Special issue on robot vision: what is robot vision?

Udo Frese · Heiko Hirschmüller

The Oxford English Dictionary defines a robot as ''a machine capable of carrying out a complex series of actions automatically […]''. So the key point of a robot is that it can act. However, to be able to act, a robot must first perceive: What is around me? Where am I? And where are things I want to act on?

Since the early days of robotics, the prime perception channel to answer these questions is vision—robot vision. Nowadays, this is even more true than before, since RGBD cameras provide direct geometrical perception. Remarkably, these devices provide ''vision'' in two senses. Seen from the outside, they extend the classical RGB images, where geometry must be indirectly inferred from with a direct sense of depth of scene geometry. Seen from the inside, they are actually vision sensors based on a conventional camera and structured light.

As in robotics the objective is to make robots act, there are two other points related (at least to some degree) to robot vision worth stating here: first, the final output is usually metric 3D information, which is needed as input for robot motion. And second, since motion involves time, robot perception and robot vision need to be accomplished in real-time, which is the theme of the JRTIP journal.

This special issue on Robot Vision aims at reporting on recent progress made to use real-time image processing towards addressing the above three questions of robotic perception. The call for papers of this special issue received a total of 26 manuscripts. Based on thorough reviews conducted by three reviewers per manuscript, seven high-quality papers for inclusion in this issue were selected which are briefly mentioned below.

The first two papers aim at perceiving what is around a robot. In their paper 'Dense real-time mapping of object-class semantics from RGB-D', Stückler et al. reconstruct a 3D environment model from a moving RGBD-camera and simultaneously label parts of this model with semantic categories such as ground, furniture, palette, or human. Kriegel et al. focus on specific unknown objects in the environment in their paper 'Efficient next-best-scan planning for autonomous 3D surface reconstruction of unknown objects', which is a good example regarding how the capability of robots is used in perception.

The question 'Where am I?' is addressed by Asadi et al. in their paper 'Delayed fusion for real-time vision-aided inertial navigation'. It discusses that a moving robot has already moved further than that shown in the image once that image has been processed by computer vision and how this aspect is incorporated into sensor-fusion algorithms.

Three papers aim at finding objects around a robot and determining their pose, thereby answering the third question. Orts-Escolano et al. concentrate on the performance of the first feature extraction stages and speed them up using GPGPU processing in their paper 'Real-time 3D semi-local surface patch extraction using GPGPU'. Wang et al. propose a novel global object descriptor that combines color and shape in their paper 'Textured/textureless object recognition and pose estimation using RGB-D image'. Finally, in their paper 'Advances in real-time object tracking—extensions for robust object tracking with a monte carlo particle filter' Mörwald et al. emphasize the tracking view on object pose estimation and propose how the tracker can improve reliability and accuracy of the determined pose.

U. Frese (✉)
University of Bremen, Enrique-Schmidt-Str. 5, 28215 Bremen, Germany
e-mail: ufrese@informatik.uni-bremen.de

H. Hirschmüller
German Aerospace Center (DLR), Institute of Robotics and Mechatronics, Oberpfaffenhofen, Germany

Finally, Miramond et al. discuss and investigate the strict separation in perception and action prevalent in many robotic systems and argue for a tightly integrated control loop in the tradition of bio-inspired robotics. They address the resulting demand for high frame rates with a specifically optimized bio-inspired real-time architecture in their paper 'Embedded and real-time architecture for bio-inspired vision-based robot navigation'.

Can we identify an overarching trend in robot vision from these articles? One trend is clearly the use of RGBD data with its direct geometric information that is very valuable, in particular for robotic applications. Whereas early approaches often relied on the depth data alone, more methods are being investigated that combine RGB and D in a smart way. The second trend is to address real-time performance using massive parallelism by GPU computing. This is remarkable as parallel computing has been a research topic for a long time. The early 90s were a very active period in this research field, where an outstanding example was Dickmann's pioneering work on autonomous driving, which used 60 transputers to achieve real-time performance [1]. However, parallelism did not become a widespread performance paradigm because single-core performance was still growing fast. This changed after the introduction of general purpose GPU-computing (and a variety of other strands to achieve real-time throughput) around 2005 and since then one sees an increasing use of parallelism in real-time robot vision.

The guest editors wish to thank the authors of this special issue for their contributions and the reviewers for their insightful reviews and wish the readers of this special issue a pleasant and informative reading.

Reference

1. Dickmanns, E.D.: Dynamic vision for perception and control of motion. Springer, ISBN 978-1-84628-638-4 (2007)