Preferred Spatial Frequencies for Human Face Processing Are Associated with Optimal Class Discrimination in the Machine

Matthias S. Keil1*, Agata Lapedriza2, David Masip3, Jordi Vitria4

1 Basic Psychology Department, Faculty for Psychology, University of Barcelona (UB), Barcelona, Spain, 2 Computer Vision Center (CVC), Computer Science Department, Autonomous University of Barcelona, Bellaterra, Spain, 3 Department of Computer Science, Universitat Oberta de Catalunya, Barcelona, Spain, 4 Department of Applied Mathematics and Analysis (MAiA), University of Barcelona (UB), Barcelona, Spain

Abstract
Psychophysical studies suggest that humans preferentially use a narrow band of low spatial frequencies for face recognition. Here we asked whether artificial face recognition systems have an improved recognition performance at the same spatial frequencies as humans. To this end, we estimated recognition performance over a large database of face images by computing three discriminability measures: Fisher Linear Discriminant Analysis, Non-Parametric Discriminant Analysis, and Mutual Information. In order to address frequency dependence, discriminabilities were measured as a function of (filtered) image size. All three measures revealed a maximum at the same image sizes, where the spatial frequency content corresponds to the psychophysical found frequencies. Our results therefore support the notion that the critical band of spatial frequencies for face recognition in humans and machines follows from inherent properties of face images, and that the use of these frequencies is associated with optimal face recognition performance.

Introduction
Accumulating evidence supports the view that the processing of sensory information in the brain has adapted to statistical properties of sensory stimuli e.g., [15,22–24,26]. In this way, in principle the highest possible amount of information about the signal is encoded in the neuronal response [2,21]. In reality, however, signal coding is subject to constraints, that include, for example, minimizing energy expenditure [3,17,19,20], minimizing wiring costs between processing units [18], or reducing spatial and temporal redundancies in the input signal [1,2,4,14,29].

In a recent study, Keil [16] examined the statistical properties of a large number of face images by analyzing their amplitude spectra. The spectra were transformed such that the distribution of amplitudes versus spatial frequencies had maximum entropy (“whitening”). Whitened spectra revealed amplitude maxima at around 10 cycles per face, but only for the spectra of face images without external face features (i.e., hair, shoulder). This result compares well with corresponding psychophysical data, which suggest that humans process face identity preferentially in a narrow band of spatial frequency band (about 2 octaves) from 8 to 16 cycles per face [5–7,12,25,27,28,30]. The study of Keil [16] thus suggests that the processing of face identity in humans adapted to the statistical properties of face stimuli. The psychophysical results, on the other hand, suggest that face identification is best at spatial frequencies around 10 cycles per face. Given this link between stimulus statistics and psychophysics, we reasoned that also artificial face recognition systems should show an optimal recognition performance at spatial frequencies situated around 8 to 16 cycles per face.

In this work we compare the quality of the different spatial frequencies to perform subject recognition task in the machine. The problem of subject recognition in computer vision consists on automatically assigning to a face image a label corresponding to the identity of the person that appears in the image. For this aim we usually have a set of training data from where we learn this task. Thus, the training face images are labelled according to the subject, belonging to the same class all the images obtained from the same person. This study aims to satisfy three goals: (i) To analyze the data distribution of the different spatial frequencies representations and find out if there exists a relationship between the most suitable representation in the machine and the results obtained by the psychophysical studies; (ii) to give a statistical interpretation of the human visual system procedure for recognizing faces (iii) to study which is the minimal resolution that preserves the relevant information of a face to perform computational subject recognition.

In section “Materials and Methods” we justify that the best option to evaluate features quality is using discriminability measures, which will return large values when the data is appropriately distributed to perform subject recognition and low values otherwise. Thus, to perform this study we evaluated three class discriminability measures as a function of the spatial frequency content of face images to find out if there is a maximum in the same representation found with the psychophysical studies. The obtained results suggest that artificial face recognition systems
should have an optimal performance when the original face images contain spatial frequencies at around 16 cycles per degree, coinciding with the stimulus statistics and psychophysics.

