Biologically Inspired Methods for Model Matching and Object Recognition

Rolf P. Würtz

Institut für Neuroinformatik, Ruhr-Universität Bochum, D-44780 Bochum, Germany
Email: Rolf.Wuertz@neuroinformatik.ruhr-uni-bochum.de
http://www.neuroinformatik.ruhr-uni-bochum.de/ini/PEOPLE/rolf/

Abstract

I will present various systems for the recognition of human faces and technical objects. They consist of three steps: feature extraction, solving the correspondence problem, and the actual comparison with stored models. Feature extraction is based on detailed models of cells in the primary visual cortex. The matching will be done by methods based on the Dynamic Link Architecture, which is currently gaining much biological plausibility.

1 Introduction

The major problem in face and object recognition is establishing a correspondence map between a stored model and a presented image. This means that pairs of points in model and image must be found which are images of the same point on the physical face. This is not trivial and has acquired the name *correspondence problem*.

The difficulty of the correspondence problem depends on the choice of features. If, e.g., grey values of pixels are taken as local features, there is a lot of ambiguity, i.e., many points from very different locations share the same pixel value without being correspondent points. A possible remedy to that consists in combining local patches of pixels, which of course reduces this ambiguity. If this is done too extensively, i.e., if local features are influenced by a large area, the ambiguities disappear if identical images are used, but the features become more and more sensitive to distortions and changes in background.

2 Features

2.1 Gabor jets

The processing of a retinal grey-level image in *simple cells* of the primary visual cortex can be modeled by a wavelet transform based on complex-valued Gabor functions [3, 8]. The single wavelet is parameterized by its two-dimensional spatial frequency vector. The responses of all spatial frequencies of some fixed length form a *frequency level*, which assigns a small feature vector to all image points on an appropriate sampling grid. These features have turned out to be a good compromise in the dilemma discussed above. Furthermore, the complex numbers invite a splitting into *modulus* and *phase*, which is very convenient for matching purposes. The modulus of the Gabor transform constitutes a good model for *complex cells*, which are also found in the visual cortex. The systems I am describing differ in the *sampling* of this transform. The ones in [3, 7] need a sampling grid which is uniformly dense for all spatial frequencies, the one in [8, 9] uses a pyramidal arrangement.

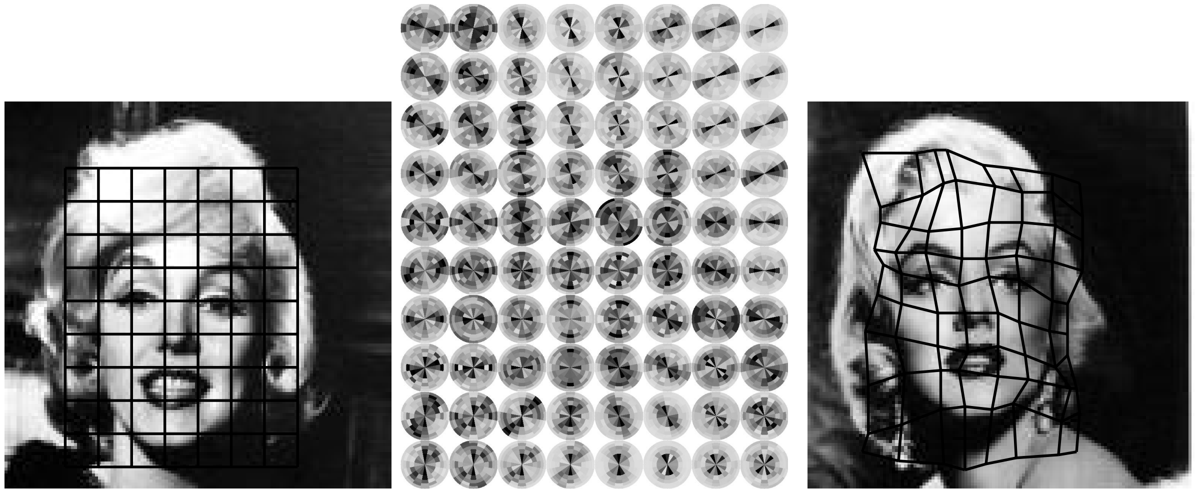

Figure 1: Elastic graph matching. The left hand side shows the model graph, the center a visualization of the Gabor features, and the right hand side the graph with maximal similarity to the one on the left.

