

## **Symmetry and Beauty of Human Faces**

Teresa Breyer<sup>1</sup>  
Department of Mathematics  
University of Wisconsin  
Madison, WI 53706, U.S.A.  
And Technical University of Vienna, Austria  
E-mail: breyerteresa@hotmail.com

### **Abstract**

In this project we study the correlation between symmetry and attractiveness of faces through statistical pattern recognition and neural networks. We design an intelligent agent that learns to use a given set of standards to distinguish between beautiful and ordinary faces. We represent the faces as feature vectors in a vector space in order to measure the degree of symmetry. To implement a larger data set, we use 2-dimensional wavelet representation to improve the performance.

### **1 Introduction**

There have been many studies trying to figure out whether there is a correlation between facial symmetry and attractiveness. Below we briefly mention some related results:

In 1878 [9] Francis Galton found out, that composite faces (faces that are a combination of several faces) were considered more beautiful than the original faces. In 1990 Langlois and Roggman confirmed this. They argued that increasing averageness also increases attractiveness. Of course these composite faces were also more symmetric.

Conversely the findings of Grammer and Thornhill in 1994 [8] suggested that there was a negative correlation between attractiveness and fluctuating asymmetry (a measure for the deviation from bilateral symmetry used by biologists).

Rotem Kowner did a study on "Facial Asymmetry and Attractiveness Judgement" in 1996. In one of his experiments he took pictures of young children, young adults and old people. He used the original images as well as symmetrical composites, formed by taking one hemisphere and flipping it vertically. But symmetry only had a positive influence on older people. Kowner [1] suggests "that facial asymmetry has a curvilinear effect on facial attractiveness, as both extreme...asymmetry ...and extreme symmetry...somewhat diminish attractiveness". The proposed neural network in essence learns from examples to approximate Kowner's conjectural curvilinear function.

Our approach to discuss the correlation between symmetry and attractiveness of faces is new in the sense that we use an intelligent agent for the study of a cultural issue. Our learning-machine, a multi-layer perceptron network, which provides a simplified imitation of information processing in the human brain, can learn to distinguish between beautiful and ordinary faces. This is the first attempt to explore the underlying mathematical structure and patterns that are related to human emotion, preferences, perception of beauty, etc.

---

<sup>1</sup> This research was done with supervision of Amir Assadi as part of the NSF Project Symmetry supported in part by NSF-HER-DUE-CCD and UW-Madison Office of the Provost

One approach to conceptualizing the human brain, where the basic units, the neurons, are connected in average to 10 000 other neurons, is via artificial neural networks. In some models for artificial neural networks, one models parts of the cortex by layers of neural nodes, each node, also called perceptron, being connected to all nodes of the two adjacent layers, as in Figure 1. In the human brain, the relevant cortical neurons respond to input stimuli by “spikes,” that is, they pass on signals, if the input they have accepted exceeds a certain level, called threshold. An artificial network is modeled as a graph whose nodes correspond to neurons and whose edges measure neuronal connectivity strength. We assign values, called *weights*, to the edges connecting nodes. These pass the signals received from the nodes of the preceding layer on to all nodes of the following layer after multiplying them with the weights of the corresponding edge. At the initial stage, these weights are set to random values. Learning means giving the input-layer some input, comparing the output calculated by the network to the desired output. The objective is to construct an error function from measuring the difference between the desired output and network’s output as a function of the network weights, and to modify the weights so as to minimize the error function. One achieves this by propagating the error back and adjusting the weights according to the so-called “back-propagation algorithm.” By repeating this procedure with a large set of data, the error made by the network will consistently diminish, provided that the data members form clusters or any other appropriate configuration in the given n-dimensional space. This is related to the convergence properties of the network. A network with only one layer of weights would only be able to separate classes with linear hyperplanes. Our network consists of two layers of weights. This second layer makes it possible to classify with convex hulls [4].

