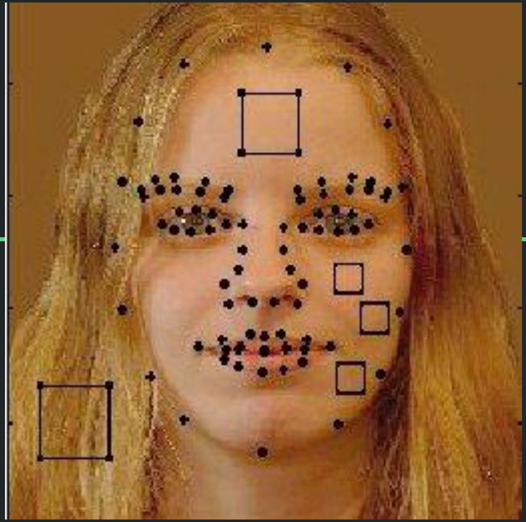
A Humanlike Predictor of Facial Attractiveness

Video Presentation



Shreya Pagaria (B20CH042)

What is the paper about?

(a) (b) (c)

The human face has been a great source of interest to scientists, researchers and philosophers because of the well developed ability of human to process, recognize and extract information from other's faces. Although, it is a matter of subjective opinion that people might share a common taste for facial attractiveness. The paper presents a method based using supervised learning and different tried and tested methods to predict facial attractiveness like a human would. The task of evaluating human attractiveness ratings adds the notion of *judgment of taste* to the previous achievements in machine perception of faces and it could also be the demonstration of computer's ability to succeed in a quantitative task, human judgement task.

This paper highlights many facial features to be considered for a building a method so that the computer can train and predict like us, humans describing facial geometry, skin color, texture, age, averageness of the face, hair color etc.

Different Hypothesis:

Shedding light on the different hypotheses we have; Rubenstein,Langlois and Roggman who was inspired by sir Francis Galton’s photographic method of composing faces and created averaged faces by combining multiple images together and suggested that averageness is the answer for facial attractiveness. On the other hand, Grammar and Thornhill concluded symmetry being more important than averageness of the face but then what if some faces have extreme features like sexually dimorphic traits, they than might be more attractive than average faces. Since, some facial qualities are universally attractive to people so it’s hard to conclude that whether the role is more of symmetry or averageness or any other feature.

Conventional thinking suggests that standards of beauty are formed from a gradually learned subjective process that is a product of the media; however, research conducted over the past decades contradicts these widely held beliefs. Theoretical and empirical work has attempted to understand face attractiveness through evolutionary models of signaling.

According to that view, facial traits imply mate quality and imply chances for reproductive success and parasite resistance, some say that the “good taste” in itself is an evolutionary adaptation. Another mechanism project attractiveness through a cognitive theory i.e. a preference for attractive faces might be induced because of general perception and or recognition mechanisms, attractive faces are pleasant to look at. Another view suggests that facial attractiveness originates in a social mechanism, depending on the learning history of the person and his/her goals. Many studies have tried to use computational methods to analyze the relationship between age, facial attractiveness and averageness. Another approach was made using a genetic algorithm which reflects the process of natural selection where the fittest individuals are selected for reproduction in order to produce offspring of the next generation.

A machine was tested using this algorithm and achieved a correlation of 0.6 with average human ratings showing that facial beauty can be learned by a machine.

GOAL OF THE PAPER?

We all know the challenge of constructing a facial attractiveness machine with human level accuracy has remained open. The primary goal of this paper is to excel the results in developing a machine which obtains human level performance in predicting facial attractiveness. The second goal is to conduct a series of simulated psychophysical experiments and study the resemblance between humans and machine judgements.

The latter task has two main potential rewards:
1. Determining if the machine can aid in understanding the psychophysics of human facial attractiveness.

2. To study whether learning an explicit operational rating prediction task also entails learning implicit humanlike biases, at least for the case of facial attractiveness.