2.2 Corners and edges

Beside simple and complex cells a third important type of cells in the primary visual cortex are *end-stopped cells*.In [2] a model is presented of how the responses of these cells can be understood as the output of a feedforward neural network that takes its inputs from the complex cells mentioned above. These cells are a natural candidate for corner detection. However, they are also influenced by noise and texture. As has been shown in [10, 11], a robust corner detector can be constructed by combining the outputs of end-stopped cells over a range of scales.

Edge detection is a problem that has haunted computer vision from the very beginning. The complex cells are edge detectors with fair performance, but a suitable combination of scales is also required. In the system described in section 3.4 a top-down approach will be employed by following lines from corner to corner on a trajectory that yields maximal evidence for the presence of an edge.

3 Topology

In order to overcome the feature ambiguity the *relative position* of the features must be taken into account. Four different ways to do this will be presented. First comes the *elastic graph* approach. The second and third will be coarse-to-fine matching, first in a technical implementation and then in a detailed neuronal model based on the Dynamic Link Architecture [5, 6]. finally, I will describe a matching system based on *edge graphs*.

3.1 Elastic graphs

In the system described in [3, 7] the model faces are represented by sparse graphs which are vertex labeled with the Gabor features and edge labeled with the distance vector of the connected vertices. Matching is done by first optimizing the similarity of an undistorted copy of the graph in the input image and then optimizing the individual locations of the vertices. This results in a distorted graph whose vertices are at corresponding locations to the ones in the model graph. Rectangular model graphs arrangement has been chosen in [3], in [7] the vertices have been carefully placed on salient points, thus yielding a larger recognition rate.

