Huffman Compression: An Analysis of the Compression Effectiveness

**An Analysis of Compression Effectiveness**

Using this compression algorithm, we were able to compress practically any file that could be stored on a computer. An interesting question can then be posed by this wide-range of possible file types; do certain types of files compress better than others? In order to address this problem most analytically, we chose to assess the differences in compression between .tiff and .txt files.

We found that .txt files actually compress better than .tiff files on average. Figure 1 shows the spectrum of compression degree as the number of characters (out of the 256 possible for the 8 bits per letter that we used) increases. Whenever the files use a similar number of the binary sequences out of the possible 256. As competent reader might notice, the low letter count TIFF files compress about as much as the average TXT files, but most TIFF files are high letter count files, so the average compression rate is much lower than the average TXT file compression rate. Table 1 outlines the basic statistics that were calculated during the analysis of the two different file formats.

Figure : Comparison of Compression as a Function of Letter Count

|  |  |  |
| --- | --- | --- |
|  | TXT Files | TIFF Files |
| total % compression: | 43.241 | 18.137 |
| Compress bits / sec | 158037 | 384205 |

*Table 1: Comparison of Compression Statistics Between File Formats*