Results

In the experiments, extrinsic face features (e.g., hair) were suppressed by centering a Blackman-Harris (B.H.) window at the nose (Fig. 1A and methods). To make computations feasible, spatial frequency content of face images was selected by decreasing the size of face images and applying high-pass filtering, respectively, rather than performing naive low-pass and band-pass filtering, respectively (see methods). The mentioned class discriminability measures were then computed for the down-sized images (corresponding to low-pass filtered original images), and their high-pass filtered versions (corresponding to band-pass filtered original images).

The dependency of FLD, NDA, and MI, respectively (see Methods), on spatial frequencies (or image size) is shown in Fig. 2. Each of the three measures reveals a distinct maximum at approximately the same image size (around 37 × 37 pixels), what corresponds to approximately 16 cycles per face width, as illustrated by Fig. 1C. The discriminability measures have very similar dependencies on image size irrespective of applying high-pass filtering. Thus, our results suggest that class discriminability is band-pass, meaning that the lower spatial frequencies do not contribute to a good separation of classes (which can be conceived as clouds of points produced from one individual). Adopting a different viewpoint, one can also argue that decreasing image size is equivalent to reducing dimensionality, and class separation collapses beyond a certain dimension.

Discussion

Psychophysical studies suggest that for face recognition, human observers make use of a narrow band at low spatial frequencies (8 to 16 cycles per face, bandwidth two octaves). Here we measured class discriminability, using Fisher Linear Discriminant Analysis and Non-Parametric-Discriminant-Analysis, and computing Mutual Information as a function of image size (and thus spatial frequency). These measures are used to quantify the efficiency of the different face representations to perform subject recognition in general, without depending on a specific implementation of a classifier. All three measures gave similar results for the high-pass filtered and the unfiltered face images, and revealed an unimodal distribution with a maximum at about 16 cycles per face width, which is close to the psychophysically found frequency optimum. Our results therefore support the conclusion that face representation to perform subject recognition task is optimal within a narrow band of spatial frequencies. Moreover, the presence of low spatial frequencies does not seem to compromise recognition performance.

Specifically, FLD and NDA reveal narrow peaks, which is compatible with the fact that human face discriminability of different subjects performance is best within a small band of spatial frequencies (bandwidth around two octaves, e.g., [25]). Nevertheless, MI shows a broad maximum, what may be interpreted as that recognition would still work if critical frequencies were not available. Similar observations were made in psychophysical studies [27], where it has been reported that face recognition is suboptimal in the absence of the critical frequencies. In this context, “suboptimal” means that it takes more time for subjects to recognize face identity, presumably due to a decreased signal-to-noise ratio [27].

The present study lends further support to the findings of Keil [16] in that the stimuli (i.e., face images) provide the explanation of the preference of a narrow spatial frequency band for both human and artificial face recognition. As a consequence, it is reasonable that artificial face recognition systems focus on these frequencies to achieve an optimal recognition performance, given that they are the most effective in terms of class discriminability. Because these critical spatial frequencies correspond to small image patches, a further advantage emerges through an economic use of resources for both processing and storing faces.

Figure 1. Illustration of processing steps. (A) External features are suppressed by centering a Blackman-Harris window at the face center \((x_n, y_n)\) (indicated by a cross-hair; see methods). In this way the windowed image is obtained as shown. (B) The central region of each windowed face image (dark-shaded) is maintained for further processing (note that the original face image is shown here only for illustration). In this way an image with an initial size (or equivalently dimensionality) of \(64 \times 64\) pixels is obtained. (C) Class discriminability measures are evaluated at each image size from the initial size down to \(10 \times 10\) pixels. Optimal recognition performance (i.e., highest class discriminability, see Fig. 3) is obtained for images of about \(37 \times 37\) pixels (here shown magnified), what corresponds to ca. 16 cycles per face width.