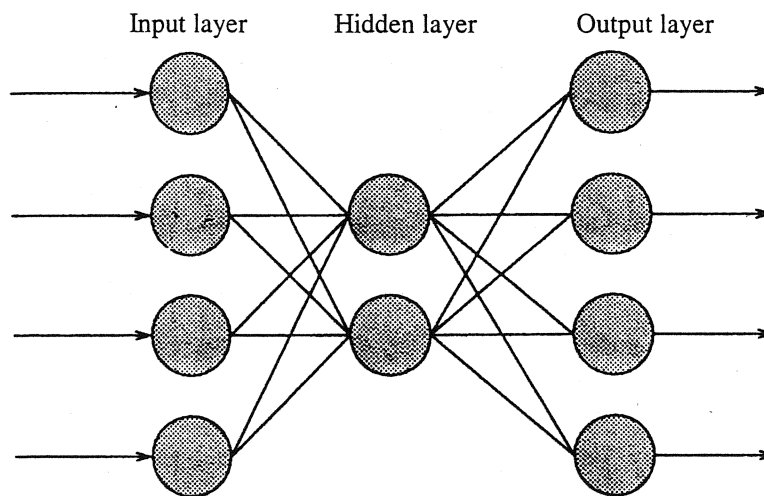


Figure 1: A Multi layer Artificial Neural Network [4]

According to Kowner [1], perfectly symmetric faces as well as extremely asymmetric ones are more likely to be evaluated as less attractive. So if the information representing the level of symmetry is presented to the perceptron network, one expects that the algorithm should converge. We distinguish only between two classes: the class of data members that we assume or simply agree as representing “beautiful faces” and the class of data members representing ordinary faces, by some reasonable standards. The points in pattern space representing the latter class should result in a cluster for perfectly symmetric faces, surrounded by an area of less density of points of this class, where the points representing beautiful faces outweigh. Moving farther away from the center there should be an increase in the density of members of the class of ordinary faces caused by the points representing faces with rather extreme asymmetries again. In this project we run a test to validate Kowner’s hypotheses.

As mentioned before this artificial network needs a set of data to learn, i.e. to adjust its weights. So the data would include information about the face itself as well as whether it is considered as beautiful. For this purpose we have to rely on the perception of beauty of a comparatively small group of people. Consequently the intelligent agent will learn to judge according to the opinion of these people. What they consider beautiful will again strongly depend on their cultural and ethnic background. Still we can assume that this group's judgment, determined by averaging the individual judgments, is a good representative of what the part of society they belong to consider beautiful.

This artificial neural network will be fast at adding up and multiplying all the inputs and at giving us its judgment whether a face is beautiful or not, but it still depends on us providing it with the applicable type of data. Even if we use more sophisticated methods to represent faces than simple measurements of the degree of symmetry, at some point the machine will fail to recognize a face as such. We just have to think of smiling and crying faces, images of faces taken from the front and profiles, faces covered by a scarf or smoking a cigarette. Our brain won't have any difficulty recognizing these images as faces, but training a computer to fulfill this task is a much more complex issue. Trying to understand better how our brain works, in our example how it perceives images, will help achieve these goals. But up to now, many more questions regarding the human brain remain to be fully explored.

**Remarks.**<sup>2</sup> 1. The reasons for using only female faces in our study are: (a) mixing female and male images in one study complicates the analysis due to differences in weights of parameters entering in judgements of beauty in different genders. There are more subtle correlations between the two types that are as yet not well-understood. (b) It was simply easier to collect appropriate data for female faces.  
2. We have not included any data that would compare systematically the response of human subjects versus the network, simply due to lack of time and resources, such as special permission for human subject studies. However, it is important and interesting to conduct such a study.  
3. There is considerable anatomical evidence to support the hypothesis that the brain of human and other primates has specialized networks for face recognition (localized in the inferotemporal cortex IT). This ties the computational studies of symmetry in human faces to the biology more readily than symmetry of other objects. For us, this is another reason to favor faces versus other objects.

## 2 Measurements of facial degree of asymmetry

I wanted to find out if there is a correlation between symmetry and attractiveness of faces. So I used Kownner's [1] measurements of the facial degree of asymmetry, which is a modification of the method used by Grammer and Thornhill [8].

There are more sophisticated ways to represent faces. We will come across 2 other methods using deformable templates respectively wavelet analysis later on. These are much more complicated and I wanted to concentrate on the aspect of symmetry first.

First Krownner [1] defines 12 specific points in a face:

- "The outermost and innermost eye corners",
- "The right and left junctions where the lower part of the ear touches the head"
- "The rightmost and leftmost points of the nose in the lower nose region"
- "The top leftmost and rightmost points of the chin"
- "The rightmost and leftmost points of the mouth"

He connects each pair of points with a straight line and calculates a midpoint with the following formula: " $(\text{left point-right point})/2+\text{right point}$ "[1].

