Organic Computing for Video Analysis

Rolf P. Würtz
Institute for Neurocomputing
Ruhr-University Bochum, Germany
rolf.wuertz@organic-computing.org

1. Introduction

Information technology (IT) is plagued with the rapidly increasing complexity of systems to be deployed. In the world of living beings on the other hand, it can be observed that extremely complex systems function in a robust, fault-tolerant, flexible, adaptive, self-organizing way, and apparently goal-directed way [21].

It is therefore intriguing to identify strategies by means of which these properties are achieved by living systems. This "Learning from Nature" is the founding idea of *Organic Computing*. Earlier IT applications include *Neural Networks* and *Evolutionary Computation*. The current interest in Organic Computing is also sustained by a notion of "Organic" which relates to the user rather than the developer. In that aspect, Organic Computing requires that the interface between the IT system and the user be organic, intelligible, and friendly. This again imposes constraints on the user interfaces, which can only be partly fulfilled by current technology.

1.1. Self-organization

The introduction is followed by some facts and theories about self-organizing systems including a short description of current research projects and open issues. One project develops flexible control of traffic for the city of Hanover, which tries to optimizes the overall flow without relying on centralized controllers [16]. Within the automobile, the exploding number of components and interactions and the combinatorics of possible models has prompted the development of an evolutionary architecture which self-organizes according to a goal description and reorganizes in the presence of partial failure [6]. In an ongoing project, the organization of varying office users and people looking for them within the building is handed over to an organic system [18].

1.2. Organic computing for computer vision

The application domain I will present in more detail is computer vision [20] and user interaction [7]. Here, the complexity is dictated by the difficulties posed by noisy and highly variable camera data.

About a third of the human neocortex is dedicated to visual processing. This is an estimated 10^9 neurons with some 10^{12} interation synapses, each of which is to be represented by a floating point number, according to the modelling approach used in Artificial Neural Networks.

On an abstract level, a recognition systems 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 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.

0-7695-2946-1/07 $25.00 © 2007 IEEE
DOI 10.1109/HIS.2007.29

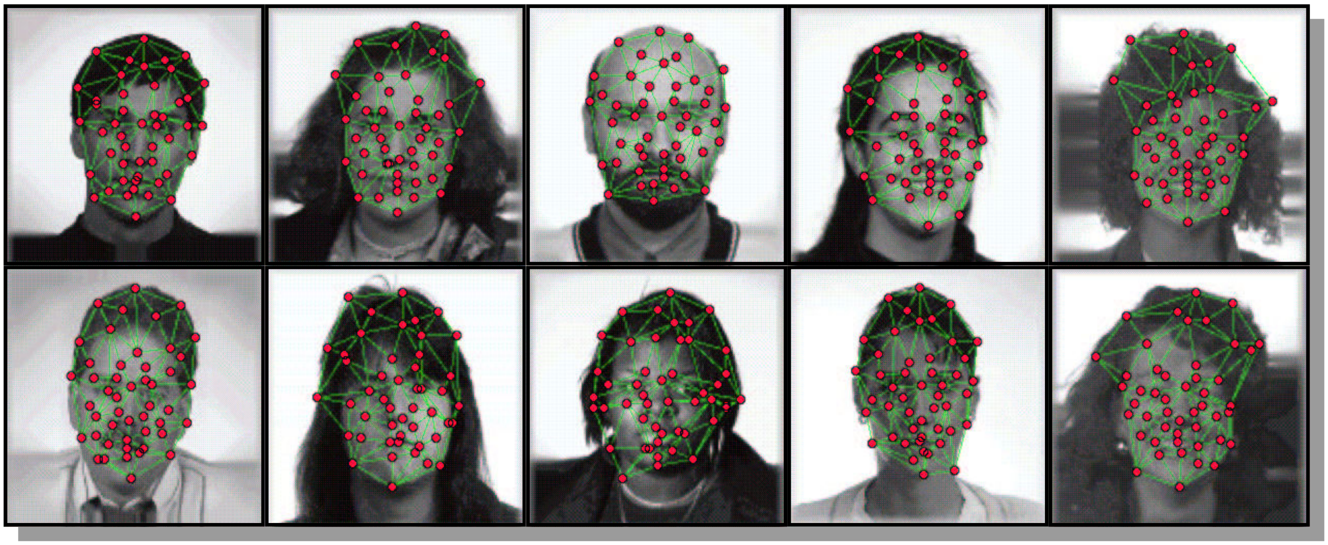

Figure 1. Bunchgraph Matching [27]. This figure shows a variety of images with the graphs corresponding to a fixed bunch graph overlayed. It can be seen that the positions of many landmarks are estimated well. These graphs can also be used to extend the original bunch graph.

2. Face recognition

First a system for automatic face recognition is described which has been constructed according to neurobiological findings and a theory of self-organizing neural networks.

