Human and Machine Recognition of Fourier-Bessel Filtered Face Images

Yossi Zana
Roberto M. Cesar-Jr
Jesus P. Mena-Chalco

Dept. of Computer Science - IME
University of São Paulo, Brazil
(zana, cesar, jmena)@vision.ime.usp.br

Abstract

Motivated by a recently proposed biologically-inspired face recognition approach, psychophysical experiments have been carried out. We measured recognition performance of polar frequency filtered face images using an 8-alternatives forced-choice method. Test stimuli were generated by converting the images from the spatial to the polar frequency domain using the Fourier-Bessel transformation (FBT), filtering of the resulting coefficients with band-pass filters, and finally taking the inverse FBT of the filtered coefficients. We also evaluated an automatic FBT-based face recognition model. Contrast sensitivity functions of the human observers peaked in the 8-11.3 radial and angular frequency range, with higher peak sensitivity in the former case. The automatic face recognition algorithm presented similar behavior. These results suggest that polar frequency components could be used by the human face processing system and that human performance can be constrained by the polar frequency information content.

1. Introduction

Knowledge of which visual features are used for the recognition of different types of objects is crucial for understanding of human visual processing and can indicate useful features for automatic face recognition systems. Face recognition is one of the best understood cognitive tasks [14], due in part to the identification of several critical spatial components, although the way these components are integrated is still a controversial issue [2]. However, the known studies looked for Cartesian-defined spatial components, usually using Fourier-filtered face images. These studies and the resulting theoretical models did not take into account physiological and psychophysical evidence that suggests the existence of mechanisms for visual analysis in polar coordinates (see Section 2). In order to fill this gap, a computationally successful biologically-inspired approach to face recognition using polar domain representation has been recently reported [19].

In the current study we investigate the possibility that spatial polar-defined components are used selectively in human face processing and report here the first results. Moreover, we compare the performance of human observers with that of the proposed face recognition model. The main contribution of this study is in demonstrating, for the first time, that human visual face processing could involve selective use of polar frequency components.

This paper is organized as follows. In the next section we present a brief review of the literature relevant to face recognition and spatial frequency analysis. In Section 3 we detail our experimental design and stimuli generation and finally we describe our results and discuss their implications in Section 4.

2 Selective spatial frequency usage in face recognition

In classical studies of the human visual system, the luminance of test stimuli is modulated by a sine function in Cartesian coordinates. This choice is based on the shape of the receptive fields, sensitivity of retinal ganglion cells as well as cells in area V1 of the brain [4]. In accordance with this view, all the previous studies (to the best of our knowledge) searched for the fundamental components of human face processing in the Cartesian frequency domain. Such experiments typically employed face images whose spatial frequency content were manipulated using band-pass Fourier filters. Most of these studies confirmed that face

recognition is sensitive to the spatial frequency content of the images, and concluded that the mid-range spatial frequencies, between 10 and 20 cycles per face, are the most important for this task [3, 7, 11, 15]. This knowledge was essential to a comprehensive understanding of the cognitive function, since it delimited the quantity and quality of the information available in higher level stages.

However, more recent physiological and psychophysical studies gathered evidence on the tuning of visual cells to stimuli defined in coordinate systems other than the Cartesian. Sensitivity to complex shapes, like stars, rather than to simple Cartesian stimuli (bars), was observed in several cells in the visual area V4 of macaque monkeys by [9]. At the same time, [5, 6] probed cells in area V4 with Cartesian, polar, or hyperbolic gratings and showed specificity to these types of stimuli. A few years later, [10] extended the study to lower processing levels of the visual pathway and found that populations of cells in areas LGN, V1 and V2 are also tuned to these types of stimuli. The physiological evidences on the specificity of cells to non-Cartesian stimuli were further supported by psychophysical experiments using Glass patterns. The stimuli used by [16, 17] consisted of a pattern of random dots, presented within a circular window, that generated a percept of global structure of Cartesian, concentric, radial, and hyperbolic patterns. Detection threshold was measured by degrading the patterns through the addition of noise. It was found that threshold decreases from Cartesian, to hyperbolic, radial and concentric patterns. Measurements of the thresholds, as function of the stimulated area, showed a 3-4° visual degrees global pooling of orientation information in the detection of radial and concentric patterns, but only local pooling in the detection of parallel patterns. Similar results were obtained when subjects had to judge which of two square arrays of Gabor contained global structures, with higher sensitivity found to concentric than to radial patterns [1].

