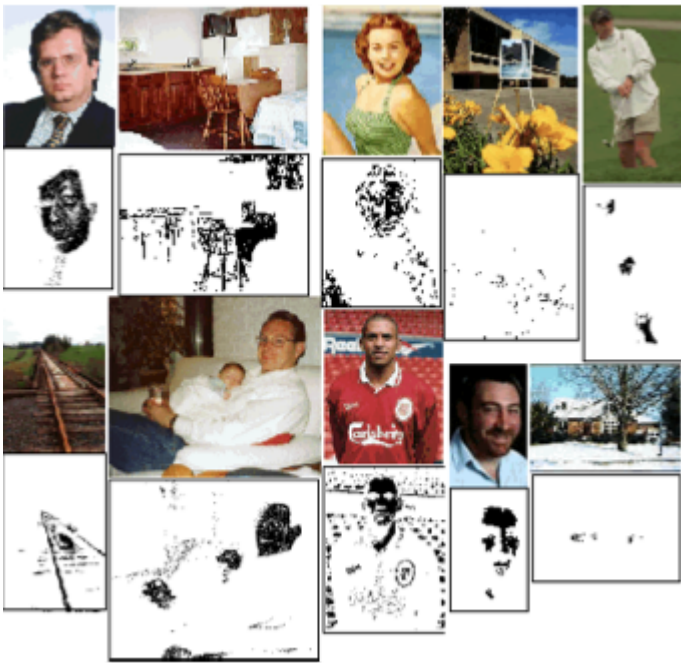
CSCI 4831/5722 Computer Vision, Spring 2019
Instructor: Fleming
Homework 6 - Extra Credit
Due Thursday, May 1st, by 11:55pm

Skin Detection

In this homework assignment, you will implement the skin detection algorithm outlined by Michael J. Jones and James M. Rehg in *Statistical Color Models with Application to Skin Detection,* 1998



Provided files:
A database of training labeled images is provided.

Source: the database was used in
A Fusion Approach for Efficient Human Skin Detection
W.R. Tan, C.S. Chan, Y. Pratheepan and J. Condell
IEEE Transactions on Industrial Informatics, vol.8(1):138-147 ***(T-II 2012)***

There are 3 folders:

1. **`Pratheepan_Dataset:`** contains color images grouped into 2 folders:
    a. **`FacePhoto:`** contains color pictures which portray (mostly) a single person on a simple background. Total images (.jpg): 25
    b. **`FamilyPhoto:`** contains color pictures of multiple subjects on complex background. Total images (.jpg): 40

2. **`Ground_truth:`** contains 65 binary (black-and-white) images representing labeling of skin and non-skin pixels. They are also grouped into 2 folders, with the same names as the original images, but of .png type.
3. **`Testing_data:`** contains 13 images of one or multiple subjects.


What You Have to Do:

Task 1 (20 points): Histogram of training data

Write a function `train_skin_data` that takes a path to the training data and the bin size, and returns **two histograms** based on RGB values for skin and non-skin pixels. Each histogram is a 3D matrix. For example, if the *bin size* value is 1, then each histogram will be a `256 x 256 x 256` matrix. If the *bin size* value iz 16, then each matrix will be of size `16 x 16 x 16.`

Note: Use vectorization as much as possible to identify and count the occurrence of every RGB triple. *Points will be deducted if no vectorization is used.* Recommended MATLAB function: `find.`

Task 2 (25 points): Testing data

The testing data consists of 13 color images. Write a function `test_skin_data` that takes a new color image, and two RGB histogram matrices (one for skin pixels and one for non-skin pixels), and returns a binary image that labels each pixel in the

input image as skin pixel (white) or non-skin pixel (black). The size of the histogram images should be taken into account in order to compute the bin size used in training.

Follow the algorithm outlined in the Jones & Rehg paper, and also the theory and example in the lecture slides. A pixel of color *rgb* should be labeled at "skin" if

```
P(skin|rgb) > P(non-skin|rgb)
```

Note: Use vectorization as much as possible to identify the occurrence of every RGB triple. *Points will be deducted if no vectorization is used.* Recommended MATLAB function: `find`.

Task 3 (5 points): Script

a. **(5 points)** Write a script that calls the function `train_skin_data` with a path to the training data and a bin size of 1. Use the resulting histograms and the `test_skin_data` function to generate the binary label image for each of the 13 images in the testing set. Save the binary images as .png files.
b. **(5 points)** Write another script and repeat the training function call but with a bin size of 16. Use the resulting histograms and the `test_skin_data` function to generate the binary label image for each of the 13 images in the testing set. Save the binary images as .png files.
c. Load the binary label images images generated at a. and b. and write a script to produce a difference image between the pair of binary images for each testing image. The resulting values for each pixel location should be either 0 (if the label is the same) or 1 (if the labels are different). Save the diff images as .png

Submitting the assignment:
Make sure each script or function file is well commented and it includes a block comment with your name, course number, assignment number and instructor name. Zip all the .m files, together with the .png files (there should be 39 binary images) and submit the resulting *.zip* file through Moodle as Homework 6 by Thursday, May 1st, by 11:55pm.

Total points: 50 for both undergraduate and graduate students.