25

Facing Visual Tasks Based on Different Cognitive Architectures

Marcos Ruiz-Soler and Francesc S. Beltran
Universidad de Málaga and Universitat de Barcelona
Spain

1. Introduction

Today’s technology has produced machines that imitate and even exceed many human abilities. Nobody is surprised when a calculator does highly complex mathematical computations or an electronic chess program beats a renowned chess master. Any computer can store and retrieve detailed information about very different topics and establish a multitude of complex relations. However, in spite of recent spectacular advances, robotics has not yet been able to reproduce with the same efficiency some basic tasks every human being can do effortlessly, such as understanding contextual images and moving in complex physical spaces.

The apparent simplicity of understanding images and walking may be an obstacle when it comes to judging the real complexity of these tasks. But even now, after many ingenious attempts to solve the problems inherent to these perceptual and motor processes, technology has still not been able to recreate levels similar to those of a human being. We therefore think it is of interest to review what we know about human beings and try to learn about this very efficient biological system. This chapter will examine the results of research on humans to come up with some valuable suggestions for designs of artificial systems for face recognition.

We will begin with a quick review of the contributions made in two main areas of face-recognition research in the last thirty years: image properties and perceptual tasks. The analysis of both will lead us to explore some internal characteristics of the system (cognitive architecture) that are not usually considered: the representational format of visual information and kinds of flow processing. Based on these factors, we will make some suggestions about the direction future research efforts should take in the field of face recognition.

2. Looking outside: from image properties to visual tasks

2.1 Image properties and spatial frequencies

Given the fact that any image, whether of a human face or any other visual object, can be described in terms of spatial frequencies (SFs) (i.e. it can be described as the sum of a set of sinusoidal grids with different frequencies and orientations), psychophysical research into contrast detection and adaptation to specific SFs has proven that our perceptual system analyses visual input on multiple scales and frequencies (see De Valois & De Valois, 1988;

Source: Face Recognition, Book edited by: Kresimir Delac and Mislav Grgic, ISBN 978-3-902613-03-5, pp.558, I-Tech, Vienna, Austria, June 2007

Graham, 1989). It is therefore generally agreed that spatial filtering is the basic mechanism for extracting visual information from luminance contrasts in early visual processes (see Legge & Gu, 1989; Marr & Hildreth, 1980; Marshall et al., 1996; and Morgan, 1992). In light of all this, one of the main approaches in face-perception research involves manipulating the SF bands in the luminance spectrum of images and observing how these changes affect the performance of visual tasks.

Two main questions were asked when investigating the role of SFs in face perception: (1) What range of SFs is necessary to recognize a face? and (2) in what order are low spatial frequencies (LSFs) and high spatial frequencies (HSFs) integrated in face perception and how does this order affect recognition? Studies done to answer the first question mainly used a *masking approach*, while studies to answer the second question used a *microgenetic approach*. Unfortunately, definitive results were not found because the results obtained to answer the first question showed that an extensive range of SFs seems to play a role in recognition; and the results obtained to answer the second question showed that the order of integration does not always point towards the same length of time or order of integration.

The results of the studies designed to determine what range of SFs is necessary to recognize a face indicated that recognition decreases when images contain only SFs below about 8 cycles/fw (between 6 and 9 cycles/fw), and that the elimination of the SF range between 8 and 16 cycles/fw produces greater disruption than the elimination of SFs outside this range. Hence, the information contained in a small medium range of SFs contributes more to the face-recognition process than the information contained in all the other SFs (Costen et al., 1994, 1996; Näsänen, 1999; Parker & Costen, 1999). However, though all these results indicated that privileged information can be found in medium-range SFs, the role of the SFs outside that range should not be overlooked. The same studies that identified the optimal medium range of SFs also showed acceptable performance by subjects when SFs above and below the medium range were used. Images of faces made with SFs centred at 50.15 cycles/fw or 2.46 cycles/fw (which is extraordinarily far from the medium range) showed a recognition efficiency only 15% lower than the efficiency when recognizing images of faces made with medium-range SFs (Parker & Costen, 1999). Moreover, the tails obtained in the sensitivity function for images of faces indicated that an extensive range of SFs contributes to recognition (Näsänen, 1999). Given all these results, the conclusion was reached that the idea of a “critical range” of SFs for face recognition should be replaced with the notion of an “optimal range” of SFs for face recognition: a preferred, but not exclusive, tendency to use the information contained in a given range of SFs.