Inspired by these latter studies, we first determined the contrast sensitivity functions to fundamental patterns defined in polar Cartesian [18] and later developed an automatic face recognition system based on polar frequency features, as extracted by Fourier-Bessel transformation (FBT), and dissimilarly representation [19, 21]. This system was thoroughly tested on large datasets and achieved state of the art performance when compared to previous algorithms. The current study is the first step in determining the relation between the proposed face recognition model and human behavior. At this stage, we limited ourselves to comparison of the human contrast sensitivity functions with those obtained with the simplest possible version of the model, i.e. using only the polar components.

3 Material and methods

3.1 Polar frequency analysis

Let $f(x, y)$ be the region of interest in the image. FBT analysis starts by converting the image coordinates from Cartesian (x, y) to polar (r, θ) domain. Let (x_0, y_0) be the origin of the Cartesian image. The polar coordinates necessary to obtain the new image representation $f(r, \theta)$ are defined as $\theta = \tan^{-1}(y - y_0 / x - x_0)$ and $r = \sqrt{(x - x_0)^2 + (y - y_0)^2}$.

The $f(r, \theta)$ function is represented by the two-dimensional FB series, defined as [19]

$$f(r, \theta) = \sum_{i=1}^{\infty} \sum_{n=0}^{\infty} A_{n,i} Jn(\alpha_{n,i}r) \cos(n\theta)$$
$$+ \sum_{i=1}^{\infty} \sum_{n=0}^{\infty} Bn, i Jn(\alpha_{n,i}r) \sin(n\theta), \quad (1)$$

where J_n is the Bessel function of order n, $f(R, \theta) = 0$ and $0 \le r \le R$. $\alpha_{n,i}$ is the ith root of the J_n function, i.e. the zero crossing value satisfying $J_n(\alpha_{n,i}) = 0$ is the radial distance to the edge of the image. The orthogonal coefficients $A_{n,i}$ and $B_{n,i}$ are given by

$$A_{0,i} = \frac{1}{\pi R^2 J_1^2(\alpha_{n,i})} \int_0^{2\pi} \int_0^R f(r, \theta) r J_n(\frac{\alpha_{n,i}}{R}r) dr d\theta \quad (2)$$

if $B_{0,i} = 0$ and $n = 0$;

$$\left[\begin{array}{c} A_{n,i} \\ B_{n,i} \end{array} \right] = \frac{2}{\pi R^2 J_{n+1}^2(\alpha_{n,i})}$$
$$\int_0^{2\pi} \int_0^R f(r, \theta) r J_n(\frac{\alpha_{n,i}}{R}r) \left[\begin{array}{c} \cos(n\theta) \\ \sin(n\theta) \end{array} \right] dr d\theta \quad (3)$$

if $n > 0$.

In the current study, we transformed the images up to the 30th Bessel order and root with angular resolution of 3°, thus obtaining 1830 coefficients[1], each one being described by a Bessel root and order number. In the polar frequency domain, the Bessel root is related to the radial frequency (number of cycles along the image radius) while the Bessel order is related to the angular frequency (number of cycles around the center of the image). Figure 1 shows plots of

