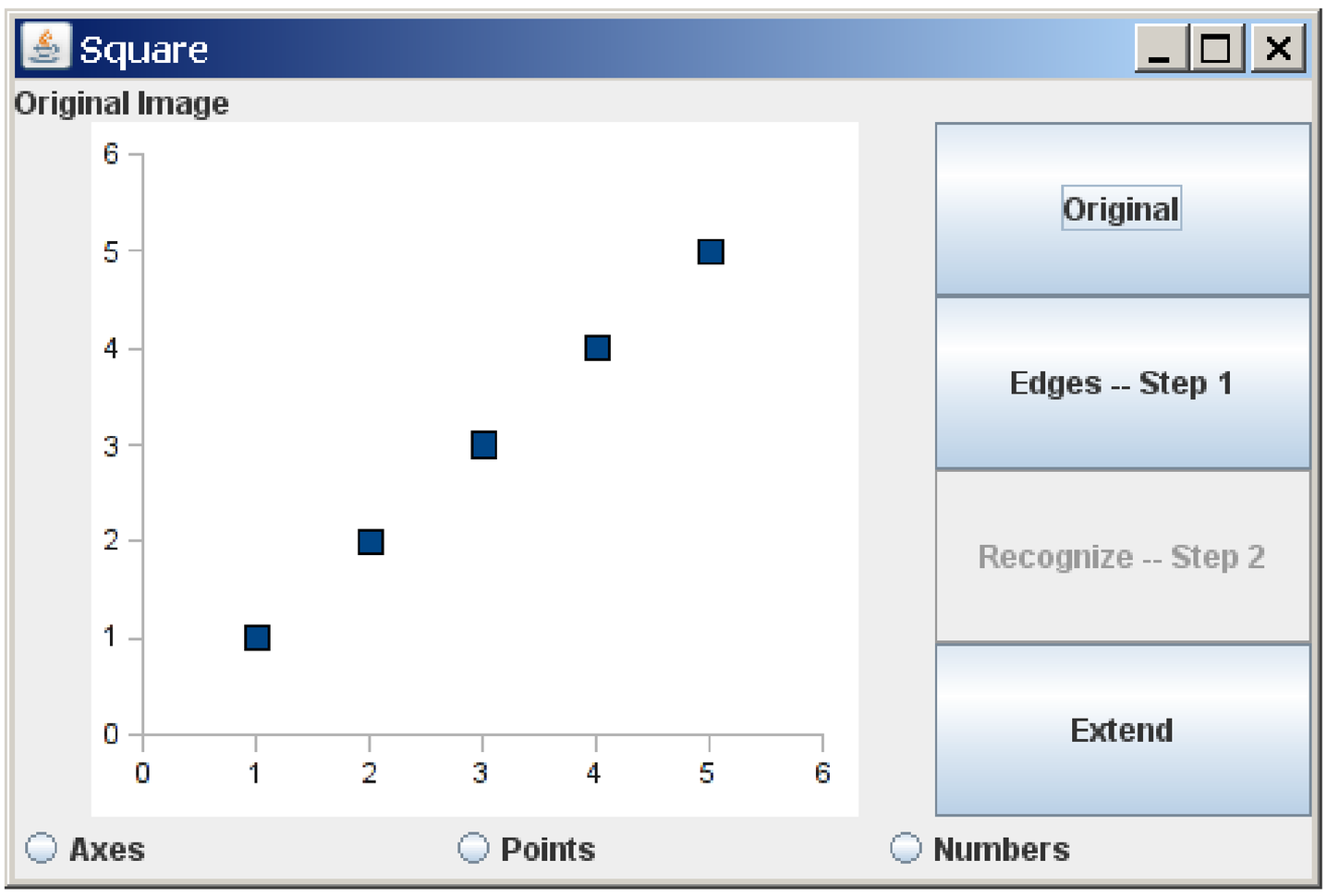
Reverse Engineering Graphs: Obtaining Data Points from Scatter Plots

By: Maya Wei

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.



Background: The project deals in basic methods of image analysis and shape differentiation. It relies much on edge detection, which can be from very basic techniques to very advanced algorithms. In addition, factors such as connectivity and image recognition are and will be used. The graph used was created in Open Office Calc; there are no gridlines and no colored background. The graph itself is very minimalist. There are points located at (1, 1), (2, 2), (3, 3), (4, 4), and (5, 5). The image is saved as .png.

Methodology: Currently, all the "edges" of the images are found – edges are defined as pixels different from the background color which are adjacent to at least one pixel of negative space. This list of edges are then sorted into connected points – for example, the edges around (1, 1) would be connected to each other. These connected points are then sorted into three categories: axes (subsets: x-axis, y-axis, tick marks), points, and numbers.

Taking the center of each connected point "edge," we can extend the points to the axis, from which we can then calculate the point's relation to the graph.

Reading the scale of the scatter plot has yet to be implemented.

Expected Results: With the point's relation to the axis discovered, the next step is to read the axis scale; subsequently, the output generated should ideally be a returned array of [(1,1), (2,2), (3,3), (4,4), (5,5)].