International Journal of Computer Vision 72(1), 5–7, 2007

DOI: 10.1007/s11263-006-8889-2

Editorial: ECOVISION: Challenges in Early-Cognitive Vision

First online version published in June, 2006

This special issue contains selected work from the early cognitive vision workshop which took place from May 29 until June 1 2004 on the Isle of Skye in Scotland. The event was organised and financed by the European project ECOVISION (Artificial systems based on early-cognitive cortical processing) (Wörgötter et al., 2004) that ran from January 2002 to December 2004.

The scope of this workshop was biologically motivated vision. A number of leading scientists covering neurophysiology (U. Eysel, G. Orban), neuronal modelling (Ch. v.d. Malsburg, K.A. Boahen) and computer vision (J.O. Eklundh, S. Sarkar, Y. Aloimonos, D. Fleet) participated. Selected papers presented at this workshop are published in this special issue as well as a special issue in 'Network: Computation in Neural Systems' (Goodhill, 2005).

The focus of the ECOVISION project was on *early cognitive vision*. This term expresses a specific understanding of visual mechanisms as outlined here. It is generally agreed that visual processing in the mammalian brain is governed by three structural (anatomical) principles: It is parallel within regions, hierarchical along the afferent pathways, and it contains substantial feedback from higher processing areas to lower ones (Felleman and Van Essen, 1991; Young, 1992). At the core of the ECOVISION project was the question of how these principles can lead to advanced image analysis, starting from early visual features, and arriving at more complex image descriptors. The central assumption is that in early visual processing there exists two stages. Different aspects of visual information in terms of visual modalities such as edge and junction information, colour, optic flow or stereo are computed by *local* filter operations at the first stage of *early vision* (Marr, 1997). At the consecutive stage of *early cognitive vision* these different visual modalities become processed and integrated by *global* mechanisms. Where we had coined the term "early cognitive vision" to indicate that at this stage the first meaningful visual entities are thought to arise. These structures on the one hand give rise to the Gestalt laws known from Psychology (see, e.g., Ellis, 1938), but still clearly stop short of truly cognitive aspects (e.g. object categorization or action planning). Recent psychological evidence suggests that the transformation to cognitive vision (or cognition) might then be defined by the interaction of these visual processes with other inputs and with the actual motor outputs by which objects are manipulated (Hommel, 2001). Thereby the vision domain is finally left and cognition has to be addressed.

To subserve higher processes, early cognitive vision needs to rely on two properties:

Rich and generic representations: Visual information is represented in a format that is rich enough and generally usable for a large number of cognitive tasks. This notion has been motivated by the fact that there exists two large areas in the human visual system (V1 and V2) that encode early visual modalities, providing input to areas associated with high level tasks (Kandell et al., 2000). Such a shared use of early processing stages makes sense for species that deal with a significant number of different tasks that rely on visual information. These include navigation, object recognition and categorisation, object manipulation and event anticipation, which themselves can be separated into further subtasks of substantial complexity.

Disambiguation by recurrent loops realizing contextual predictions: Based on early vision, different visual modalities are processed in a recurrent and interactive way, gated by constraints (such as attention, motivation) as well as cognitive tasks (such as scene understanding, decision making, and planning). Locally computed and therefore inherently ambiguous information in the different visual modalities becomes structured and disambiguated in processes that make use of *regularities* in visual data (see (Aloimonos and Shulman, 1989)). These regularities cover deterministic dependencies such as the motion of rigid and non-rigid bodies (Faugeras, 1993; Hartley and

Zisserman, 2000; Horn, 1994), as well as statistical dependencies such as collinearity, parallelism and symmetry (Krüger and Wörgötter, 2004; Sarkar and Boyer, 1994). The problems addressed at this stage such as motion estimation, segmentation, grouping, detection of independently moving objects are rather complex and deeply intertwined (Norman et al., 2004). It is probably this fact that has lead to the development of the existing recurrent (in contrast to sequential) processing architecture in the visual system.

Based on this understanding, the ECOVISION consortium has developed a precursor of a real-time vision machine described in Wörgötter et al. (2004). Furthermore, one application in the area of driver assistance systems has been developed as an industrial prototype (Mota et al., 2004).

The two aspects mentioned above were at the core of many workshop contributions as well as the contributions to this special issue.

The modeling of interacting early vision processes is addressed by Ogale and Aloimonos (2006) and Bayerl and Neumann (2006). Ogale and Aloimonos (2006) describe early vision as a chicken and egg problem in which different visual modalities interact to disambiguate visual information. In their model, the correspondence problem becomes embedded by recurrent loops into a framework that addresses depth segmentation, shape estimation and occlusion detection at the same time leading to an algorithm that extracts visual information in a reliable and robust way. Bayerl and Neumann (2006) introduce a model for the integration of feedforward and feedback processes between areas V1 and MT. In their model, locally computed erroneous information computed at edges and junctions—in particular T-junctions indicating occlusions—becomes disambiguated by a reccurrent process including image feature extraction, optic flow computation and segmentation.

