ELSEVIER

Cognitive Systems Research 1 (2002) 000–000

Cognitive Systems RESEARCH

www.elsevier.com/locate/cogsys

Book review

Review of: Dynamic Vision – From Images to Face Recognition

Action Editor: Stefan Wermter

Eric Postma

IKAT/Computer Science, Universiteit Maastricht, PO Box 616, 6200 MD, Maastricht, The Netherlands

Accepted 9 January 2002

The artificial recognition of faces still poses a major challenge to present-day researchers in cognitive science and related disciplines. Gong, McKenna and Psarrou (GMP) took the bold step of writing a book on the current state of the art in face recognition. Although face recognition is a specialized field, the developments over the last decade have been tremendous. Covering all developments in a single volume is therefore not an easy task. In the preface of their book, GMP remark that although many books on face recognition have been written, a ''book that provides a coherent and unified treatment of the issue from a computational and systems perspective'' (p. xvi) does not yet exist. GMP express the hope that their book does provide such a unified treatment.

One of the most interesting aspects of vision research is its multidisciplinary nature. New insights from psychological and biological studies fuel the development of new models and techniques. For instance, recent studies by Leopold et al. (2001) suggest human face recognition to be based on adaptive prototypes. Conversely, novel schemes for recognition developed in, for instance, computer science may provide new theoretical frameworks to steer research in psychology and biology. The power of cross-fertilization is evident from *Dynamic Vision*. Since the book emphasizes the engineering aspects of face recognition it does not address the biological plausibility of the approaches to full extent. However, at the end of each chapter interesting parallels with biological visual systems are drawn that aid the reader in appreciating the biological plausibility of the approaches discussed.

The book is divided in five main parts. Below I discuss each part separately.

Part I – Background

The first part of *Dynamic Vision* presents background information on issues relevant to dynamic face recognition. Chapter 2, on perception and representation, is of particular interest to readers unfamiliar with image recognition. GMP discuss the three main types of representations used in vision research: three-dimensional representations (Marr, 1982), two-dimensional (view-based) representations (e.g., Ullman, 1996), and iconic (template-based) representations (e.g., Rao & Ballard, 1995). Present-day face recognition systems rely heavily on automatic learning techniques. Chapter 3 gives a concise introduction to automatic learning techniques covering, amongst others, statistical learning theory, Bayesian inference, dimensionality reduction, clustering, and support vector machines. The theory and

1389-0417/02/$ – see front matter
PII: S1389-0417(02)00038-4

techniques covered provide the reader with sufficient background knowledge for the subsequent chapters.

Part II – From Sensory to Meaningful Perception

The second part of the book follows the ''feedforward route'' towards face recognition by focusing on perceptual grouping, attention, face detection, and face tracking, respectively. In human vision, perceptual grouping is assumed to be a pre-attentive process that groups features on the basis of Gestalt laws and related principles (see, e.g., Palmer, 1999). The pre-attentive process provides cues for directing attention to parts of the visual scene. Chapter 4 describes how motion and color cues are particularly effective in determining salient locations for attention. GMP conclude the chapter by stating that motion and color cues alone are insufficient to determine the location of the face within a complex scene. They claim that ''explicit knowledge about *what faces look like*'' (p. 79) is required. The word ''explicit'' should not be taken to mean that the underlying representations are explicit (e.g., rule based). As is evident from subsequent chapters, the knowledge about what faces look like is often obtained by means of statistical learning techniques that result in implicit representations. Chapter 5 focuses on determining the (near-) frontal views of faces using template-based representations. The chapter uncovers many practical considerations in dealing with real-world classification problems such as detecting a face in a cluttered dynamic environment. In chapter 6, pose-invariant and pose estimation are considered. The limited performance of existing computational techniques leads quite naturally to considering the spatio-temporal properties of moving faces. In chapter 7, various techniques such as Kalman filters and hidden Markov models are evaluated on their ability to predict the spatio-temporal patterns of facial movements.

