ROBOTICS AND MACHINE PERCEPTION

NEWSLETTER NOW AVAILABLE ON-LINE

Technical Group members are being offered the option of receiving the Robotics and Machine Perception Newsletter electronically. An e-mail is being sent to all group members with advice of the web location for this issue, and asking members to choose between the electronic and printed version for future issues. If you are a member and have not yet received this message, then SPIE does not have your correct e-mail address.

To receive future issues electronically, please send your e-mail address to: `spie-membership@spie.org` with the word `Robotics` in the subject line of the message and the words `electronic version` in the body of the message.

If you prefer to receive the newsletter in the printed format, but want to send your correct e-mail address for our database, include the words `print version preferred` in the body of your message.

Where to look next: gaze control for humanoid walking

Autonomous navigation of a walking robot requires close coordination of perception and locomotion, i.e. perception must be controlled to assure that the humanoid is always provided the largest possible amount of task-relevant information about the walking scenario. Though perception-guided robot navigation has been an active research topic for decades, navigation strategies reported were primarily designed for wheeled robots. This was due to the unsatisfactory state of legged robots at the time. In the meantime, major problems of mechanical design and stabilization of humanoids have been solved and aspects of perception-based biped locomotion have recently attracted interest. Obviously, a key issue in attaining autonomous locomotion is the development of new navigation strategies, or the adaptation of existing ones to the peculiar properties of legged robots.

For a humanoid to execute a given locomotion task, the guidance system must generate appropriate step sequences allowing the robot to navigate and reach the specified goal position in spite of obstacles in the walking trail (see Figure 1). Information about these obstacles is provided by an active vision system including a gaze control scheme. This selects the orientation of the sensor so that the relevant objects of the scene are kept within the sensor's limited field of view. Here, we propose a modular task-oriented and situation-dependent gaze control architecture (see Figure 2),
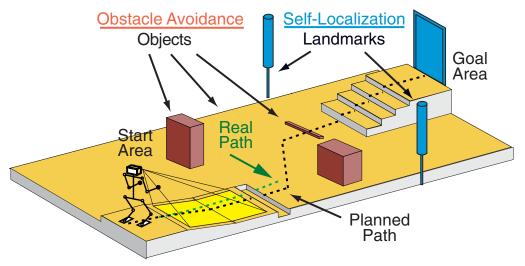
Figure 1. Humanoid navigation separated into self-localization and obstacle-avoidance tasks. The biped robot is shown in a prototypical scenario with objects and landmarks.

Editorial

Integrative perspectives: the cognitive robot, the cognitive environment and the cognitive brain

Welcome to this issue of the R&MP newsletter. When preparing these editorials, I like to read through the articles for the issue at my leisure noting phrases or keywords that highlight key aspects of each. When done I then look for patterns that relate the articles to each other and to the wider robotics and machine perception enterprise. The current issue illustrates this process very nicely.

On first scan there are seven articles on very different topics. On a second scan one can see some similarities. For example, the articles by Tian, Li & Chen, and Seara & Schmidt, all work with computer vision; the first for the recognition of facial expressions; the second for robot navigation, specifically for a walking robot; the third for building a model of an object or scene. The paper by Mundhenk et al could be included as it incorporates visual analysis for identifying salient visual features in a room populated with cameras. The articles by Haidacher & Hirzinger and Okamura & Abbott could be collocated under the topic of manipulation; the first is concerned with techniques for guiding a human operator during a minimally invasive surgical operation while the

second is concerned with retrieving the pose of an object held in a gripper. The article by Lewis et al is much more 'technology', focusing on devices that incorporate visual intelligence.

If we look a little harder, focusing on integrative perspectives rather than comparative features, an interesting pattern emerges. The articles by Seara & Schmidt and Li & Chen incorporate different perspectives on the world. In the first the perceiver, the 'walking' robot, is moving through an object, namely the environment, whereas in the second the perceiver is circling the object; the relative position of the perceiver and object are transposed. In the paper by Mundhenk et al, in contrast, the perceiver is embedded within the object, the room, with many simultaneous viewpoints. The three papers collectively contrast the cognitive robot with the cognitive environment; the intelligent robot versus embedded robotic intelligence.