---

[1]Each Bessel mode is represented by two coefficients, except those of order zero that are represented by a single coefficient.
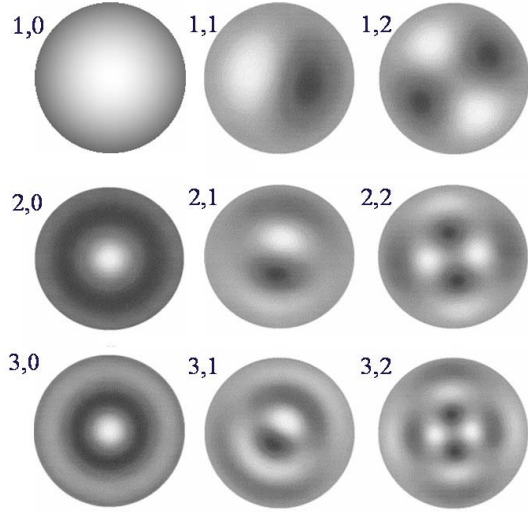
Figure 1. Spatial representation of FB coefficients. The pairs of numbers indicate the Bessel root and order (adapted from [8]).

a few FB patterns. Thus, these coefficients correspond to a frequency range of up to 30 cycles/image of angular and radial frequency. This frequency range was selected since perceptually it preserved most of the original image information.

3.2 Observers and equipment

Two of the authors participated in the tests. Observer S2 had no previous experience in psychophysical experiments, while observer S1 had a few years of experience. However, both were practiced with the unmanipulated stimuli prior to data collection until recognizing with ease all subjects' images. The stimuli were generated on a Philips 2020p color monitor and the graphics board was set at a resolution of 1024 x 768 pixels with a frame rate of 85 Hz. Viewing was binocular from a distance of 75 cm. The average luminance of the display was 10 cd/m^2 in an otherwise dark environment. To increase the number of luminance levels available from 256 to 4096, the red and blue color channels of the graphics board were combined in a resistance network [12]. The combined signal was connected to the green input in the monitor and gamma corrected to produce a linear luminance-modulated image.
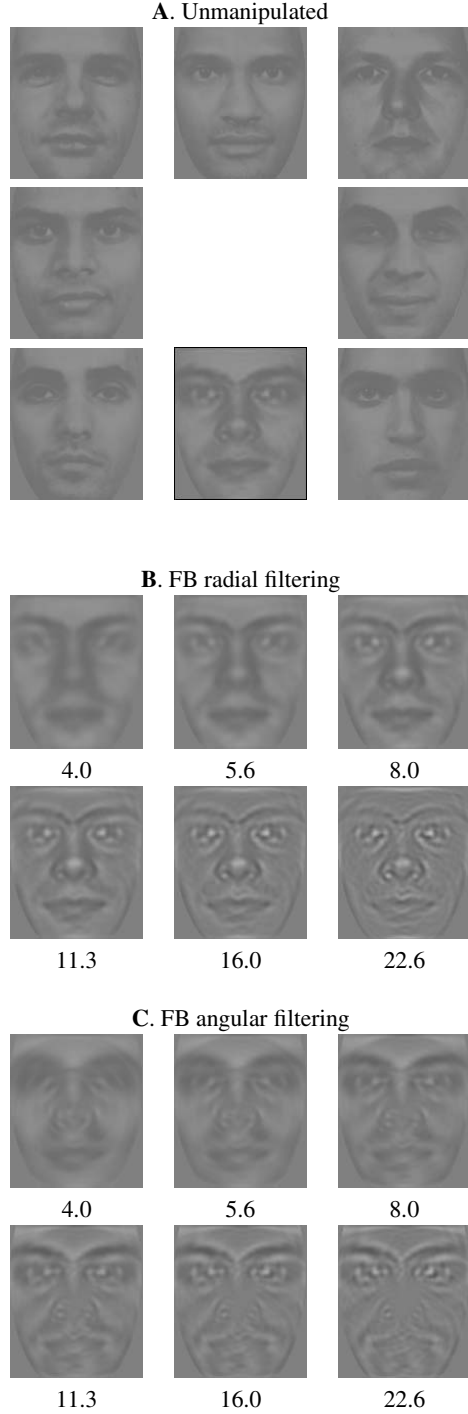
3.3 Stimuli

We used eight face images from the FERET database [13]. The criteria for the selection were: male gender, age between 20 and 40 years, neutral expression, Caucasian race, and absence of any special marks like beard, glass, etc. Using the ground-truth eyes coordinates, we translated, rotated, and scaled the images so that the eyes were registered at specific locations. Next, the images were cropped to 130 x 150 pixels size and a mask (zero value) was applied to remove most of the hair and background. The unmasked region was histogram equalized and normalized to zero mean (Fig. 2A). From the viewing distance, the images subtended 2.9° of horizontal visual degrees. Signal strength was defined as the image contrast variance [7]. Signal strength was manipulated by multiplying the image data by an appropriate constant and converting of the contrast values to luminance values.