Part III – Models of Identity

The problem of identifying a given face is addressed in the third part of *Dynamic Vision*. Chapter 8 addresses identification and generalization from a single view. Amongst others, the use of Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA) is discussed. Shortcomings of both techniques are addressed. In Chapter 9, multi-view identification of static faces is considered. GMP explore the use of shape models for representing the different poses of a particular face. In addition, the generalization from a single view to multiple views is treated. Then, in Chapter 10, the identification of moving faces is discussed. Here, the use of temporal information becomes pivotal to deal with the dynamics of faces. Existing representation schemes fall short in accounting for temporal variations. GMP propose the use of ''spatio-temporal signatures'' for identifying dynamic faces. Their proposal generalizes naturally to the identification of arbitrary objects.

Part IV – Perception of Context

Perceptual integration is discussed in chapter 11. In particular, GMP go into the problem of how bottom-up and top-down processes are integrated, where bottom-up processes are general sensory-based processes and top-down processes more specific knowledge-based processes. The chapter is interesting but lacks the rigor and clarity of the earlier chapters, because it addresses theoretical and philosophical issues that are beyond the (mainly engineering) scope of the book. Part IV ends with a discussion on the broader area of recognizing objects in dynamic scenes. Chapter 12 points at some interesting future applications such as visually mediated interaction, immersive virtual reality, and visual database screening.

Part V – Appendices

The final part of the book contains some really useful appendices giving an overview of facial image databases (A), commercial systems (B), and mathematical background on the main techniques employed in the book (principal component analysis, linear discriminant analysis, Gaussian mixture esti-

mation, Kalman filters, Bayesian belief networks, hidden Markov models, and Gabor wavelets).

Evaluation

Dynamic Vision is a unique book. To my knowledge, there is no comparable book that covers the broad and complex domain of adaptive visual recognition in such a readable way. The clear presentation style helps the reader to appreciate the painstaking work involved in making the automatic recognition of faces possible. The emphasis on the engineering aspects of face recognition is complementary to the good many books that emphasize biological vision at the cost of the engineering aspects (e.g., Bruce & Young, 1998; Zeki, 1993). My only point of critique concerns the title of the book, which I consider to be slightly misleading. Rather than being about ''dynamic'' face recognition, the book is about the recognition of static and dynamic faces. In many of the approaches described in the book, it is the face (or person) that is dynamic rather than the recognition method. Truly dynamic vision would require an approach based on what O'Regan & Noë (in press) call sensorimotor contingencies (see also Pfeifer & Scheier, 1999).

Notwithstanding this minor point, my overall evaluation of the book is very positive. After reading *Dynamic Vision*, the reader is acquainted with most state-of-the-art approaches to face (and object) recognition. I therefore conclude that the authors were successful in providing ''a coherent and unified treatment of the issue from a computational and systems perspective'' and highly recommend the book to any researcher interested in face recognition or visual recognition in general.

References

Bruce, V. & Young, A. W. (1998). *In the Eye of the Beholder: The Science of Face Perception*. Oxford: Oxford University Press.

Leopold, D. A., O'Toole, A. J., Vetter, T., & Blanz, V. (2001). Prototype-referenced Shape Encoding Revealed by High-level Aftereffects. *Nature Neuroscience*, *4*(1), 89–94.

Marr, D. (1982). *Vision*. Freeman.

O'Regan, J. K. & Noë, A. (in press) A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24 (5).

Palmer, S. (1999). *Vision Science: Photons to Phenomenology*. Cambridge: MIT Press.

Pfeifer, R. & Scheier, C. (1999). *Understanding Intelligence*. Cambridge, MA: MIT Press.

Rao, R. P. N., & Ballard, D. H. (1995). An Active Vision Architecture based on Iconic Representations. *Artificial Intelligence*, *78*, 461–505.

Ullman, S. (1996). *High-level Vision: Object Recognition and Visual Cognition*. Cambridge, MA: MIT Press.

Zeki, S. (1993). *A Vision of the Brain*. Oxford: Blackwell Scientific.