© EYEWIRE

SIFT-ing Through Features with ViPR

Application of Visual Pattern Recognition to Robotics and Automation

The first IEEE-IFR joint forum on Innovation and Entrepreneurship in Robotics and Automation took place on 10 April 2005 in Barcelona, Spain. The following article was presented at the forum as one of the three nominees selected for the Innovation and Entrepreneurship in Robotics and Automation Award. For further information, please visit http://teamster.usc.edu/~iera05/index.htm.

Hadi Moradi
Associate Vice President
RAS Industrial Activity Board

Visual recognition of patterns is a basic capability of most advanced species in nature. A big percentage of the human brain is devoted to visual processing, with a substantial portion of this area used for pattern recognition (see references [1] and [2] for an in-depth treatment of the subject). Visual pattern recognition enables a variety of tasks, such as object and target recognition, navigation, and grasping and manipulation, among others. Advances in camera technology have dramatically reduced the cost of cameras, making them the sensor of choice for robotics and automation. Vision provides a variety of cues about the environment (motion, color, shape, etc.) with a single sensor. Visual pattern recognition solves some fundamental problems in computer vision: correspondence, pose estimation, and structure from motion. Therefore, we consider visual

BY MARIO E. MUNICH, PAOLO PIRJANIAN, ENRICO DI BERNARDO, LUIS GONCALVES, NIKLAS KARLSSON, AND DAVID LOWE

1070-9932/06/$20.00©2006 IEEE

pattern recognition to be an important, cost-efficient primitive for robotics and automation systems.

Recent advances in computer vision have given rise to a robust and invariant visual pattern recognition technology that is based on extracting a set of characteristic features from an image. Such features are obtained with the scale invariant feature transform (SIFT) [5], [6], which represents the variations in brightness of the image around the point of interest. Recognition performed with these features has been shown to be quite robust in realistic settings. This article describes the application of this particular visual pattern recognition technology to a variety of robotics applications: object recognition, navigation, manipulation, and human-machine interaction. The following sections describe the technology in more detail and present a business case for visual pattern recognition in the field of robotics and automation.

Visual Pattern Recognition (ViPR)

The *visual pattern recognition* system developed by Evolution Robotics is versatile and works robustly with low-cost cameras. The underlying algorithm addresses an important challenge of all visual pattern recognition systems: performing reliable and efficient recognition in realistic settings with inexpensive hardware and limited computing power.

ViPR can be used in applications such as manipulation, human-robot interaction, and security. It can be used in mobile robots to support navigation, localization, mapping, and visual servoing. It can also be used in machine vision applications for object identification and hand-eye coordination. Other applications include entertainment and education since ViPR enables the automatic identification and retrieval of information about a painting or sculpture in a museum.

Based on the work of David Lowe [5], [6], ViPR represents an object as a set of SIFT features extracted from one or more images of such an object. SIFT features are highly localized visual templates, which are invariant to changes in scale and rotation and partially invariant to changes in illumination and viewpoint. Each SIFT feature is described by its location in the image (at subpixel accuracy), its orientation, scale, and a keypoint descriptor that has the above-mentioned invariance properties. The key components of ViPR are both the particular choice of features to be used (SIFT features) and the very efficient way of organizing and searching through a database of hundreds of thousand of SIFT features.

On a 640 × 480 image, the detector would typically find about 2,000 such features. In the training phase, the features are added to the model database and labeled with the object name associated with the training image. In the matching phase, a new image is acquired, and the algorithm searches through the database for all objects matching subsets of the features extracted from the new image using a bottom-up approach. Euclidean distance in the SIFT descriptor space is used to find similar features. A greedy version of a k-d tree allows for efficient search in this very high-dimensional space. A voting technique is then used to consolidate information coming from individual feature matches into a global match hypothesis. Each hypothesis contains features that match the same object and the same change in viewpoint for that particular object. The last step of the process refines the possible matches by computing an affine transformation between the set of features and the matched object(s) so that the relative position of the features is preserved through the transformation. The residual error after the transformation is used to decide whether to accept or reject the match hypothesis.

Recent advances in computer vision have given rise to a robust and invariant visual pattern recognition technology based on extracting a set of characteristic features from an image.

Figure 1 shows the main characteristics of the recognition algorithm. The first row displays the two objects to be recognized, and the other rows present recognition results under different conditions. The main characteristics of the algorithm are summarized in the following.

Invariant to Rotation and Affine Transformations

ViPR recognizes objects even if they are rotated upside down (rotation invariance) or placed at an angle with respect to the optical axis (affine invariance). See the second and third rows of Figure 1.

