Organic Computing Methods for Face Recognition

Methoden des Organic Computing zur Gesichtserkennung

Rolf P. Würtz, Ruhr-Universität Bochum

Summary Automatic face recognition is a slowly maturing and economically highly important technology, which still falls short of the high expectations set on it. The variation in images taken from the same person makes face recognition systems difficult to design – it is impossible to explicitly code all variability. Successful systems have relied heavily on the principle of self-organizing many fragile cues to arrive at a robust decision and have been built by learning from biological systems. The paper describes the techniques of elastic bunch graph matching as a hierarchical integration of image pixels into Gabor responses, jets, graphs, and bunch graphs. Beside face recognition, these concepts are used for face classification learned from examples. It is attempted to develop them further to reach a complete parameterization of all faces. Methods for more general object recognition using the same Organic Computing principles are outlined and include the concept of end-stopped cells as corner detectors.

►►► **Zusammenfassung** Automatische Gesichtserkennung ist eine Technik von immensem wirtschaftlichen Interesse, die aber die hohen Erwartungen noch nicht erfüllen kann. Die Schwierigkeit besteht darin, dass die Variation der Bilder ein und derselben Person nicht explizit kodiert werden kann. Erfolgreiche Systeme sind auf dem Prinzip der Selbstorganisation vieler unzuverlässiger Hinweise zu einer robusten Entscheidung aufgebaut und sind von biologischen Systemen inspiriert. Die Arbeit beschreibt die Technik des Bündelgraphenmatching als hierarchische Integration von Pixeln zu Filterantworten, Jets, Graphen und Bündelgraphen. Außer der Personenerkennung werden diese Konzepte auch zur Klassifizierung von Gesichtseigenschaften verwendet, die vollkommen aus Beispielen gelernt ist. Sie sollen zu einer kompletten Parameterisierung aller Gesichter weiterentwickelt werden. Methoden zur allgemeinen Objekterkennung, die auf denselben Prinzipien des Organic Computing beruhen, werden angedeutet, z. B. das Konzept der End-stopped cells zur Eckendetektion.

KEYWORDS I.4.7 [Feature Measurement] I.4.10 [Image Representation] I.5.1 [Models] I.5.4 [Applications – Computer Vision]

1 Introduction

Automation has a huge demand for interpreting the data acquired from sensors. This is problematic because the environment is generally much too complicated to allow for a good computational model. The most informative sensors are cameras and the most important task for understanding their output is the identification of known objects. Computer vision is expected to gain substantially from Organic Computing, because the complicated models must self-organize to a large degree.

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. The subsystems to perform these tasks can be designed by using the Organic Computing principles of "Learning from nature" and "Self-organization".

Meaningful objects occupy an extended part of the visual field, but the data format in a camera (or on the first layer of the retina) is that of unrelated pixels which must be actively organized into a coherent whole if they belong to one object. In the visual system of the brain this is done by a complicated and 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.

Comparing given images requires a process which can *register* images in the sense that only image points which belong to the same physical point on the object are compared. Finding these points in a given image is known as the *correspondence problem*, which is at the heart of many computer vision problems, but general solutions have not yet been found. We have used self-organizing dynamics on various levels of abstraction to solve it well enough for recognition purposes.

2 Jets and model graphs

The data format for the first integration step has been gleaned from neurobiology. The first step of processing in the visual cortex of the brain can be regarded as filtering by a huge set of neurons, each specialized to a local part of the image (its receptive field), a preferred orientation and a spatial frequency (or scale). The mathematical modeling of these cells' responses is done by complex-valued Gabor functions [2]. As an interesting aside, these response profiles are themselves a result of self-organization. When neurons are shown many real images and given the constraint that their activity should be sparse, they develop precisely the properties of simple cells [9]. Therefore, they are well adapted to the task of processing natural images.