Contrasting perspectives can also be founding in the article by Haidacher & Hirzinger, where a grasped object can be viewed from the perspective of the robot's

fingers or the physical object; the two perspectives are combined to provide the object's posture relative to the hand. Okamura & Abbot include a perspective that is one removed from this, namely of the arm that carries a grasped object, but in support of the human operator; or rather, the cognitive brain. The ability to understand facial expressions, the topic of Tian's paper, is a basic capability that benefits humans, and will surely benefit the cognitive robot and environment as they attempt to understand and interact with humans. Finally, the article by Lewis et al rather nicely offers technology that can support the cognitive robot, the cognitive environment, and the cognitive brain.

It is rewarding to see and discover patterns. I recommend that you read through all of the articles and see for yourself what patterns emerge. The effort will help to broaden and deepen your understanding of robotics and machine perception.

Dr Gerard McKee
Technical Group Chair
The University of Reading, UK.

Where to look next: gaze control for humanoid walking

Continued from cover.

composed of three major modules: (I) *Information Management*, (II) *Task-Specific Gaze Evaluation*, and (III) *Decision Scheme*.[1,2]

Our investigations consider an information-management scheme, shown in Figure 2(I), that registers and administers the uncertainties associated with the different processes involved in perception-based locomotion.[1] A coupled, hybrid EKF (computer module) is employed for information/uncertainty management. Based on this scheme, a description of the accumulated information, combined with the uncertainties in the measurements, provides a basis for the task-oriented gathering of the relevant information, i.e. task-specific gaze evaluation (see Figure 2(II)).

The proposed approach towards information quantification is based on Shannon's information theory, which states that information is a measure of the decrement of uncertainty. Since information must be redefined in the context of biped walking, the information content of a view situation is herein defined as a measure of the degree to which perception under these terms is capable of reducing relevant task-dependent uncertainties. For the formal description of which uncertainties are related to a specific task, the term *incertitude* is proposed. High-precision estimates are not required throughout the locomotion process, i.e. not all uncertainties have to be minimized simultaneously (even if that would be desirable if possible). Depending on the task, some uncertainties may become critical and others irrelevant. The term incertitude indicates which uncertainties, i.e. estimates and measurements, are directly involved in the current task and must therefore be considered as relevant or critical.

With the concepts defined here, we have developed a biologically-inspired approach to predictive gaze control for an active vision system based on the maximization of the predicted visual information. This general approach takes into account the accumulated available information about both the scene and current task to predict the optimal pose of the visual sensor for a future view situation. An optimal sensor pose is found by selecting the state of maximum predicted information content among a set of predicted states.

One of the main goals while navigating is to follow a planned path and arrive at a desired location without getting lost. This means that the uncertainties related to self-localization must be kept small. These describe the position of the robot in the world. In order to achieve a precise knowledge about the robot's location, the perception system must focus on objects whose position in the world is precisely known, i.e. landmarks. The second main goal of safe
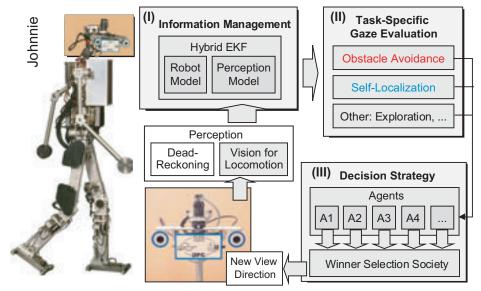

Figure 2. Shown is the humanoid Johnnie (Institute of Applied Mechanics, Technische Universität München) and the modular gaze-control scheme for vision-based humanoid navigation.

locomotion is avoiding collisions with obstacles. Given a local path through a scenario with different objects, there exist several points on this path with a high risk of collision with respect to a certain obstacle. The goal of the gaze controller is to minimize the risk of collision in these critical points.

The next step in fulfilling the requirements of autonomous locomotion is finding some means to combine the two goals. This is achieved by defining an optimal decision strategy, shown in Figure 2(III), based here on utility theory—i.e. where to look next—in order to optimally fulfill both navigation tasks depending on the current situation. The decision maker is facing an action/selection dilemma (here, the selection of view direction) with two different types of objective. The analysis of the action/selection problem showed clearly that one agent—the use of only one utility function—that could be optimal in a specific situation, could actually lead to risky states in another scenario. To safeguard the flexibility of the decision maker, the principal objective is to find some kind of suboptimal behavior that could assure a safe and satisfactory view direction choice in a wide range of situations and scenarios. For the decision process, a winner-selection society is proposed.

