Organic Computing for face and object recognition

Rolf P. Würtz

Institut für Neuroinformatik, Ruhr-Universität Bochum

Abstract: In this paper I describe how the various subsystems for a vision system capable of recognizing faces and objects have been developed using Organic Computing methodology, namely the reliance on "learning from nature" and "self-organization". Then, extensions and refinements currently under development are discussed.

1   Introduction

Automatization has a huge demand for interpreting the data that can be acquired from sensors, and the goal of human-friendly interfaces also requires that the machine understand something of its environment. This is difficult, because most environments are much too complicated to allow for good computational models. The most informative sensors are cameras, and the most important task for understanding their output is the identification of known objects. In [vdM04] the need for Organic Computing to solve the vision problem is discussed in general.

On an abstract level, a recognition system requires a *data format* for stored objects, a method to *compare given images* with those objects, and a method to turn the resulting similarities into a *decision* about the object's identity. For efficient application to real world data, the *efficient organization* of the object database and methods to *presegment* images and image sequences become important. Our choices for subsystems to perform these tasks are guided by the Organic Computing principles of "learning from nature" and "self-organization".

Meaningful objects occupy an extended part of the visual field. The data format of the sensor is that of unrelated pixels or photoreceptor activities, which must be organized into a coherent whole. In the visual system this is done by a highly recurrent network, whose global properties are still poorly understood. In our endeavor to build a technical system for face recognition we have followed the concept of hierarchical self-organization from the image pixels up to a decision about the identity of the observed person.

Image comparison requires a process which can *register* the images such that only points stemming describing the same physical point compared by their visual appearance. Finding these point pairs in a given and a stored image, respectively, is known as the *correspondence problem*. It is at the heart of many computer vision problems, and general solutions have not yet been found. We have proposed self-organizing dynamics on various levels of abstraction to solve it well enough for recognition purposes.
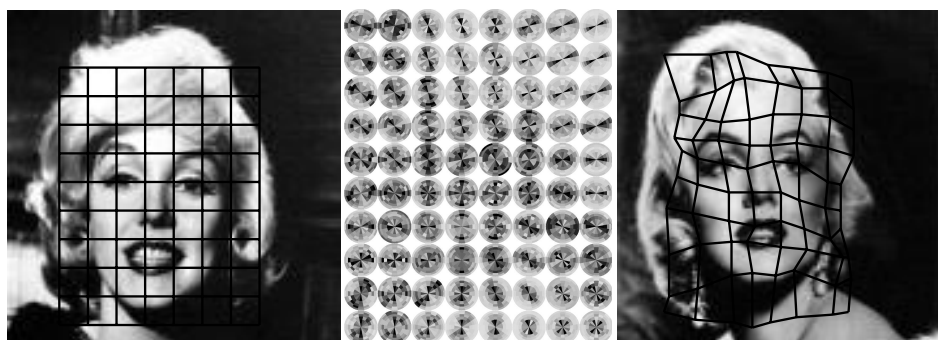
Figure 1: A basic form of elastic graph matching for face recognition. A stored face is represented by a graph (left) whose nodes are labeled with jets (center), which are organized sets of responses of feature detectors. To solve the correspondence problem a self-organized matching scheme finds the most similar graph in an input image. This allows for a comparison of the two faces, which is independent of the position and robust against other influences like 3D-movement and occlusion.

2 Jets and model graphs

For choosing the data format, existing knowledge from neurobiology has been put to use. The first step of processing in the visual cortex is filtering by a set of cells. Each of them is specialized to one point in the input field, an orientation and a scale. It combines the pixels from an area surrounding this point by means of a weighted sum, which can be described by Gabor functions.

The next integration is done by combining all the cells specialized to the same point but to different scales and orientations. The resulting vector is called a *jet* and describes an image patch surrounding the given point. The description becomes less accurate with the distance to the point. These jets are further organized into spatial arrangements, which are formally coded as labeled graphs.