The next integration step is done by combining the responses of all the cells specialized to the same location but at different scales and orientations. The resulting vector is called a *jet* and describes the image patch in the area. For a global object description, 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 a labeled graph the correspondence problem can be solved by the self-organization of a neural net with rapidly modifiable synaptic connections [4; 14]. The architecture of the net consists of an input layer which contains the transformed image data and as many model layers as there are faces to be recognized. The layers are interconnected recurrently such that each pair of image and model location has a link with a strength that can be interpreted as a likelihood for being corresponding. This is initialized by the similarities between the jets. Then rapid learning dynamics start, which support the growth of such links that connect similar jets and of link combinations that preserve the rough neighborhood relationships (graph edges) between points in image and model graph. Additionally, there is strong competition between all links originating at a single cell and all links targeting a single cell. This supports the development of one-one connections out of a state with undetermined correspondences. Furthermore, competition between models enforces convergence to the recognition of only one model [19].

The simulation of this self-organizing process on a sequential computer is rather time consuming – in the range of minutes on a standard PC. Recently, a self-organizing network based on minicolumn dynamics has been formulated, which has much higher intrinsic parallelism by employing more cells [8]. An alternative approach is to in-

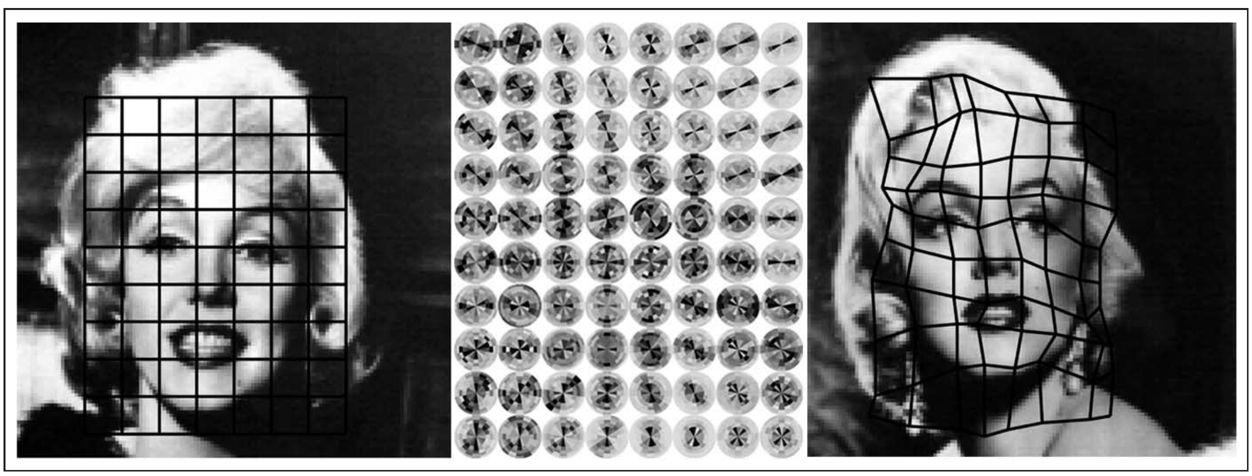

Figure 1 A basic form of elastic graph matching for face recognition. A stored face is represented by a graph (left), whose nodes are labeled with jets (center), which are organized sets of responses of feature detectors. Each disk represents a jet attached to the respective node on the graph. The disk segments stand for cells selective for orientation (arranged angularly) and spatial frequency (arranged radially). The gray shades inside the segments show the activation of the cells. To solve the correspondence problem a self-organized matching scheme finds the most similar graph in an input image, the comparison of those graphs is independent of the position and robust against other influences like 3D-movement and occlusion.

troduce special map-coding neurons (maplets) [22].