The developed gaze-control scheme has been validated by its integration into a guidance system for visually-guided biped robots. This system comprises, in addition to the gaze controller, appropriate scene analysis algorithms and a step-sequence planning module that trans-

forms the gathered information into a step sequence guaranteeing the safe locomotion of the humanoid. The performance of the proposed guidance architecture has been demonstrated in several experimental campaigns with different locomotion platforms. Together with this guidance system, the humanoid robot Johnnie (shown in Figure 2), was able to perform vision-guided navigation: following a pre-planned path including curves, avoiding obstacles by walking around or stepping over them, finding and walking towards a goal position, and even climbing stairs.

J. F. Seara and G. Schmidt
Institute of Automatic Control Engineering
ISAC Group
Technische Universität München
Munich, Germany
E-mail: javier.fernandez.seara@ei.tum.de

References:
1. J. F. Seara, K. H. Strobl, E. Martin, and G. Schmidt, *Task-Oriented and Situation-Dependent Gaze Control for Vision Guided Autonomous Walking,* **Proc. of the 3rd IEEE-RAS Int'l Conf. on Humanoid Robots (Humanoids),** Munich and Karlsruhe, Germany, October 2003.
2. J. F. Seara, K. H. Strobl, and G. Schmidt, *Path-Dependent Gaze Control for Obstacle Avoidance in Vision Guided Humanoid Walking,* **Proc. of the IEEE Int'l Conf. on Robotics and Automation (ICRA),** pp. 887-892, Taipei, Taiwan, September 2003.

Automatic facial-expression analysis systems

Facial expression is one of the most powerful, natural, and immediate means for human beings to communicate their emotions and intentions. An automatic facial-expression analysis (AFEA) system includes both measurement of facial motion and recognition of expression. The general approach to such systems consists of three steps (see Figure 1): face acquisition, facial-feature extraction and representation, and facial-expression recognition. We will discuss these in the following sections.

Figure 1. Basic structure of a facial-expression analysis system.

Face acquisition

This processing stage involves automatically finding the face region for the input images or sequences. It can be designed to detect a face in each frame or just in the first frame and then track the face through the remainder of the video sequence. Head finding, head tracking, and pose estimation can all be applied to a facial-expression analysis system in order to handle large head motion.

We developed a 2D-image-based method of detecting head position and pose. The head detection uses the smoothed silhouette of the foreground object as segmented using background subtraction and computing the negative curvature minima points of the silhouette. After the head has been located, the head image is converted to grayscale, histogram equalized, and resized to the estimated resolution. Then, a three-layer neural network is employed to estimate the head pose. Currently our system outputs three head poses: frontal or near frontal view, side view or profile, and others (such as back of the head or occluded face).

Facial feature extraction and representation

After the face has been located, the next step is to extract and represent the facial changes caused by facial expressions. In facial feature extraction for expression analysis, there are mainly two types of approaches: geometric-feature- and appearance-based methods. The geometric facial features present the shape and locations of facial components (including mouth, eyes, brows, nose etc.). The facial components or facial feature points are extracted to form a feature vector that represents the face geometry. In appearance-based methods, im-
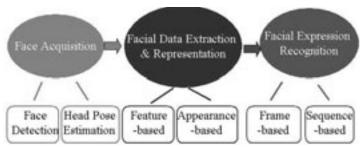

Figure 2. An example of head detection, feature extraction, and expression recognition.

age filters (such as Gabor wavelets) are applied to either the whole face or specific regions in a face image to extract a feature vector. Depending on the different facial-feature extraction methods, the effects of in-plane head rotation and different scales of the faces can be eliminated, either by face normalization before the feature extraction or by feature representation before the step of expression recognition.