The results of the studies designed to determine in what order low spatial frequencies (LSFs) and high spatial frequencies (HSFs) are integrated in face perception and how this order affects recognition appeared to contradict each other: some favoured the hypothesis of anisotropic integration, whereas others pointed to a third interaction factor that might explain why one order of integration is used instead of another. This factor could be the focus of attention and/or the complexity of the stimulus (Bachmann & Kahusk, 1997; Hoeger, 1997, respectively). In summary, all these results indicated that the critical question for predicting subjects’ performance, after the first integrative stage from LSFs to HSFs, is: which SFs provide the information required to solve the on-going task?

2.2 Visual tasks and face perception

Research into face perception using spatial filtering has shown that one of the aspects most analysed are the physical properties of images. In the *masking approach*, the spatial effects of face representation were the main ones studied, whereas in the *microgenetic approach*, the focus was primarily on the temporal effects of face representation. But, as discussed above, no conclusive results were found. This may have been due to the different tasks used in face-perception research. It is therefore necessary to differentiate between them as a first step towards clarifying research results. All of them can easily be grouped into five categories:

1. *Detection*. This consists of distinguishing between face visual stimuli and similar visual stimuli. A detection task asks the viewer of a visual stimulus: "Is *x* a face?" (e.g. Kuehn & Jolicoeur, 1994; Purcell et al., 1996).
2. *Discrimination*. This consists of distinguishing perceptually between pairs of faces, either following a holistic or analytical strategy. A discrimination task asks the following question: "Is *x* the same face as *y*?" The level of complexity in this task depends on the level of similarity of the faces compared, one or more components of which are usually manipulated by computer software (such as the eyes, mouth, nose, hair, chin, etc.) or orientation (frontal, profile or ¾). Examples of this can be found in Bradshaw & Wallace (1971) or Sergent (1984).
3. *Categorization*. This consists of answering the question: "Does this face belong to the category *x*?" It is a classification with two modalities: automatic and controlled. *Automatic categorization* involves classifying a face into a well-learned conceptual category, which demands very little effort. Categorization by sex or race belongs to this group. It has been employed in research about perceptual discrimination based on gender (Bruce et al., 1993; Burton et al., 1993; Brown & Perrett, 1993; Bruce & Langton, 1994; Chronicle et al., 1995;) and to study the so-called "race effect". *Controlled categorization* involves classifying a face into a major category for the subject's goals; it is conscientiously carried out and could admit very different levels of complexity. This category can include judgements about facial emotions, dispositional attributions (e.g. he/she looks intelligent) and situational attributions (e.g. she/he looks doubtful). This task has been used in research into social cognition.
4. *Recognition*. This consists of deciding if a face has been seen before. It is assumed that any known or familiar face will be recognized, and that other faces shown during a controlled projection will also be recognized. Therefore, a task like this demands an answer to the question: "Have you seen face *x* before?".
5. *Identification*. This involves establishing a biunivocal assignation between one face and one specific person. An identification task asks the question: Who does face *x* belong to? (or simply: "Who is he/she?"). The identification task is usually carried out by naming, but an answer such as "It is the face of the president's wife" is also a form of identification. This is the most specific form of face perception.

From a general perspective, all these tasks can be considered specific cases of categorization, ranging from the broadest category ("It is a face") to the most specific one ("It is Marc's face"). Therefore, the cognitive resources required are very different, depending on the level required by the task. As a result, Morrison & Schyns (2001) pointed out that the mechanisms of categorization can modulate the use of different scales, depending on the presence of task-dependent, diagnostic information.