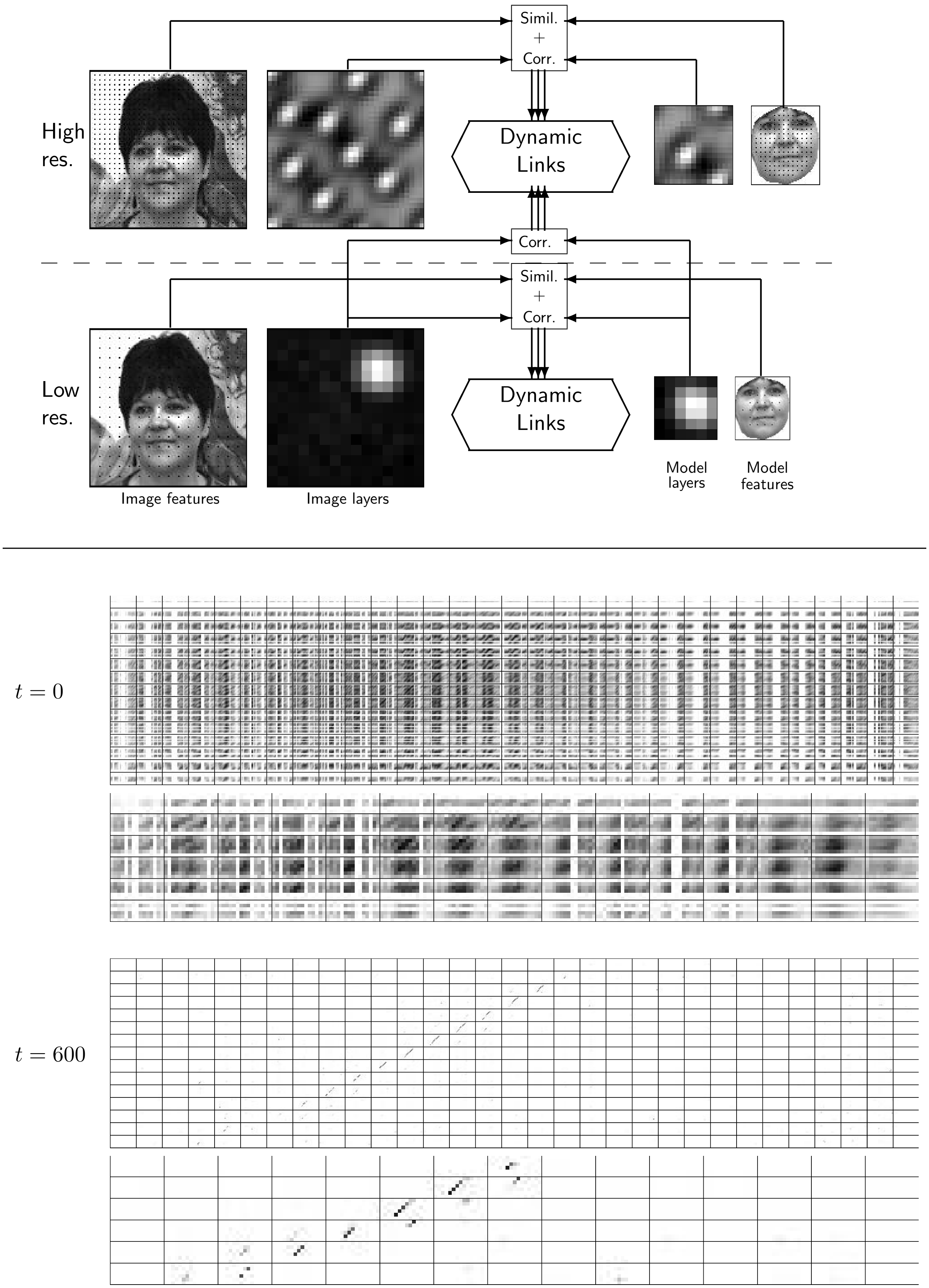

Figure 2: Hierarchical Dynamic Link Matching. Above: The setup of the matching structure. Below: The development of the link distribution is from the highly ambiguous feature similarities to a clear one-one-mapping

3.2 Coarse to fine matching

The choice of features in the face recognition system from [3, 7] has two major disadvantages. At any given point, all spatial frequencies must be present, which results in undersampling for high, oversampling for low frequencies. Furthermore, the features at points close to the object boundary are corrupted by the background. This has been remedied in [8, 9] by introducing a pyramidal representation of model and image information.

Those representations can be matched using the following modules:

1. The *coarse localization* of the counterpart of the model in the image is done by global template matching of the vectors of Gabor amplitudes on the lowest frequency level. This is not very expensive, because the resolution is low, and yields a first rough correspondence mapping.
2. Mappings acquired using only the amplitudes of the Gabor responses are not very precise, because the fine geometrical information resides in the phases. On the other hand, the phases or the full complex responses are not suitable for template matching because they depend strongly on the sampling grid. Therefore, a *local phase matching* has been implemented that enhances the accuracy of amplitude-based mappings. This can be done in parallel on all model points.
3. In order to cope with occlusion problems, it must be possible to *exclude points* from the mapping. This is done on the basis of poor similarity, which is also possible in parallel on all image points.
4. Finally, any mapping can be refined by local template matching with amplitudes from the next higher frequency level. For this, the model is split up into several patches that independently search for correspondences in an area defined by the coarse mapping already known.

The pyramidal arrangement has the advantage that all responses in the stored model which are influenced by the background can be discarded, but the object (which here is the face without hair) still represented well enough for recognition. The structure of image and model features is the same as in the feature layers shown in figure 2. Note that the image to be analyzed consists of a full pyramid, a presegmentation is not required. The resulting invariance under changes in hairstyle and background constitutes an important advantage of this system as compared with the ones in [3, 7]. For more details on the coarse-fine system see [8, 9].

3.3 Coarse-to-fine matching with neuronal dynamics

All matching methods presented in this paper are inspired by a biological model based on the Dynamic Link Architecture [5, 6]. The system from the previous subsection can be implemented in neuronal dynamics as follows.

