In this section, the preprocessing of the images are shown. The preprocessing is preformed to enable easier classification for the classifier. Several preprocessing tools were explored to make the character easier to recognize for the classifier. The first preprocessing tool applied was total-variation denoising on n-dimensional images [**?**]. This technique reduces the total variation of an image. This is useful to avoid classifying noise, as the picture will now contain fewer components with high spacial frequency. An example of this denoising technique is shown in the image Figure 1b.

To increase contrast, we used Otsu's method, which is a way to perform image thresholding based on clustering [**?**] automatically. After thresholding, the image was morphologically closed. This is a standard technique in image processing to remove small wholes in binary images. The result is shown in Figure 1c. All though the result was visually pleasing; it did not decrease the error rate. Therefore, this method was discarded from further analysis.

The last preprocessing tool applied before dimention reduction was Histogram of Oriented Gradients. It is a feature descriptor uses local object appearences and shapes and describe them as a distribution of intensity gradients. It can be interpreted as the spacial derivative of the intencity of the image. The purpose of this step is to format the image for easier detection later as each element now carry more information than the pixels in the original.

0.0.1 Principal component analysis

It is a substantial challenge and computationally heavy to perform classification on high dimensional data. Therefore, a dimension reduction tool is preferable before performing a k-Nearest Neighbor Classifier. In this project, the Principal Component Analysis procedure was used for reducing the number of correlated dimensions down to a set of the linearly uncorrelated principal component. The optimal condition for the k-Nearest Neighbor Classifier is when PCA is used to reduce the dimension down to

figures/pp/raw.pdf

(a) Raw image.

figures/pp/denoise.pdf

(b) Denoised image.

figures/pp/img_to_bool.pdf

(c) Binary image.

figures/pp/hog.pdf
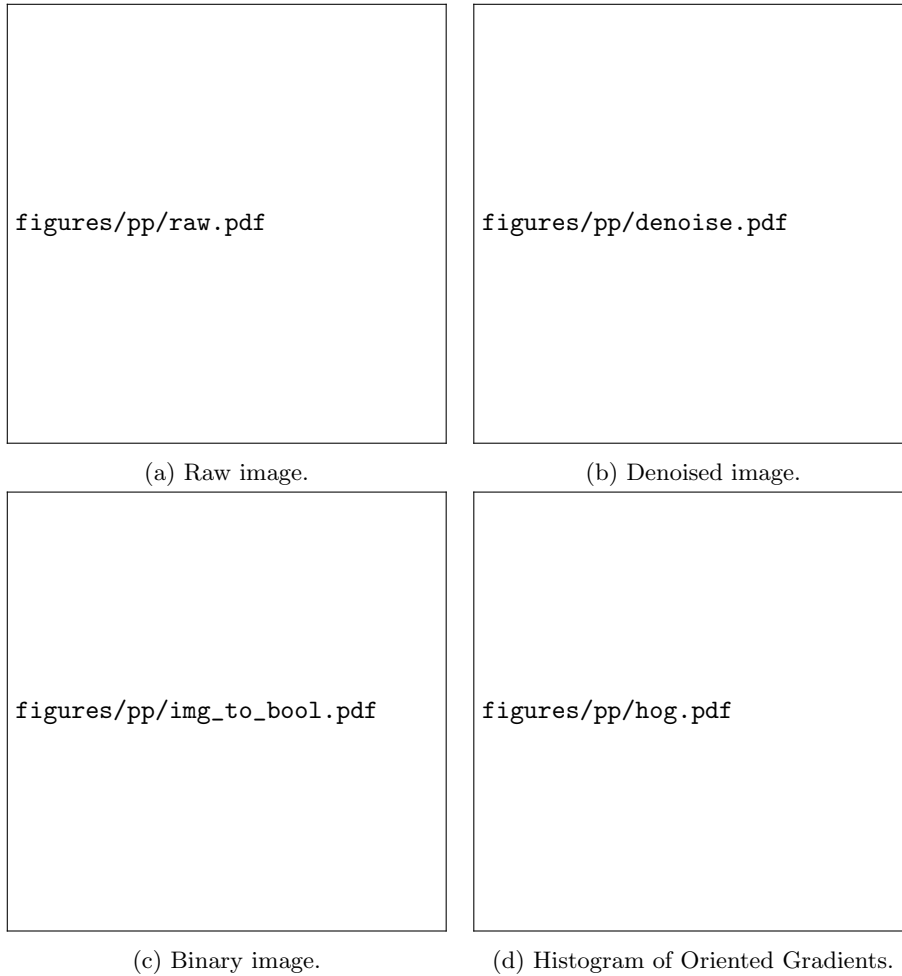
(d) Histogram of Oriented Gradients.

Figure 1: Example of random selected image before and after preprocessing. The colors has been altered using a color map to visually enhance variance.