# Vision for Robotics

# Vision for Robotics

**Danica Kragic**

*School of Computer Science and Communication, KTH*
*Stockholm 10044*
*Sweden*
*dani@kth.se*

**Markus Vincze**

*Technische Universitat Wien*
*Vienna*
*Austria*
*vincze@acin.tuwien.ac.at*

the essence of knowledge

Boston – Delft

**Foundations and Trends® in Robotics**

*Published, sold and distributed by:*
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
USA
Tel. +1-781-985-4510
www.nowpublishers.com
sales@nowpublishers.com

*Outside North America:*
now Publishers Inc.
PO Box 179
2600 AD Delft
The Netherlands
Tel. +31-6-51115274

The preferred citation for this publication is D. Kragic and M. Vincze, Vision for Robotics, Foundation and Trends® in Robotics, vol 1, no 1, pp 1–78, 2010

ISBN: 978-1-60198-260-5

**Foundations and Trends® in Robotics**

Volume 1 Issue 1, 2010

# Editorial Scope

**Foundations and Trends® in Robotics** will publish survey and tutorial articles in the following topics:

- Mathematical modelling
- Kinematics
- Dynamics
- Estimation Methods
- Robot Control
- Planning
- Artificial Intelligence in Robotics
- Software Systems and Architectures
- Mechanisms and Actuators
- Kinematic Structures
- Legged Systems
- Wheeled Systems
- Hands and Grippers
- Micro and Nano Systems
- Sensors and Estimation
- Force Sensing and Control
- Haptic and Tactile Sensors
- Proprioceptive Systems
- Range Sensing
- Robot Vision
- Visual Servoing
- Localization, Mapping and SLAM
- Planning and Control
- Control of manipulation systems
- Control of locomotion systems
- Behaviour based systems
- Distributed systems
- Multi-Robot Systems
- Human-Robot Interaction
- Robot Safety
- Physical Robot Interaction
- Dialog Systems
- Interface design
- Social Interaction
- Teaching by demonstration
- Industrial Robotics
- Welding
- Finishing
- Painting
- Logistics
- Assembly Systems
- Electronic manufacturing
- Service Robotics
- Professional service systems
- Domestic service robots
- Field Robot Systems
- Medical Robotics

## Information for Librarians

Foundations and Trends® in Robotics, 2010, Volume 1, 4 issues. ISSN paper version 1935-8253. ISSN online version 1935-8261. Also available as a combined paper and online subscription.

Foundations and Trends® in
Robotics
Vol. 1, No. 1 (2010) 1–78

DOI: 10.1561/2300000001

# Vision for Robotics

**Danica Kragic[1] and Markus Vincze[2]**

[1] *Centre for Autonomous Systems, Computational Vision and Active Perception Lab, School of Computer Science and Communication, KTH, Stockholm, 10044, Sweden, dani@kth.se*

[2] *Vision for Robotics Lab, Automation and Control Institute, Technische Universitat Wien, Vienna, Austria, vincze@acin.tuwien.ac.at*

## Abstract

Robot vision refers to the capability of a robot to visually perceive the environment and use this information for execution of various tasks. Visual feedback has been used extensively for robot navigation and obstacle avoidance. In the recent years, there are also examples that include interaction with people and manipulation of objects. In this paper, we review some of the work that goes beyond of using artificial landmarks and fiducial markers for the purpose of implementing vision-based control in robots. We discuss different application areas, both from the systems perspective and individual problems such as object tracking and recognition.

# Contents

# 1

# Introduction

For many living species, not least in the case of humans, visual perception plays a key role in their behavior. *Hand–eye coordination* ability gives us flexibility, dexterity, and robustness of movement that no machine can match yet. To locate and identify static, as well as moving objects, to determine how to grasp and handle them, we often rely strongly on our visual sense. One of the important factors is our ability to *track* objects, that is, to maintain an object in the field of view for a period of time using our oculomotor system as well as head and body motions. Humans are able to do this quickly and reliably without much effort. It is therefore natural to expect that the artificial cognitive systems we aim at developing will, to a certain extent, be able to demonstrate similar capabilities.

Robot vision refers to the capability of a robot to visually perceive the environment and interact with it. Robot vision extends methods of computer vision to fulfill the tasks given to robots and robotic systems. Typical tasks are to navigate toward a given target location while avoiding obstacles, to find a person and react to the person‘s commands, or to detect, recognize, grasp and deliver objects.

Thus, the goal of robot vision is to exploit the power of visual sensing to observe and perceive the environment and react to it. This follows the example of humans. It has been found that more than half of the human sensory cortex is attributed to seeing. Computer vision attempts to achieve the function of understanding the scene and the objects of the environment. With the increasing speed of processing power and progress in computer vision methods, making robots see became a main trend in robotics.

There, however, remains a fundamental difference between computer vision and robot vision. Computer vision targets the understanding of a scene mostly from single images or from a fixed camera position. Methods are tailored for specific applications and research is focused on individual problems and algorithms. On the other hand, robot vision requires to look at the system level perspective, where vision is one of several sensory components that work together to fulfill specific tasks. This property of the robotic system is also referred to as embodiment, where similar to biological systems the properties of the body shape the tasks of perception. Vision is used as a mean for the robot to act in and interact with the world–a robot system perceives to act and acts to perceive. Hence, visual processing is not an isolated entity, but part of a more complex system.

The future expectation is that robots will become ubiquitous. To robustly and safely interact with the world, robots need to perceive and interpret the environment so as to achieve context awareness and act appropriately. In general, we want to equip robots with minimal information in advance and get them to gather and interpret the necessary information required for execution of new tasks through interaction and on-line learning. This has been a long-term goal and one of the main drives in the field of artificial cognitive systems development. As an example, for a service robot that is to perform tasks in a human environment, it has to be able to learn about objects and object categories. However, the robots will not be able to form useful categories or object representations by being a passive observer of the environment. They should, like humans, learn about objects and their representations through interaction.

Vision has been used in robotic applications for more than three decades. Examples include applications in industrial settings, service, medical, and underwater robotics, to name some. In this paper we review some of the aspects of robot vision from early beginnings to more recent works. We concentrate in particular on attempts of developing active vision systems and examples where visual processing is considered as a primary aspect of the work rather than just a necessary input to the control loop.

There are many characteristics in common in computer vision research and vision research in robotics. For example, the Structure-and-Motion problem in vision has its analog of SLAM (Simultaneous Localization and Mapping) in robotics, visual SLAM being one of the important topics. Tracking is another area seeing great interest in both communities, in its many variations, such as 2D and 3D tracking, single and multi-object tracking, rigid and deformable object tracking. Other topics of interest for both communities are object and action recognition. In the subsequent sections, we will discuss the differences in more detail.

