Name(s) **Jacob D. Andreas**	**Project Number** # S1201

Project Title

The Optimum Aspect Ratio for Compressing Image Files of Text

Abstract

Objectives/Goals

This experiment measured whether non-square compressions were more effective than square ones at compressing image files of text. As digital storage of documents grows, a highly efficient image format specifically designed for text would be especially useful.

Methods/Materials

52 samples of letters in two different fonts, each 100*100 pixels, was prepared.
A Java scaling utility was used to compress these samples to various rectangles of the same area and then decompress them. An OCR (Optical Character Recognition) program was then used to measure the effectiveness of the compression by testing recognizeability of each sample.

Results

It was predicted that a non-square compression would be more effective than a square one. This was found to be false when first tested (reduction to 400 pixels), but then found to be true under a more extreme compression (reduction to 144 pixels). An 18*8 compression was most effective for the 144 pixel reduction, and a 20*20 compression was most effective for the 400 pixel reduction.

Conclusions/Discussion

Compressions reducing the image to 4% of its original size were found to be most effective when square; however, when the images were compressed to 1.44% of their original size several non-square compressions were more effective. This likely arises because rectangular compressions are better than square ones at preserving most of the vertical and horizontal strokes that make up letters. While the overall effectiveness of a compression depends in part on how the quality is judged, non-square compressions tend to be most effective when the degree of compression is extreme.
The success of this experiment in showing the effectiveness of non-square compressions suggests expanding research on this topic, including implementation of other OCR programs and samples in more fonts.

Summary Statement

This experiment tested whether the linear characteristics of letters can be exploited in image compressions that use non-square compressions instead of relying on averaging pixels in a square area, as most do now.

Help Received

Some help with analysis of data from Greg Andreas and Dr. Eric Neufeld.