2.3 Interaction between images and tasks: the diagnosticity approach

The varying importance of SFs depending on task demands was described by Schyns (1998). It is well known that one object can be put into different categories, depending on the categorization criteria used. For example, a car can be categorized by trademark, model, power, colour, etc.; and a human face can be categorized by sex, race, expression, attraction, etc. According to Schyns' proposal, the information required to place the same object in one category or another will change depending on the categorization criterion chosen or, in other words, categorization/recognition processes can be characterized as an interaction of task constraints and object information. Task constraints are related to the information needed to place the perceptual object in the category required by the task. For example, given the question: "Is this object a car?", it will be necessary to find certain visual information, such as wheels, rear-view mirrors, a steering wheel, etc., before providing an answer. Object information is related to the informative-perceptual structure available for placing the perceptual object in the category demanded by the task. If it is possible to observe wheels, rear-view mirrors, a steering wheel, etc. in the image of the object, the necessary information is available for categorization and to answer the question. Therefore, given a specific perceptual task, a group of visual characteristics of the object becomes particularly useful (diagnostic), since it provides the information required to place the object in the category that resolves the task.

Information about objects is organized in categories, which are then organized in a hierarchy where it is possible to distinguish three levels (Rosch et al., 1976): *the basic level* (e.g. a car or a face), *the subordinate level* (e.g. a BMW Z8 or Claudia Schiffer) and *the superordinate level* (e.g. a vehicle or a head), where the basic level plays a role of *primal access* (Biederman, 1987) or *entry point* (Jolicoeur et al., 1984) in the hierarchical system. The categorization process at the *superordinate level* requires more functional information than perceptual information, while at the *subordinate level* it requires supplementary perceptual information. Thus, the subordinate level represents maximum informativity and minimum distinctiveness, while the subordinate level represents maximum distinctiveness and minimum informativity. The *basic level* is on an intermediate level between informativity and distinctiveness, and this provides a compromise solution between accuracy in categorization at a more general level and predictive power at a more specific level (Murphy & Lassaline, 1998), which explains its critical role as primal access in the hierarchy. Nevertheless, requirements of informativity and distinctiveness are not uniform for every category, but depend on the subject's level of expertise and history of learning. Therefore, in categorization processes where the subject's expertise skills are at a maximum, as in the case of face recognition, perceptual cues must be diagnostic for the task (sufficient), they cannot overlap with other categories (unique) and they must have sufficient perceptual salience (significant). Therefore, the information I perceive when I see a face will be very different if I have to recognize the face of someone of a different race among people attending a conference, or if I have to recognize the face of a family member among a group of people, or if I have to recognize my partner's face in a shopping centre. In the first case, the colour of the skin or the shape of the eyes can be maximally diagnostic, while in the second and third cases, the configurational properties will probably be maximally diagnostic for recognition.

Oliva & Schyns (1997) found that when the already integrated early perceptual representation is formed, it may be used flexibly in a top-controlled manner permitting selective use of LSFs or HSFs depending on how "diagnostic" they are for the task. Taking this into account, although the possible importance of task demands in face perception has been explicitly affirmed by several researchers (e.g., Costen et al., 1996; McSorley & Findlay, 1999; and Sergent, 1986, 1994), we suggest that a key question for determining the role SFs play in face perception is not really which SFs are necessary or in which sequential order they are integrated, but rather how LSFs and HSFs are made use of in face perception depending on the demands of the task involved. Therefore, the role of different SFs is critically modulated by the subject's visual task and it is only when there is no specific visual task that the mandatory aspects of SF processing work by default. When the results of the research in face perception carried out in the last thirty years are examined from the *diagnosticity approach,* it is possible to see that some contradictions disappear. And this is due to the fact that the questions "Which SFs are critical?" and "Which SFs are integrated?" lose their meaning in an isolated context and have to be considered within the frame of the demands of the task at hand. The questions must then be transformed into "which SFs are diagnostic for recognition/identification of an image?" (Ruiz-Soler & Beltran, 2006).

3. Looking inside: the importance of the functional cognitive architecture

How can it be explained that the same visual task can be solved using different SFs? The observed fact that certain perceptual tasks can be solved using different SFs (Sergent, 1985) makes it necessary to include another factor to explain these data. We believe that, together with image properties and task demands, we must include another explanatory factor: the subject's characteristics (observer), characteristics that affect individual differences in two areas: (1) the mental representation for faces (something conditioned by the familiarity level or expertise level in relation to them) and (2) the preferential strategy for visual processing (something conditioned by the subject's hemispheric dominance or cognitive style).

