

Multiscale keypoint hierarchy for Focus-of-Attention and object detection J Rodrigues, J M H du Buf* (Escola Superior de Tecnologia, ADEE, Campus da Penha; *Vision Laboratory, FCT-DEEI, Campus de Gambelas; University of Algarve, P 8000-810, Faro, Portugal; e-mail: jrodrig@ualg.pt; <http://w3.ualg.pt/dubuf/vision.html>)

Hypercolumns in area V1 contain frequency- and orientation-selective simple and complex cells for line (bar) and edge coding, plus end-stopped cells for keypoint (vertex) detection. A single-scale (single-frequency) mathematical model of single and double end-stopped cells on the basis of Gabor filter responses was developed by Heitger et al. (1992 Vision Research 32 963-981). We developed an improved model by stabilising keypoint detection over neighbouring micro-scales. Because of the many filter scales represented by simple and complex cells, it is likely that, apart from a multi-scale line/edge representation, the visual cortex also constructs a multi-scale keypoint representation over multiple frequency octaves. Simulations with many different objects showed that, at very coarse scales, keypoints are found near the centre (centroid) of the objects. At medium scales, keypoints are detected at important parts of objects, for example the "fingers" of plant leaves, whereas at finest scales they are found at points of high curvature on the contour. In other words, the multi-scale keypoint representation offers a hierarchical structure in terms of object, sub-objects and contour. In addition, a retinotopic summation of all detected keypoints over all scales provides one map with peaks caused by keypoints that are stable over many scales, and this map can be used as a saliency map for Focus-of-Attention. Further experiments showed that, for example, face detection can be achieved by grouping keypoints at expected positions (eyes, nose, mouth), taking into account symmetries and distances, and by combining suitable scales. Hence, position, rotation and scale invariant face detection may be achieved by embedding the multi-scale keypoint representation, in addition to the line/edge representation, into feedforward and feedback streams to/from higher areas V2, V4 and IT (what or parvo system), whereas the saliency map for FoA interacts with short-term memory via areas PP and MT (where or magno system).

[Supported by PRODEP III Medida 5-Accção 5.3]