From: Proceedings of ESANN '97, Bruges, Belgium, 16–18 April 1997, (Michel Verleysen, ed.), D facto, Brussels, 1997

Neuronal theories and technical systems for face recognition

Rolf P. Würtz

Computing Science, University of Groningen, The Netherlands
Email: rolf@cs.rug.nl

Abstract. I present various systems for the recognition of human faces. They consist of three steps: feature extraction, solving the correspondence problem, and the actual comparison with stored faces. Two of them are implemented in the Dynamic Link Architecture and are, therefore, close to biological hardware, the others are more technical in nature but also have some biological plausibility. At the end, I will briefly discuss the coherence with the results of psychophysical experiments on human face recognition.

1. Introduction

The major problem in face recognition is establishing a correspondence map between a stored model face 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

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, 7]. 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

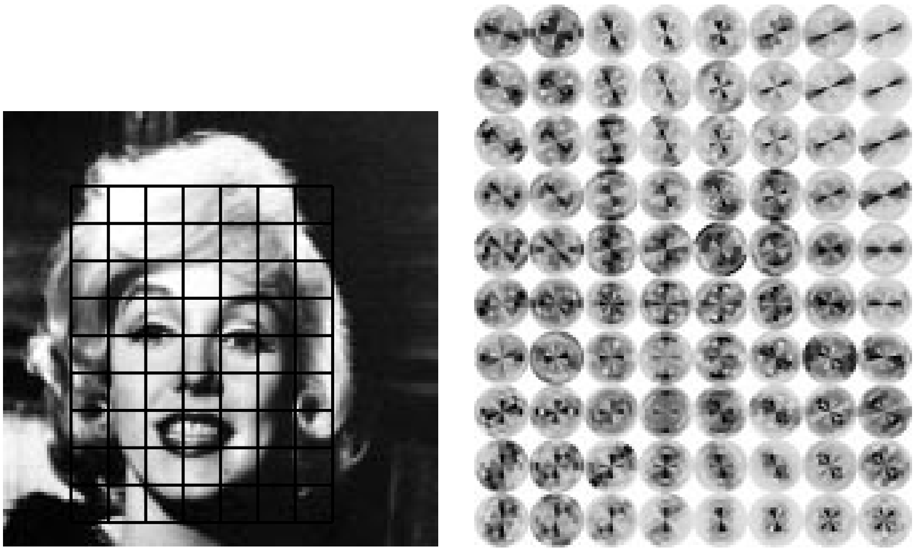

Figure 1: Visualisation of Gabor features with frequency-independent sampling. Each point in the grid is assigned a little frequency space, which is visualized on the right hand side. The grey values of the little segments show the moduli of the Gabor responses as a function of the (two-dimensional) spatial frequency.

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 systems I am describing differ in the *sampling* of this transform. The ones in [3, 5] need a sampling grid which is uniformly dense for all spatial frequencies, the one in [7, 8] uses a pyramidal arrangement.

3. Topology

In order to overcome the feature ambiguity the *relative position* of the features must be taken into account. Three different ways to do this will be presented.

3.1. Elastic graphs

In the system described in [3, 5, 2] the model faces are represented by sparse graphs (see figure 1) 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 invdividual locations of the vertices. This results in a distorted graph whose vertices are at corresponding locations to the ones

