

Preface

This thesis is published in partial fulfillment of the requirements for the degree of ‘Doktor der Naturwissenschaften’ (Dr. rer. nat.) at the Department of Mathematics and Computer Science of Freie Universität Berlin. Prof. Dr. Raúl Rojas (FU Berlin) and Prof. Dr. Volker Sperschneider (Osnabrück) acted as referees. The thesis was defended on November 27, 2002.

Summary of the Thesis

Human performance in visual perception by far exceeds the performance of contemporary computer vision systems. While humans are able to perceive their environment almost instantly and reliably under a wide range of conditions, computer vision systems work well only under controlled conditions in limited domains.

This thesis addresses the differences in data structures and algorithms underlying the differences in performance. The interface problem between symbolic data manipulated in high-level vision and signals processed by low-level operations is identified as one of the major issues of today’s computer vision systems. This thesis aims at reproducing the robustness and speed of human perception by proposing a hierarchical architecture for iterative image interpretation.

I propose the use of hierarchical neural networks for representing images at multiple abstraction levels. The lowest level represents the image signal. As one ascends these levels of abstraction, the spatial resolution of two-dimensional feature maps decreases while feature diversity and invariance increase. The representations are obtained using simple processing elements that interact locally. Recurrent horizontal and vertical interactions are mediated by weighted links. Weight sharing keeps the number of free parameters low. Recurrence allows for integration of bottom-up, lateral, and top-down influences.

Image interpretation in the proposed architecture is performed iteratively. An image is interpreted first at positions where little ambiguity exists. Partial results then bias the interpretation of more ambiguous stimuli. This is a flexible way to incorporate context. Such a refinement is most useful when the image contrast is low, noise and distractors are present, objects are partially occluded, or the interpretation is otherwise complicated.

The proposed architecture can be trained using unsupervised and supervised learning techniques. This allows one to replace manual design of application-specific computer vision systems with the automatic adaptation of a generic network. The task to be solved is then described using a dataset of input/output examples.

Applications of the proposed architecture are illustrated using small networks. Furthermore, several larger networks were trained to perform non-trivial computer vision tasks, such as the recognition of the value of postage meter marks and the binarization of matrixcodes. It is shown that image reconstruction problems, such as super-resolution, filling-in of occlusions, and contrast enhancement/noise removal, can be learned as well. Finally, the architecture was applied successfully to localize faces in complex office scenes. The network is also able to track moving faces.

Acknowledgements

My profound gratitude goes to Prof. Raúl Rojas, my mentor and research advisor, for guidance, contribution of ideas, and encouragement. I salute Raúl's genuine passion for science, discovery and understanding, superior mentoring skills, and unparalleled availability.

The research for this thesis was done at the Computer Science Institute of the Freie Universität Berlin. I am grateful for the opportunity to work in such a stimulating environment, embedded in the exciting research context of Berlin. The AI group has been host to many challenging projects, e.g., to the RoboCup FU-Fighters project and to the E-Chalk project. I owe a great deal to the members and former members of the group. In particular, I would like to thank Alexander Gloye, Bernhard Frötschl, Jan Dösselmann, and Dr. Marcus Pfister for helpful discussions.

Parts of the applications were developed in close cooperation with Siemens ElectroCom Postautomation GmbH. Testing the performance of the proposed approach on real-world data was invaluable to me. I am indebted to Torsten Lange, who was always open for unconventional ideas and gave me detailed feedback, and to Katja Jakel, who prepared the databases and did the evaluation of the experiments.

My gratitude goes also to the people who helped me to prepare the manuscript of the thesis. Dr. Natalie Hempel de Ibarra made sure that the chapter on the neurobiological background reflects current knowledge. Gerald Friedland, Mark Simon, Alexander Gloye, and Mary Ann Brennan helped by proofreading parts of the manuscript. Special thanks go to Barry Chen who helped me to prepare the thesis for publication.

Finally, I wish to thank my family for their support. My parents have always encouraged and guided me to independence, never trying to limit my aspirations. Most importantly, I thank Anne, my wife, for showing untiring patience and moral support, reminding me of my priorities and keeping things in perspective.

Berkeley, June 2003

Sven Behnke