Invariant to Changes in Scale

Objects can be recognized at different distances from the camera, depending on the size of the objects and the camera resolution. Recognition works reliably from distances of several meters. See the second and third rows of Figure 1.

Invariant to Changes in Lighting

ViPR handles changes in illumination ranging from natural to artificial indoor lighting. The system is insensitive to artifacts caused by reflections or backlighting. See the fourth row of Figure 1.

Invariant to Occlusions

ViPR reliably recognizes objects that are partially blocked by other objects and objects that are partially in the camera's view. The amount of allowed occlusions is typically between 50–90%, depending on the object and the camera quality. See the fifth row of Figure 1.

Reliable Recognition

ViPR has an 80–95% success rate in uncontrolled settings. A 95–100% recognition rate is achieved in controlled settings.

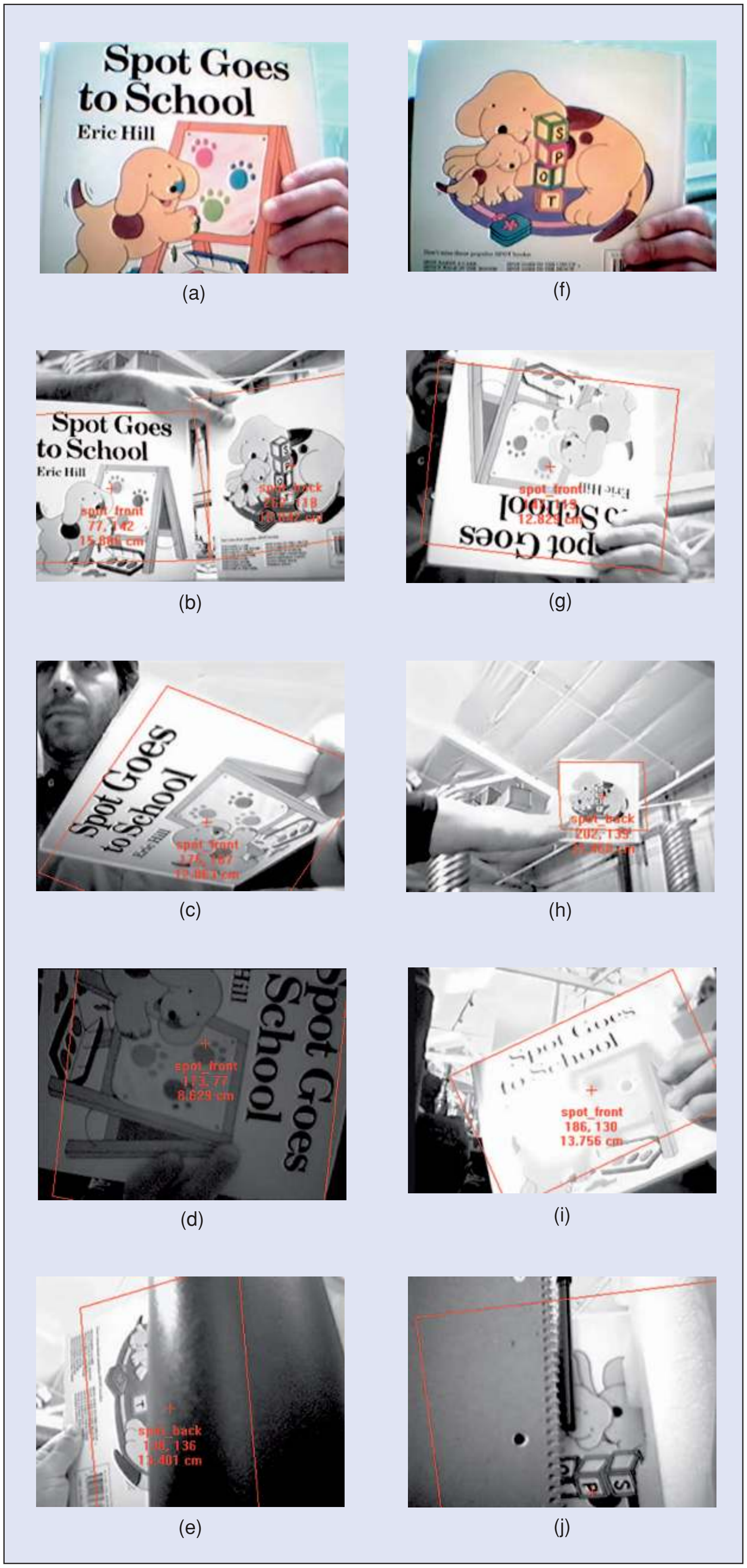


Figure 1. Examples of ViPR working under a variety of conditions. The first row presents the two objects to be recognized. The other rows display recognition results [the bounding box of the matched object is shown in red (dark gray)].