Disambiguation in the context of an attentional vision system is also at the core of the contribution of Elder et al. (2006) and Pauwels et al. (2006). Elder et al. (2006) describe a system in which multiple weak cues (skin colour, motion and foreground extraction) become combined in a Bayesian framework. This integration of cues allows for a reliable detection of persons in a wide-field view integrated in an active foveated vision system. The use of attention mechanisms for optic flow disambiguation is discussed in Pauwels et al. (2006). Motivated by human gaze control, they fixate in a foveated vision framework as a means to improve optic flow computation and ego-motion computation.

Finally, Duits et al. (2006) approach the construction of and reconstruction from orientation scores in the context of wavelet theory and relate this to early visual processing in striate cortex. Furthermore, they apply channel theory (Granlund, 2000) in the context of robust estimation of local orientation.

References

Aloimonos, Y. and Shulman, D. 1989. *Integration of Visual Modules—An extension of the Marr Paradigm*. Academic Press, London.

Bayerl, P. and Neumann, H. 2006. Disambiguating visual motion by form-motion interaction—A computational model. *International Journal of Computer Vision* (Special Issue).

Duits, R., Felsberg, M., Granlund, G., and Romeny, B. 2006. Image analysis and reconstruction using a wavelet transform constructed from a reducible representation of the Euclidian motion group. *International Journal of Computer Vision* (Special Issue).

Elder, J.H., Prince, S.J.D., Hou, Y., Sizinstev, M., and Olevskiy, E. 2006. Pre-attentive and attentive detection of humans in wide-field scenes. *International Journal of Computer Vision* (Special Issue).

Ellis, W.D. (Ed.) 1938. *Gestalt Theory, A Source Book for Gestalt Psychology*.

Faugeras, O.D. 1993. *Three-Dimensional Computer Vision*. MIT Press.

Felleman, D.J. and Van Essen, D.C. 1991. Distributed hierarchical processing in primate cerebral cortex. *Cerebral Cortex*, 1:1–47.

Goodhill, G. (Ed.) 2005. *Network: Computation in Neural Systems* (Special Issue). Taylor and Francis, 16(4).

Granlund, G.H. 2000. An associative perception-action structure using a localized space variant information representation. In *Proceedings of Algebraic Frames for the Perception-Action Cycle (AFPAC)*, Kiel, Germany.

Hartley, R.I. and Zisserman, A. 2000. *Multiple View Geometry in Computer Vision*. Cambridge University Press.

Hommel, B. 2001. Eventfiles: Feature binding in and across perception and action. *Trends in Cognitive Sciences* 8:494–500.

Horn, B.K.P. (Ed.) 1994. *Robot Vision*. MIT Press.

Kandell, E.R., Schwartz, J.H., and Messel, T.M. 2000. *Principles of Neural Science (4th edition)*. McGraw Hill.

Krüger, N. and Wörgötter, F. 2004. Statistical and deterministic regularities: Utilisation of motion and grouping in biological and artificial visual systems. *Advances in Imaging and Electron Physics* 131:82–147.

Marr, D. 1977. Vision: *A Computational Investigation into the Human Representation and Processing of Visual Information.* Freeman.

Norman, J., Todd, J., and Orban, G. 2004. Perception of three-dimensional shape from specular highlights, deformations of shading, and other types of visual information. *Psychological Science* 15:565–570.
Ogale, A.S. and Aloimonos, Y. 2006. A roadmap to the integration of early visual modules. *International Journal of Computer Vision*.
Pauwels, K., Lappe, M., and Van Hulle, M. 2006. Fixation as a mechanism for stabilization of short image sequences. *International Journal of Computer Vision* (Special Issue).
Mota S., Ros E., Ortigosa E.M., and Pelayo F.J. 2004. Bio-inspired motion detection for blind spot overtaking monitor. *International Journal of Robotics and Automation*, 4(19):190–196.
Sarkar, S. and Boyer, K.L. 1994. *Computing Perceptual Organization in Computer Vision.* World Scientific.
Wörgötter, F., Krüger, N., Pugeault, N., Calow, D., Lappe, M., Pauwels, K., Van Hulle, M., Tan, S., and Johnston, A. 2004. Early cognitive vision: Using gestalt-laws for task-dependent, active image-processing. *Natural Computing* 3(3):293–321.
Young, M.P. 1992. Objective analysis of the topological organization of the primate cortical visual system. *Nature*, 358:152–155.

Norbert Krüger
Florentin Wörgötter
Marc M. van Hulle
marc@neuro.kuleuven.be