Test stimuli were generated by first FB transforming the original images from the spatial to the polar frequency domain, as described in Section 3.1. The resulting coefficients were filtered by three-octave Gaussian band-pass filters centered at frequencies of 4, 5.6, 8, 11.3, 16 and 22.6 (half-octave steps). The final images were obtained by taking the inverse FBT of the filtered coefficients. Radial and angular filtering was achieved by multiplying the Gaussian filters along the Bessel root or Bessel order dimension, respectively. Examples of radial and angular filtering are shown in Figure 2. Unmanipulated original images and unfiltered FBT images were also tested to establish a baseline performance.

3.4 Procedure

Identification thresholds were determined using a single-interval eight-alternatives forced-choice paradigm. Stimulus contrast variance was manipulated using the method of constant stimuli. Stimuli were filtered off-line and presented at a pseudo-random contrast level. At least 20 trials were shown at each of six contrast levels, as determined in pilot studies. The presentation order of the stimuli was randomized. Observers were thoroughly familiarized with the unmanipulated images. At the start of a trial, a fixation cross appeared for 500 ms at the center of the screen, and a brief tone indicated the presentation of the target stimulus. The target image was exposed for 1500 ms and followed by a set of eight unmanipulated images that appeared around the region where the target had been displayed (see Fig. 2A for the images layout). Observers identified the target image by pressing one of eight keys on the computer keypad. Decision time was unlimited (usually less than four seconds). Auditory feedback was given for an incorrect response. Threshold estimations at the probability level of 0.67 were calculated by fitting the data to Weibull functions.

A. Unmanipulated

B. FB radial filtering

| 4.0 | 5.6 | 8.0 |
| 11.3 | 16.0 | 22.6 |

C. FB angular filtering

| 4.0 | 5.6 | 8.0 |
| 11.3 | 16.0 | 22.6 |

Figure 2. Face stimuli used in the experiments. All images are set to the same mean luminance and contrast variance. (A) The original normalized face images in the spatial layout displayed to the observers. Radial (B) and angular (C) filtering of the black-squared image in A. Numbers below the images indicate the respective filters central frequencies.
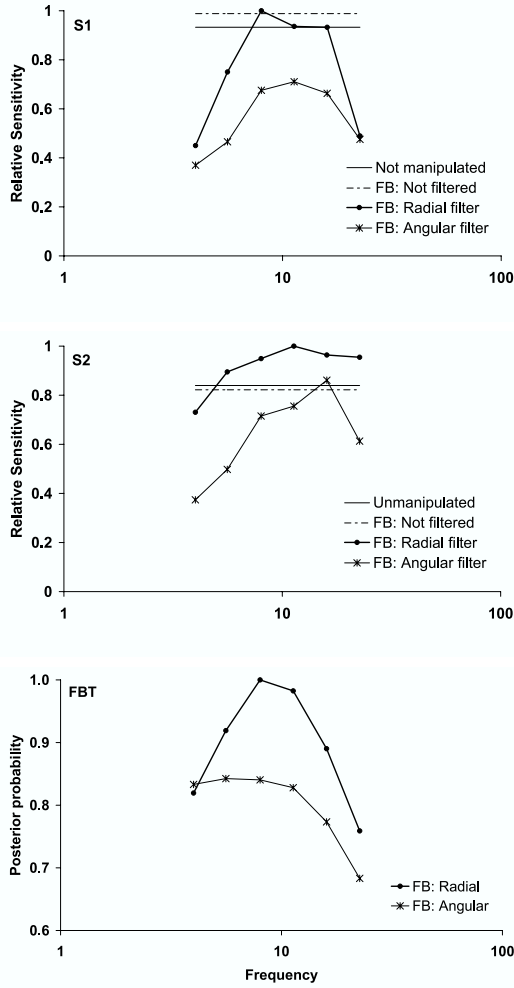
3.5 Automatic FBT-based face recognition