3 Self-organized correspondence finding

Given the data structure of an elastic graph the correspondence problem can be solved by the self-organization of a neural net with rapidly modifiable synaptic connections [vdM88, LVB⁺93]. The basic mechanism is a dynamic system which starts from a state in which all point correspondences have equal likelihood coded into a synaptic strength. The dynamics support the growth of such links that connect similar features (jets) and of link combinations that preserve the rough neighborhood relationships (graph edges) between points in image and stored graph. Like many self-organizing processes, this dynamics is time consuming. In recent work [Zh04] the situation could be improved by introducing special map-coding neurons.

For technical purposes this dynamical system has been further abstracted into a hierarchi-

cal optimization in various ways [LVB$^+$93, Wü97]. The detailed neuronal dynamics have been replaced by a potential function, which has the desired correspondence structure as its optimum. For finding this optimum a hierarchical optimization scheme is used which decomposes the search space into the geometrically most probable degrees of freedom and optimizes them hierarchically. When good correspondences are established, the similarities between the local features are added up to a robust similarity measure between the images and finally lead to a recognition decision. After careful optimization on the algorithmic and software levels, correspondence finding remains the most time consuming part of face and object recognition.

4 Bunch graphs

In the case of face recognition the situation can be greatly improved by calculating the correspondences between all stored faces offline and storing many faces (called candidates) in one data structure called a *bunch graph* [WFKvdM97] with the same graph topology as a model graph but with the corresponding jets of all candidates as vertex label.

Bunch graph matching can proceed in two modes. In *recognition mode* it works as a set of model graphs and similarities are evaluated for each person in the graph. However, the time consuming matching is required only for one graph (with more complicated features), instead of one graph per person. In *finding mode* the local jets of different candidates are compared for maximal similarity independently. This works in situations where the person in the image is not part of the bunch graph. This mode has also been called *general face knowledge*, because of its potential to describe all possible faces by combinatorics, once enough faces are part of the bunch graph. The algorithm has a strong self-explanatory component in the sense that the information of which facial parts resemble which of the candidates stored in the bunch graph is readily available.

This property of bunch graph matching has been further exploited by attaching personal properties to the candidates. Simple examples include "gender", "beardedness" and "wearing spectacles". These can be attached to the candidates in a supervised manner and are inherited by all jets of this candidate. Applied to an unknown face, the locally best fitting jets can make a majority decision (jet voting) about the global property of the face [Wi97]. This decision is purely learned from examples, rules like the (obvious) constraint that eye jets have nothing to say about "beardedness" need not be specified, but are learned implicitly. This implicit learning has been applied successfully in cases where the rules are far from obvious, e.g., the classification of rare genetic diseases, which influence the facial appearance [LWW$^+$03]. Given the choice between five such diseases performance was close to that of human experts.

For many applications like video phones or facial gesture recognition it is important that facial points be tracked reliably in a video sequence. This is only possible if tracking is constrained by model knowledge about the object to be tracked. In [WvdMW03] these constraints could be learned from the displacement fields encountered during bunch graph matching to a large dataset of persons.

For still images taken under controlled conditions the above described system has performed very well in the FERET and Face Recognition Vendor tests [PMRR00, PGM$^+$03]. Especially the good performance in difficult situations as compared to the mathematically inspired eigenface method is a strong example for the success of Organic Computing methodology.

5 Segmentation

Although powerful, the self-organized matching scheme alone cannot cope with arbitrarily cluttered images. Therefore, before recognition of a person can proceed, areas with a high probability of containing a face must be extracted. This is done by dynamic combination of cues like motion, predicted position, rough facial shape, skin color, stereo depth, etc by a process that self-organizes the actual relative weighting of the different cues, whose reliability may vary strongly in time [vdM04]. The whole system is capable of recognizing persons from a database in real time while walking towards a camera [SEN98, Lo00].

6 Ongoing projects

The most pressing problems for the extension of the bunch graph technique for face recognition concern illumination changes, which must be learned from examples given the very complicated reflection properties of human skin. Bunch graphs must be empowered to acquire new knowledge in a self-organized way, and the need for human interaction and correction during their creation (still required because of errors in correspondence finding) must be further reduced. In the long run, a bunch graph should become an active entity, able to decide when information needs to be added or reorganized.