Training Database: Rating Facial Attractiveness

The dataset consists of 91 facial images of American Females, all frontal color photographs of young Caucasian females with a neutral expression. They were of similar age, skin color and gender, had no distracting items or accessories like jewellery. All the 91 images were rated by 15 males and 13 females on a 7-Likert scale which is a scale used to represent people's attitudes to a topic.
Every rater was asked to view the entire dataset and there was no time limit for judging the attractiveness of each sample and they could even go back and adjust the ratings of already rated samples.
The final attractiveness rating of each sample was its mean rating across all raters. To validate that the number of ratings collected adequately represented the ``collective attractiveness rating" we randomly divided the raters into two disjoint groups of equal size. For each facial image, we calculated the mean rating on each group, and calculated the Pearson correlation between the mean ratings of the two groups. This process was repeated 1,000 times. Shedding some light on the Pearson correlation coefficient : it is the measure of the linear correlation between two sets of data. The ratio between the covariance of two variables and the product of their standard deviation thus it is essentially a normalized measurement of the covariance such that the result is between -1 and 1.
The mean correlation between the groups was 0.92. Final attractiveness ratings range from 1.42 to 5.75 and the mean rating was 3.33

Data preprocessing

The most effective method to capture the facial attractiveness is to measure proportions, distances and angles of faces. An automatic engine was developed to extract facial features which is capable of identifying eyes, nose, lips, eyebrows and head contour leading to 84 coordinates describing the locations of those facial features. The facial coordinates are used to create a distances-vector of all 3,486 distances between all pairs of coordinates in the complete graph created by all coordinates. For each image, all distances are normalized by face length. In a similar manner, a slopes-vector of all the 3,486 slopes of the lines connecting the facial coordinates is computed.

Combining the distances-vector and the slopes-vector yields a vector representation of 6,972 geometric features for each image. Since strong correlations are expected among the features in such representation, principal component analysis (PCA) was applied to these geometric features, producing 90 principal components which span the sub-space defined by the 91 image vector representations. The geometric features are projected on those 90 principal components and supply 90 orthogonal eigenfeatures representing the geometric features. Eight measured features were not included in the PCA analysis, including CFA, smoothness, hair color coordinates (HSV) and skin color coordinates. These features are assumed to be directly connected to human perception of facial attractiveness and are hence kept at their original values. These 8 features were added to the 90 geometric eigenfeatures, resulting in a total of 98 image-features representing each facial image in the dataset.

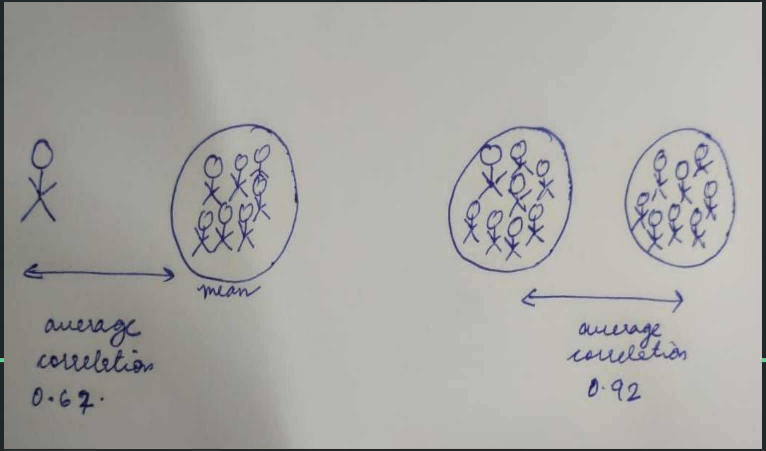
Experimentation

They experiment with Simple Linear Regression,Least Squares Support Vector Machine (LS-SVM) (both linear as well as non-linear) and Gaussian Processes (GP). However, as the LS-SVM and GP showed no substantial advantage over Linear Regression. Simple Linear Regression is a type of Regression algorithms that models the relationship between a dependent variable and a single independent variable. A key ingredient in our methods is to use a proper image-features selection strategy.

Predictors and validation.

