Reviewer: Robert Collins, September 9, 2007

Citation: Michael J. Jones, James M. Rehg, "Statistical Color Models with Application to Skin Detection", International Journal of Computer Vision, 46(1):81-96, January 2002.

Quick Summary: The authors learn a nonparametric model of skin color using over 13,000 hand-labeled images from on the web. Pixels in new images are labeled using a likelihood ratio test based on learned empirical distributions of skin and not-skin. The resulting classifier proved to be surprisingly effective (detection rate of 88% at false alarm rate of 11%) at quickly labeling skin pixels in arbitrary web images.

Ideas/Approach/Results: Histograms in RGB space were accumulated for skin pixels and for non-skin pixels, providing empirical probability distributions for each class. To classify pixels in a new image, each pixel is treated independently and labeled as skin or not-skin based on a likelihood ratio test using the learned distributions. The computed likelihood ratio value is compared against a threshold that includes prior information about the probability of a pixel being skin, and application-dependent costs associated with getting false positives and false negatives.

For comparison, the authors also used the EM algorithm to learn a 16-component mixture of Gaussian distribution for both the skin and non-skin cases. Using a likelihood ratio test with the MoG models did not yield any significant difference in classification performance as compared to using the raw RGB histogram models, although the storage size of the MoG descriptors was smaller.

The authors go on to use the pixel-level classifications in new images to try to build two higher-level classifiers, one that determines if there is a person in the image, and one to determine whether an image contains adult content. Both use simple features such as number of skin pixels, size of largest connected component, number of connected components, etc. The adult image classifier achieved detection performance of 86% with 7% false positives. When color features were combined with text features (using text from the web page on which an image was extracted), that rate increased to 94% with 8% false positives. These rates are significant given that no analysis of shape was performed to identify body parts, as others had done previously (e.g. Forsyth&Fleck).

Comparison/Strengths/Weaknesses: Although this paper shows an example of using EM to learn MoG distributions in RGB space, it is a straightforward application of the standard method of estimation (no novelty on the EM/MoG side).

An important lesson from this paper is that given a large amount of training data (nearly 1 billion hand-labeled pixels), a simple classifier based on likelihood ratio of empirical distributions can work surprisingly well to capture the properties of skin versus not-skin in images with unconstrained lighting conditions.

Questions/Issues: I am curious how a k-nearest neighbor classifier would work in this scenario. Given a large amount of training data, the k-NN classifier is known to yield a good bound on the underlying Bayes error rate for a problem.

Basing an adult image classifier on percentage of skin pixels in an image and size of largest connected component seems very weak, since any portrait photo of a person's face will rank high in both. Perhaps we can combine with the Viola&Jones face classifier (not available at the time this paper was written) to say that large face images are OK.