<sup>2</sup> We thank the referee for helpful comments. These remarks are in response to referee's suggestions.

To represent the faces as vectors, I chose these midpoints as coordinates and the distance between the outermost left and right eye corners for standardizing.

### 3 Another method to describe faces

Alan L. Yuille [2] describes “an approach for extracting facial features from images and for determining the spatial organization between these features using the concept of deformable templates”. He describes two different approaches. In the first one he uses global templates consisting of the basic features connected with springs. He adjusts these to the faces by changing the parameters of the template, which are the locations of the features. He defines a function containing of two parts: The first part is a measure for how well the single features actually fit. The second part is a cost function for the springs and is responsible for the spatial relations. Yuille maximizes this function and uses the values of parameters, at which the maximum is achieved, to represent the face.

In the second approach he uses templates for the eyes, the mouth, etc. He represents these features before concentrating on the spatial relations.

In both approaches the templates only explore the specified features and ignore the rest. An advantage is that this method can be made robust, so that for example also a mouth smoking a cigarette can be recognized as a mouth.

### 4 Data and Principal Component Analysis

As data I took 16 faces from one ethnic group. I chose pictures of African American women from the December/January 2000 edition of the magazine Black Hair. All of them look straightforward; their mouths are closed and not smiling. Each face is represented by a vector, which consists of the measurements of the facial asymmetry degree as in Kowner[1].

I performed *principal* component analysis (PCA) [3] to find out which linear combination of the measured features carry the most statistically significant information and to exclude less important information. Principal components for a data set given by a collection of feature vectors are the unit eigenvectors of the covariance matrix of the data. PCA in this case could be interpreted as transforming the standard coordinate system into the one spanned by the *principal components* of the matrix consisting of the face vectors. The origin of this new coordinate system is the sample mean of the data. The covariance matrix is given by the formula  $BB^T/(N-1)$  where B denotes the matrix consisting of the original vectors from which the sample mean is subtracted and N the dimension of the vector space. The first principal component is the eigenvector corresponding to the largest eigenvalue, which has largest variance. Up to now the data has not been compressed and no information is lost. In this new coordinate system the coefficients of the coordinates corresponding to the last few principal components will be negligible and we are going to suppress them.

This results in a new coordinate system of smaller dimension encoding the features that is more efficient than the original representation.

A Matlab function (prcoan.m) prepares the data as explained and then does the principal component analysis. First it calculates the covariance matrix. Then it carries out the singular value decomposition. After putting the eigenvectors in order according to the size of the associated eigenvalue, starting with the one belonging to the largest eigenvalue, it transforms the original data into the new coordinate system. Then it also calculates the percentage of the information contained by the principal components. In my example the first PC

contains 87.7%, the second one 10.79 % and the remaining 3 each only less than 1% of the information. So I transformed my data into a plane spanned by the first two principal components.

I presented the network with these components, carrying 98.49% of the information, after adding another coordinate of the value 1 indicating that the faces used for the measurements and PCA were considered as beautiful. My analysis at this point is based on the assumption that the faces of models in fashion magazines are selected typically from beautiful women. If we change this assumption by having another collection of faces representative of perceived bias towards beauty, then the results may differ, while the procedures and algorithms remain the same.

## 5 Neural Network-Matlab implementation

The input layer and the first layer, which is a hidden layer, have as many nodes as the data dimensions. The second layer, which already is the output layer, consists of just one node. Only the values 1 and 0 are admissible. The target output tells if a face is considered beautiful (1) or ordinary (0).

The program consists of one main function `neuraln.m`. It calls the functions `forward2.m`, which multiplies for each node in each layer the inputs with the weights, sums these and passes these as new inputs to the next layer, and the function `pback2.m`. This function does the backwards calculation, which means it distributes the mistake made in one layer on the preceding layer using the weights and adjusts the weights according to this. I applied a modified version of the formula in Beale and Jackson [4] (Chapter 4.5, p.73-74) by simply using the value of the output of the corresponding node instead of the final output for the error terms for hidden units. A momentum term is also introduced. It controls the changes in the weights. As long as the error is large, bigger changes are made, but as the error decreases also the changes become smaller.