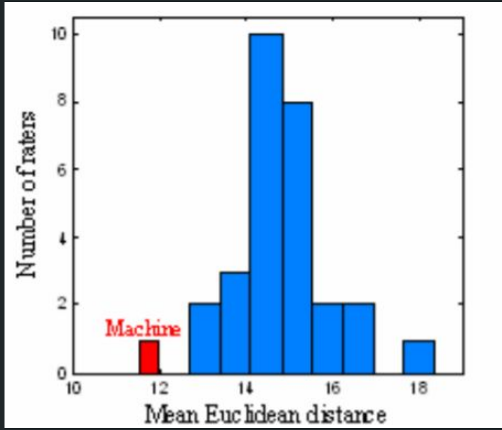
To check the performance of the whole dataset, they removed on sample from the whole dataset and the removed sample then served as a test set. Now, for each left out sample, the optimal number of image-features by performing leave-one-out-cross-validation (LOOCV) on the remaining samples and selecting the number of features that minimizes the absolute difference between the algorithm's output and the targets of the training set i.e. the score for a test set was predicted using a single model based on the training set. This process was repeated 91 times and then the predictions were compared with the true targets. These scores are found to be in a high Pearson correlation of 0.82 with the mean ratings of humans (P-value $< 10^{-23}$), which corresponds to a normalized Mean Squared Error of 0.39.

We know that for correlation, a p-value tells us the probability that randomly drawn dot will result in a similarly strong relationship or stronger. Thus, the smaller the p-value, the more confidence we have in the predictions we make.

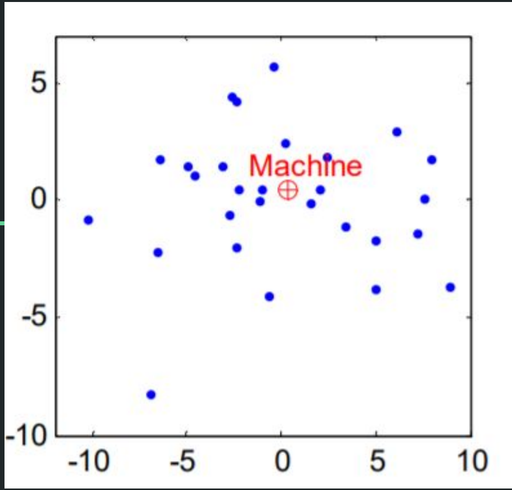


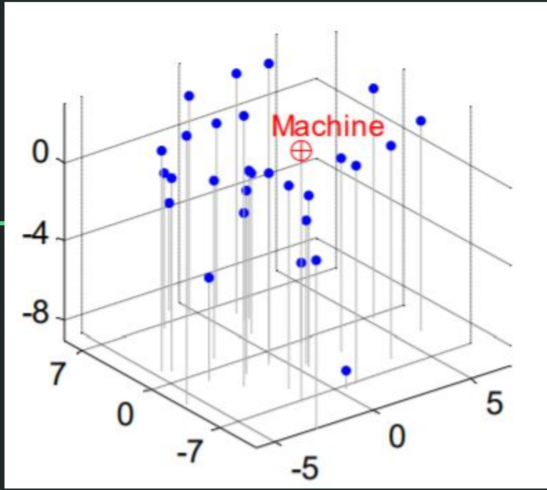
It should be noted that we tried to use this feature selection and training procedure with the original geometric features instead of the eigen features, ranking them by their correlation to the targets and selecting up to 300 best ranked features. This, however, has failed to produce good predictors due to strong correlations between the original geometric features (maximal Pearson correlation obtained was 0.26)

Similarly in judgments : Each rater has a 91 dimensional rating vector that can be presented in a 91 dimensional ratings space. The euclidean distance is computed and to verify machine ratings are not outliers they surrounded each of the rating vectors in the ratings space with multidimensional spheres of several radius sizes.



