ine Institute of Neuromorphic Engineering

THE NEUROMORPHIC ENGINEER
A PUBLICATION OF INE-WEB.ORG

10.2417/1200712.0052

Biologically inspired mobile-robot self localization

Christian Siagian and Laurent Itti

Using both global and local visual features, Beobot *can find its own position in the environment.*

The problem of localization is central to that of endowing mobile machines with intelligence. Vision, the human's main perceptual system for localization, is attractive because of its effectiveness in various kinds of environment. Range sensors such as sonar and ladar are, in contrast, not particularly robust when used outdoors due to their need for structural regularities such as flat walls and narrow corridors. Also, global positioning systems (GPS) are not suitable for environments where there is no satellite visibility, such as underwater, in caves, or on Mars.

What is critical in implementing a vision-based localization system is the reliability of the visual cues/features it uses. These cues can be categorized somewhere in between the two extremes of local and global features. At one end of the spectrum, local features (such as scale invariant feature transform, SIFT[1], keypoints and Harris corners[2]) are computed over a limited area of the image. At the other, global features (such as color[3] or texture[4]) may pool information over the entire image into, e.g., histograms. In general, global features are more robust because histograms average out local properties. However, these holistic approaches, for the most part, are limited to roughly classifying places rather than providing very accurate localization.

Today, with so many human vision studies available for consideration, we have a unique opportunity to develop systems that take inspiration from neuroscience. For example, even in the initial viewing of a scene, the human visual-processing system already guides its attention to visually interesting regions within its field of view using perceptual saliency.[5] This highlights a limited number of possible points of interest in the image. Concurrent with mechanisms of saliency, humans also exhibit the ability to rapidly get the 'gist' of a scene.[6] Human subjects are consistently able to answer basic questions including those on general semantic classification (indoors vs. outdoors or room types such kitchen, office, etc.) and rough visual-feature distributions in a scene[7,8] after briefly glancing (50–100ms) at an image.

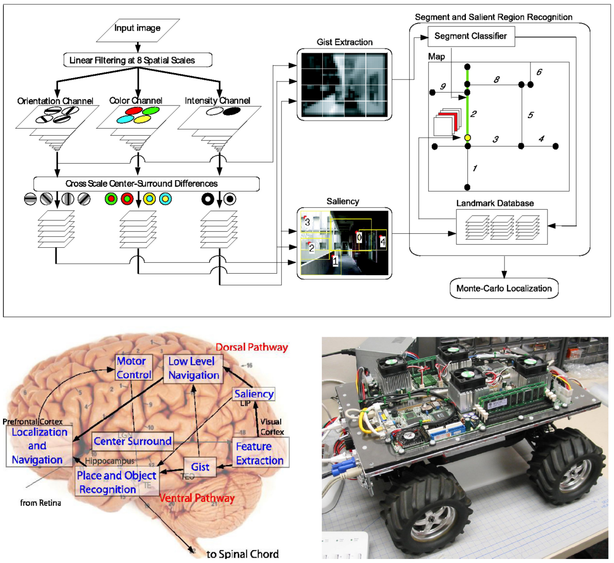

Figure 1. *The top diagram shows how the system works. The bottom left image shows a brain mapping of the different parts of the system modules. Bottom right is the picture of the system platform, called* Beobot.

The concurrent use of gist and saliency is a major trait of biological vision systems, which usually consider multiple abstractions (or, using the terminology above, both local and global features) to form a well-rounded understanding. The use of saliency[5] and gist[9]are also at the core of our system,[10] illustrated in Figure 1, which works by having the two competences complementing each other to form a multi-expert recognition system that localizes at two levels. Specifically, gist, which is a global feature, tries to recognize places called segments: continuous paths in the environment such as sections of hallway, alley, or road. The saliency combined with the SIFT keypoints (all local features) form a salient region to further refine the result to the coordinate location using a back-end Monte-Carlo localization. Figure 2 displays (and explains) how the system works in a park full of trees at one time step.

In order to take full advantage of the processing power of our robot, the *Beobot* (a BeoWulf mobile cluster) gist and saliency are