The recognition speed is a logarithmic function of the number of objects in the database; i.e., the recognition speed is proportional to log (N), where N is the number of objects in the database. The object library can store hundreds or even thousands of objects without a significant increase in computational requirements. The recognition frame rate is proportional to CPU power and image resolution. For example, the recognition algorithm runs at 14–18 fps (frames per second) at an image resolution of 208 × 160 on a 1,400-MHz Pentium IV processor, 5 fps at 208 × 160 on a 600-MHz MIPS-based 64-b RISC processor and 7 fps at 320 × 240o n a 400-MHz processor.

Reducing the image resolution decreases the image quality and, ultimately, the recognition rate. However, the object recognition system allows for graceful degradation with decreasing image quality. Each object model requires about 40 kB of memory.

The ViPR algorithm was initially developed at the laboratory of Prof. David Lowe at the University of British Columbia. Evolution Robotics faced the challenge of productizing and commercializing ViPR. The first stage in the process was the implementation of ViPR as a reliable piece of code. A first level of quality assurance (QA) was performed at this stage. QA was designed in order to ensure that the algorithm performed robustly and reliably in completely unknown and unstructured environments. The next stage was documentation and usability. A powerful, yet simple to use set of APIs was designed for ViPR, sample code examples and documentation were written, and a second level of QA was performed. At this point in time, a graphical user interface (GUI) application was developed in order to simplify customer evaluation of ViPR. Optimization of the code was also started at this point; we achieved about a 15 times improvement in computation while keeping the same recognition performance. This optimization of ViPR involved a lot of code reorganization, assembly implementation of the most time-consuming portions of the code, and modifications of the original algorithm. The final stage of the process has been the delivery and integration of the code in actual robotics and automation products.

A Business Case for ViPR in Robotics and Automation

In this section we analyze the different aspects of ViPR that make it a significant innovation for the robotics industry.

Novelty

The ViPR algorithm presented in the previous section is the first algorithm to date that has shown recognition rates above 90% in real-world conditions (undefined and unknown environments, variable lighting, etc.). ViPR also provides the basis for the visual simultaneous localization and mapping (vSLAM) system [3], [4], the first simultaneous localization and mapping (SLAM) system that enables low-cost and robust navigation in cluttered and populated environments. SLAM is one of the most fundamental, yet most challenging, problems in mobile robotics. To achieve full autonomy, a robot must possess the ability to explore its environment without user intervention, build a reliable map, and localize itself in the map. In particular, if global positioning sensor (GPS) data and external beacons are unavailable, the robot must determine autonomously what are appropriate reference points (or landmarks) on which to build a map. ViPR provides the capability for robust and reliable recognition of visual landmarks.

Figure 2 shows the result of vSLAM after the robot has traveled in a typical two-bedroom apartment. The robot was driven along a reference path (this path is unknown to the SLAM algorithm). The vSLAM algorithm builds a map consisting of landmarks marked with blue circles in the figure. The corrected robot path, which uses a combination of visual features and odometry, provides a robust and accurate position determination for the robot as seen by the red path in the figure.

The green path (odometry only) is obviously incorrect, since, according to this path, the robot is traversing through walls and furniture. The red path (the vSLAM corrected path), on the other hand, is consistently following the reference path.

Based on experiments in various typical home environments on different floor surfaces (carpet and hardwood floor), and using different image resolution (low: 320 × 280, high: 640 × 480), robot localization accuracy was achieved as shown in Table 1.

