License information, and how to attribute this work

This work is licensed under the Creative Commons Attribution-NonCommercial-ShareAlike License. To view a copy of this license, visit http://creativecommons.org/licenses/by-nc-sa/3.0/. Please pass this license along if you redistribute this code. And if you use any of this for published work, please cite:

Said and Todorov (2011). A Statistical Model of Facial Attractiveness. Psychological Science.

General information

I have tried to document the code as much as possible. Let me know if there are any problems or errors, but please understand that I won't always have time to give detailed technical support.

The characters "FrM" refer to male faces rated by female observers. The characters "MrF" refer to female characters rated by male observers. Also, what I refer to as the "reflectance" dimensions are called "texture" dimensions in FaceGen.

Key files

arrow.m
This matlab function can be found on http://www.mathworks.com/matlabcentral/fileexchange/278-arrow-m. It was written by Erik A. Johnson and distributed with a BSD license. Please refer to the license at the end of this document for fair use instructions. Required to run data_analysis.m.

averagefemale.fg and **averagemale.fg**
.fg files can only be viewed in FaceGen. These are the averages of the FaceGen real subject samples. Extracted from the .ctl file. For more details, see the Methods section of Said and Todorov (2011). For information about how these coordinates were extracted, see the readme in the FaceGen SDK.

data_analysis.m
This is the main Matlab function for running the analysis.
        It reads in the normalized attractiveness ratings and face space coordinates (all in .csv files) of the training sets and testing sets, and spits out some text files containing the coordinates for many of the faces shown in the publication.
        It also computes a variety of statistics reported in the publication, although these are not written to any file. They are just variables in Matlab. You can see their values by placing 'keyboard' commands in appropriate places in the code.
        This function follows the ordering convention: male faces then female faces. As such, the figures produced by this function have male faces on the left, and female faces on the right. We reversed this order in the publication.

exag_...txt
These text files contain the coordinates for certain faces output by data_analysis.m. The first 25 values are the first 25 shape values. The next 25 values are shape values for higher (basically negligible) dimensions not varied in the face set. The next 50 values are analogous, but for reflectance.

facemaker_general.py
Python file to creates facegen .fg format faces from .txt coordinate files. This code was heavily inspired by some code written by Ron Dotsch (http://ron.dotsch.org/). See also http://ron.dotsch.org/scripts/.

　　Using this script will really require getting you hands dirty! It requires the FaceGen SDK (http://facegen.com/sdk.htm). You will have to pay for this, but it definitely gets my endorsement for fast, automated manipulation of FaceGen files. It also requires Python and the NumPy package. I am using Python because I only had the SDK compiled for Windows, and I don't have Matlab for Windows. If you don't use Python, you could probably rewrite this script in Matlab for Windows using the system.m function. And maybe it is possible to obtain or compile the SDK for Mac. Not sure.

　　This script sets values for the principal components underlying the FaceGen model, and therefore requires a different si.ctl file from the default. Please contact me (cps6@nyu.edu) if you need this.

faces_training and **faces_testing**
These directories contain the .fg files used to train the regression, and the .fg files used to test the model after it was built.

faces_training_jpg and **faces_testing_jpg**
Same thing, but in .jpg format.

fig_...eps
These files are figures produced by data_analysis.m, and correspond to figures in the publication. In the publication, we put females on the left and males on the right, which is the opposite of how data_analysis.m works.

FrM_attractivenessratings_formatlab.csv
MrF_attractivenessratings_formatlab.csv
FrM refers to male faces rated by females. MrF is the reverse. These files contain data for the *training* set of faces. Each row is a face. Unlike in the data files for the testing faces, below, the first row corresponds to face 000. The first column of each of these files contain the normalized attractiveness ratings. The next 25 columns are the values or the first 25 shape dimensions. The next 25 columns are the values for the first 25 reflectance dimensions. These files are read by data_analysis.m and are used for the main regression.

readFG.m
Reads FG files into Matlab. Inspired by Ron Dotsch. http://ron.dotsch.org/scripts/

separatecomponents...txt
These text files contain the coordinates for certain faces output by data_analysis.m. The first 25 values are the first 25 shape values. The next 25 values are shape values for higher (basically negligible) dimensions not varied in the face set. The next 50 values are analogous, but for reflectance.

validationresultsFrM_formatlab.csv
FrM refers to male faces rated by females. MrF is the reverse. These files contain data for the *testing* set of faces. Each row is a face. *Unlike in the data files for the training faces, the first row corresponds to the average face. The second row corresponds to face 000.* The first column of each of these files contain the normalized attractiveness ratings. The next 50 columns are the values of the first 50 shape dimensions. The next 50 columns are the values for the first 50 reflectance dimensions.

      In both the training set and the testing set, the shape dimensions 26-50 and reflectance dimensions 26-50 were set to zero. These higher dimensions are pretty negligible. It is just historical accident that the zeros are listed in the testing set .csv file but not the training set .csv file. Data_analysis.m takes care of this.