doi:10.1371/journal.pone.0002590.g001
Materials and Methods

Face Images
We used 868 female face images, and 868 male face images from the Face Recognition Grand Challenge database (FRGC, www.frvt.org/FRGC or www.bee-biometrics.org, Fig. 3) belonging to 55 different persons. We have selected all the subjects that have more than 20 images to obtain more accurate estimators of the discriminability measures. Original images (1704 × 2272 pixels, 24-bit true color) were adjusted for horizontal alignment of eyes, before they were down-sampled to 256 × 256 pixels and converted into 8-bit gray-scale. The positions of left eye (xle, yle), right eye (xre, yre), and mouth (xmo, ymo), respectively, were used to approximate the position of each face center (=nose) as

\[
\begin{align*}
x_{no} &= \text{rnd}(x_{le} + x_{re})/4 + x_{mo}/2 \\
y_{no} &= 0.95 \times \text{rnd}(y_{le} + (y_{mo} - (y_{le} - y_{re})/2)/2)/2
\end{align*}
\]

where \(\text{rnd}(x)\) denotes rounding to the nearest integer value.

Windowing of face images
Let the features which are not part of the actual face be denoted by external features (e.g., shoulder region or hair). On the other hand, internal features refer to the eyes, the mouth, and the nose. The presence of external features in our face images may distort recognition performance. It is thus desirable to compare results without the presence of external features. We found that a good suppression of external features could be achieved by centering a minimum 4-term Blackman-Harris window [11] at (xmo, ymo). The procedure is illustrated with Fig. 1A.

Varying spatial frequency content
We adopted the following procedure to assess the frequency-dependence of face recognition. Each image was resized to continuously smaller sizes, starting with an initial size of 64 × 64 pixels (see Fig. 1B). We used a bilinear interpolation scheme with the Matlab function “resize” to this end (Matlab version 7.1.0.183 R14 SP3 Image Processing Toolbox, see www.mathworks.com). A down-sized image is equivalent to its low-pass filtered original image, with a cut-off frequency equivalently to the Nyquist frequency (half of pixel width or height in cycles per image). This means that the smaller image contains all spatial frequencies of the original image which are smaller or equal than the Nyquist frequency. We subsequently performed high-pass filtering of the smaller images. The latter procedure is equivalent to band-pass filtering or the original image with a narrow filter bandwidth. Notice that down-sizing reduces the dimensionality of the feature space, and saves computational time when compared to naive low-pass and high-pass filtering, respectively.

Evaluation of Recognition Performance
The best criterion to evaluate the effectiveness of a features set to perform a concrete classification task is the Bayes error [38]. The Bayes error corresponds to the minimal probability of
Để làm cho các samples từ FRGC database. FRGC database chứa ảnh mặt của nam và nữ người trưởng thành từ các chiều khác nhau, với nhiều hình ảnh khác nhau cho từng chủ thể, khác nhau nhau, và khác nhau. Các hình ảnh được hiển thị theo phương pháp fronto-parallel, mặc dù một số hình ảnh đã thay đổi theo độ cao. Tất cả các hình ảnh được hiển thị đối diện với một nền vàng, và điều kiện ánh sáng được đặt ra ổn định.

Để làm cho các samples từ FRGC database. FRGC database chứa ảnh mặt của nam và nữ người trưởng thành từ các chiều khác nhau, với nhiều hình ảnh khác nhau cho từng chủ thể, khác nhau nhau, và khác nhau. Các hình ảnh được hiển thị theo phương pháp fronto-parallel, mặc dù một số hình ảnh đã thay đổi theo độ cao. Tất cả các hình ảnh được hiển thị đối diện với một nền vàng, và điều kiện ánh sáng được đặt ra ổn định.

Các samples do được lấy từ các danh sách khác nhau, với nhiều hình ảnh khác nhau cho từng chủ thể, khác nhau nhau, và khác nhau. Các hình ảnh được hiển thị theo phương pháp fronto-parallel, mặc dù một số hình ảnh đã thay đổi theo độ cao. Tất cả các hình ảnh được hiển thị đối diện với một nền vàng, và điều kiện ánh sáng được đặt ra ổn định.

Các samples do được lấy từ các danh sách khác nhau, với nhiều hình ảnh khác nhau cho từng chủ thể, khác nhau nhau, và khác nhau. Các hình ảnh được hiển thị theo phương pháp fronto-parallel, mặc dù một số hình ảnh đã thay đổi theo độ cao. Tất cả các hình ảnh được hiển thị đối diện với một nền vàng, và điều kiện ánh sáng được đặt ra ổn định.