As shown in figure 2, each frequency level of image and model, respectively, is represented by a feature layer and a topology layer. The topology layers are interconnected by matrices of dynamic links. The internal wirings of the topology layers support a single blob (see [1]) on the lowest level and a structure of coherently moving blobs on the higher levels.

On the lowest frequency level a single blob of activity moves across the image and model layer, respectively. Dynamic links between these layers are initialized to the (highly ambiguous) feature similarities. Links grow or decline according to a combination of feature similarity and correlated activation. This enforces correct neighborhood relationships in addition to feature similarity. On the higher levels the established correspondences are refined by several blobs in parallel. The lower half of figure 2 shows that the system converges to a state where basically only the one link per topology neuron is active that is connected with the correct correspondence. For more details see [8, 12]

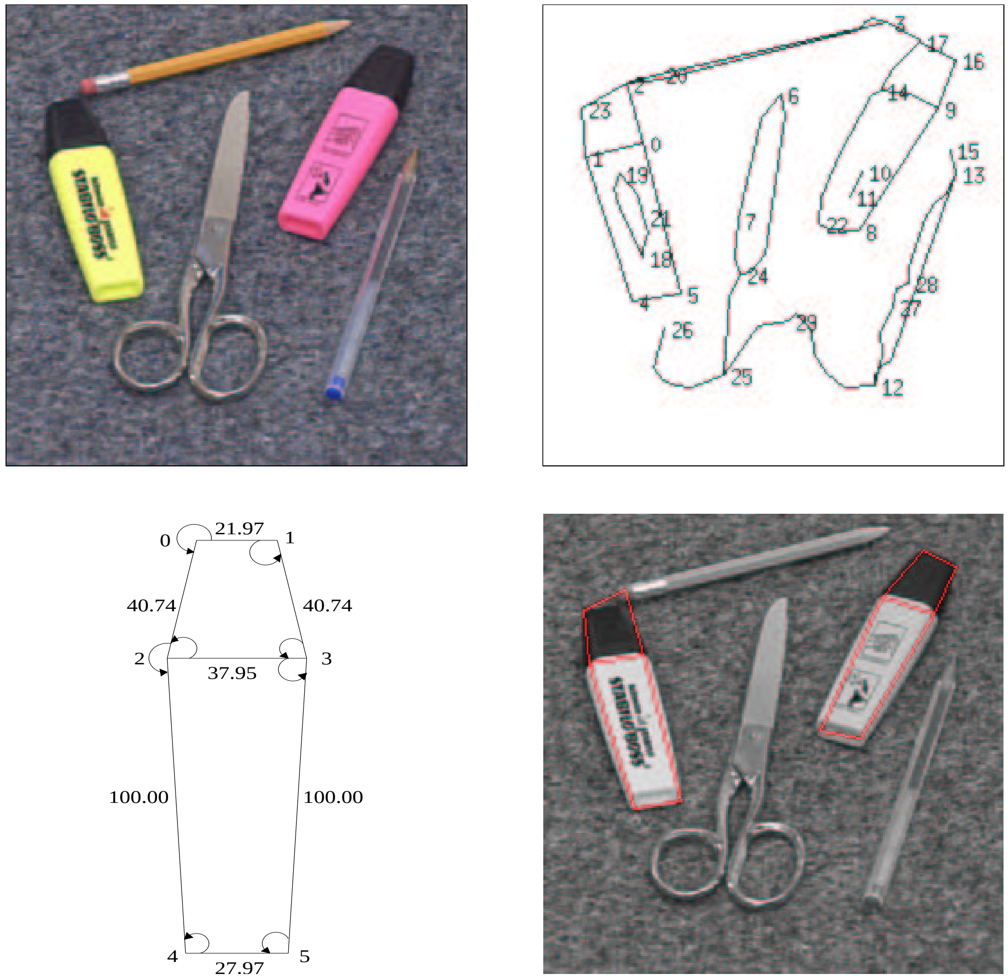

Figure 3: Edge graph matching. The lower left image shows a scene of objects. From this, a graph is extracted which has image corners as vertices and image edges as edges (upper right). The model graphs (lower left) are vertex-labeled with angles and edge-labeled with relative lengths. The resulting matches are shown in the lower right image.