## 1.1 Scope and Outline

Visual feedback enables robots to interact with the environment in various ways. In some cases, visual feedback is used for navigation and obstacle avoidance, while more complex examples include interaction with the user and manipulation of objects. The simplest interaction that can occur between a robot and an object may be to, for example, push an object in order to retrieve information about the size or weight of the object. Here, simple visual cues providing approximate 3D position of the object may be sufficient. A more complex interaction may be to grasp the object for the purpose of gaining the physical control over the object. Once the robot has the object in its hand, it can perform further actions on it, such as examining it from other views. Information obtained during interaction can be used to update the robots representations about objects and the world.

In cases where visual feedback is input for robot localization, mapping, or obstacle avoidance algorithms, extraction of low level visual features such as corners, interest features such as SIFT [132], or optical flow may be sufficient. Hence, visual feedback facilitates only state estimation step and no advanced reasoning is needed to explain what is really happening in a video sequence.

For the applications we envision in the future, this is not enough. We need vision systems that are able to provide adequate information no matter if the system is to manipulate an object or interact with a human. We need systems that understand what they "see" according to known or autonomously acquired models: these systems must perceive to act and act to perceive. An example may be a robot that enters a room, detects a table from a few meters distance, localizes a number of objects on it, and shifts it gaze toward each of them to obtain a more detailed foveal view of the whole or parts of an object. This information can then be used to either approach an object for picking it up or for storing the information about typical object positions in the environment. The processes that are necessary here are figure-ground segmentation and attention–these are commonly not considered in specific applications of object tracking or recognition.

Thus, the nature and level of detail of the extracted visual information depends on several factors: (i) the task a robot system is required to accomplish, (ii) number and position of visual sensors, (iii) required processing rate and (iv) indoor/outdoor environment, to name some. In this paper, we discuss different applications of visual input, both from the systems perspective and individual problems such as object tracking and recognition. This is structured as follows.

The discussion starts with Chapter 2, where we give an overview of methods from the early days and the use of vision in industrial applications (Section 2.1) to more recent trends in robot vision taking into account findings from biology, neuroscience, and cognitive science (Section 2.2). As last part of this section we stress the importance of considering not only individual functions in robot vision but also robot vision *systems*.

A tentative model of a robot vision system is shown in Figure 1.1. The overview aims at indicating that, at this rather abstract level of description, a robot vision system fulfills three major functions: navigation, grasping, and Human Robot Interaction (HRI). The interplay of these functions depends on the task. For example, navigation is today

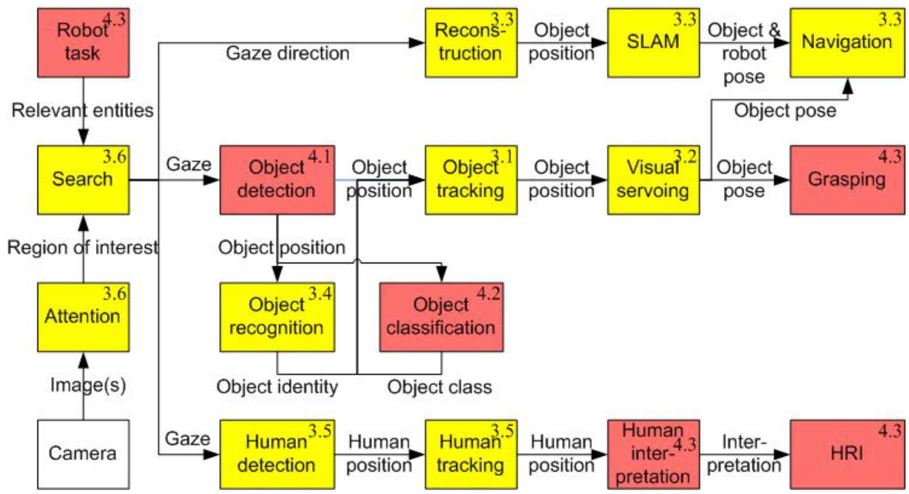


Fig. 1.1 Block diagram of the main tasks of a robot vision system: navigation, grasping and Human Robot Interaction. The numbers refer to Sections. Yellow indicates Chapter 3 “What works” and red indicates Chapter 4 “Open challenges”. Please see text for more details.

considered a largely solved problem with methods suitable for applications and advanced topics open to research. Thus, in Chapter 3 we present aspects of robot vision for which robust performance has been achieved. This is indicated by boxes colored in yellow in Figure 1.1. In Chapter 4 we review the open challenges that are still considered unsolved (indicated in red) and more related to formalizing the semantics of robot tasks and binding them to grasping and HRI. Finally, the review ends with a discussion and a short outlook in Chapter 5.

We note that the strict sequence of functions in Figure 1.1 is only for clarity. There are several approaches that combine functions and establish direct links that are not shown. Other functions, such as adaptation of functions to specific tasks or learning are also not explicitly given, may apply to several of the blocks, and will be mentioned when appropriate.

# References


[1] A. Albu-Schaffer, W. Bertleff, B. Rebele, B. Schafer, K. Landzettel, and G. Hirzinger, "Rokviss–robotics component verification on ISS current experimental results on parameter identification," in *ICRA*, p. 38793885, 2006.

[2] R. Alferez and Y. Wang, "Geometric and illumination invariants for object recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, no. 6, pp. 505–536, 1999.

[3] Y. Aloimonos and D. Shulman, *Integration of Visual Modules.* Academic Press, Inc., 1989.

[4] Y. Aloimonos, I. Weiss, and A. Bandyopadhyay, "Active vision," in *Proceedings of the DARPA Image Understanding Workshop*, pp. 552–573, 1987.

[5] R. T. amd D. Stoianovici, "Medical robotics in computer-integrated surgery," *IEEE Transactions on Robotics and Automation*, vol. 19, no. 5, pp. 765–781, 2003.

[6] O. Amidi, T. Kanade, and R. Miller, *Vision-Based Autonomous Helicopter Research at CMU.* in [6], 2000.

[7] R. Bajcsy, "Active perception," *in Proceedings of the IEEE*, vol. 76, no. 8, pp. 996–1005, 1988.

[8] D. H. Ballard, "Animate vision," *Artificial Intelligence*, vol. 48, no. 1, pp. 57–86, 1991.

[9] E. Baseski, N. Pugeault, S. Kalkan, D. Kraft, F. Wrgtter, and N. Krger, "A scene representation based on multi-modal 2d and 3d features," in *ICCV*, pp. 1–7, 2007.

[10] G. Bekey and J. Yuh, "The status of robotics," *IEEE Robotics & Automation Magazine*, vol. 15, no. 1, pp. 80–86, 2008.

[11] S. Benhimane and E. Malis, "Homography-based 2d visual tracking and servoing," *International Journal of Robotic Research (Special Issue on Vision and Robotics Joint with the International Journal of Computer Vision)*, vol. 26, no. 7, pp. 661–676, 2007.

[12] M. Bertero, T. Poggio, and V. Torre, "Ill-posed problems in early vision," in *Proceedings of the IEEE*, pp. 869–889, 1988.

