



Auto-digitizer for Graph to Data Converting

Sarah Radzihovsky, Deepti Mahajan, Fiona Wang:

sradzi13@stanford.edu, dmahaj@stanford.edu, chwang9@stanford.edu

EE368/CS232:

Winter 2018

Introduction

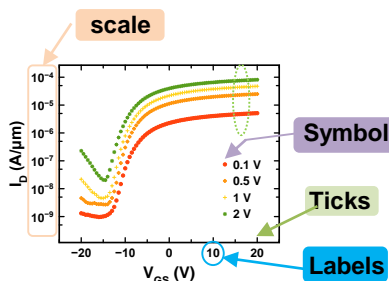
All available digitizers require manual selection for defining axis coordinate and data points. This manual work is very tedious and time-consuming.

Therefore, we focused on developing an auto-digitizer that can process most figures from literature without manual handling.

Dataset

Generated datasets with different...

- **Scale:** log/linear y-axis
- **Symbols:** different shapes
- **Ticks:** major/minor
- **Labels:** format of number



Approach

Axis Segmentation

- Find x and y axis with horizontal/vertical Sobel filters

Label Recognition

- Auto-detect label number with OCR (optical character recognition) function in Matlab

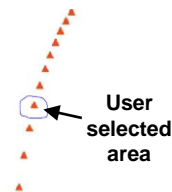
Assign data value to the ends of x and y axis

- Find label box center on the axis and assign its value

Pixel and x-y Value Mapping (for both linear and log scale)

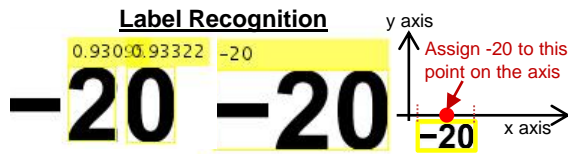
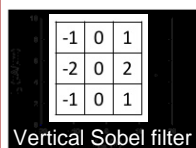
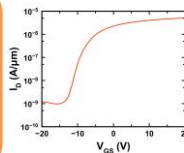
Symbol Recognition

- User manually circles the targeted symbol
- Extract the symbol from the drawing area
- Slightly erode the selected symbol
- Process "Close" operation with the plot
- Erode plot with extracted symbol



Line recognition

- Crop image to remove tick marks
- Find linearly spaced columns at desired resolution
- Threshold and search for nonzero values along each column



Experiment Results

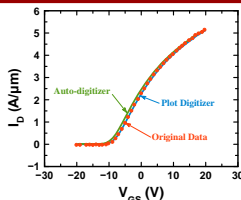
Plot Digitizer



Runtime: 80 s

Our Digitizer

Runtime 20 seconds (Only need one manual crop)



- The whole process costs < 20 s
- Much more efficient for massive data processing

Requirements:

- Assume ticks at center of label
- Limited background noise
- Linear x-axis

Conclusion and Future Work

Compared to the available digitizers for graph to data conversion, our auto-digitizer...

- Auto-locates the center of the symbols
- Recognizes label value without user input
- Does not need manual clicking for min/max x & y value

Future work:

- Calibrate the small offset in data mapping
- Increase robustness by eliminating any background noise

Acknowledgements: We would like to thank Prof. Girod, Jayant Thatte, and Jean-Baptiste Boin for their guidance this quarter.