Các samples do được lấy từ các danh sách khác nhau, với nhiều hình ảnh khác nhau cho từng chủ thể, khác nhau nhau, và khác nhau. Các hình ảnh được hiển thị theo phương pháp fronto-parallel, mặc dù một số hình ảnh đã thay đổi theo độ cao. Tất cả các hình ảnh được hiển thị đối diện với một nền vàng, và điều kiện ánh sáng được đặt ra ổn định.

Các samples do được lấy từ các danh sách khác nhau, với nhiều hình ảnh khác nhau cho từng chủ thể, khác nhau nhau, và khác nhau. Các hình ảnh được hiển thị theo phương pháp fronto-parallel, mặc dù một số hình ảnh đã thay đổi theo độ cao. Tất cả các hình ảnh được hiển thị đối diện với một nền vàng, và điều kiện ánh sáng được đặt ra ổn định.

Các samples do được lấy từ các danh sách khác nhau, với nhiều hình ảnh khác nhau cho từng chủ thể, khác nhau nhau, và khác nhau. Các hình ảnh được hiển thị theo phương pháp fronto-parallel, mặc dù một số hình ảnh đã thay đổi theo độ cao. Tất cả các hình ảnh được hiển thị đối diện với một nền vàng, và điều kiện ánh sáng được đặt ra ổn định.

Các samples do được lấy từ các danh sách khác nhau, với nhiều hình ảnh khác nhau cho từng chủ thể, khác nhau nhau, và khác nhau. Các hình ảnh được hiển thị theo phương pháp fronto-parallel, mặc dù một số hình ảnh đã thay đổi theo độ cao. Tất cả các hình ảnh được hiển thị đối diện với một nền vàng, và điều kiện ánh sáng được đặt ra ổn định.

Các samples do được lấy từ các danh sách khác nhau, với nhiều hình ảnh khác nhau cho từng chủ thể, khác nhau nhau, và khác nhau. Các hình ảnh được hiển thị theo phương pháp fronto-parallel, mặc dù một số hình ảnh đã thay đổi theo độ cao. Tất cả các hình ảnh được hiển thị đối diện với một nền vàng, và điều kiện ánh sáng được đặt ra ổn định.

Các samples do được lấy từ các danh sách khác nhau, với nhiều hình ảnh khác nhau cho từng chủ thể, khác nhau nhau, và khác nhau. Các hình ảnh được hiển thị theo phương pháp fronto-parallel, mặc dù một số hình ảnh đã thay đổi theo độ cao. Tất cả các hình ảnh được hiển thị đối diện với một nền vàng, và điều kiện ánh sáng được đặt ra ổn định.

Các samples do được lấy từ các danh sách khác nhau, với nhiều hình ảnh khác nhau cho từng chủ thể, khác nhau nhau, và khác nhau. Các hình ảnh được hiển thị theo phương pháp fronto-parallel, mặc dù một số hình ảnh đã thay đổi theo độ cao. Tất cả các hình ảnh được hiển thị đối diện với một nền vàng, và điều kiện ánh sáng được đặt ra ổn định.

Các samples do được lấy từ các danh sách khác nhau, với nhiều hình ảnh khác nhau cho từng chủ thể, khác nhau nhau, và khác nhau. Các hình ảnh được hiển thị theo phương pháp fronto-parallel, mặc dù một số hình ảnh đã thay đổi theo độ cao. Tất cả các hình ảnh được hiển thị đối diện với một nền vàng, và điều kiện ánh sáng được đặt ra ổn định.

Các samples do được lấy từ các danh sách khác nhau, với nhiều hình ảnh khác nhau cho từng chủ thể, khác nhau nhau, và khác nhau. Các hình ảnh được hiển thị theo phương pháp fronto-parallel, mặc dù một số hình ảnh đã thay đổi theo độ cao. Tất cả các hình ảnh được hiển thị đối diện với một nền vàng, và điều kiện ánh sáng được đặt ra ổn định.

Các samples do được lấy từ các danh sách khác nhau, với nhiều hình ảnh khác nhau cho từng chủ thể, khác nhau nhau, và khác nhau. Các hình ảnh được hiển thị theo phương pháp fronto-parallel, mặc dù một số hình ảnh đã thay đổi theo độ cao. Tất cả các hình ảnh được hiển thị đối diện với một nền vàng, và điều kiện ánh sáng được đặt ra ổn định.