[13] F. Berton, G. Sandini, and G. Metta, "Anthropomorphic visual sensors," in *In Encyclopedia of Sensors*, American Scientific Publishers, 2005.

[14] I. Biederman, "Recognition-by-components: A theory of human image understanding," *APA Journal; Psychological Review*, vol. 94, no. 2, pp. 115–147, 1987.

[15] G. Biegelbauer, A. Pichler, M. Vincze, C. Nielsen, H. Andersen, and K. Haeusler, "The inverse approach of flexpaint [robotic spray painting]," *IEEE Robotics & Automation Magazine*, vol. 12, no. 3, pp. 24–34, 2005.

[16] G. Biegelbauer, M. Vincze, and W. Wohlkinger, "Model-based 3d object detection–Efficient approach using superquadrics," *Machine Vision Applications,* vol. accepted, 2008.

[17] H. Bischof, H. Wildenauer, and A. Leonardis, "Illumination insensitive eigenspaces," in *IEEE International Conference Computer Vision ICCV*, pp. 233–238, 2001.

[18] B.J.Kuipers, "The cognitive map: Could it have been any other way?," in *Spatial Orientation: Theory, Research, and Application*, (H. L. Pick, Jr., and L. P. Acredolo, eds.), pp. 345–359, New York: Plenum Press, 1983.

[19] M. Björkman and D. Kragic, "Combination of foveal and peripheral vision for object recognition and pose estimation," *in Proceedings of the IEEE International Conference on Robotics and Automation, ICRA'04*, vol. 5, pp. 5135–5140, April 2004.

[20] R. Bolles and P. Horaud, "3dpo: A three-dimensional part orientation system," *The International Journal of Robotics Research*, vol. 5, no. 3, pp. 3–26, 1986.

[21] H. Borotschnig, L. Paletta, M. Prantl, and A. Pinz, "Appearance based active object recognition," *International Journal of Image and Vision Computing*, vol. 18, no. 9, pp. 715–728, 2000.

[22] S. Brandt, C. Smith, and N. Papanikolopoulos, "The Minnesota robotic visual tracker: A flexible testbed for vision-guided robotic research," in *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics, "Humans, Information and Technology"*, vol. 2, pp. 1363–1368, 1994.

[23] C. Brautigam, J. Eklundh, and H. Christensen, "A model-free voting approach for integrating multiple cues," in *ECCV*, pp. 734–750, 1998.

[24] R. Brooks, "Model-based 3d interpretation of 2d images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 5, no. 2, pp. 140–150, 1983.

[25] M. Brown, D. Burschka, and G. Hager, "Advances in computational stereo," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 993–1008, p. 8, 2003.

[26] J. Byne and J. Anderson, "A CAD-based computer vision system," *Image and Vision Computing*, vol. 16, pp. 533–539, 1998.

[27] J. A. Castellanos and J. D. Tardós, *Mobile Robot Localization and Map Building: A Multisensor Fusion Approach*. Kluwer Academic Publishers, 1999.

[28] L. Y. Chang, N. S. Pollard, T. M. Mitchell, and E. P. Xing, "Feature selection for grasp recognition from optical markers," in *IEEE International Conference on Intelligent Robots and Systems*, 2007.

[29] F. Chaumette, "Potential problems of stability and convergence in image-based and position-based visual servoing," in *The Confluence of Vision and Control,* no. 237 in Lecture Notes in Control and Information Sciences, pp. 66–78, Springer-Verlag, 1998.

[30] F. Chaumette and S. Hutchinson, "Visual servo control I: Basic approaches," *IEEE Robotics and Automation Magazine*, vol. 13, no. 4, pp. 82–90, 2006.

[31] Y. Cheng, M. Maimone, and L. Matthies, "Visual odometry on the mars exploration rovers–A tool to ensure accurate driving and science imaging," *IEEE Robotics & Automation Magazine*, vol. 13, no. 54-62, p. 2, 2006.

[32] S. Choi, S. Ban, and M. Lee, "Biologically motivated visual attention system using bottom-up saliency map and top-down inhibition," *Neural Information Processing–Letters and Review*, vol. 2, pp. 19–25, 2004.

[33] H. I. Christensen and H.-H. Nagel, eds., *Cognitive Vision Systems: Sampling the Spectrum of Approach*. Springer Verlag, Lecture Notes in Computer Science, pp. 3948, 2006.

[34] J. Clark and A. Yuille, *Data Fusion for Sensory Information Processing Systems*. Kluwer Academic Publisher, 1990.

[35] A. Comport, E. Marchand, and F. Chaumette, "A real-time tracker for markerless augmented reality," in *IEEE International Symposium on Mixed and Augmented Reality*, pp. 36–45, 2003.

[36] A. Comport, M. Pressigout, E. Marchand, and F. Chaumette, "A visual servoing control law that is robust to image outliers," in *Proceedings of the 2003 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 492–497, Las Vegas, Nevada, October 2003.

[37] S. Coradeschi and A. Saffiotti, "An introduction to the anchoring problem," *Robotics and Autonomous Systems, Special Issue on Perceptual Anchoring*, vol. 43, no. 2–3, pp. 85–96, 2003.

[38] P. Corke and S. Hutchinson, "A new partitioned approach to image-based visual servo control," *IEEE Transactions on Robotics and Automation*, vol. 17, pp. 507–515, 2001.

[39] P. I. Corke, *Visual Control of Robots: High Performance Visual Servoing*. Research Studies Press, John Wiley, 1996.

[40] J. Crowley and H. Christensen, *Vision as Process*. Springer Verlag, 1995.

[41] L. Crowley, J. Coutaz, and F. Bérard, "Things that see: Machine perception for human computer interaction," *Communications of the A.C.M.*, vol. 43, pp. 54–64, March 2000.

[42] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *CVPR*, pp. 886–893, 2005.

[43] B. Dariush, "Human motion analysis for biomechanics and biomedicine," *Machine Vision and Applications*, vol. 14, pp. 202–205, 2003.

[44] A. J. Davison, "Real-time simultaneous localisation and mapping with a single camera," in *ICCV*, 2003.

[45] M. Deans and M. Hebert, "Experimental comparison of techniques for localization and mapping using a bearings only sensor," in *ISER'00, Seventh International Symposium on Experimental Robotics*, December 2000.

[46] L. Deng, W. J. Wilson, and F. Janabi-Sharifi, "Characteristics of robot visual servoing methods and target model estimation," in *Proceedings of the 2002 IEEE International Symposium on Intelligent Control*, pp. 684–689, Vancouver, Canada, October 27–30 2002.

[47] S. Dickinson, D. Wilkes, and J. Tsotsos, "A computational model of view degeneracy," *IEEE Transactions on PAMI*, vol. 21, no. 8, pp. 673–689, 1999.

