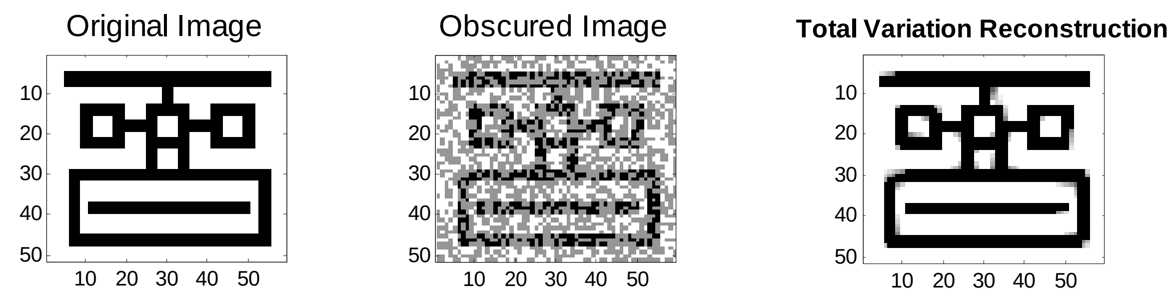
Christopher Tsai
June 6, 2008
Honors Thesis Abstract

Learning to Understand

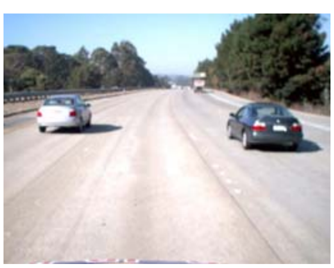
Mathematical Methods of Training a Computer to Extract Visual Information

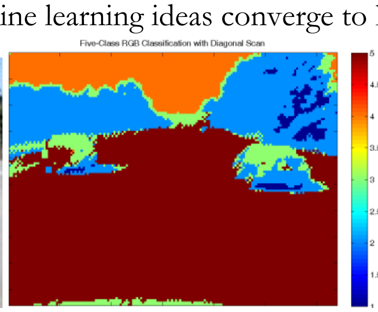
How much information does an image contain? The entropy of the image, often used to determine image compressibility, may quantify the randomness or variability of image intensity, but it seldom parallels human intuition gained from viewing an image, particularly when the underlying distribution is not mathematically but rather intuitively ordered. We would like mathematical quantitative information to match intuitive qualitative information, but the computer has no inherent framework to interpret images like we do. This thesis will explore and expound various techniques and heuristics that we might use to translate our mode of extracting information from images into computer algorithms. We will juxtapose not only the merits of many mathematical methods but also alternative forms of representation, such as the 2D Fourier transform, the discrete cosine transform (DCT), the 2D wavelet transform, eigenimages (principal components), the Hough transform, and Kolmogorov complexity.

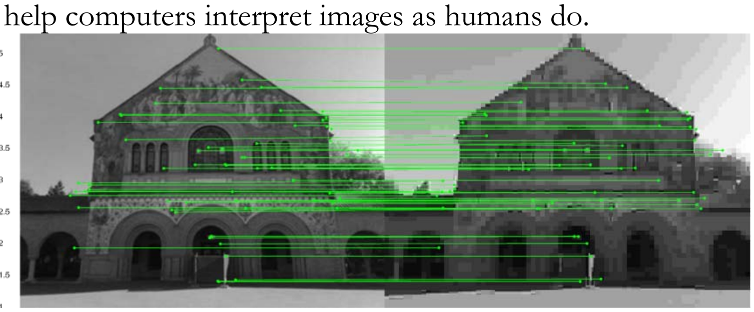
First of all, we consider image denoising. How can a filter distinguish noise from image detail? Can we impose a Wold decomposition on an image, cleanly separating unpredictable random noise (a regular signal) from repeatable content (a strongly predictable signal)? Can computers *classify* noise? Would a frequency domain or wavelet coefficient representation help us denoise an image? How might we improve the results of Kalman filtering? We assay a number of traditional denoising filters, formulate denoising as a convex optimization problem, and apply a number of denoising heuristics, such as DUDE, cubic splines, and total variation image interpolation.



Unsupervised learning also represents a family of techniques designed to extract underlying order from an array of pixel intensities. Computer scientists literally *train* computers to classify blocks of pixels in an image based only on observable features such as color and texture. Can a computer possibly resolve class structure better than humans? How might we best prepare a computer to segment and separate recognizable objects as humans would, into categories we find as intuitive as manmade-vs.-nature or land-vs.-sky? We investigate how traditional image processing tools and machine learning ideas converge to help computers interpret images as humans do.







Finally, we investigate identification in a variety of contexts, including (but not limited to) automatic content-based image retrieval from a database of CD covers, segmentation and recognition of famous artwork, emotion detection on human faces, and feature extraction from planetary surfaces. We evaluate a number of techniques: exploiting statistical redundancy as in JPEG-DCT or wavelet-based image compression; convolving images with several known edge detection filters such as the Sobel and Canny; morphologically processing images with informative structuring elements; applying SIFT and SURF feature matching followed by geometrical consistency checking à la RANSAC; implementing forward recursion and particle filtering. Our quest for knowledge will take us through a myriad of approaches: Fourier optics, super-resolution interpolation, focused SAR imaging, alternative color spaces, Kolmogorov complexity, *k*-means, vector quantization in a hidden Markov model, histogram equalization, the Radon transform, projection-slice, eigenimages and principal component analysis, singular value decomposition, face tracking, multiresolution analysis, Karhunen-Loève transforms, Wiener filtering, apodization, and much more.