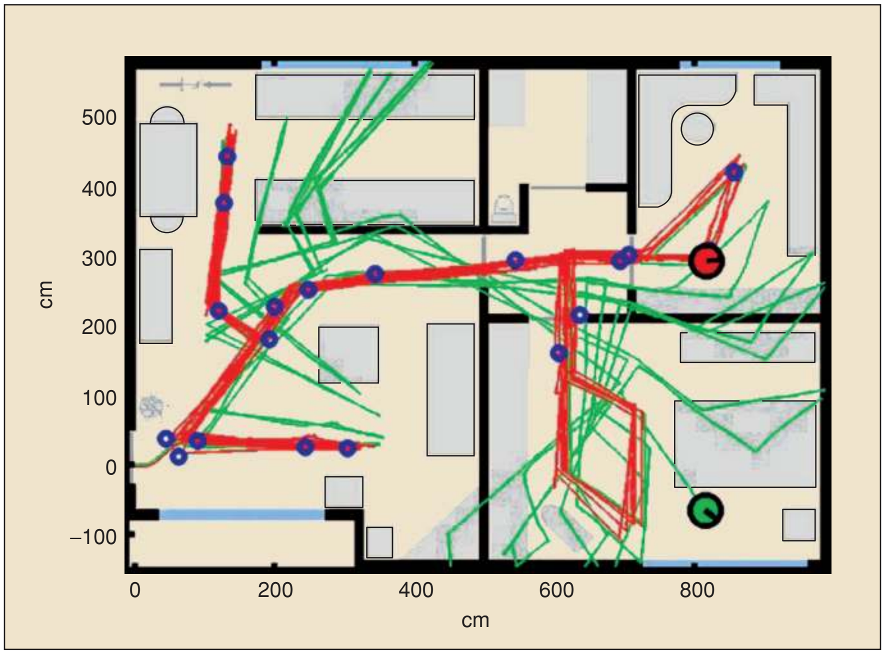


Figure 2. Example result of SLAM using vSLAM. Red path: vSLAM estimate of robot trajectory. Green path: odometry estimate of robot trajectory. Blue circles: vSLAM landmarks created during operations.

Potential for Commercialization

ViPR has already been integrated into a series of commercial products and prototypes. ViPR is also part of the Evolution Robotics Software Platform (ERSP) and is being evaluated by a number of companies at the moment. Some commercial products that currently use ViPR are presented in Figure 3 and highlighted in the following.

Sony's AIBO ERS-7

AIBO is an entertainment robot that has incorporated ViPR for supporting two critical features: reliable human-robot interaction and robust self-charging. AIBO recognizes a set of commands issued with a set of predefined, feature-rich cards that are reliably matched with ViPR. AIBO's charging station has a characteristic pattern that is recognized with ViPR, supporting localization of the station and robust self-docking.

> ***The key components of ViPR are both the particular choice of features to be used and the very efficient way of organizing and searching through a database of hundreds of thousand of SIFT features.***

Yaskawa's SmartPal

SmartPal [7] is a prototype robot that is capable of recognizing a set of objects placed on a table understanding a user request

Table 1. Robot localization accuracy.