[48] E. Dickmanns, *Dynamic Vision for Perception and Control of Motion*. Springer, 2007.

[49] E. D. Dickmanns and V. Graefe, "Dynamic monocular machine vision," *Machine Vision and Applications*, vol. 1, pp. 223–240, 1988.

[50] R. Dillmann, "Teaching and learning of robot tasks via observation of human performance," *Robotics and Autonomous Systems*, vol. 47, pp. 109–116, 2004.

[51] G. Dissanayake, P. Newman, S. Clark, H. Durrant-Whyte, and M. Corba, "A solution to the slam building problem," *IEEE TRA*, vol. 17, no. 3, pp. 229–241, 2001.

[52] T. Drummond and R. Cipolla, "Real-time visual tracking of complex structures," *IEEE Transactions on PAMI*, vol. 24, no. 7, pp. 932–946, 2002.

[53] C. Eberst, M. Barth, K. Lutz, A. Mair, S. Schmidt, and G. Farber, "Robust vision-based object recognition integrating highly redundant cues for indexing and verification," in *IEEE ICRA*, pp. 3757–3764, 2000.

[54] A. Edsinger and C. C. Kemp, "Manipulation in human environments," in *IEEE/RSJ International Conference on Humanoid Robotics*, pp. 102–109, 2006.

[55] Eklundh, Bjorkman, and Hayman, "Object appearance from integration of 3d and 2d cues in real scenes," *Journal of Vis.*, vol. 3, pp. 646–646, October 2003.

[56] S. Ekvall, P. Jensfelt, and D. Kragic, "Object detection and mapping for service robot tasks," *Robotics*, vol. 25, no. 2, pp. 175–188, 2007.

[57] S. Ekvall and D. Kragic, "Interactive grasp learning based on human demonstration," in *IEEE International Conference on Robotics and Automation, ICRA'04*, 2004.

[58] S. Ekvall, D. Kragic, and P. Jensfelt, "Object detection and mapping for service robot tasks," *Robotica*, vol. 25, pp. 175–187, 2007.

[59] B. Espiau, "Effect of camera calibration errors on visual servoing in robotics," in *3rd International Symposium on Experimental Robotics*, Kyoto, Japan, October 1993.

[60] B. Espiau, F. Chaumette, and P. Rives, "A new approach to visual servoing in robotics," *IEEE Transactions on Robotics and Automation*, vol. 8, pp. 313–326, June 1992.

[61] M. Everingham, "Overview and results of the classification challenge," *http: pascallin.ecs.soton.ac.uk challenges VOC voc2008 workshop everingham_cls.pdf*.

[62] A. H. Fagg and M. A. Arbib, "Modeling parietal–premotor interactions in primate control of grasping," *Neural Networks*, vol. 11, no. 7–8, pp. 1277–1303, 1998.

[63] J. Feddema and C. Lee, "Adaptive image feature prediction and control for visual tracking with a hand–eye coordinated camera," *IEEE Transactions on Systems, Man and Cybernetics*, vol. 20, no. 5, pp. 1172–1183, 1990.

[64] W. Feiten, B. Magnussen, J. Bauer, G. Hager, and K. Toyama, "Modeling and control for mobile manipulation in everyday environments," in *8th International Symposium on Robotics Research*, 1998.

[65] R. Fergus, P. Perona, and A. Zisserman, "Object class recognition by unsupervised scale-invariant learning," in *CVPR*, pp. 264–271, 2003.

[66] V. Ferrari, T. Tuytelaars, and L. Gool, "Simultaneous object recognition and segmentation from single or multiple model views," *International Journal of Computer Vision*, vol. 67, no. 2, pp. 159–188, 2006.

[67] P. Fitzpatrick, G. Metta, L. Natale, S. Rao, and G. Sandini, "Learning about objects through action–Initial steps towards artificial cognition," in *IEEE International Conference on Robotics and Automation*, pp. 3140–3145, 2003.

[68] J. Folkesson, J. P, and H. I. Christensen, "Vision slam in the measurement subspace," in *IEEE ICRA05*, 2005.

[69] S. Frintrop, "VOCUS: A visual attention system for object detection and goal-directed search," *Lecture Notes in Computer Science, Springer*, vol. 3899, 2006.

[70] A. Gee, D. Chekhlov, A. Calway, and W. Mayol-Cuevas, "Discovering higher level structure in visual slam," *IEEE Transactions on Robotics*, vol. 24, no. 5, pp. 980–990, 2008.

[71] J. J. Gibson, *The Ecological Approach to Visual Perception.* Lawrence Erlbaum Associates, 1987.

[72] L. Goncavles, E. di Bernardo, D. Benson, M. Svedman, J. Ostrovski, N. Karlsson, and P. Pirjanian, "A visual front-end for simultaneous localization and mapping," in *IEEE ICRA*, pp. 44–49, 2005.

[73] M. Goodale and A. Milner, "Separate visual pathways for perception and action," *Trends Neuroscience*, vol. 15, no. 1, p. 205, 1992.

[74] A. Gopalakrishnan and A. Sekmen, "Vision-based mobile robot learning an navigation," in *IEEE International Workshop on Robot and Human Interactive Communication, RO-MAN'05*, pp. 48–53, 2005.

[75] W. Grimson, *Object Recognition by Computer: The Role of Geometric Constraints.* MIT Press, 1990.

[76] J. Gutmann and K. Konolige, "Incremental mapping of large cyclic environments," in *IEEE International Symposium on Computational Intelligence in Robotics and Automation*, pp. 318–325, 1999.

[77] G. Guy and G. Medioni, "Inferring global perceptual contours from local features," *International Journal of Computer Vision*, vol. 20, no. 1–2, pp. 113–133, 1996.

[78] G. Hager, "A modular system for robust positioning using feedback from stereo vision," *IEEE Transactions on Robotics and Automation*, vol. 13, no. 4, pp. 582–595, 1997.

[79] G. Hager and P. Belhumeur, "Real-time tracking of image regions with changes in geometry and illumination," in *Proceedings of the Computer*

*Society Conference on Computer Vision and Pattern Recognition, CVPR'96*, pp. 403–410, 1996.

[80] G. Hager and P. Belhumeur, "Efficient region tracking with parametric models of geometry and illumination," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 10, pp. 1025–1039, 1998.

[81] G. Hager and K. Toyama, "The XVision system: A general-purpose substrate for portable real-time vision applications," *Computer Vision and Image Understanding*, vol. 69, no. 1, pp. 23–37, 1996.

[82] C. Harris, "Tracking with rigid models," in *Active Vision*, (A. Blake and A. Yuille, eds.), pp. 59–73, MIT Press, 1992. ch. 4.

[83] H. Harzallah, C. Schmid, F. Jurie, and A. Gaidon, "Classification aided two stage localization," *http://pascallin.ecs.soton.ac.uk/challenges/VOC/voc2008/workshop/harzallah.pdf*.