In order to compare the human performance with that of a FBT-based model [19], we built a simple nearest neighbor classifier. The classifier output quantifies the similarity between the test and training objects. When this output is normalized so that it sums to one, the individual values can be considered as the posterior probability (or confidence value) estimation that the two objects are of the same class. The training objects were the unfiltered FB representations, while the filtered versions were used for testing. The posterior probability value that a test object belongs to the training objects was computed for each test (filtering) condition. Recognition level at a each test condition was defined as the average posterior probability value across the eight test objects. For example, if the posterior probabilities between the training and 4-cycles radial filter images formed an 8-by-8 matrix, performance level would be the mean value of the principal diagonal.

4 Results and discussion

Figure 3 shows the face recognition performance of the two subjects and of the automatic FBT-based algorithm. The contrast sensitivity functions of observer S1 had a bell-shape and peaked in the 8 and 11.3 frequency with radial and angular filtering, respectively. However, peak sensitivity with the angular filtering was only two-third that with radial filters. The close match between the performance with the unmanipulated and unfiltered FBT images was expected, as they are perceptually very similar, but we were surprised to find out that contrast sensitivity could be higher with radially filtered images than with unfiltered images. We discuss this finding below.

Observer S2 showed somewhat different contrast sensitivity functions. His curves peaked at slightly higher frequencies, 11.3 and 16 cycles with radial and angular filters, respectively. The function to angularly filtered images had also a bell shape, but the radial curve had a much wider band-width. As with observer S1, sensitivity with angular filters was considerably lower than that with radial filters, and had higher sensitive to radially filtered images than to unfiltered images.

Although the differences between the results of the two observers do not permit drawing a definitive quantitative evaluation, several observations are clear. First, face recognition is better tuned to mid-range frequency angular components. Second, sensitivity to images filtered in the angular frequency domain is lower than to images filtered in the radial frequency domain. The mid-frequency radial components are also the most important for face recognition, especially in the case of the more experienced observer S1. The fact that sensitivity can be higher to radially filtered

Figure 3. Normalized face recognition performance. Top and middle panels: Contrast sensitivities of subjects S1 and S2. Bottom panel: Posterior probabilities of the FBT-based computational model.

images than to unfiltered images is interesting and is being investigated specifically in our ongoing work. One possible explanation is that unfiltered and filtered images could have the same (global) contrast variance, but differ still in their local contrast, and observers relay more on information from these higher contrast regions. A related effect is a possible "exaggeration" of certain facial features (local and/or global) that can helps in the recognition.

The higher sensitivity to mid-range frequency components could reflect lack of information in the images or internal constraints in the face processing. The results of the automatic FBT-based face recognition algorithm support the first explanation. The function curves of the model were similar to those of the human observers in regard to the peak sensitivity location (8 cycles) with both radial and angular filters, and the relatively lower sensitivity with angular filters (15%). However, it should be noted that the model curves are closer in shape to those of observer S1 and that the sensitivity function with angularly filtered images has more a low-pass filter shape than a bell shape. Even after considering these differences, the similarities found between the human and simple model behaviors are surprising and suggest that humans can use polar analysis for faces identification. Moreover, they establish for the first time a direct relation between human face recognition and a polar-frequency based model. Correlating more complex aspects of the complete model to human behavior, like error analysis, is the next natural step in this research line.

The current results are not in contradiction to previous studies which indicated that face recognition is tuned to the mid-range Cartesian frequency band, and do not imply that this task relies exclusively on fundamental FB spatial patterns. In fact, judging by the appearance of the image stimuli, we expect to find a positive correlation between the polar and Cartesian filtered images. Such correlation can be measured by analyzing the Fourier amplitude spectrum of the FB filtered images. However, the differences in working coordinates and base functions turn this analysis too complex and beyond the scope of this paper.