For technical purposes the system described above has been further abstracted into a hierarchical optimization in various ways [4; 20]. The detailed neuronal dynamics have been replaced by a potential function, which has the desired correspondence structure as its optimum, together with a hierarchical optimization scheme, which decomposes the search space into the geometrically most probable degrees of freedom. Once good correspondences are established, similarities between local features are added up to a robust similarity measure between the images and finally lead to a recognition decision. Even 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 special case of face recognition the situation can be greatly improved by storing many candidate faces together with the correspondences between them in one data structure called a *bunch graph* [18]. It has the same graph topology as a model graph, and the nodes are labeled with jets of corresponding points in all faces. There is one node for the right eyes of all faces, one for all nose tips, etc. This is the fourth integration step in the hierarchy started at the pixels.

Bunch graph matching can basically proceed in two modes. In *recognition mode* it simply works as a set of model graphs and similarities are evaluated for each person in the graph. Point correspondence is a transitive relation, and therefore the internally stored correspondence structure allows to restrict the time consuming matching to one global graph. The part with linear complexity in the number of candidates is thus reduced to the very rapid evaluation of jet similarities between readily matched nodes.

In the more interesting *finding mode* the local jets of different candidates are compared for maximal similarity independently of each other and thus allow application to situations where the person in the image is not part of the bunch graph. This mode of bunch graph matching has also been called *general face knowledge*, because of its potential to describe all possible faces as combinations of known patches, once enough faces are part of the bunch graph. It has turned out that about 100 faces are sufficient to code for all possible faces. 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. Current implementations on standard PCs (3 GHz dual Xeon processor, 2 GB RAM) can recognize a person out of a database of 1000 in about 6 seconds.

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”. Attached to all candidates in a supervised manner they are inherited by all their respective jets. Applied to an unknown face, the locally best fitting jets can make a majority decision (jet voting) about the global property of the face [17]. This decision is purely learned from examples. Rules like the constraint that eye jets are irrelevant to “beardedness” need not be specified. As a more complicated example, the method has been applied to the classification of rare genetic diseases which influence the facial appearance [6]. 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 are tracked reliably in a video sequence. This is only possible if tracking is constrained by model knowledge about the object to be tracked. In [16] these constraints could be learned from the displacement fields encountered during bunch graph matching to a large dataset of persons.

For images taken under controlled conditions the above described system has performed very well in the FERET and Face Recognition Vendor tests (see http://www.frvt.org/), demonstrating that Organic Computing methods are competitive with more mathematically inspired ones. A fair comparison with other face recognition methods is far beyond the scope of this article. Alternative approaches include pattern recognition on pixel values, eigenfaces [11; 13], local feature analysis [10], and morphable models [1]. From these, local feature analysis relies strongly on self-organization.

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 high probability of containing a face must be identified. This is done by dynamic combination of cues like motion, predicted position, rough facial shape, skin color, stereo and depth, by a process that self-organizes the actual relative weighting of the different cues, whose reliability may vary strongly in time. The whole system is capable of recognizing persons from a database in real time while walking towards a camera [5]. The self-organized weighting of different cues is called democratic integration [12] and is well suited for robust segmentation of moving objects, with much higher accuracy than a comparable probability-based method [3]. Both Organic Computing methods receive their robustness from self-organization.

6 Future challenges

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 reflectance 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 (required because of erroneous correspondences) will be further reduced. In the long run, a face representation will become an active entity, able to decide on the basis of real-world data and user reinforcement 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 [15] 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 so-called endstopped cells. Integrating those over a range of scales yields robust corner detectors [21]. They are well suited as a basis for object matching [7]. Features combining neighboring Gabor responses are better suited for background suppression than jets [20]. The shown examples are only a small subset of feasible feature combinations. Selection of more complicated ones, which are appropriate for special visual decision tasks, can hardly be driven by intuition alone but must be guided by self-organization from natural image and video data.

Acknowledgements

Partial funding by the European Commission in the RTN MUHCI (HPRN-CT-2000-00111) and by the Deutsche Forschungsgemeinschaft (WU 314/2-1) is gratefully acknowledged.