[84] E. Hayman and J. Eklundh, "Probabilistic and voting approaches to cue integration for figure-ground segmentation," in *ECCV, Springer LNCS 2352*, pp. 469–486, 2002.

[85] E. Hayman and J.-O. Eklundh, "Statistical background subtraction for a mobile observer," in *Proceedings of the Ninth IEEE International Conference on Computer Vision*, pp. 67–74, 2003.

[86] S. Helmer and D. G. Lowe, "Object class recognition with many local features," in *CVPR GMBV Workshop on Generative-Model Based Vision*, 2004.

[87] J. Hill and W. Park, "Real time control of a robot with a mobile camera," in *Proceedings of the 9th ISIR*, pp. 233–246, 1979.

[88] D. Hoffman and W. Richards, "Parts of recognition," *Cognition*, vol. 18, pp. 65–96, 1996.

[89] A. Hoover, G. Jean-Baptiste, X. Jiang, and et al., "An experimental comparison of range image segmentation algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 7, pp. 673–689, 1996.

[90] B. Horn, *Robot Vision*. MIT Press, 1986.

[91] K. Huebner, M. Bjorkman, B. Rasolzadeh, M. Schmidt, and D. Kragic, "Integration of visual and shape attributes for object action complexes," in *6th International Conference on Computer Vision Systems (ICVS'08), Lecture Notes in Artificial Intelligence*, vol. 5008, pp. 13–22, D-69121 Heidelberg, Germany: Springer-Verlag, 2008.

[92] S. Hutchinson, G. D. Hager, and P. I. Corke, "A tutorial on visual servo control," *IEEE Transactions on Robotics and Automation*, vol. 12, pp. 651–670, October 1996.

[93] M. Isard and A. Blake, "CONDENSATION-Conditional density propagation for visual tracking," *International Journal of Computer Vision*, vol. 29, no. 1, pp. 5–28, 1998.

[94] L. Itti, "Models of bottom-up and top-down visual attention," PhD thesis, California Institute of Technology, 2000.

[95] L. Itti, C. Koch, and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254–1259, 1998.

[96] A. Jaklic, A. Leonardis, and F. Solina, *Segmentation and Recovery of Superquadrics*. Kluwer Academic Publishers, 2000.

[97] P. Jensfelt, J. Folkesson, D. Kragic, and H. I. Christensen, "Exploiting distinguishable image features in robotic mapping and localization," in *1st European Robotics Symposium (EUROS-06)*, (H. I. Christensen, ed.), Palermo, Italy, March 2006.

[98] P. Jensfelt, D. Kragic, J. Folkesson, and M. Björkman, "A framework for vision based bearing only 3D SLAM," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA'06)*, Orlando, FL, May 2006.

[99] F. Jurie, "Robust hypothesis verification: Application to model-based object recognition," *Pattern Recognition*, vol. 32, no. 6, 1999.

[100] F. Jurie and M. Dhome, "Real time tracking of 3D objects: An efficient and robust approach," *Pattern Recognition*, vol. 35, pp. 317–328, 2002.

[101] N. Karlsson, E. di Bernardo, J. Ostrowski, L. Goncalves, P. Pirjanian, and M. Munich, "The vSLAM algorithm for robust localization and mapping," in *International Conference on Robotics and Automation*, pp. 24–29, Barcelona, Spain, April 18–22 2005.

[102] D. Katsoulas, C. Bastidas, and D. Kosmopoulos, "Superquadric segmentation in range images via fusion of region and boundary infromation," in *IEEE Transactions on PAMI*, vol. 30, no. 5, pp. 781–795, 2008.

[103] T. Kawanishi, H. Murase, and S. Takagi, "Quick 3D object detection and localization by dynamic active search with multiple active cameras," in *IEEE International Conference on Pattern Recognition, ICPR'02*, pp. 605–608, 2002.

[104] H. Kjellstrom, J. Romero, and D. Kragic, "Visual recognition of grasps for human-to-robot mapping," in *IROS*, 2008.

[105] G. Klein and T. Drummond, "Robust visual tracking for non-instrumented augmented reality," in *Proceedings of the 2nd IEEE and ACM International Symposium on Mixed and Augmented Reality*, pp. 113–122, 2003.

[106] G. Klein and D. Murray, "Improving the agility of keyframe-based slam," in *ECCV*, 2008.

[107] C. Koch and S. Ullman, "Shifts in selective visual attention: Towards the underlying neural circuitry," *Human Neurobiology*, vol. 4, pp. 219–227, 1985.

[108] T. Koike and J. Saiki, "Stochastic guided search model for search asymmetries in visual search tasks," *Biologically Motivated Computer Vision*, pp. 408–417, 2002.

[109] D. Koller, K. Daniilidis, and H. Nagel, "Model-based object tracking in monocular image sequences of road traffic scenes," *International Journal of Computer Vision*, vol. 10, no. 3, pp. 257–281, 1993.

[110] S. Kovacic, A. Leonardis, and F. Pernus, "Planning sequences of views for 3-d object recognition and pose determination," *Pattern Recognition*, vol. 31, no. 10, pp. 1407–1417, 1998.

[111] D. Kraft, E. Baseski, M. Popovic, N. Krüger, N. Pugeault, D. Kragic, S. Kalkan, and F. Wörgötter, "Birth of the object: Detection of objectness and extraction of object shape through object action complexes," *International Journal of Humanoid Robotics*, vol. 5, pp. 247–265, 2008.

[112] D. Kragic, "Visual servoing for manipulation: Robustness and integration issues," PhD thesis, CVAP, Royal Institute of Technology, Stockholm, Sweden, 2001.

[113] D. Kragic, M. Bjorkman, H. Christensen, and J.-O. Eklundh, "Vision for robotic object manipulation in domestic settings," *Elsevier; Robotics and Autonomous Systems*, vol. 52, no. 1, pp. 85–100, 2005.

[114] D. Kragic and H. Christensen, "Cue integration for visual servoing," *IEEE Transactions on Robotics and Automation*, vol. 17, no. 1, pp. 18–27, 2001.

[115] J. Krivic and F. Solina, "Art-level object recognition using superquadrics," *Elsevier; Computer Vision and Image Understanding*, vol. 95, no. 1, pp. 105–126, 2004.

[116] G.-J. M. Kruijff, H. Zender, P. Jensfelt, and H. I. Christensen, "Clarification dialogues in human-augmented mapping," in *Proceedings of the 1st Annual Conference on Human-Robot Interaction, HRI'06*, Salt Lake City, UT, March 2006.

[117] A. Krupa, J. Gangloff, C. Doignon, M. de Mathelin, G. Morel, J. Leroy, L. Soler, and J. Marescaux, "Autonomous 3-d positioning of surgical instruments in robotized laparoscopic surgery using visual servoing," *IEEE Transactions on Robotics and Automation*, vol. 19, no. 5, pp. 842–853, 2003.