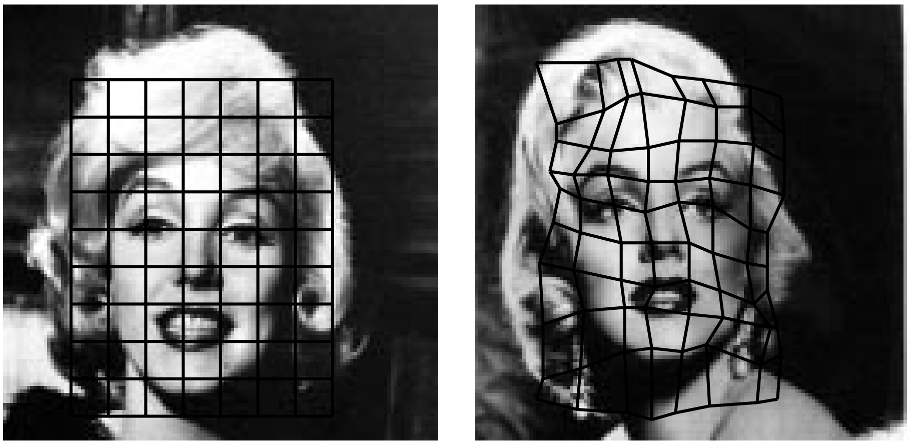

Figure 2: The result of labeled graph matching. The right hand side shows the graph with maximal similarity to the one on the left.

in the model graph (see figure 2). The rectangular model graph arrangement has been chosen in [3], in [5] the vertices have been carefully placed on salient points, thus yielding a larger recognition rate.

3.2. Coarse to fine matching

The pyramidal representations of image and model [7, 8] 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.

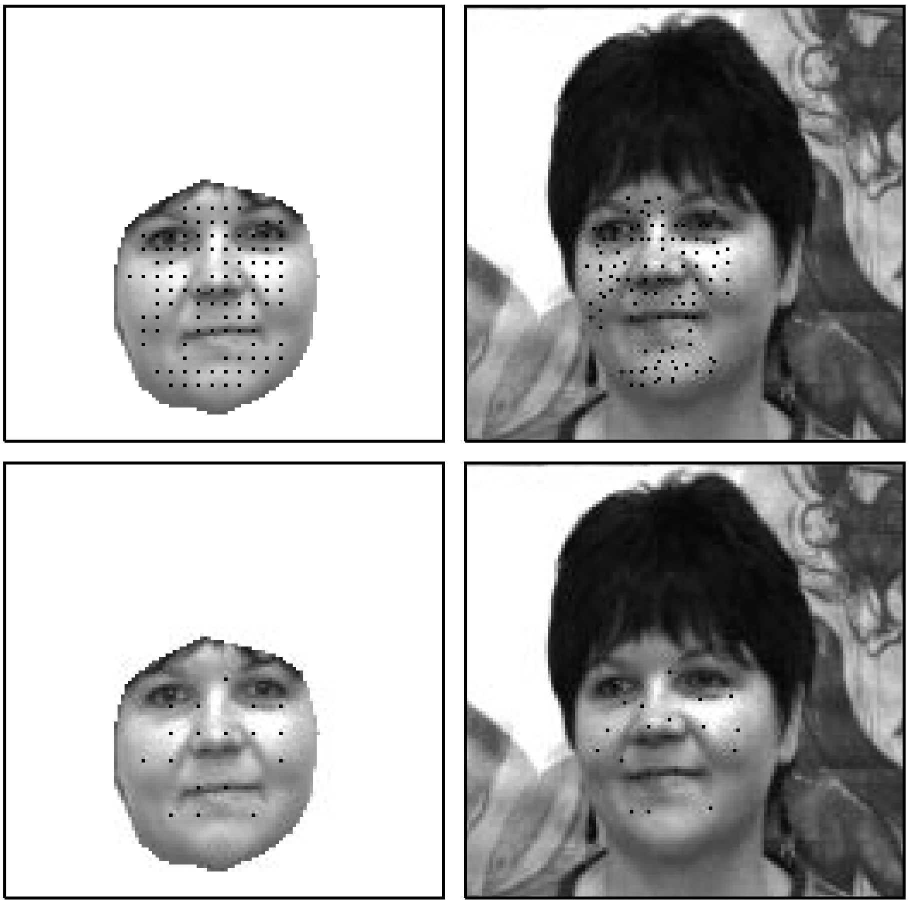

Figure 3: Course-to-fine matching. The left column shows the model representation on the lowest and next higher frequency levels, the right column the matching points in the image. Due to phase matching correspondences are fairly accurate on the lowest level already.

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. See figure 3 for an illustration. 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] and [5].