In our system, two types of facial features—location and shape—are extracted only for the face in the frontal or near-frontal view. In total, six location features are extracted for expression analysis. They are eye centers, eyebrow inner endpoints, and corners of the mouth. After extracting these, the face can be normalized to a canonical face size based on two of the features, i.e., the eye-separation after the line connecting two eyes (eye-line) is rotated to horizontal. The extracted features are transformed into five parameters for expression recognition. These parameters are the distances between the eye-line and the corners of the mouth, the distances between the eye-line and the inner eyebrows, and the width of the mouth (the distance between two corners of the mouth). Another type of distinguishing feature is the shape of the mouth. In order to extract the mouth-shape features, an edge detector is applied to the normalized face to get an edge map. This is divided into 3×3 zones. The size of the zones is selected to be half of the distance between the eyes. The mouth-shape features are computed from zonal shape histograms of the edges in the mouth region, and are represented as a feature vector of 12 components.

Facial expression recognition

Facial expression recognition is the last stage of AFEA systems to identify facial changes as facial action coding system action units[1] or prototypical emotional expressions. We used a neural-network-based recognizer. The inputs to the network were the five location features and the 12-zone components of shape features of the mouth. The outputs were a set of prototypical emotional expressions: neutral, smile, angry, surprise, fear, sad, and disgust. Figure 2 shows an example of the detected heads, extracted facial features, and recognized expressions[2] on ICVS-PETS (IEEE workshops on Performance Evaluation of Tracking and Surveillance) datasets which were provided by FGnet project.

Ying-Li Tian
Exploratory Computer Vision Group
IBM T. J. Watson Research Center
Yorktown Heights, USA
E-mail: yltian@us.ibm.com

References
1. P. Ekman and W. Friesen. **The Facial Action Coding System: A Technique For The Measurement of Facial Movement,** Consulting Psychologists Press, Inc., San Francisco, CA, 1978.
2. Y-L. Tian, L. Brown, A. Hampapur, S. Pankanti, A. W. Senior, and R. M. Bolle, *Real World Real-time Automatic Recognition of Facial Expressions.* **IEEE Workshop on Performance Evaluation of Tracking and Surveillance,** Graz, Austria, 31 March 2003.

An intelligent room based on the operation of the brain

Understanding of the human brain has recently lead to exciting advances in computer vision. For instance, knowing how neurons in visual cortex extract features from a visual scene has allowed advances in scene analysis[1] and face recognition/tracking.[2] This comes as no surprise since the challenges of scene understanding, visual analysis, and action creation in the real world have been efficiently addressed by the human brain. The primary difficulty has been the limitations on our understanding, but recent advances in brain imaging, micro-cellular recording, and neurochemistry have pushed these back.

Our aim has been to exploit these new insights as they become available for real-world applications. Our current goal is the realization of a smart room that can use multiple cameras and distributed computation in a way that is behaviorally flexible and robust in the same way that biological organisms are. For instance, by exploiting biological models of saliency (which guides visual attention) we can limit the scope of visual search such that performance is not strictly bound to any specific domain: this is because we do not create hard constraints or heuristics. As such, saliency can be applied to outdoor, room, and other scenes with different content and arrangement of objects without the need for *a priori* constraints. That is, we can limit the amount of input analyzed in a general biological manner, which allows us to apply more sophisticated algorithms in post processing. The visual saliency program can then produce sets of visual features. This is because, while analyzing the scene, it has to find what the image features are in order to determine saliency. This then allows for fast rudimentary object identification.[3]

The saliency code itself is run, load-balanced, on several Beowulf nodes (Figure 1). Since only a few processes are CPU intensive, this allows us to use the nodes for additional purposes.[4] For instance, cameras can be hooked up to individual nodes, which then act as separate agents: each CPU then acts as both an in-

Figure 1. Shown is a basic schematic of the smart room, which we currently call the iRoom. Six robotic cameras work together to gather visual input and feed it into a Beowulf cluster for image processing. Each camera is connected to its controlling CPU by firewire for video and category 5 cable that carries servo control and power. High speed ethernet bonding connects the CPUs, which need not even be in the same room to work together.
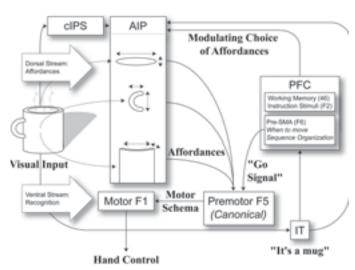

Figure 2. The figure (modified from Reference 5) illustrates current understanding about how the human brain processes visual information to create action. A dorsal stream (top of the brain) extracts features of objects using cIPS (caudal Intraparietal Sulcus) and AIP (Anterior-intra Parietal area), while a ventral stream (bottom of the brain) performs object identification through IT (Inferotemporal cortex). These centers then act reciprocally with PFC (Prefrontal Cortex) to modulate scene understanding and action selection.

dividual, and as part of a whole in analyzing a room scene. For instance, each node has a notion of what it wants to track, but it gains information about what it wants to track by distributing the workload of visual processing across the cluster.