Các samples do được lấy từ các danh sách khác nhau, với nhiều hình ảnh khác nhau cho từng chủ thể, khác nhau nhau, và khác nhau. Các hình ảnh được hiển thị theo phương pháp fronto-parallel, mặc dù một số hình ảnh đã thay đổi theo độ cao. Tất cả các hình ảnh được hiển thị đối diện với một nền vàng, và điều kiện ánh sáng được đặt ra ổn định.

Các samples do được lấy từ các danh sách khác nhau, với nhiều hình ảnh khác nhau cho từng chủ thể, khác nhau nhau, và khác nhau. Các hình ảnh được hiển thị theo phương pháp fronto-parallel, mặc dù một số hình ảnh đã thay đổi theo độ cao. Tất cả các hình ảnh được hiển thị đối diện với một nền vàng, và điều kiện ánh sáng được đặt ra ổn định.
The estimate of the between-class scatter matrix $S_B$ is found as the average of the local matrices

$$S_B = \frac{1}{N} \sum_{\sim X} \Lambda_{B}$$

(5)

The resulting $S_B$ is used in the criterion [1], while $S_D$ remains as in the first case. We will denote this second discriminability parametric implementation, without prior assumptions about class quadratic divergence measure that allows an efficient non-parametric implementation, without prior assumptions about class densities. In concrete, the Mutual Information from the data can be computed by

$$I(X,C) = I_{Y} + V_{ALL} - 2V_{BTW}$$

where

$$V_{IN} = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} G(x_{ij} - x_{ik}, 2\sigma^2 I)$$

$$V_{ALL} = \frac{1}{N} \left( \sum_{i=1}^{N} \left( \frac{x_i}{C} \right) \right) \sum_{j=1}^{N} \sum_{k=1}^{N} G(x_{ij} - x_{ik}, 2\sigma^2 I)$$

$$V_{BTW} = \frac{1}{N} \sum_{i=1}^{N} \sum_{j=1}^{N} \sum_{k=1}^{N} G(x_{ij} - x_{ik}, 2\sigma^2 I)$$

denoting a sample by one index, $x_i$ if the class is irrelevant and by two indexes, $x_{ij}$ when its class is relevant. The function $G$ is a multi-dimensional Gaussian Kernel with covariance matrix $\Sigma$.

\[ G(x, \Sigma) = \frac{1}{(2\pi)^{n/2} \det(\Sigma)^{1/2}} \exp\left(-\frac{1}{2} x^T \Sigma^{-1} x\right) \]

being $d$ the corresponding dimensionality.

Author Contributions
Conceived and designed the experiments: MK JV. Performed the experiments: AL DM. Analyzed the data: AL DM. Wrote the paper: MK AL DM JV.