2.1. 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. 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 [12]. 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.

2.2. 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 [19, 9]. 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 [28].

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 [11]. An alternative approach is to introduce special map-coding neurons (maplets) [33].

For technical purposes the system described above has

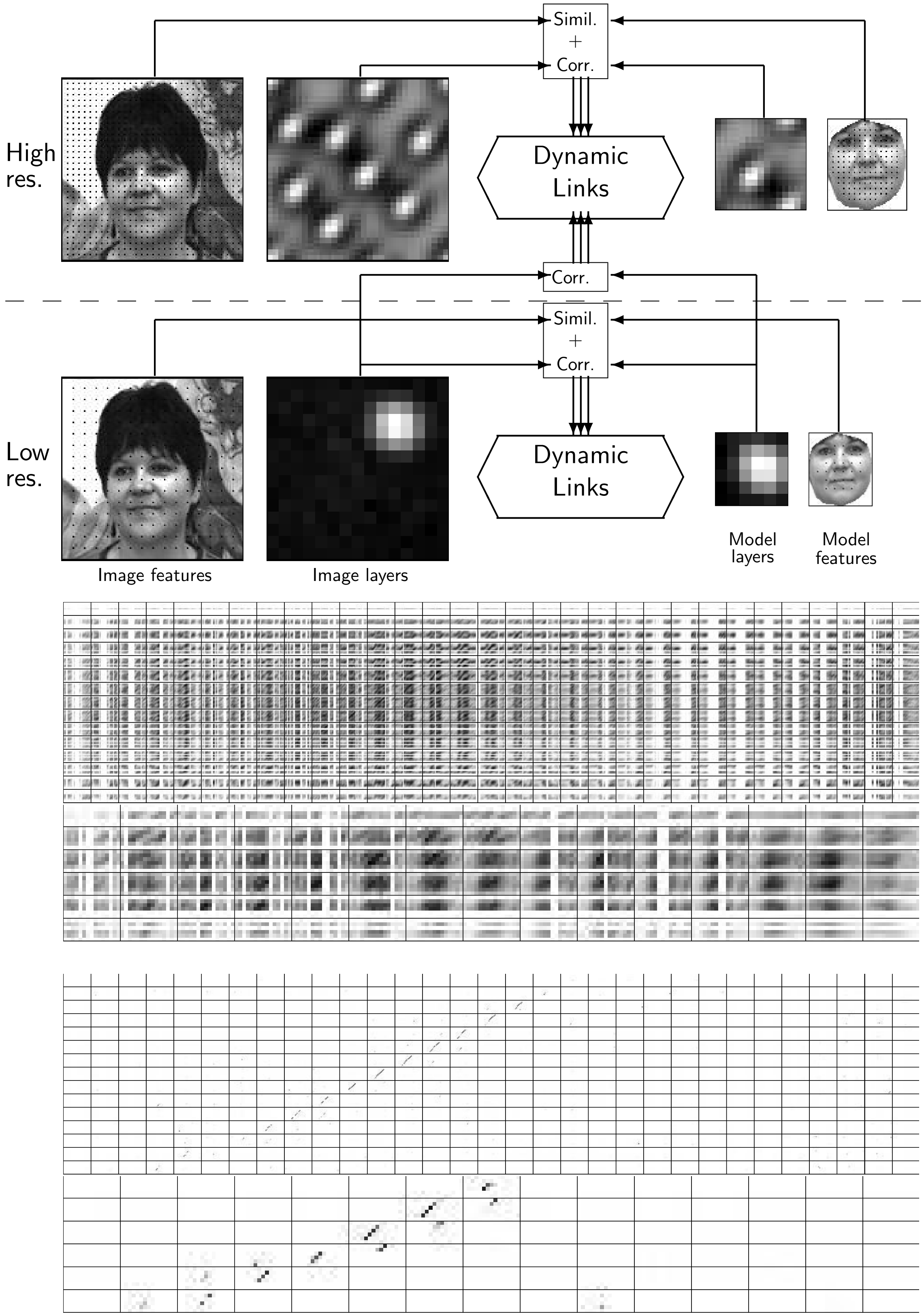

Figure 2. Neuronal Matching [29, 31]. The scheme on top shows the structure of the dynamic neuronal network for multiresolution correspondence finding. In the middle, dynamic links are initialized to feature similarities, which are highly ambiguous, illustrating the correspondence problem. Below, the links have converged to a one-one mapping with the appropriate correspondences.

been further abstracted into a hierarchical optimization in various ways [9, 30]. 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.

2.3. 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* [27]. 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 (3GHz dual Xeon processor, 2GB RAM) can recognize a person out of a database of 1000 in about 6 seconds.