The future goals of our project revolve

around understanding how to extend the capabilities of our intelligent room using brain-operating principles (BOPs) involving perception, action, and recognition. For instance the human brain performs the act of recognition in many different stages, some of which seem to run in parallel. For example, recognition of features for object affordances (which might tell us how to grasp an object with our hands) seems to run in parallel with recognition of the object itself (Figure 2). However, the act of visual saliency that comes before these processes seems to perform the task of reducing a visual scene to a smaller area of attention, thus, saliency seems to have the property of serializing some visual data. In general, we believe such knowledge, along with other data, should give us insight into practical solutions for vision, action, and scene understanding. We might, for instance, create two parallel processes for vision with one based upon feature-affordance extraction and another upon recognition. However, the input to such systems may only come from a single saliency pipeline.

Additionally, it is not only important to recognize objects, but to react to and understand them. For instance, not only might I see a man, but also I might see a man at the photocopier. Is this person supposed to be there, and how do I in fact know that a man is standing at the photocopier? Humans perform such complex analyses by taking in object and agent (the person) identification from temporal cortex areas of the brain. Identification is then combined with understanding about how the world works—in the parietal areas of the brain—and context in the hippocampal area. Pre-frontal cortex and surrounding areas help us to then combine world information into a storyboard and make both conscious and sub-conscious decisions about how to react to our world. So, for instance, building a system that can react to complex scenes may be aided by our knowledge of how the pre-frontal cortex interacts with other parts of the brain to collect information about the world and how it reciprocates by initializing

Active visual sensing for robotic perception planning

A robot often needs to perceive its environment to obtain models of the objects around it. Examples include robotic grasping/manipulation, object recognition, inspection, and construction of a model of the environment. To reconstruct a complete and accurate 3D model of an object, two fundamental issues must be addressed. The first is how to acquire the 3D data accurately and efficiently enough to reconstruct the object surfaces. Currently, laser range finders/scanners are widely used as active visual sensors for such data acquisition in industrial applications. However, due to the mechanical scanning involved, the acquisition speed is limited. To increase efficiency, we have developed an active vision system based on pattern projections.[1,2]

The second issue to be addressed is how to determine viewpoints,[3] with their corresponding sensing parameters for each visual sensing action such that all the needed information about the object surface can be acquired in an optimal way. This kind of robotic perception planning can be used both known environments (inspection[4]) and unknown environments (modeling).

An active vision system

Our active vision system consists of a light-pattern projector and a CCD camera, as illustrated in Figure 1. A projector is used to illuminate the scene to be measured with a pattern of light, e.g. a number of stripe planes. Each of the stripes intersecting with the scene produces a deformed curve of its light plane in 3D space, the pattern of which is detected via a camera. The scene can then be reconstructed in the image processing. The pattern projector is assumed to be pre-calibrated offline, whereas both the intrinsic and extrinsic parameters of the camera can be changed or even totally unknown at the beginning of a reconstruction task. This allows un-calibrated reconstruction of the 3D models and makes the vision system self-adapt to the environment in which it must work.

Although the pre-calibrated projector may seems to be a limitation, it is acceptable from an engineering point of view, in that such calibration may be expected. In many practical robotic applications, it is the camera and its relative pose that needs to be adjusted online most often: assuming a calibrated projector allows more such camera parameters to be changed and calibrated as needed.
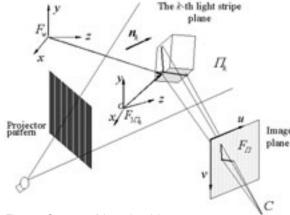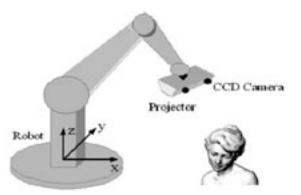

Figure 1. Structure of the active vision system.

