

Implementation and Evaluation of the Unified Visual Attention Model (UVAM)

Alexander Schlegel, Rafael Schultze-Kraft

3. Juli 2010

Introduction

Object recognition in machines faces the problem that it requires complex calculations and therefore real-time performances are difficult to achieve. Visual attention has shown to be a good mechanism to improve the runtime of object recognition by rapidly filtering interesting image regions likely to contain significant information for a detailed analysis. Thereby the need for computational resources are reduced and the object recognition procedure is speeded up.

In the present work we implemented and evaluated the Unified Visual Attention Model (UVAM), first introduced by [1]. The UVAM is based on the Shift Invariant Feature Transform (SIFT) [2], a popular approach in object-recognition based on local features. In addition the UVAM makes use of bottom-up and top-down computational models of attention to speed up the object recognition procedure. For bottom-up attention we used the saliency based computational model by Itti et al. [3], where regions of an image are marked as salient with respect to their corresponding low-level features. As a metric for top-down attention Lee et. al [1] introduced the concept of 'familiarity': Familiarity is a measure of the resemblance of local features extracted from the input image to features of trained object models stored in a database. Features of high familiarity are seen as evidence of object existence, and are used to guide attention to locations likely to contain the corresponding object. Both saliency and familiarity map are combined to the unified attention map, that is used to guide attention for detailed analysis of the input image. The outline of the proposed UVAM is shown in Fig. 1. The top-down and

bottom-up components of the UVAM can be divided into two stages: the feed-forward attention stage of the left hand side, and the attention feedback loop of the right hand side. The feed-forward attention stage provides a preliminary estimation of the location of trained objects before starting the detailed object recognition. The attention feedback loop updates this estimation later based on the results of detailed object recognition on each selected ROI. The bottom-up saliency map (S-map) is calculated once during the feed-forward attention stage. In contrast, top-down familiarity is calculated once during the feed-forward attention stage to obtain the feed-forward familiarity map (FF F-map), and then repeatedly during the attention-feedback loop to obtain the feedback familiarity map (FB F-map). The S-map and the two F-maps are combined into the unified attention map (UA-map), which is used to select the ROI for detailed object recognition.