What is the empirical evidence for considering mental representation a new explanatory factor? With regard to mental representation, memory research using faces as stimuli has reported a different codification of them depending on the previous knowledge level (Liu et al., 2000; O'Toole et al., 1992). Moreover, research into experts and novices using stimuli with perceptual characteristics very similar to faces (complex, symmetrical, 3D, intersimilars, etc.) have proved the existence of different mental representations (Coin et al., 1992; Harvey & Sinclair, 1985; Millward & O'Toole, 1986).

With regard to the processing strategy, research taking into account hemispheric cerebral dominance (Keenan et al., 1989, 1990 and, in particular, Ivry & Robertson, 1998) can be considered, as well as some other research designed to study the development of expert skills in perceptual discrimination (Gauthier & Tarr, 1997; Gauthier et al., 1998; Gauthier et al., 1999; Gauthier & Logothetis, 2000) and the reinterpretation of data from specific research in visual perception. Results point to processing linked to cognitive styles, where some subjects are basically analytical (field-independence subjects) and others are basically holistic (field-dependence subjects), a circumstance that we could re-conceptualize as subjects who preferentially process HSFs and subjects who preferentially process LSFs. Though some previous studies have not shown the relationships between these two aspects (Bruce, 1998), this is a field that we have begun to explore, after creating some procedure controls, by classifying field-dependence subjects, but not merely as those who are excluded from the group of field-independence subjects, as is usually done (Ruiz-Soler et al., 2000).

4. Looking everywhere: new directions in face-recognition research

In this chapter, we have seen how a great deal of research has shown that image properties and task requirements are two interacting factors. We have also seen that the representational format of the information and the preferential processing mode are relevant factors in face perception. What does all this contribute to the design of artificial face-recognition systems? Looking outside shows that the most important information in an image is none other than the information that is most diagnostic (sufficient, unique and significant) for the task at hand. Looking inside shows that we should probably have several representational formats (based on LSFs and HSFs) and a number of different information systems (coarse-to-fine and fine-to-coarse) to come up with a very flexible, efficient system (at least as flexible and efficient as a human being). Designing systems that access representational formats with fine information or that merely use HSFs to process tasks that do not require such fine information (e.g. detection) means having a very inefficient system because it will use much more processing resources than are strictly necessary. But designing systems that have only one representational format or a single processing mode means losing the possibility of performing many of the tasks inherent to face recognition.

5. References

Bachmann, T. & Kahusk, N. (1997). The effects of coarseness of quantization, exposure duration, and selective spatial attention on the perception of spatially quantized (“blocked”) visual images. *Perception, 26*, 1181-1196.

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review*, *94*, 115-145.

Bradshaw, J. L. & Wallace, G. (1971). Models for the processing and identification of faces. *Perception & Psychophysics, 9*, 433-448.

Brown, E. & Perrett, D. (1993). What gives a face its gender? *Perception*, *22*, 829-840.

Bruce, V. (1998, December). Personal communication.

Bruce, V., Burton, A. M., Hanna, E., Healey, P., Mason, O., Coombes, A., Fright, R., & Linney, A. (1993). Sex discrimination: how do we tell the difference between male and female faces? *Perception, 22*, 131-152.

Bruce, V., Langton, S., (1994). The use of pigmentation and shading information in recognising the sex and identities of faces. *Perception 23* (7), 803–822.

Burton, A. M., Bruce, V. & Dench, N. (1993). What’s the difference between men and women? Evidence from facial measurement. *Perception, 22*, 153-176.

Coin, Ch., Versace, R., & Tiberghien, G. (1992). Role of spatial frequencies and exposure duration in face processing: potential consequences on the memory format of facial representations. *Cahiers de Psychologie Cognitive,* 1, 79-98.

Costen, N. P., Parker, D. M. & Craw, I. (1994). Spatial content and spatial quantization effects in face recognition. *Perception*, *23*, 129-146.