Figure 2. A robot equipped with an active vision sensor.

Perception planning in known environments

Due to the limited field of view of most cameras, a vision sensor can only sample a portion of an object from a single viewpoint (see Figure 2). As a result, multiple images need to be taken from different vantage points and integrated to enable all features of interest to be sensed by the robot. It is thus of critical importance to determine the viewpoints with its viewing pose, as well as the corresponding visual-sensing parameters, to achieve full automation and high efficiency in a robotic perception task. In many practical applications—robotic assembly, inspection, object recognition—the object's geometry and a rough estimate of its pose are available. This is a robotic perception-planning problem in known environments.

When multiple features need to be observed and multiple viewpoints planned, the minimum number of viewpoints needs to be determined. To achieve high efficiency and quality, the optimal spatial distribution of the viewpoints should be determined too. In our research, we developed a method for planning model-based perception tasks, with optimal viewpoint dis-
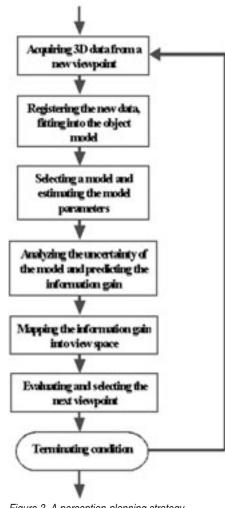

Figure 3. A perception-planning strategy.

tribution, sensing parameters, and sensing sequence. The procedures in a typical perception-planning task include the following: inputting the object's geometric information from a model database; specifying the vision tasks; generating a plan with the fewest viewpoints; searching for a shortest path for robot execution; and outputting the sensing plan with the viewing pose and corresponding sensing configuration at each view. Each viewpoint should satisfy multiple constraints due to the physical and optical properties of the sensor, scene occlusion, and robot reachability in the environment. Our method provides a stable and complete solution for model-based perception tasks, including viewpoint decision, constraint satisfaction, optimization of viewpoint distribution, planning of robot operation sequence. All these techniques are integrated into the software we

Position sensors for robotic-hand grasping

For many years now, robots have been used to perform tedious work in industrial environments: they do this more precisely than we could, and without fatigue. Consequently, in the last few years, researchers have striven to develop robots that can also perform jobs outside the well-structured, predictable industrial environment, e.g. as domestic assistants. In contrast to situations in manufacturing environments, it is not desirable to adjust the home to the particular technical needs of the robot. Also, such a machine—in order to be useful—would have to adapt to each new job it has to perform. Since its working environment was designed for humans, one obvious approach to the design of a robot assistant is to mimic both human anatomy and skills.

At the Institute of Robotics and Mechatronics at the German Aerospace Center (DLR) three generations of lightweight arms and two generations of dexterous robotic hands have been built (see Figure 1). While developing these anthropomorphic robotic devices, size and weight had to be traded-off against the machine's ability to feel and explore the environment using numerous different sensors. Due to size constraints, this was especially true for the robotic hand. As a result, during this process, we examined several different approaches to substituting information from some sensors using the advanced evaluation of other.

In particular, when delicately grasping and manipulating an object with several fingers of a robotic hand, we require the location of those points at which the individual fingers touch the object. There are two general ways of describing these points. First, from the perspective of the hand, when the position of some reference point on each finger is known from e.g. joint-angle measurements, the contact point can be described as a two-dimensional coordinate on the surface of the finger. For a spherical fingertip this could be, for example, longitude and latitude. Second—from the perspective of the object—when the position of a given reference point at the object is known, again the contact points can be described as two-dimensional coordinates, this time on the surface of the object. In the first case, the coordinates could be measured using tactile sensors.[1] In the latter case, an external camera could be used. In both, however, to save space and effort, algorithms were developed to substitute new sensors with the intelligent use of existing sensors.