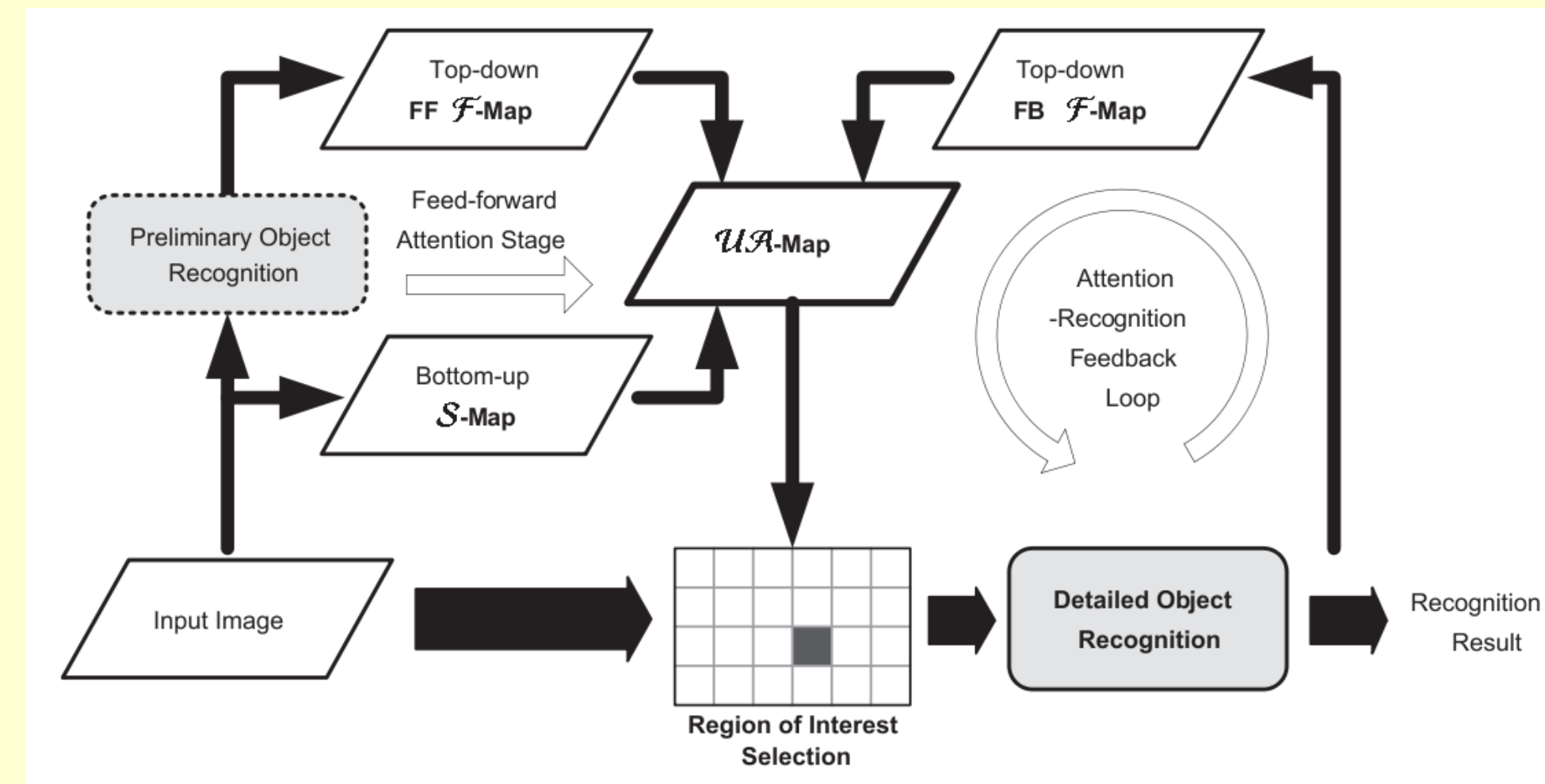


Fig.1: Outline of the UVAM algorithm.

Experiments

To evaluate the UVAM algorithm we used the Columbia Object Image Library (COIL-100, <http://www1.cs.columbia.edu/CAVE/software/softlib/coil-100.php>). Therefor we extracted the 75 first objects containing the largest amount of keypoints and neglected all others. In addition, we only used objects without rotation. These objects were embedded into backgrounds taken from the McGill Calibrated Colour Image Database (<http://tabby.vision.mcgill.ca/>). An example test image is illustrated in Fig. 2.



Fig. 2: An example of a test image using COIL-100 objects embedded into a background.

To assess the performance of the UVAM algorithm and compare it, our testing procedure contained four different steps: First, we tested speed performance without the use of neither bottom-up saliency nor familiarity maps. Second, we incorporated only saliency maps into the algorithm. Third, we used only familiarity maps to test the performance. And finally, we used the actual UVAM algorithm (including both saliency and familiarity maps) to assess the speed performance of the object recognition process.

Results

Conclusion

References

- [1] S. LEE, K. KIM, J.-Y. KIM, M. KIM, H.-J. YOO, "Familiarity based unified visual attention model for fast and robust object recognition", *Pattern Recognition* 43, 2010.
- [2] D. G. LOWE, "Distinctive Image Features from Scale-Invariant Keypoints", *International Journal of Computer Vision* 60, 2004.
- [3] L. ITTI, C. KOCH & E. NIEBUR, "A Model of Saliency-Based Visual Attention for Rapid Scene Analysis", *IEEE Transactions on Pattern Analysis and Machine Intelligence* 20, 1998.