To test whether a face is considered beautiful by the network, you just have to transform the vector representing this face into the same coordinate system determined through PCA and rerun `pforward2.m` with the already found weight matrices. The output will be a number between 0 and 1.

## 6 Results

After the network learns from data, which means letting it adjust the weights, it should be able to recognize faces considered as beautiful by the hairstyle magazine as such. We presented it with the measurements of a completely symmetric face, the null vector transformed into the new coordinate system, and got an output of 0.9888. This emphasizes that a symmetric face is considered as beautiful by this network and consequently also following the trend of this magazine. But according to Kowner [4] perfect symmetry should have had a negative effect on facial attractiveness.

Analyzing the distribution of the points in the face vector space more carefully we might be able to detect a lower density around the point representing a perfectly symmetric face. The network's output value at that point still could be slightly smaller than the ones of surrounding points. In our example we only have 20 points in a plane, which won't display a very sophisticated pattern. Enlarging the data set might improve the result. At least it would narrow the ambiguity of possible interpretations. Using another way of representing faces than these measurements of the degree of asymmetry is another alternative.

We have to keep in mind that our method of representing faces only concentrated on a few measurements of bilateral symmetry; and so the sample size and the details of recording features are not representative of the realistic complexity of faces. It ignores a lot of other aspects as shading and proportions, just to mention a few.

When we look at a face we see much more than the features recorded above. We associate our experiences with people and their facial expressions, and these might have a positive or negative effect on the way we perceive them. If we meet a person that has similarities with somebody we didn't get along with in the past, we will probably judge him as less attractive. To mention another point, we performed our test on the ideally symmetric face with the measurements of a face that doesn't exist in real world. At least we haven't found it in our face collection and we probably haven't seen too many perfectly symmetric faces in our lives. Similarly, Kowner created synthetic symmetric faces by taking one hemisphere and flipping it vertically instead of using faces that really occur in nature, that we are confronted with in real life. In both cases, we have left unanswered the real question of how *natural* symmetry (as opposed to *synthetic* symmetry) and perception of beauty in the human society are combined.

## 7 Work in progress

To improve the results and to implement a larger data set, we are going to use 2-dimensional wavelet representation. Again we transform our data and, to compress it, we try to store only those coefficients that include most of the information about the images. These wavelet methods are similar to *Fourier* transformation.

In Fourier analysis [6] we describe an  $L^2$ -function with its Fourier series representation. Since the periodic functions  $\{\sin(nx), \cos(nx), n \in \mathbb{N}\}$  form a basis of  $L^2$ , we can represent the functions with the coefficients according to this new basis. The strength of this method lies in its ability to capture frequency and its weakness is in providing an efficient representation when the function varies non-periodically.

In wavelet analysis [6], we construct an orthonormal system spanning the whole space by taking one block-function, the mother wavelet, and translating and dilating it. The simplest one-dimensional mother wavelet is the *Haar function*,

$$\psi(x) = \begin{cases} 1, & 0 \leq x < 1/2 \\ -1, & 1/2 \leq x < 1 \\ 0, & \text{otherwise.} \end{cases}$$

$L^2$ -functions can be approximated as precisely as necessary with a finite number of block functions of the derived orthonormal system.

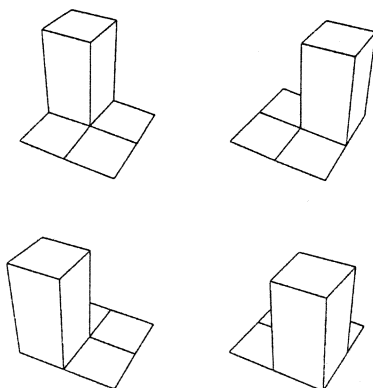


Figure 2: The original block function basis [5]

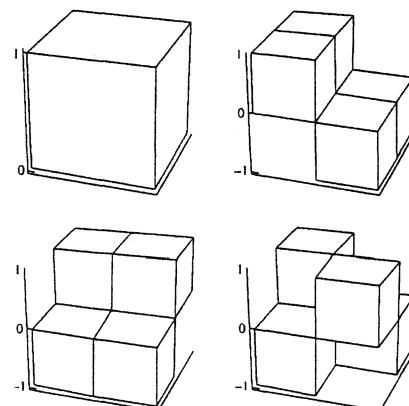


Figure 3: Tensor-Product Haar Wavelet [5]

In our case, we want to transform 2-dimensional matrices, whose entries measure the gray-scale of the pixels of our images. For simplicity of exposition, we mention only the simplest two-dimensional wavelets, namely, the *Tensor-Product Haar Wavelets*. They consist of the tensor products of the function  $\psi(x)$  from above and the function  $\phi(x)$ , which equals 1 on its support, the interval  $[0,1)$ .