Continued on next page

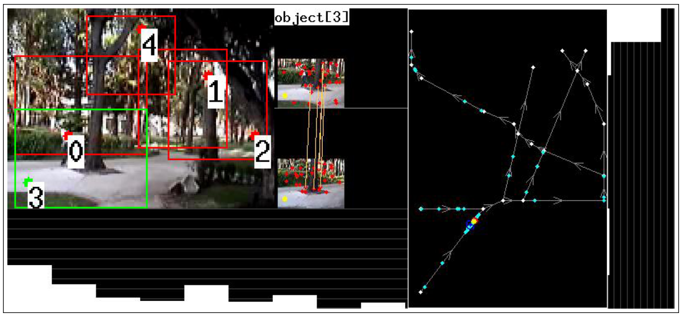

Figure 2. *A snapshot of a test run of the system. The main image shows the salient region windows. A green window means a database match, while red means 'not found'. An object match is displayed to the right of the main image. Bottom left is the segment estimation vector derived from gist alone (there are nine possible segments in the environment): a higher bar corresponds to a higher belief that the robot is on a given segment. The dotted paths towards the right of the image are the projections of the robot state onto a map: cyan disks are the different hypotheses the system considers, the yellow disk is the location of the matched database item, the blue disk is the most likely location, while the red disk is the ground truth. The radius of the blue circle is five feet. The right-most histogram shows the number of particles at each segment. Here, the robot believes that it is towards the end of the first segment, which is correct within a few feet.*

computed using the same raw features from various visual domains (color, intensity, and orientation). These are extracted in parallel as, studies suggest, happens in the human visual cortex. In addition, the lengthy salient-region matching process (time consuming because it requires a recall of stored information from a database of the environment, usually obtained during some kind of training) is also performed in parallel by having different processors search different parts of the database.

We tested the system in three outdoor environments: a building complex with an area of 82×55m^2, for which we got a 99cm location error; a 82×110m^2 park (2.63m error); and a 137×178^2 open-field area (3.46m). The position disparities occur because a substantial number of the landmarks used are far away from the robot, which makes more accurate visual deduction difficult. Nevertheless, the system is able to accurately and efficiently localize itself by extracting both global and local visual information (gist and saliency, respectively) thanks to the design choices we made in both software and hardware. The next step in the research is to incorporate navigation so that the robot can follow a command to go to a goal location.

We gratefully acknowledge the contribution of the National Science Foundation, Human Frontier Science Program, National Geospatial-Intelligence Agency, and the Defense Advanced Research Projects Agency, and thank Rob Peters and Lior Elazary for their valuable time and advice.

Author Information

Laurent Itti and Christian Siagian
Computer Science Department
University of Southern California (USC)
Los Angeles, CA
http://ilab.usc.edu
http://ilab.usc.edu/siagian

Laurent Itti received his PhD in Computation and Neural Systems from California Institute of Technology in 2000. He is now an associate professor of Computer Science, Psychology, and Neuroscience at USC. His research interests are in biologically-inspired computational vision, in particular in the domains of visual attention, gist, saliency, and surprise, with technological applications to video compression, target detection, and robotics.

Christian Siagian is currently working towards a PhD in computer science. His research interests include robotics and computer vision: particularly biologically inspired techniques.

References

1. S. Se, D. G. Lowe, and J. J. Little, *Vision-based global localization and mapping for mobile robots*, **IEEE Trans. Robotics 21** (3), pp. 364–375, 2005.
2. S. Frintrop, P. Jensfelt, and H. Christensen, *Pay attention when selecting features*, **Int'l Conf. on Pattern Recognition**, August 2006.
3. I. Ulrich and I. Nourbakhsh, *Appearance-based Place Rcognition for Topological Localization*, **IEEE Int'l Conf. Robotics and Automation (ICRA)**, pp. 1023–1029, April 2000.
4. A. Torralba, K. P. Murphy, W. T. Freeman, and M. A. Rubin, *Context-based vision system for place and object recognition*, **IEEE Int'l Conf. Computer Vision (ICCV)**, pp. 1023–1029, October 2003.
5. L. Itti, C. Koch, and E. Niebur, *A Model of Saliency-Based Visual Attention for Rapid Scene Analysis*, **IEEE Trans. Pattern Analysis and Machine Intelligence 20** (11), pp. 1254–1259, Nov 1998.
6. I. Biederman, *Do background depth gradients facilitate object identification?*, **Perception 10**, pp. 573–578, 1982.
7. T. Sanocki and W. Epstein, *Priming spatial layout of scenes*, **Psychol. Sci. 8**, pp. 374 – 378, 1997.
8. R. A. Rensink, *The Dynamic Representation of Scenes*, **Visual Cognition 7**, pp. 17–42, 2000.
9. C. Siagian and L. Itti, *Rapid Biologically-Inspired Scene Classification Using Features Shared with Visual Attention*, **IEEE Trans. Pattern Analysis and Machine Intelligence 29** (2), pp. 300–312, Feb 2007.
10. C. Siagian and L. Itti, *Biologically-Inspired Robotics Vision Monte-Carlo Localization in the Outdoor Environment*, **Proc. IEEE Int'l Conf. Intelligent Robots and Systems (IROS)**, Oct 2007.