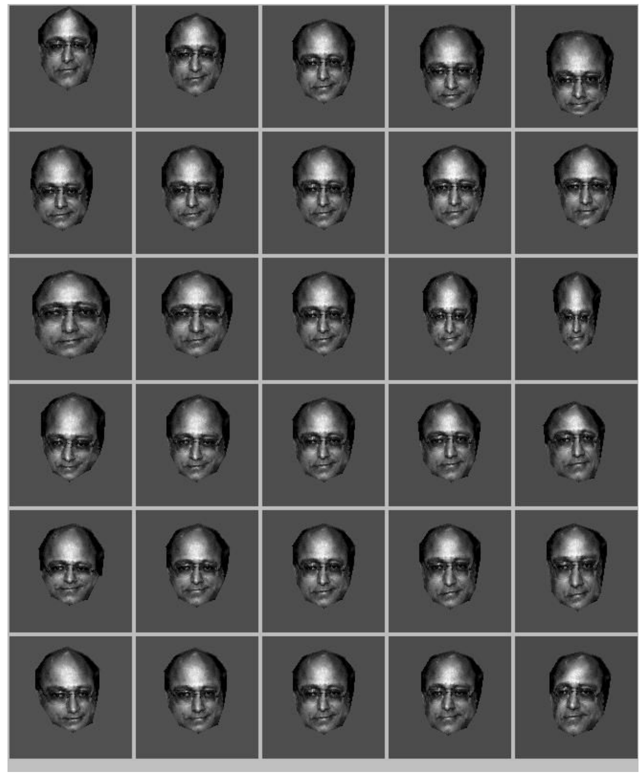

Figure 3. Textured principal components of correspondence fields: The first six PCs principal (top to bottom) of the feature point locations are illustrated here in terms of the mapping they perform on the standard gray value image shown in the central column. Each row shows the deformation from the mean along one principal component by -4,-2,0,2 and 4 standard deviations, respectively.

2.4. Learning facial attributes

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 [26]. 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 syndromes which influence the facial appearance [10]. Given the choice between five such syndromes performance was close to that of human experts, and acceptable classification rates could be achieved recently on 14 syndromes [3].

2.5. Face tracking

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 can only be done robustly when constrained by model knowledge about the object to be tracked. In [24, 25] these constraints could be learned from the displacement fields encountered during bunch graph matching to a large dataset of persons looking relatively straight into the camera (see figure 1). Interestingly, the principal components of the correspondence fields over all images captured the local 3D geometry of the faces (figure 3)

For still images taken under controlled conditions the above described system has performed very well in the FERET and Face Recognition Vendor tests [14, 15], 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, but almost all winning commercial systems in [15] are based to some extent on graph matching with Gabor wavelets.

It is also an example for the hierarchical self-organization of elementary feature detectors into structures of higher and higher complexity. Detailed self-organizing neuronal dynamics are presented as well as the techniques of pyramid matching [30] and Elastic Graph Matching [9], the latter being more efficient on digital computers. The basic matching mechanism is extended to the *bunch graph* data format and recognition procedure, which has made this technology one of the leading methods for facial identification [27].

3. Analysis of human motion

The extension from faces to body gestures is more complicated and partly subject of ongoing research. I present the method of "democratic integration"[17], which allows for flexible integration of many fragile cues into a robust decision. This is used for user interaction with a robot [1] and for interpretation of user gestures [7, 8]. This will be extended to a system that can learn constraints for body tracking analogously to the face tracking system.

4. Object recognition

Concluding the tutorial I present a system for the self-organization of a recognition memory for everyday objects. Again, the focus is on automatic learning from examples [13, 23, 22]. In that work we proposed a form of graph dynamics which proceeds in three steps. In the first step position-invariant feature detectors, which decide whether a feature is present in an image, are set up from training images. For processing arbitrary objects these features are small regular graphs, termed parquet graphs, whose nodes are attributed with Gabor amplitudes. Through combination of these classifiers into a linear discriminant that conforms to Linsker's infomax principle a weighted majority voting scheme is implemented. This network, termed the preselection network, is well suited to quickly rule out most irrelevant matches and only leaves the ambiguous cases, so-called model candidates, to be processed in a third step using a rudimentary version of elastic graph matching, a standard correspondence-based technique for face and object recognition. To further differentiate between model candidates with similar features it is asserted that the features be in similar spatial arrangement for the model to be selected. Model graphs are constructed dynamically by assembling model features into larger graphs according to their spatial arrangement. The model candidate whose model graph attains the best similarity to the input image is chosen as the recognized model.

5. Closing remarks

As this paper could give only a very brief overview of the topics covered in the tutorial the reader is referred to the website http://www.neuroinformatik.ruv.de/VDM/PUBLIST/newlist/newlist.html, from which the full papers by the Institut für Neuroinformatik at Ruhr-University Bochum can be obtained. For a broader overview of Organic Computing projects and ideas the upcoming book [32] is recommended.