3.3. Point distribution models

In addition to these advantages, the pyramidal system yields fairly dense and very accurate correspondence maps. The accuracy is mainly due to the phase adjustment. These can be used for tracking facial points. In this case, the coarse to fine matching becomes unpractical, and it has been replaced by a learned shape model in [4]. This sort of tracking can eventually lead to automatic measurement of emotions.

4. Recognition

Once a mapping is established the feature similarities of corresponding point pairs are added up (or averaged) over the whole mapping. This yields a similarity value for each stored model, and the highest similarity belongs to the recognized person. A measure for the reliability of the recognition is derived by a simple statistical analysis of the series of all similarity values.

5. Neural models

All the systems discussed here have been inspired by a biologically plausible framework, C. von der Malsburg's Dynamic Link Architecture. The basic idea is as follows. Neurons that have physical connections (with a long-term strength) have the possibility may also have a dynamic link between them, i.e. a connection which is modifiable on a very short timescale by a combination of Hebbian learning and competition.

This can be used to solve the correspondence problem as follows. Two layers of neurons that represent the image space in model and image, respectively, are fully interconnected by dynamic links. They have an internal wiring that supports moving localized blobs of activity. The development of links is supported by feature similarity and synchronous activation of the connected neurons. The link dynamics then converge to a correspondence mapping. In [6] this has been extended by a competition between a multitude of model layers to a full-blown neural face recognition system. In [7, 9] this system is sped up by a coarse to fine strategy working on the Gabor pyramid. The speedup (in principle) is due to the fact that all refinement steps can be done in parallel. That system also shows background invariance.

6. Links to human perception

Humans do very poorly in recognizing faces from photographic negatives. The system from [3] has no problem with this, because Gabor amplitudes are identical for positive and negative images. The recognition rate of the pyramidal system [7, 8] drops to zero when negatives are presented.

The study [1] has shown that the similarity values found by the system from [5] are correlated with human judgements of similarity both in faces with and without hair. The authors conclude that this system captures essential *face* features better than eigenface-approaches, which seem to be superior in capturing essential properties of a single *image*. Comparisons with the system from [7, 8] are currently in progress.

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, Wolfgang Konen, Shaogang Gong, and Stephen McKenna for excellent cooperation.

7. References

[1] Peter J.B. Hancock, Vicky Bruce, and A. Mike Burton. A comparison of two computer-based face identification systems with human perception of faces. *Vision Research*, 1997. Submitted.

[2] W. Konen and E. Schulze-Krüger. ZN-Face: A system for access control using automated face recognition. In Martin Bichsel, editor, *Proceedings of IWAFGR95 (International Workshop on Automatic Face- and Gesture-Recognition), Univ. of Zürich, June 1995*, pages 18–23, 1995.

[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] Stephen J. McKenna, Shaogang Gong, Rolf P. Würtz, Jonathan Tanner, and Daniel Banin. Tracking facial feature points with Gabor wavelets and shape models. In *Proceedings of the First International Conference on Audio- and Video-based Biometric Person Authentication Crans-Montana, Switzerland, 12-14 March 1997*, 1997. In press.

[5] Laurenz Wiskott, Jean-Marc Fellous, Norbert Krüger, and Christoph von der Malsburg. Face recognition and gender determination. In Martin Bichsel, editor, *Proceedings of IWAFGR95 (International Workshop on Automatic Face- and Gesture-Recognition), Univ. of Zürich, June 1995*, pages 92–97, 1995.

[6] Laurenz Wiskott and Christoph von der Malsburg. Face recognition by dynamic link matching. In Joseph Sirosh, Risto Miikkulainen, and Yoonsuck Choe, editors, *Lateral Interactions in the Cortex: Structure and Function*. The UTCS Neural Networks Research Group, Austin, TX, Electronic book, ISBN 0-9647060-0-8, http://www.cs.utexas.edu/users/nn/web-pubs/htmlbook96, 1996.

[7] 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.

[8] Rolf P. Würtz. Object recognition robust under translations, deformations and changes in background. *IEEE Transactions on Pattern Recognition and Machine Intelligence*, 1995. Submitted.

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