The contact points of the fingers can be computed from their constrained motion when they cooperatively grasp and move an object:[2] the fingers have to maintain stable contact and are hence restricted in their mobility. For example, a spherical fingertip can only roll along a planar surface in two directions in the presence of sufficient friction. A sliding motion would break this contact, in a similar way to a lift off the surface or a twist around the surface normal. Mathematical models of contact have been established and can be used to describe these constraints.[3] Also, the motion itself depends on the location of the contact point. A hand with several fingers makes contact at several locations simultaneously. The motion of all these can be detected using the position sensors usually available in robotic hands. Observing several directions of motion constitutes an overdetermined system of equations for the object velocity. Using a least-squares approach allows those contact points at the fingers to be found that best match the measurements of motion and the constraints of mobility through contact.

The determination of the location of the contact points on the object is related to the computation of the position and orientation of the object itself. In a multi-fingered grasp, the locations of contact from the perspective of the fingers can be computed using the algorithm described above. The geometry of the surface of a fingertip is usually known from design. Hence, with knowledge of the contact points on the surface of the fingertip, the direction normal to the surface at this particular point can also be determined. Position and normal of the point of contact can be compared to a geometrical model of the surface of an object. Although numerous possibilities to describe the surface of an object are available, an approximation with polygons is most suitable in this case: polygonal descriptions are widely used, easy to obtain, can describe any kind of object, do not need to include unimportant parts of the object, and best describe what can be detected by contact (namely position and normal of that particular polygon). A complete comparison, however, would be of combinatorial order. By

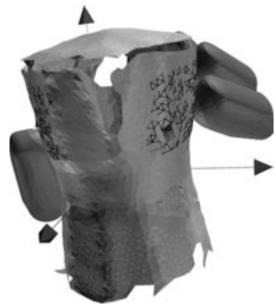*Figure 1. The DLR Hand II and Light-Weight Arm III playing the piano.*

Figure 2. Detection of the location of a grasped object.

Neuromorphic color processing

Why color?

Color carries useful information about the world around us. In a natural environment it might help us pick out danger—as in the case of a colorful but poisonous snake or insect—or signal a source of food such as berries or fruit in a tree. Red, the color of blood, raises one's blood pressure and may lead to a heightened sense of awareness. This idea is not lost on the designers of traffic signs. This mechanism, which evolved for use in natural situations, has been capitalized on by marketers. They have found out by intuition and empirical testing that colorful objects can alter a consumer's preference to buy certain items, and can also assist in product branding and the establishment of a product identity.
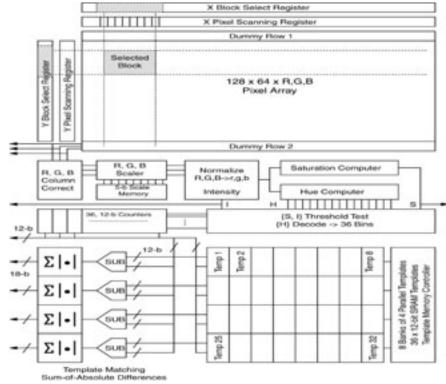
Color is used to identify players in sports teams or citizens of a country through their national flag. It is used in dress to appear attractive, in homes to lead to a more aesthetically pleasing environment. Color can be used to find an item of clothing heaped together with others in a pile. Hyper color sensitivity, i.e. the ability to identify more than the normal three spectral bands (as in the visual system of shrimp), can be used to break camouflage. In short, the color content of the environment is an important quality that is separate from shape and form. From the perspective of the brain, color processing is performed by specialized cells and regions. The ability to sense, perceive, and use color information is important in a wide range of the activities performed by intelligent beings.

Understanding the environment

The identification of a single color alone in a scene is a non-trivial problem that is not soluable under all lighting conditions. Some researchers have looked at ways of finessing the problem of color identification over varying illumination. One approach is to normalize colors by the aggregate color content of an image. Another is to assume that all light sources illu-

Figure 1. Overview of the ColorCam chip.

minating a surface are black-body radiators. This can lead to certain assumptions about form of the spectral distribution of illumination. A trivial and robust approach is simply to provide the ambient illumination as well.

Using this simple idea, Iguana Robotics has developed a device called the *ColorStick*. About the size of a fat pen or small flashlight, it provides multi-spectral illumination to a surface. Interestingly, the device classifies the surface into color, or even color families, and reports the color via sound (a voice shouting "Green!"), a light screen or even, potentially, a haptic display for the blind. This device will find use as a toy, for children learning color. It will also find a use for the hundreds of millions blind, colorblind, and aesthetically-impaired people in the world who would like to dress themselves in a color-coordinated way. It can also be used by home decorators to match colors in their home versus fabrics they may purchase on-line.