Acknowledgements

Funding from the Deutsche Forschungsgemeinschaft (WU 314/2-2, MA 697/5-1, WU 314/5-2) is gratefully acknowledged. I was supported and encouraged by the IEEE task force on Organic Computing to give this tutorial.

References

[1] M. Becker, E. Kefalea, E. Maël, C. von der Malsburg, M. Pagel, J. Triesch, J. C. Vorbrüggen, R. P. Würtz, and S. Zadel. GripSee: A gesture-controlled robot for object perception and manipulation. *Autonomous Robots*, 6(2):203–221, 1999.

[2] K. Bellman, P. Hofmann, C. Müller-Schloer, H. Schmeck, and R. Würtz, editors. *Organic Computing – Controlled Emergence*, number 06031 in Dagstuhl Seminar Proceedings. Internationales Begegnungs- und Forschungszentrum (IBFI), Schloss Dagstuhl, Germany, 2006.

[3] S. Böhringer, T. Vollmar, C. Tasse, R. Würtz, G. Gillessen-Kaesbach, B. Horsthemke, and D. Wieczorek. Syndrome identification based on 2D analysis software. *European Journal of Human Genetics*, 14(10):1082–1089, 2006. Advance online publication, 14 June 2006 (DOI 10.1038/sj.ejhg.5201673).

[4] 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 Procedeedings, Innsbruck Austria, March 14–17*. VDE Verlag, Berlin, Offenbach, 2005.

[5] P. Dadam and M. Reichert, editors. *Informatik 2004*, volume 2. Gesellschaft für Informatik, 2004.

[6] P. E. H. Hofmann and S. F. Leboch. Evolutionäre Elektronikarchitektur für Kraftfahrzeuge. *it - Information Technology*, 47(4):212–219, 2005.

[7] M. Krüger, A. Schäfer, A. Tewes, and R. P. Würtz. Communicating agents architecture with applications in multimodal human computer interaction. In Dadam and Reichert [5], pages 641–645.

[8] M. Krüger, C. von der Malsburg, and R. P. Würtz. Self-organized evaluation of dynamic hand gestures for sign language recognition. In Würtz [32]. Submitted.

[9] 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 Transactions on Computers*, 42(3):300–311, 1993.

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

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

[12] 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.

[13] G. Peters, T. Leopold, C.-P. Alberts, M. Briese, S. Entian, C. Gabriel, Z. Gao, A. Klandt, J. Schultze, J. Spiegel, J. Thyen, M. Vaupel, P. Voß, and Q. Zhu. Adaptive object acquisition. In Brinkschulte et al. [4], pages 117–122.

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

[15] P. Phillips, W. T. Scruggs, A. J. O'Toole, P. J. Flynn, K. W. Bowyer, C. L. Schott, and M. Sharpe. FRVT 2006 and ICE 2006 large-scale results. Technical Report NISTIR 7408, http://www.frvt.org/, 2007. http://www.frvt.org/FRVT2006/docs/FRVT2006andICE2006LargeScaleReport.pdf.

[16] F. Rochner and C. Müller-Schloer. Adaptive decentralized and collaborative control of traffic lights. In Dadam and Reichert [5], pages 595–599.

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

[18] W. Trumler, F. Bagci, J. Petzold, and T. Ungerer. Towards an organic middleware for the smart doorplate project. In Dadam and Reichert [5], pages 626–630.

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

[20] C. von der Malsburg. Vision as an exercise in Organic Computing. In Dadam and Reichert [5], pages 631–635.

[21] C. von der Malsburg. The organic future of information technology. In Würtz [32]. Submitted.

[22] G. Westphal, C. von der Malsburg, and R. P. Würtz. Feature-driven emergence of model graphs for object recognition and categorization. In Bellman et al. [2]. Online publication http://drops.dagstuhl.de/opus/volltexte/2006/575/pdf/06031.WuertzRolf.Paper.575.pdf.

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

[24] J. Wieghardt. *Learning the Topology of Views: From Images to Objects*. Shaker Verlag, Aachen, 2001.

[25] 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, editors, *Dynamic Perception*, pages 121–126. infix Verlag/IOS press, 2002.

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

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

[28] 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.

[29] R. P. Würtz. *Multilayer Dynamic Link Networks for Establishing Image Point Correspondences and Visual Object Recognition*. Verlag Harri Deutsch, Thun, Frankfurt am Main, 1995.

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

[31] R. P. Würtz. Biologically inspired methods for model matching and object recognition. In Y. Sawada, editor, *Proceedings of the 2nd R.I.E.C. International Symposium on Design and Architecture of Information Processing Systems Based on the Brain Information Principles, in Sendai, Japan, March 16–18, 1998*, pages 101–106, 1998. Invited talk.

[32] R. P. Würtz, editor. *Organic Computing*. Springer, 2007. In preparation.

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