Costen, N. P., Parker, D. M. & Craw, I. (1996). Effects of high-pass and low-pass spatial filtering on face identification. *Perception & Psychophysics*, *58*, 602-612.

Chronicle, E. P., Chan, M. Hawkings, C., Mason, K., Smethurst, K., Stallybrass, K., Westrope, K. & Wright, K. (1995). You can tell by the nose -- judging sex from an isolated facial feature. *Perception, 24*, 969-973.

De Valois, R.L. & De Valois, K.K. (1988). *Spatial vision.* New York: Oxford University Press

Gauthier, I. & Logothetis, N. (2000). Is face recognition not so unique after all?. *Journal of Cognitive Neuropsychology.*

Gauthier, I. & Tarr, M.J. (1997). Becoming a "Greeble" expert: Exploring mechanisms for face recognition. *Vision Research, 37* (12), 1673-1682.

Gauthier, I., Tarr, M.J., Anderson, A.W., Skudlarski, P., & Gore, J.C. (1999). Activation of the middle fusiform "face area" increases with expertise in recognizing novel objects. *Nature Neuroscience, 2* (6), 568-573.

Gauthier, I., Williams, P., Tarr, M.J., & Tanaka, J.W. (1998). Training "Greeble" experts: A framework for studying expert object recognition processes. *Vision Research, 38*, 2401-2428.

Graham, N.V.S. (1989). *Visual pattern analyzers*. New York: Oxford University Press.

Harvey, L.O. & Sinclair, G.P. (1985). On the quality of visual imagery. *Investigative Ophthalmology & Visual Science*, *Suppl. 26*, 281

Hoeger, R. (1997). Speed of processing and stimulus complexity in low-frequency and high-frequency channels. *Perception*, *26*, 1039-1045.

Ivry, R.B. & Robertson, L. (1998). *The two sides of perception*. Massachusetts, NJ: The MIT Press.

Jolicoeur, P., Gluck, M., & Kosslyn, S. M. (1984). Pictures and names: Making the connexion. *Cognitive Psychology*, *19*, 31-53.

Keenan, P.A., Witman, R. & Pepe, J. (1989). Hemispheric assymetry in the processing of high and low spatial frequencies: A facial recognition task. *Brain & Cognition, 11*, 229-237.

Keenan, P.A., Whitman, R. & Pepe, J. (1990). Hemyspheric assymetry in the processing of high and low spatial frequencies: A facial recognition task erratum. *Brain & Cognition, 13*, 130

Kuehn, S.M. & Jolicoeur, P. (1994). Impact of quality of image, orientation, and similarity of the stimuli on visual search for faces. *Perception, 23*, 95-122.

Legge, G.E. & Gu, Y. (1989). Stereopsis and contrast. *Vision Research, 29*, 989-1004.

Liu, Ch-H., Collin, Ch., Rainville, S. & Chaudhuri, A. (2000). The effects of spatial frequency overlap on face recognition. *Journal of Experimental Psychology: Human Perception and Performance, 26,* 956-979.

Marr, D. & Hildreth, E.C. (1980). Theory of edge detection. *Proceedings of the Royal Society of London B, 207,* 187-217.

Marshall, J.A., Burbeck, C.A., Ariely, J.P., Rolland, J.P. & Martin, K.E. (1996). Occlusion edge blur: A cue to relative visual depth. *Journal of the Optical Society of America A, 13,* 681-688.

McSorley, E. & Findlay, J. M. (1999). An examination of a temporal anisotropy in the visual integration of spatial frequencies. *Perception, 28*, 1031-1050.

Millward, R.B. & O'Toole, A. (1986). Recognition memory transfer between spatial-frequency analyzed faces. In H.D. Ellis, M.A. Jeeves, F. Newcombe & A. Young (Eds.), *Aspects of face processing* (pp. 34-44). Dordrech, The Netherlands: Martinus Nijhoff.

Morgan, M.J. (1992). Spatial filtering precedes motion detection. *Nature, 355,* 344-346.

Morrison, D.J. & Schyns, P.G. (2001). Usage of spatial scales for the categorization of faces, objects, and scenes. *Psychonomic Bulletin & Review, 8* (3), 454-469.