The  $4 \times 4$  matrix associated with a square-step function can be expressed in terms of the latter tensor products. Originally the entries of the matrix, the values of a pixel, represent the height of the block in the corresponding quadrant. The entries of the transformed matrix are a certain linear combination of all four blocks (see Figure 2 and 3) and allow the following interpretation:

The upper left-hand entry measures the average of the original entries.

The upper right-hand entry measures the horizontal change in the original entries.

The lower left-hand entry measures the vertical in the original entries.

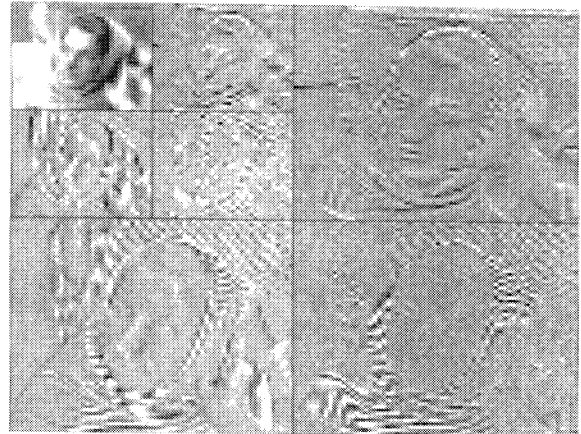
The lower right-hand entry measures possible diagonal edges.

This transformation does not result in any loss of information.

To transform the matrix of our image we split it into  $4 \times 4$  matrices and apply this method to them. Then we save the upper right-hand values of all the  $4 \times 4$  sub-matrices in the upper right-hand corner of the original matrix, the upper left-hand values in the upper left-hand corner, etc. Then we continue with the upper left-hand quarter of the matrix, which now consists of the image with only  $1/4$  of the pixel values. We can repeat this procedure, as in Figure 3, until we are down to one average value.



(a): Original Image



(b): Haar Wavelet Decomposition at level 2

Figure 4

After each step we gain a new matrix without any loss of information, but instead of giving only the values of the matrix we encode information about the “patterns of change” [5].

In our work in progress, we compare different wavelet decompositions for the images, to determine the criteria that selects the optimal wavelet family to save the most significant features, while it compresses the data through a minimal representation.

---

**References**

- [1] Rotem Kowner, *Facial Asymmetry and Attractiveness Judgment in Developmental Perspective*, Journal of Experimental Psychology: Human Perception and Performance, Vol. 22, No.3, pp. 662-675. 1996.
- [2] Alan L. Yuille, *Deformable Templates for Face Recognition*, Journal of Cognitive Neuroscience, Vol. 3, No. 1. 1989 – 1992.
- [3] Lay, *Linear Algebra, Chapter 8: Symmetric Matrices and Quadratic Forms*, Addison-Wesley, 1997.
- [4] Beale, *Neural Networks and Pattern Recognition in Human-Computer Interaction*, New York: Ellis Horwood. 1992.
- [5] Yves Nievergelt, *Wavelets Made Easy*, Birkhaeuser, 1999.
- [6] R. Todd Ogden, *Essential Wavelets for Statistical Applications and Data Analysis*, Birkhaeuser, 1997.
- [7] David Beymer, Tomaso Poggio, *Image Representations for Visual Learning*, Science, Vol. 272, pp. 1905-1909. 1996.
- [8] K. Grammer, R.Thornhill, *Human facial attractiveness and sexual selection: The role of averageness and Symmetry*, Journal of Comparative Psychology, Vol. 108, pp. 233-242. 1994.
- [9] F. J. Galton, *Composite portraits*, Nature, Vol. 18, pp. 97-100. 1878