References

[1] V. Blanz and T. Vetter. Face recognition based on fitting a 3d morphable model. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 25(9):1063–1074, 2003.

[2] J. Jones and L. Palmer. An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex. *J. of Neurophysiology*, 58(6):1233–1258, 1987.

[3] O. Kähler and J. Denzler. Self-organizing, adaptive data fusion for 3d object tracking. In U. Brinkschulte, J. Becker, D. Fey, C. Hochberger, T. Martinetz, C. Müller-Schloer, H. Schmeck, T. Ungerer, and R. Würtz, editors, *ARCS 2005 – System Aspects in Organic and Pervasive Computing – Workshops Proc., Innsbruck Austria, March 14–17*, pages 109–116. VDE Verlag, Berlin, Offenbach, 2005.

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

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

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

[7] T. Lourens and R. P. Würtz. Extraction and matching of symbolic contour graphs. *Int'l J. of Pattern Recognition and Artificial Intelligence*, 17(7):1279–1302, 2003.

[8] J. Lücke. *Information Processing and Learning in Macrocolumn Networks.* PhD thesis, Physics Dept., Univ. of Bochum, Germany, Nov. 2004.

[9] B. A. Olshausen and D. J. Field. Wavelet-like receptive fields emerge from a network that learns sparse codes for natural images. *Nature*, 381:607–609, 1996.

[10] P. S. Penev and J. J. Atick. Local feature analysis: a general statistical theory for object representation. *Network*, 7(3):477–500, 1996.

[11] L. Sirovich and M. Kirby. Low-dimensional procedure for the characterization of human faces. *J. of the Optical Society of America A*, 4:519–524, 1987.

[12] J. Triesch and C. von der Malsburg. Democratic integration: Self-organized integration of adaptive cues. *Neural Computation*, 13(9):2049–2074, 2001.

[13] M. A. Turk and A. P. Pentland. Face recognition using eigenfaces. In *Proc. of CVPR'91*. IEEE Press, 1991.

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

[15] G. Westphal and R. P. Würtz. Fast object and pose recognition through minimum entropy coding. In *17th Int'l Conf. on Pattern Recognition (ICPR 2004), Cambridge*, volume 3, pages III–53–III–56. IEEE Press, 2004.

[16] J. Wieghardt, R. P. Würtz, and C. von der Malsburg. Gabor-based Feature Point Tracking with Automatically Learned Constraints. In R. P. Würtz and M. Lappe (eds.) *Dynamic Perception,* pages 121–126. infix/IOS Press, 2002.

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

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

[19] L. Wiskott and C. von der Malsburg. Face recognition by dynamic link matching. In J. Sirosh, R. Miikkulainen, and Y. 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.

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

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

[22] J. Zhu and C. von der Malsburg. Maplets for correspondence-based object recognition. *Neural Networks*, 17:1311–1326, 2004.

Dr. rer. nat. Rolf P. Würtz obtained his diploma in mathematics from the University of Heidelberg, Germany in 1986. After that, he was research assistant at the Max-Planck-Institute for Brain Research in Frankfurt, Germany. In 1990, he joined the Institute for Neurocomputing at the University of Bochum, Germany, where he received his Ph.D. from the Physics Department in 1994. Until 1997, he was a postdoctoral researcher at the Department of Computing Science at the University of Groningen, The Netherlands. He is currently a scientific staff member at the Institute for Neurocomputing in Bochum. Research interests include neuronal models and efficient algorithms for object recognition, hand-eye coordination, integration of visual and tactile information, and links to higher cognition. General organization problems and solution methods inspired by living systems are studied under the framework of "Organic Computing".
Address: Ruhr-Universität Bochum, Institut für Neuroinformatik, Universitätsstraße 150, 44801 Bochum,
Tel.: +49-234-32-27994,
Fax: +49-234-32-07994,
E-Mail: rolf.wuertz@neuroinformatik.rub.de