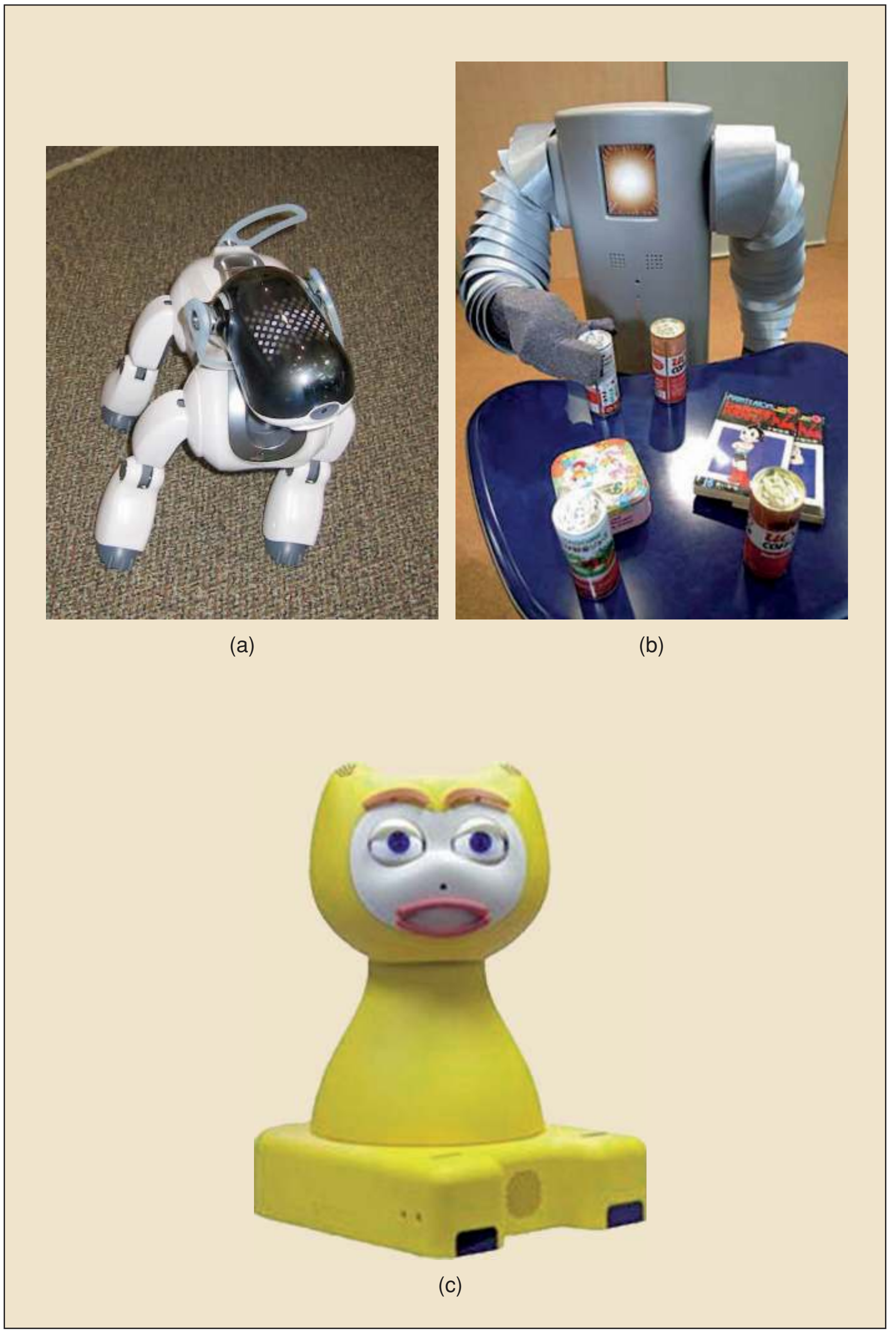
Floor Surface	Image Resol Ution	Median Error (cm)	Std Dev Error (cm)	Median Error (deg)	Std Dev Error (deg)
Carpet	Low	13.6	9.4	4.79	7.99
Carpet	High	12.9	10	5.97	7.03
HW	Low	12.2	7.5	5.28	6.62
HW	High	11.5	8.3	3.47	5.41

for one of the objects and then grabbing it. SmartPal uses ViPR for object recognition and distance calculation in order to perform manipulation and grasping of the object.

To achieve full autonomy, a robot must possess the ability to explore its environment without user intervention, build a reliable map, and localize itself in the map.

Phillips' iCat

The iCat [8], [9] is an experimentation platform for studying human-robot interaction. The iCat uses ViPR to recognize cards with a particular picture/symbol that represents a particular genre of music (e.g., rock, pop, romantics, or children's music). Once the type of music is recognized, the iCat selects music of the desired genre from a content server (PC) and sends it to a music renderer (Philips Streamium device).


(a)

(b)

(c)

Figure 3. Robots using ViPR: (a) Sony's AIBO ERS-7, (b) Yaskawa's SmartPal, and (c) Phillips' iCat.

Evolution Robotics' LaneHawk

LaneHawk is a ViPR-based system developed for the retail market that helps in detecting and recognizing items placed in the bottom of the basket of shopping carts. A camera flush-mounted in the checkout lane is continuously checking for the presence of the cart and for the existence of items in the bottom of the basket. When ViPR recognizes an item, its UPC information is passed to the point of sale (POS). The cashier verifies the quantity of items that were found under the basket and continues to close the transaction. LaneHawk provides a solution for the problem of bottom-of-basket (BOB) losses (items that are overlooked or forgotten at checkout) and also improves productivity by increasing the checkout speed of the lanes.

Economic Viability

ViPR requires two pieces of hardware to work: a camera and a CPU. ViPR is able to perform at recognition rates better than 90% using a US$40–80 Web camera or a US$4–9 sensor (640 × 480 pixels resolution). We have implemented ViPR in a variety of computing platforms, ranging from Pentium-based systems to embedded processors, like the PMC Sierra family of processors, to DSP chips, like the 200-MHz TI TMS320C6205. A complete ViPR system composed of a custom camera, a TI DSP, and a FPGA to interface the camera and the DSP will cost of about US$21 (US$9 + US$5

+ US$7) in large volumes. We have also implemented vSLAM in two different platforms: a Pentium-based system and a custom-embedded board composed of a 400-MHz PMC Sierra Processor and a TI DSP that handled the image-filtering blocks of ViPR. The embedded implementation of vSLAM was developed for the robotic vacuum cleaner market, targeting vacuum devices that would cost about US$500. The embedded board also includes a module for motor control of the robot for about US$100 in large volumes. The embedded implementation of vSLAM provides all the computing power needed for the vacuum cleaner at a reasonable percentage of the estimated retail price of the product. Therefore, we have a suite of implementations of ViPR that would enable robotic applications, ranging from the ones that already have a sizable computing platform, such as AIBO, to the ones that have a minimal computing platform (or none), such as simple robotic toys or robotic vacuum cleaners.