[118] N. M. Kwok, G. Dissanayake, and Q. Ha, "Bearing only SLAM using a SPRT based Gaussian sum filter," in *IEEE ICRA05*, 2005.

[119] V. Kyrki and D. Kragic, "Integration of model-based and model-free cues for visual object tracking in 3d," in *IEEE International Conference on Robotics and Automation, ICRA'05*, pp. 1566–1572, 2005.

[120] V. Kyrki and D. Kragic, "Tracking unobservable rotations by cue integration," in *IEEE International Conference on Robotics and Automation 2006. ICRA'06*, pp. 2744–2750, Orlando, Florida, 2006.

[121] V. Kyrki, D. Kragic, and H. Christensen, "Measurement errors in visual servoing," *Robotics and Autonomous Systems*, vol. 54, no. 10, pp. 815–827, 2006.

[122] V. Kyrki and K. Schmock, "Integation methods of model-free features for 3d tracking," in *Scandinavian Conference on Image Analysis*, pp. 557–566, 2005.

[123] L. Ladicky, P. Torr, and P. Kohli, "Object-class segmentation using higher order CRF," *http://pascallin.ecs.soton.ac.uk/challenges/VOC/voc2008/workshop/ladicky.pdf*.

[124] K. Lee, H. Buxton, and J. Feng, "Selective attention for cueguided search using a spiking neural network," in *Proceedings of the International Workshop on Attention and Performance in Computer Vision*, pp. 55–62, Graz, Austria, July 2003.

[125] S. Lee, D. Jang, E. Kim, S. Hong, and J. Han, "A real-time 3d workspace modeling with stereo camera," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2140–2147, August 2005.

[126] T. Lemaire, S. Lacroix, and J. Solà, "A practical 3D bearing-only SLAM algorithm," in *IEEE/RSJ IROS*, pp. 2757–2762, 2005.

[127] A. Leonardis and H. Bischof, "Robust recognition using eigenimages," *Computer Vision and Image Understanding: CVIU*, vol. 78, no. 1, pp. 99–118, 2000.

[128] Z. Li, "A saliency map in primary visual cortex," *Trends in Cognitive Sciences*, vol. 6, no. 1, pp. 9–16, 2002.

[129] G. Lopez-Nicolas, C. Sagues, J. Guerrero, D. Kragic, and P. Jensfelt, "Nonholonomic epipolar visual servoing," in *In IEEE International Conference on Robotics and Automation 2006. ICRA'06*, pp. 2378–2384, Orlando, Florida, 2006.

[130] D. G. Lowe, "Robust model-based motion tracking through the integration of search and estimation," *International Journal of Computer Vision*, vol. 8, no. 2, pp. 113–122, 1992.

[131] D. G. Lowe, "Object recognition from local scale-invariant features," in *Proceedings of the International Conference on Computer Vision*, pp. 1150–1157, 1999.

[132] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[133] D. G. Lowe, "Three-dimensional object recognition from single two-dimensional images," *Artificial Intelligence*, vol. 31, pp. 355–395, March 1987.

[134] E. Malis and F. Chaumette, "Theoretical improvements in the stability analysis of a new class of model-free visual servoing methods," *IEEE Transactions on Robotics and Automation*, 2002.

[135] E. Malis, F. Chaumette, and S. Boudet, "2-1/2-d visual servoing," *IEEE Transactions on Robotics and Automation*, vol. 15, pp. 238–250, April 1999.

[136] G. Mariottini, D. Prattichizzo, and G. Oriolo, "Epipole-based visual servoing for nonholonomic mobile robots," in *ICRA*, 2004.

[137] D. Marr, *Vision*. San Francisco: W. H. Freeman and Company, 1982.

[138] L. Masson, F. Jurie, and M. Dhome, "Robust real time tracking of 3d objects," *International Confrence on Pattern Recognition*, pp. 23–26, 2004.

[139] D. Meger, P.-E. Forssen, K. Lai, S. Helmer, S. McCann, T. Southey, M. Baumann, J. J. Little, and D. G. Lowe, "Curious george: An attentive semantic robot," *Robotics and Autonomous Systems* 2008 (in submission).

[140] G. Metta and P. Fitzpatrick, "Better vision through experimental manipulation," in *2nd International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems*, vol. 11, pp. 109–128, 2002.

[141] G. Metta and P. Fitzpatrick, "Early integration of vision and manipulation," *Adaptive Behavior*, vol. 11, no. 2, pp. 109–128, 2003.

[142] A. Mian, M. Bennamoun, and R. Owens, "Three-dimensional model-based object recognition and segmentation in cluttered scenes," *IEEE Transactions on PAMI*, vol. 28, no. 10, 2006.

[143] B. Micusik, H. Wildenauer, and J. Kosecka, "Detection and matching of rectangular structures," in *IEEE CVPR*, 2008.

[144] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *IEEE Transactions on PAMI*, vol. 27, no. 10, pp. 1615–1630, 2005.

[145] L. Montesano, M. Lopes, A. Bernardino, and J. Santos-Victor, "Learning object affordances: From sensorymotor coordination to imitation," *IEEE Transactions on Robotics*, vol. 24, no. 1, pp. 15–26, 2008.

[146] J. Moren, A. Ude, A. Koene, and G. Cheng, "Biologically-based top-down attention modulation for humanoid interactions," *International Journal of Humanoid Robotics*, pp. 3–24, 2008.

[147] A. C. Murilo, J. Kosecka, J. J. Guerrero, and C. Sagues, "Visual door detection integrating appearance and shape cues," *Robotics and Autonomous Systems*, 2008.

[148] A. Namiki, K. Hashimoto, and M. Ishikawa, "A hierarchical control architecture for high-speed visual servoing," *International Journal of Robotic Research*, vol. 22, no. 10–11, pp. 873–888, 2003.

[149] V. Navalpakkam and L. Itti, "Sharing resources: Buy attention, get recognition," in *International Workshop on Attention and Performance in Computer Vision*, 2003.

[150] E. S. Neo, T. Sakaguchi, K. Yokoi, Y. Kawai, and K. Maruyama, "Operating humanoid robots in human environments," in *Workshop on Manipulation for Human Environments, Robotics: Science and Systems*, 2006.

[151] P. Newman and K. Ho, "SLAM-loop closing with visually salient features," in *IEEE ICRA*, pp. 644–651, 2005.

[152] K. Ogawara, J. Takamatsu, K. Hashimoto, and K. Ikeuchi, "Grasp recognition using a 3D articulated model and infrared images," in *IEEE International Conference on Intelligent Robots and Systems*, vol. 2, pp. 1590–1595, 2003.

[153] A. Oliva, A. Torralba, M. Castelhano, and J. Henderson, "Top-down control of visual attention in object detection," in *International Conference on Image Processing*, pp. 253–256, 2003.