In the more general field of object recognition the unaltered bunch graph concept is not successful because of the very different geometrical structure of objects. Correspondence finding between different views of the same object works fairly well, and a very efficient neural network for object learning [WW04] has also been implemented. A convincing integration of correspondence-based comparison and fast retrieval is currently pursued.

The organization of Gabor functions into jets is only one possible way of useful feature combinations. Another neurobiology-based one is the combination to cells which are sensitive to terminating visual contours, so-called endstopped cells. Combining those over a range of scales yields robust corner detectors [WL00]. Combination of neighboring Gabor responses yields an alternative feature combination, which is better suited for background suppression [Wü97]. The shown examples are only a small subset of feasible feature combinations. Selection of more complicated ones can hardly be driven by intuition alone but must be guided by self-organization from natural image and video data.

References

[Lo00]   Loos, H. S.: Suchbilder – Computer erkennt Personen in Echtzeit. *c't*. (15):128–131. 2000.

[LVB⁺93]   Lades, M., Vorbrüggen, J. C., Buhmann, J., Lange, J., von der Malsburg, C., Würtz, R. P., and Konen, W.: Distortion invariant object recognition in the dynamic link architecture. *IEEE Transactions on Computers*. 42(3):300–311. 1993.

[LWW⁺03]   Loos, H. S., Wieczorek, D., Würtz, R. P., von der Malsburg, C., and Horsthemke, B.: Computer-based recognition of dysmorphic faces. *European Journal of Human Genetics*. 11:555–560. 2003.

[PGM⁺03]   Phillips, P., Grother, P., Micheals, R., Blackburn, D., Tabassi, E., and Bone, J.: FRVT 2002: Overview and summary. Technical report. 2003. http://www.frvt.org/DLs/FRVT_2002_Overview_and_Summary.pdf.

[PMRR00]   Philips, P. J., Moon, H., Rizvi, S. A., and Rauss, P. J.: The FERET evaluation methodology for face-recognition algorithms. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 22(10):1090–1104. 2000.

[SEN98]   Steffens, J., Elagin, E., and Neven, H.: PersonSpotter – fast and robust system for human detection, tracking, and recognition. In: *Proc. 3rd Intl. Conf. on Face and Gesture Recognition, Nara, Japan, April 1998*, pages 516–521. 1998.

[vdM88]   von der Malsburg, C.: Pattern recognition by labeled graph matching. *Neural Networks*. 1:141–148. 1988.

[vdM04]   von der Malsburg, C.: Vision as an exercise in Organic Computing. In: *This workshop*. 2004.

[WFKvdM97]   Wiskott, L., Fellous, J.-M., Krüger, N., and von der Malsburg, C.: Face recognition by elastic bunch graph matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 19(7):775–779. 1997.

[Wi97]   Wiskott, L.: Phantom faces for face analysis. *Pattern Recognition*. 30(6):837–846. 1997.

[WL00]   Würtz, R. P. and Lourens, T.: Corner detection in color images through a multiscale combination of end-stopped cortical cells. *Image and Vision Computing*. 18(6-7):531–541. 2000.

[Wü97]   Würtz, R. P.: Object recognition robust under translations, deformations and changes in background. *IEEE Transactions on Pattern Analysis and Machine Intelligence*. 19(7):769–775. 1997.

[WvdMW03]   Wieghardt, J., von der Malsburg, C., and Würtz, R. P.: Automatic learning of constraints for tracking facial feature points. *Computer Vision and Image Understanding*. 2003. Submitted.

[WW04]   Westphal, G. and Würtz, R. P.: Fast object and pose recognition through minimum entropy coding. In: *Proceedings of ICPR 2004, Cambridge*. 2004. In press.

[Zh04]   Zhu, J.: *A dynamic method to reduce the search space for visual correspondence problems*. PhD thesis. University of Southern California. May 2004.