On the industrial automation side, we have fully productized LaneHawk. The system consists of a camera, a lighting system, and a complete computer, and it sells for about US$3,000. Losing about US$20 per lane per day in a typical store (10–15 lanes) represents US$50,000 of annual lost revenue. A LaneHawk installation would pay for itself in less than 12 months, assuming a conservative BOB detection rate of 50%.

Conclusions

ViPR is a basic primitive for robotics systems. It has been used for human-robotic interaction, localization and mapping, navigation and self-charging, and manipulation. ViPR is the first algorithm of its kind that performs robustly and reliably using low-cost hardware. ViPR could be integrated as an add-on hardware component in its DSP implementation or as a pure software component, giving robotics developers a variety of possibilities and great flexibility for designing their applications. Therefore, we postulate that ViPR could potentially be a very valuable component for the robotics industry.

Keywords

Visual pattern recognition, SIFT, vSLAM, SLAM, vision-based robotics, vision-based automation.

References

[1] F. Fang and S. He, “Cortical responses to invisible objects in the human dorsal and ventral pathways,” *Nature Neurosci.*, vol. 8, no. 10, pp. 1380–1385, 2005.

[2] D.J. Felleman and D.C. Van Essen, “Distributed hierarchical processing in the primate cerebral cortex,” *Cerebral Cortex*, vol. 1, no. 1, pp. 1–47, 1991.

[3] L. Goncalves, E. Di Bernardo, D. Benson, M. Svedman, J. Ostrowski, N. Karlsson, and P. Pirjanian, “A visual front end for simultaneous localization and mapping,” in *Proc. Int. Conf. Robotics Automation (ICRA)*, 2005, pp. 44–49.

[4] N. Karlsson, E. Di Bernardo, J. Ostrowski, L. Goncalves, P. Pirjanian, and M. Munich, “The vSLAM algorithm for robust localization and mapping,” in *Proc. Int. Conf. Robotics Automation (ICRA)*, 2005, pp. 24–29.

[5] D. Lowe, “Local feature view clustering for 3-D object recognition,” in *Proc. 2001 IEEE Conf. Computer Vision Pattern Recognition*, vol. 1, 2001, pp. 682.

[6] D. Lowe, “Distinctive image features from scale-invariant keypoints,” *Int. J. Comput. Vis.*, vol. 60, no. 2, pp. 91–110, 2004.

[7] K. Matsukuma, H. Handa, and K. Yokoyama, “Vision-based manipulation system for autonomous mobile robot ‘smartpal’,” in *Proc. Japan Robot Association Conf.*, Yaskawa Electric Corporation, Sept. 2004.

[8] A.J.N. van Breemen, icat community Web site [Online]. Available: http://www.hitech-projects.com/icat

[9] A.J.N. van Breemen, “Animation engine for believable interactive user-interface robots,” in *Proc. Int. Conf. Intelligent Robots Systems (IROS)*, vol. 3, pp. 2873–2878, 2004.

Mario E. Munich is vice president of engineering and principal scientist at Evolution Robotics. He holds a Ph.D. in electrical engineering from the California Institute of Technology.

Paolo Pirjanian is president and chief technology officer at Evolution Robotics. He holds a Ph.D. from the University of Aalborg, Denmark.

Enrico Di Bernardo is a senior research scientist at Evolution Robotics. He holds a Ph.D. in electrical engineering from the University of Padova, Italy.

Luis Goncalves is vice president of resesarch and development at Evolution Robotics Retail. He holds a Ph.D. in computational and neural systems from the California Institute of Technology.

Niklas Karlsson is a senior research scientist at Evolution Robotics. He holds a Ph.D. in control systems from the University of California, Santa Barbara.

David Lowe is a professor in the Computer Science Department of the University of British Columbia. He is a Fellow of the Canadian Institute for Advanced Research and a member of the Scientific Advisory Board of Evolution Robotics.

Address for Correspondence: Mario E. Munich, 130 W. Union St., Pasadena, CA 91103 USA. Phone: +1 626 535 2871. E-mail: mario@evolution.com.