[154] B. Olshausen, C. Anderson, and D. van Essen, "A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information," *Journal of Neuroscience*, vol. 13, pp. 4700–4719, 1993.

[155] PCCV. Performance Characterization in Computer Vision website, http://peipa.essex.ac.uk/benchmark/ index.html.

[156] M. Pechuk, O. Soldea, and E. Rivlin, "Learning function-based object classification from 3d imagery," *Computer Vision and Image Understanding*, vol. 110, no. 2, pp. 173–191, 2008.

[157] L. Petersson, P. Jensfelt, D. Tell, M. Strandberg, D. Kragic, and H. I. Christensen, "Systems integration for real-world manipulation tasks," in *Proceedings of the IEEE International Conference on Robotics and Automation, ICRA 2002*, vol. 3, pp. 2500–2505, 2002.

[158] J. Piazzi and D. Prattichizzo, "An auto-epipolar strategy for mobile robot visual servoing," in *IEEE/RSJ International Conference on Intelligent Robots and Systems*, vol. 2, pp. 1802–1807, 2003.

[159] M. Piccardi, "Background subtraction techniques: A review," in *IEEE International Conference on Systems, Man and Cybernetics, vol. 4*, 2004.

[160] N. Pinto, D. D. Cox, and J. J. DiCarlo, "Why is real-world visual object recognition hard?," *PLoS Computational Biology*, vol. 4, p. e27, January 2008.

[161] A. Pinz, "Object categorization," *Foundations and Trends in Computer Graphics and Vision*, vol. 1, no. 4, pp. 255–353, 2006.

[162] P. Pirjanian, H. Christensen, and J. Fayman, "Application of voting to fusion of purposive modules: An experimental investigation," *Robotics and Autonomous Systems*, vol. 23, no. 4, pp. 253–266, 1998.

[163] M. Pressigout and E. Marchand, "Real-time hybrid tracking using edge and texture information," *International Journal of Robotics Research*, vol. 26, no. 7, pp. 689–713, 2007.

[164] P. Prokopowicz, R. Swain, and R. Kahn, "Task and environment-sensitive tracking," in *Proceedings of 1994 IEEE Symposium on Visual Languages*, pp. 73–78, 1994.

[165] V. Raos, M. Umilta, A. Murata, L. Fogassi, and V. Gallese, "Functional properties of grasping-related neurons in the ventral premotor area F5 of the macaque monkey," *Journal of Neurophysiology*, vol. 95, no. 2, pp. 709–729, 2006.

[166] C. Rasmussen and G. Hager, "Probabilistic data association methods for tracking complex visual objects," *IEEE Transactions on PAMI*, vol. 23, no. 6, pp. 560–576, 2001.

[167] G. Rizzolatti, L. Fadiga, M. Matelli, V. Bettinardi, E. Paulesu, D. Perani, and F. Fazio, "Localization of grasp representations in humans by PET: 1. Observation versus execution," *Experimental Brain Research*, vol. 111, no. 2, pp. 246–252, 1996.

[168] G. Rizzolatti, L. Fogassi, and V. Gallese, "Parietal cortex: From sight to action," *Current Opinion in Neurobiology*, vol. 7, pp. 562–567, 1997.

[169] G. Rizzolatti, G. Luppino, and M. Matelli, "The organization of the cortical motor system: New concepts," *Electroencephalography and Clinical Neurophysiology*, vol. 106, no. 4, pp. 283–296, 1998.

[170] J. Rosen and B. Hannaford, "Doc at a distance," *IEEE Spectrum*, vol. 43, no. 10, pp. 34–39, 2006.

[171] C. Sagüés and J. Guerrero, "Visual correction for mobile robot homing," *Robotics and Autonomous Systems*, To appear, 2005.

[172] A. Sanderson and L. Weiss, "Image-based visual servo control using relational graph error signals," in *Proceedings of the IEEE*, pp. 1074–1077, 1980.

[173] A. Saxena, J. Driemeyer, and A. Y. Ng, "Robotic grasping of novel objects using vision," *International Journal of Robotics Research*, vol. 27, no. 2, pp. 157–173, 2008.

[174] S. Schaal, "Is imitation learning the route to humanoid robots?," *Trends in Cognitive Sciences*, vol. 3, no. 6, pp. 233–242, 1999.

[175] C. Schmid, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *CVPR*, pp. 2169–2178, 2006.

[176] Semantic Robot Vision Challenge, http://www.semantic-robot-vision-challenge.org/.

[177] T. Serre, L. Wolf, S. Bileschi, M. Riesenhuber, and T. Poggio, "Recognition with cortex-like mechanisms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, no. 3, pp. 411–426, 2007.

[178] A. Shahdi and S. Sirouspour, "Adaptive/robust control for time-delay teleoperation," *IEEE Transactions on Robotics*, vol. 25, no. 196-205, p. 1, 2009.

[179] P. Sharkey and D. Murray, "Delays versus performance of visually guided systems," in *IEE Proceedings of Control Theory & Applications*, vol. 143, no. 5, pp. 436–447, 1996.

[180] J. Shi and C. Tomasi, "Good features to track," in *Proceedings of the IEEE Computer Vision and Pattern Recognition, CVPR'94*, pp. 593–600, 1994.

[181] T. Shipley and P. Kellman, "Advances in psychology: Form fragments to objects," *Elsevier Science B.V.*, vol. 130, 2001.

[182] Y. Shirai and H. Inoue, "Guiding a robot by visual feedback in assembling tasks," *Pattern Recognition*, vol. 5, pp. 99–108, 1973.

[183] Y. Shirai, R. Okada, and T. Yamane, "Robust visual tracking by integrating various cues," in *Robust Vision for Manipulation*, (M. Vincze and G. Hager, eds.), pp. 53–66, Spie/IEEE Series, 2000.

[184] H. Sidenbladh, D. Kragic, and H. I. Christensen, "A person following behaviour for a mobile robot," in *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 670–675, 1999.

[185] R. Sim, P. Elinas, M. Griffin, and J. J. Little, "Vision-based slam using the rao-blackwellised particle filter," in *IJCAI Workshop on Reasoning with Uncertainty in Robotics*, July 2005.

[186] J. Sivic, B. C. Russell, A. A. Efros, A. Zisserman, and W. T. Freeman, "Discovering objects and their localization in images," in *International Conference on Computer Vision*, pp. 370–377, 2005.

[187] J. Sivic and A. Zisserman, "Video google: A text retrieval approach to object matching in videos," in *International Conference on Computer Vision*, pp. 1470–1477, 2003.

[188] S. Smith, "ASSET-2 - Real-Time Motion Segmentation and Object Tracking," Tech. Rep. TR95SMS2b, Oxford Centre for Functional Magnetic Resonance Imaging of the Brain (FMRIB), Department of Clinical Vision and Image Processing Group, DRA Chertsey, DERA, UK, 1995.