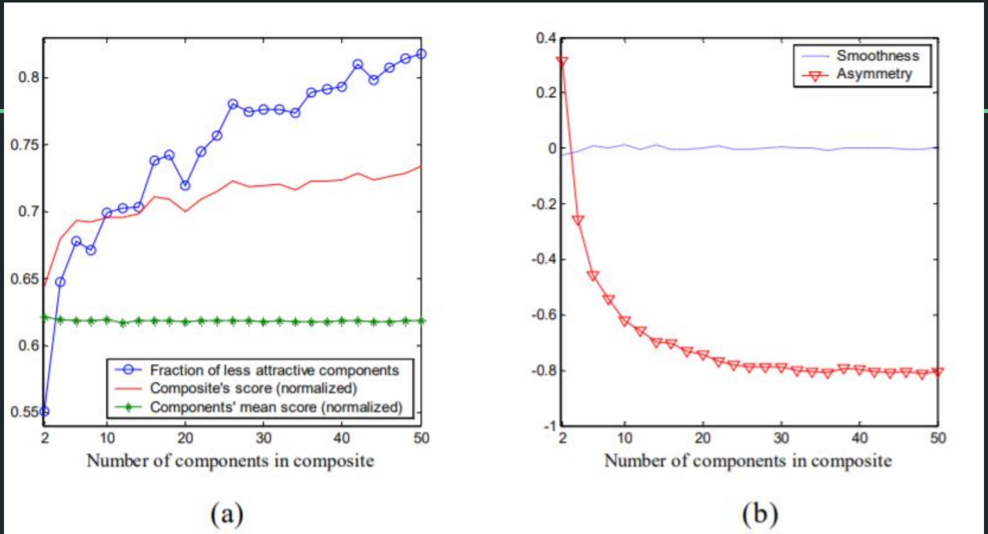
For display from the 91 dimensional vector to 2D we use PCA to machine and human ratings in the rating space and projected all ratings onto the resulting first 2 and 3 principal components.





Psychophysical experiments using computer simulations

Inspired by Rubenstein et al. , they used a similar technique to create averaged virtually-morphed faces with various numbers of components, nc, and have let the machine predict their attractiveness. To study the effect of nc on the attractiveness score we produced 1,000 virtual morph images for each value of nc between 2 and 50, and used our attractiveness predictor to compute the attractiveness scores of the resulting composites. In accordance with the experimental results of the machine manifests a humanlike bias for higher scores of averaged composites over their components' mean score.



In our experiment, a preference for averageness is found even though our method of virtual-morphing does not produce the smoothing effect and the mean smoothness value of composites corresponds to the mean smoothness value in the original dataset, for all nc. However, examining the correlation between the rest of the features and the composites' scores reveals that this high correlation is not at all unique to asymmetry. In fact, 45 of the 98 features are strongly correlated with attractiveness scores (|Pearson correlation| > 0.9). The high correlation between these numerous features and attractiveness scores of averaged faces indicates that symmetry level is not an exceptional factor in the machine’s preference for averaged faces. Instead, it suggests that averaging causes many features, including both geometric features and symmetry, to change in a direction which causes an increase in attractiveness. A virtual composite made of the 12 most attractive faces in the set (as rated by humans) was rated by the machine with a high score of 5.6 while 1,000 composites made of 50 faces got a maximum score of only 5.3.

Another study by Zaidel et al. examined the asymmetry of attractiveness perception and offered a relationship between facial attractiveness and hemispheric specialization. Interestingly, similar results were found when simulating the same experiment with the machine: Right-right and left-left chimeric composites were created from the extracted coordinates of each image and the machine was used to predict their attractiveness ratings

Conclusion

In this paper work, they were able to build a model that achieves humanlike performance for prediction of facial attractiveness.
The similarity between human and machine preferences has prompted us to further study the machine's operation in order to capitalize on the accessibility of its inner workings and learn more about human perception of facial attractiveness. To this end, we have found that that the machine favors averaged faces made of several component faces. While this preference is known to be common to humans as well, researchers have previously offered different reasons for favoring averageness. Our analysis has revealed that symmetry is strongly related to the attractiveness of averaged faces, but is definitely not the only factor in the equation since about half of the image-features relate to the ratings of averaged composites in a similar manner as the symmetry measure.
Overall, it is quite surprising and pleasing to see that a machine trained explicitly to capture an operational performance criteria such as rating, implicitly captured basic human psychophysical biases related to facial attractiveness.