References
1. Attick J, Redlich A (1992) What does the retina know about natural scenes? Neural Computation 4: 196–210.
2. Attneave F (1954) Some informational aspects of visual perception. Psychological Review 61: 183–193.
3. Ballard R, Abbott L, Booth M, Sengpiel F, Freeman T (1998) Responses of neurons in primary and inferior temporal visual cortices to natural scenes. Proceedings of the Royal Society of London B 264: 1775–1783.
4. Barlow H (1961) Possible principles underlying the transformation of sensory messages. In: Rosenblith W, ed. Sensory Communication. Cambridge, MA: MIT Press. pp 217–234.
5. Costen N, Parker D, Craw I (1994) Spatial content and spatial quantisation effects in face recognition. Perception 23: 129–146.
6. Costen N, Parker D, Craw I (1996) Effects of high-pass and low-pass spatial filtering on face identification. Perception and Psychophysics 58: 602–612.
7. Fiorentini A, Maffei L, Sandini G (1983) The role of high spatial frequencies in face perception. Perception 12: 195–201.
8. Fisher R (1936) The use of multiple measurements in taxonomic problems. Ann Eugenics 7: 179–188.
9. Fukunaga K (1990) Introduction to Statistical Pattern Recognition. 2nd ed. Boston, MA: Academic Press.
10. Fukunaga K, Mantock J (1983) Nonparametric discriminant analysis. IEEE Transactions on Information Theory 5: 671–678.
11. Harris F (1973) On the use of windows for harmonic analysis with the discrete Fourier transform. Proceedings of the IEEE 66: 51–84.
12. Hayes A, Morrone M, Burr D (1986) Recognition of positive and negative band-pass filtered images. Perception 15: 595–602.
13. Hellman M, Raven J (1970) Probability of error, equivocation and the Chernoff bound. IEEE Transactions on Information Theory 16: 368–372.
14. Howsa T, Baccus S, Meister M (2005) Dynamic predictive coding by the retina. Nature 436: 71–77.
15. Howe C, Purves D (2002) Range image statistics can explain the anomalous perception of length. Proceedings of the National Academy of Sciences USA 99: 13118–13122.
16. Keil M (2000) Does face image statistics predict a preferred spatial frequency for human face processing? Proceedings of the Royal Society B, doi:10.1098/rspb.2000.1486.
17. Laughlin S, de Ruyter van Steveninck R, Anderson J (1998) The metabolic cost of neural information. Nature Neuroscience 1: 36–41.
18. Laughlin S, Sejnowski T (2003) Communication in Neural Networks. Science 301: 1870–1874.
19. Levy W, Baxter R (1996) Energy-efficient neural codes. Neural Computation 8: 531–543.
20. Linsker R (1988) Self-organization in a perceptual network. IEEE Transactions on Computer 21: 105–117.
21. Long F, Yang Z, Purves D (2006) Spectral statistics in natural scenes predict hue, saturation, and brightness. Proceedings of the National Academy of Sciences USA 103: 6013–6018.
22. Lotto R, Purves D (2000) An empirical explanation of color contrast. Proceedings of the National Academy of Sciences USA 97: 12834–12839.
23. Lotto R, Williams S, Purves D (1999) Mach bands as empirically derived associations. Proceedings of the National Academy of Sciences USA 96: 5245–5250.
24. Na¨sa¨nen R (1999) Spatial frequency bandwidth used in the recognition of facial images. Vision Research 39: 3824–3833.
25. Na¨sa¨nen R (1999) Spatial frequency bandwidth used in the recognition of facial images. Vision Research 39: 3824–3833.
26. Nundy S, Purves D (2002) A probabilistic explanation of brightness scaling. Proceedings of the National Academy of Sciences USA 99: 14482–14487.
27. Ojamaa, Na¨sa¨nen R (2003) Utilisation of spatial frequency information in face search. Vision Research 43: 2505–2515.
28. Peli E, Lee E, Trenpe C, Buzney S (1994) Image enhancement for the visually impaired: the effects of enhancement on face recognition. Journal of the Optical Society of America A 11: 1929–1939.
29. Srinivasan M, Laughlin S, Dubs A (1982) Predictive coding: a fresh view of inhibition in the retina. Proceedings of the Royal Society of London B 216: 427–459.
30. Teger T, Ganz L (1979) Recognition of faces in the presence of two-dimensional sinusoidal masks. Perception and Psychophysics 26: 163–167.
31. Torkkola K (2003) Feature extraction by non-parametric mutual information maximization. J Mach Learn Res 3: 1415–1438.
32. Turk M, Pentland A (1991) Eigenfaces for Recognition, Journal of Cognitive Neuroscience 3: 41–79.
33. Duda R, Hart PE, Stork DG (2000) Pattern Classification, 2nd ed. New York: Wiley.
34. Hastie T, Tibshirani R, Friedman J (2001) The Elements of Statistical Learning. New York: Springer-Verlag.
35. Keinouike/Fukunaga, Sze-ro KHuang (1993) Bayes Error Estimation using Local Metrics. Electrical and Computer Engineering Technical Reports. Purdue University School of Electrical and Computer Engineering.
36. GuorgXnXan, ZhenpingZhang, PaeupChiai, YunQShi, DongdongFu (2005) A Feature Selection Based on Minimum Upper Bound of Bayes Error. IEEE 7th Workshop on Multimedia Signal Processing Page(s): 1–4.