Murphy, G. L., & Lasaline, M .E. (1998). Hierarchical structure in concepts and the basic level of categorization. In K. Lamberts, D. R. Shanks, et al. (Eds.) *Knowledge, Concepts and Categories: Studies in Cognition* (pp. 93-131). Cambridge, Massachussett: MIT Press.

Näsänen, R. (1999). Spatial frequency bandwidth used in the recognition of facial images. *Vision Research, 39,* 3824-3833.

Oliva, A. & Schyns, P. (1997). Coarse blobs or fine edges? Evidence that information diagnosticity changes the perception of complex visual stimuli. *Cognitive Psychology, 34,* 72-107.

O'Toole, A., Millward, R.B. & Anderson, J.A. (1988). A physical system approach to recognition memory for spatially transformed faces. *Neural Networks, 1,* 179-199.

Parker, D.M. & Costen, N.P. (1999). One extreme or the other or perhaps the golden mean? Issues of spatial resolution in face processing. *Current Psychology: Development, Learning, Personality, Social, 18* (1), 118-127.

Purcell, D.G., Stewart, A.L. & Skov, R.B. (1996). It takes a confounded face to pop out of a crowd. *Perception, 25*, 1091-1108.

Rosch, E., Mervis, C. B., Gray, W. D., Johnson, D. M., & Boyes-Braem, P. (1976). Basic objects in natural categories. *Cognitive Psychology, 8*, 382-352

Ruiz-Soler, M. & Beltran, F.S. (2006). Face perception: An integrative review of the role of spatial frequencies. *Psychological Research ,70,* 273-292.

Ruiz-Soler, M., López, E., Pelegrina, M., Videra, A. & Wallace, A. (2000, July). *Face recognition and response criterion: Moderating effects of cognitive style.* Paper presented at XXVII International Congress of Psychology, Stockholm, Sweden.

Schyns, P. G. (1998). Diagnostic recognition: task constraints, object information and their interactions. *Cognition, 67,* 147-179.

Sergent, J. (1986). Microgenesis in face perception. In H.D. Ellis, M.A. Jeeves, F. Newcombe & A. Young (Eds.), *Aspects of face processing* (pp. 17-73). Dordrecht, The Netherlands: Martinus Nijhoff.

Sergent, J. (1994). Brain-imaging studies of cognitive functions. *Trends in Neurosciences, 17* (6), 221-227.

Sergent, J. (1984). An investigation into component and configural process underlying face perception. *British Journal of Psychology, 75*, 221-242.

Face Recognition

Edited by Kresimir Delac and Mislav Grgic

ISBN 978-3-902613-03-5
Hard cover, 558 pages
Publisher I-Tech Education and Publishing

Published in print edition July, 2007

This book will serve as a handbook for students, researchers and practitioners in the area of automatic (computer) face recognition and inspire some future research ideas by identifying potential research directions. The book consists of 28 chapters, each focusing on a certain aspect of the problem. Within every chapter the reader will be given an overview of background information on the subject at hand and in many cases a description of the authors' original proposed solution. The chapters in this book are sorted alphabetically, according to the first author's surname. They should give the reader a general idea where the current research efforts are heading, both within the face recognition area itself and in interdisciplinary approaches.

Marcos Ruiz-Soler and Francesco S. Beltran (2007). Facing Visual Tasks Based on Different Cognitive Architectures, Face Recognition, Kresimir Delac and Mislav Grgic (Ed.), ISBN: 978-3-902613-03-5, InTech, Available from: http://www.intechopen.com/books/face_recognition/facing_visual_tasks_based_on_different_cognitive_architectures

INTECH
open science | open minds

InTech Europe
University Campus STeP Ri
Slavka Krautzeka 83/A
51000 Rijeka, Croatia
Phone: +385 (51) 770 447
Fax: +385 (51) 686 166
www.intechopen.com

InTech China
Unit 405, Office Block, Hotel Equatorial Shanghai
No.65, Yan An Road (West), Shanghai, 200040, China
中国上海市延安西路65号上海国际贵都大饭店办公楼405单元
Phone: +86-21-62489820
Fax: +86-21-62489821