3.4 Symbolic edge graphs

Finally, I briefly presented an object recognition system based on symbolic graphs with object corners as vertices and outlines as edges. Graphs are constructed by line-following between corners. Model matching is then done by finding subgraph isomorphisms in the image graph. An example is presented in figure 3. For details see [4].

We could show that the matching of edge graphs is an efficient method for model matching and object recognition. The choice of labels makes the problem tractable in spite of the NP-completeness of general graph matching. Furthermore, invariance under translation, rotation and scaling is achieved without extra computation.

Acknowledgments: Most of the work described here has been done by or in close cooperation with other people. The author wishes to thank Christoph von der Malsburg, Jan C. Vorbrüggen, Laurenz Wiskott, and Tino Lourens for excellent cooperation. Parts of the work described were supported by grants from the German Research Minister and the European Commission.

References

[1] Shun-Ichi Amari. Dynamics of pattern formation in lateral-inhibition type neural fields. *Biological Cybernetics*, 27:77–87, 1977.

[2] Friedrich Heitger, Lukas Rosenthaler, Rüdiger von der Heydt, Esther Peterhans, and Olaf Kübler. Simulation of neural contour mechanisms: from simple to end-stopped cells. *Vision Research*, 32(5):963–981, 1992.

[3] Martin Lades, Jan C. Vorbrüggen, Joachim Buhmann, Jörg Lange, Christoph von der Malsburg, Rolf P. Würtz, and Wolfgang Konen. Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computers*, 42(3):300–311, 1993.

[4] Tino Lourens and Rolf P. Würtz. Object recognition by matching symbolic edge graphs. In *Proceedings of the Third Asian Conference on Computer Vision, Hong Kong, January 8 - 11, 1998*, Lecture Notes in Computer Science. Springer Verlag, 1998. In press.

[5] Christoph von der Malsburg. The correlation theory of brain function. Technical report, Max-Planck-Institute for Biophysical Chemistry, Postfach 2841, Göttingen, FRG, 1981. Reprinted 1994 in: Schulten, K., van Hemmen, H.J. (eds.), *Models of Neural Networks*, Vol. 2, Springer.

[6] Christoph von der Malsburg. Nervous structures with dynamical links. *Ber. Bunsenges. Phys. Chem.*, 89:703–710, 1985.

[7] Laurenz Wiskott, Jean-Marc Fellous, Norbert Krüger, and Christoph von der Malsburg. Face recognition and gender determination. In Martin Bichsel, editor, *Proceedings of IWAFGR95, Zürich, June 1995*, pages 92–97, 1995.

[8] Rolf P. Würtz. *Multilayer Dynamic Link Networks for Establishing Image Point Correspondences and Visual Object Recognition*, volume 41 of *Reihe Physik*. Verlag Harri Deutsch, Thun, Frankfurt am Main, 1995.

[9] Rolf P. Würtz. Object recognition robust under translations, deformations and changes in background. *IEEE Transactions on Pattern Recognition and Machine Intelligence*, 19(7):769–775, 1997.

[10] Rolf P. Würtz and Tino Lourens. Corner detection in color images by multiscale combination of end-stopped cortical cells. In Wulfram Gerstner, Alain Germond, Martin Hasler, and Jean-Daniel Nicoud, editors, *Artificial Neural Networks – ICANN '97*, volume 1327 of *Lecture Notes in Computer Science*, pages 901–906, Berlin, Heidelberg, New York, 1997. Springer Verlag.

[11] Rolf P. Würtz and Tino Lourens. Corner detection in color images through a multiscale combination of end-stopped cortical cells. *Vision Research*, 1997. Submitted.

[12] Rolf P. Würtz and Christoph von der Malsburg. A hierarchical dynamic link network for correspondence maps between image points. 1997. In preparation.