5   Conclusions and Future Directions

It can be concluded from our preliminary results that polar frequency components could be used in human face processing and that human performance could be constrained by the polar frequency information content. In our ongoing work we are improving the experimental design by increasing the number of observers and face images, as well as by reducing the filters bandwidth. Detailed correlation analysis of the errors made by the human observers and the model can further increase our confidence in the aforementioned conclusions. Furthermore, we are considering comparison of human behavior with a local-based FBT decomposition

computational model [20].

References

[1] H. R. Achtman, R.L. and Y. Wang. Sensitivity to global shape detection. *Journal of Vision*, 3:616–624, 2003.

[2] J. D. Cohen and F. Tong. The face of controversy. *Science*, 293:2405–2407, 2001.

[3] N. Costen, D. Parker, and I. Craw. Effects of high-pass spatial filtering on face identification. *Perception and Psychophysics*, 58:602–612, 1996.

[4] R. L. DeValois and K. K. DeValois. *Spatial Vision*. Oxford Sciences Pub., 1990.

[5] J. Gallant, J. Braun, and D. VanEssen. Selectivity for polar, hyperbolic, and cartesian gratings in macaque visual cortex. *Science*, 259:100–103, 1993.

[6] J. Gallant, C. Connor, S. Rakshit, J. Lewis, and D. VanEssen. Neural responses to polar, hyperbolic, and cartesian grating in area v4 of the macaque monkey. *Journal of Neurophysiology*, 76:2718–2739, 1996.

[7] J. Gold, J. Bennet, and A. B. Sekular. Identification of bandpass filtered letters and faces by human and ideal observers. *Vision Research*, 139:3537–3560, 1999.

[8] S. Guan, C. Lai, and G. Wei. Fourier-Bessel analysis of patterns in a circular domain. *Physica D*, 151:83–98, 2001.

[9] E. Kobatake and K. Tanaka. Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex. *Journal of Neurophysiol*, 71:2269–2280, 1994.

[10] L. Mahon and R. De Valois. Cartesian and non-cartesian responses in lgn, v1, and v2 cells. *Visual Neuroscienc*, 18:973–981, 2001.

[11] R. Nasanen. Spatial frequency bandwidth used in the recognition of facial images. *Vision Research*, 39:3824–3833, 1999.

[12] D. Pelli and L. Zhang. Accurate control of contrast on microcomputer displays. *Vision Research*, 31:1337–1350, 1991.

[13] P. Phillips, H. Wechsler, J. Huang, and P. Rauss. The FERET database and evaluation procedure for face recognition algorithms. *Image and Vision Computing Journal*, 16:295–306, 1998.

[14] S. R. Schweinberger and M. A. Burton. Covert recognition and the neural system for face processing. *Cortex*, 39:9–30, 2003.

[15] T. Tieger and L. Ganz. Recognition of faces in the presence of two-dimensional sinusoidal masks. *Perception and Psychophysics*, 26:163–167, 1979.

[16] H. Wilson and F. Wilkinson. Detection of global structure in glass patterns: implications for form vision. *Vision Research*, 38:2933–2947, 1998.

[17] H. Wilson, F. Wilkinson, and W. Asaad. Concentric orientation summation in human form vision. *Vision Research*, 37:2325–2330, 1997.

[18] Y. Zana and A. Cavancanti. Contrast sensitivity functions to stimuli defined in cartesian, polar and hyperbolic coordinates. *Spatial Vision*, 18:85–89, 2005.

[19] Y. Zana and R. Cesar-Jr. Face recognition based on polar frequency features. *ACM Transactions on Applied Perception*, 3(1), 2006. 2006, to appear.

[20] Y. Zana, R. Cesar-Jr, and R. A. Barbosa. Automatic face recognition system based on local fourier-bessel features. In *XVIII Brazilian Symposium on Computer Graphics and Image Processing*, pages 233–240, 2005.

[21] Y. Zana, R. Cesar-Jr, R. Feris, and M. Turk. Face verification in polar frequency domain: a biologically motivated approach. In *Lecture Notes in Computer Science*, volume 3804, pages 138–190, 2005.