[189] F. Solina and R. Bajcsy, "Recovery of parametric models from range images: The case for superquadrics with global deformations," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 12, no. 12, pp. 131–147, 1990.

[190] T. Southey and J. J. Little, "Object discovery using motion, Appearance and shape," in *AAAI Cognitive Robotics Workshop*, 2006.

[191] M. Stark, P. Lies, M. Zillich, J. Wyatt, and B. Schiele, "Functional object class detection based on learned affordance cues," in *ICVS–International Conference on Computer Vision Systems*, pp. 435–444, 2008.

[192] A. Stoytchev, "Behavior-grounded representation of tool affordances," in *IEEE International Conference on Robotics and Automation*, pp. 3060–3065, 2005.

[193] M. Suppa, S. Kielhoefer, J. Langwald, F. Hacker, K. H. Strobl, and G. Hirzinger, "The 3d-modeller: A multi-purpose vision platform," in *Proceedings of International Conference on Robotics and Automation*, 2007.

[194] M. Sutton, L. Stark, and K. Bowyer, "Function from visual analysis and physical interaction: A methodology for recognition of generic classes of objects," *Image and Vision Computing*, vol. 16, pp. 746–763, 1998.

[195] J. Tardós, J. Neira, P. Newman, and J. Leonard, "Robust mapping and localization in indoor environments using sonar data," *IJRR*, vol. 4, 2002.

[196] M. J. Tarr and H. H. Blthoff, "Image-based recognition in man, monkey, and machine," *Cognition*, vol. 67, pp. 1–20, 1998.

[197] G. Taylor and L. Kleeman, "Fusion of multimodal visual cues for model-based object tracking," in *Australiasian Conference on Robotics and Automation*, Brisbane, Australia, 2003.

[198] G. Taylor and L. Kleeman, "Robust range data segmentation using geometric primitives for robotic applications," in *Proceedings of the 9th International Conference on Signal and Image Processing*, pp. 467–472, 2003.

[199] S. Thrun, D. Fox, and W. Burgard, "A probabilistic approach to concurrent mapping and localization for mobile robots," *Autonomous Robots*, vol. 5, pp. 253–271, 1998.

[200] S. Thrun, Y. Liu, D.Koller, A. Ng, Z. Ghahramani, and H. Durrant-White, "SLAM with sparse extended information filters," *IJRR*, vol. 23, no. 8, pp. 690–717, 2004.

[201] K. Toyama and G. Hager, "Incremental focus of attention for robust vision-based tracking," *International Journal of Computer Vision*, vol. 35, no. 1, pp. 45–63, 1999.

[202] A. Treisman and G. Gelade, "A feature integration theory of attention," *Cognitive Psychology*, vol. 12, pp. 97–136, 1980.

[203] J. Triesch and C. V. der Malsburg, "Self-organized integration of adaptive visual cues for face tracking," in *Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 102–107, 2000.

[204] M. Trivedi and T. G. J. McCall, "Looking-in and looking-out of a vehicle: Selected investigations in computer vision based enhanced vehicle safety," in *IEEE International Conference on Vehicular Electronics and Safety*, pp. 29–64, 2005.

[205] D. Tsakiris, C. Samson, and P. Rives, "Extending visual servoing techniques to nonholonomic mobile robots," in *The Confluence of Vision and Control*, vol. 1, (G. Hager, D. Kriegman, and S. Morse, eds.), Lecture Notes in Control and Information Systems, Springer-Verlag, 1999.

[206] J. Tsotsos, "On the relative complexity of passive vs active visual search," *International Journal of Computer Vision*, vol. 7, no. 2, pp. 127–141, 1992.

[207] A. Ude, C. Gaskett, and G. Cheng, "Foveated vision systems with two cameras per eye," in *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pp. 3457–3462, 2006.

[208] T. Uhlin, *Fixation and Seeing Systems.* PhD thesis, NADA, Royal Institute of Technology, KTH. 1996.

[209] L. Vacchetti, V. Lepetit, and P. Fua, "Stable real-time 3D tracking using online and offline information," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 1385–1391, 2004.

[210] K. E. A. van de Sande, T. Gevers, and C. G. M. Snoek, "Evaluation of color descriptors for object and scene recognition," in *IEEE Conference on Computer Vision and Pattern Recognition*, Anchorage, Alaska, USA, June 2008.

[211] M. Vincze, "On the design and structure of artificial eyes for tracking tasks," *Journal of Advanced Computational Intelligence and Intelligent Informatics JACIII*, vol. 9, no. 4, pp. 353–360, 2005.

[212] M. Vincze, M. Ayromlou, and W. Kubinger, "An integrating framework for robust real-time 3D object tracking," in *International Conference on Computer Vision Systems, ICVS'99*, pp. 135–150, 1999.

[213] M. Vincze and G. Hager, *Robust Vision for Vision-Based Control of Motion.* IEEE Press, 2000.

[214] M. Vincze, M. Zillich, W. Ponweiser, V. Hlavac, J. Matas, S. Obdrzalek, H. Buxton, J. Howell, K. Sage, A. Argyros, C. Eberst, and G. Umgeher, "Integrated vision system for the semantic interpretation of activities where a person handles objects," *CVIU*, vol. 113, no. 6, pp. 682–692, June 2009.

[215] M. Vinzce, M. Ayromlou, M. Ponweiser, and M. Zillich, "Edge projected integration of image and model cues for robust model-based object tracking," *International Journal of Robotics Research*, vol. 20, no. 7, pp. 533–552, 2001.

[216] I. Weiss and M. Ray, "Model-based recognition of 3d objects from single images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 2, pp. 116–128, 2001.

[217] F. Wörgötter, A. Agostini, N. Krüger, N. Shylo, and B. Porr, "Cognitive agents — A procedural perspective relying on the predictability of object–action–complexes," *Robotics and Autonomous Systems*, vol. 57, no. 4, pp. 420–432, 2009.

[218] S. Wrede, C. Bauckhage, G. Sagerer, W. Ponweiser, and M. Vincze, "Integration frameworks for large scale cognitive vision systems–an evaluative study," in *Proceedings of the 17th International Conference on Pattern Recognition ICPR*, pp. 761–764, 2004.

[219] P. Wunsch and G. Hirzinger, "Real-time visual tracking of 3-d objects with dynamic handling of occlusion," in *IEEE International Conference on Robotics and Automation, ICRA'97*, pp. 2868–2873, Albuquerque, New Mexico, USA, April 1997.

[220] M. Zillich and M. Vincze, "Anytimeness avoids parameters in detecting closed convex polygons," in *IEEE Computer Society Workshop on Perceptual Organization in Computer Vision at CVPR*, 2008.

[221] A. Zisserman, D. Forsyth, J. Mundy, C. Rothwell, J. Liu, and N. Pillow, "3D object recognition using invariance," *Artificial Intelligence*, vol. 78, no. 1–2, pp. 239–288, 1995.