We are currently developing a stand-along chip that will implement the sensing and processing in a single, neuromorphic device. Combining colors and counting the relative ratio of the color content of items adds a new dimension to vision sensing, and color-histogram-based object recognition is a well-known tech-

nique for identifying objects in the environment. Our stand-alone chip integrates color sensing, color-space transformation, histo-gramming, and histogram matching in a single device. This device runs at better frame rates, uses much less power, is considerably smaller, and is far less expensive than techniques involving conventional CMOS images and digital signal processing. We call this chip the *ColorCam*.[1]

ColorCam overview

An overview of the first generation chip is shown in Figure 1. As can be seen, the first generation device uses a 128×64 pixel imager. RGB pixel information flowing from the sensor is normalized and then the hue and saturation content of the pixel are found. Hue-angle computation requires a division. This was accomplished using a circuit with a look-up table.

If the pixel saturation exceeds a preset threshold, the hue value of the pixel is counted. If it does not reach the given threshold it is thrown away. The information is then accumulated in 36 counters that divide the hue space, uniformly or non-uniformly, into 36 regions. Once accumulation of pixel information is achieved over a region of interest, a sum-of-absolute-differences circuit finds a minimum match over a set of stored templates.

Currently, we are working on a 3rd generation device that will address, on chip, problems of scaling, light-source color changes, and other issues. We anticipate sampling the device by the end of 2004.

M. Anthony Lewis, Ralph Etienne-Cummings, and Philiand Pouliquen
Iguana Robotics, Inc. Urbana, IL
E-mail: tlewis@iguana-robotics.com
http://www.iguana-robotics.com/

References
1. R. Etienne-Cummings, P. Pouliquen, and M.A. Lewis, *A Vision Chip for Color Segmentation and Object Recognition,* **EURASIP J. Appl. Signal Processing 7,** pp. 703-712, 2003.

An intelligent room based on the operation of the brain

actions.

It is thus a challenge to understand the brain in great enough detail to apply what has been discussed to real world tasks. However, given the success by us and others at applying lessons learned from the brain to computational systems, we believe that this is an entirely fruitful approach. Additionally, thought the brain is a highly complex system, it is the only working model we have that seems to solve the problems of action and perception.

T. Nathan Mundhenk, Laurent Itti, and Michael A. Arbib
Computer Science Department
University of Southern California
Los Angeles, California, USA
E-mail: Mundhenk@usc.edu
http://ilab.usc.edu

References
1. L. Itti and C. Koch, *Computational Modeling of Visual Attention,* **Nature Reviews Neuroscience 2** (3), pp. 194-203, 2001.
2. J. Triesch and C. von der Malsburg, *Democratic integration: self-organized integration of adaptive cues,* **Neural Computation 13** (9), pp. 2049-2074, 2001.
3. T. N. Mundhenk, V. Navalpakkam, H. Makaliwe, S. Vasudevan, and L. Itti, *Biologically inspired feature based categorization of objects,* **Proc. SPIE Conf. on Human Vision and Electronic Imaging IX,** pp. 5292-29, San Jose, CA, January 2004.
4. T. N. Mundhenk, N. Dhavale, S. Marmol, E. Calleja, V. Navalpakkam, K. Bellman, C. Landauer, M. A. Arbib, and L. Itti, *Utilization and viability of biologically-inspired algorithms in a dynamic multi-agent camera surveillance system,* **Proc. SPIE 5267,** p. 281, 2003.
5. A. H. Fagg and M. A. Arbib, *Modeling parietal-premotor interactions in primate control of grasping,* **Neural Networks 11,** pp. 1277-1303, 1998.

Active visual sensing for robotic perception planning

developed: the *Viewpoint Planner*.

Perception planning in unknown environments

For unknown environments, since the number of viewpoints and their viewing pose cannot be determined prior to data acquisition, the perception task normally consists of iterative cycles of viewpoint planning, data acquisition, registration, and view integration. In this research, we developed an information-entropy-based viewpoint-planning method for the perception-planning task. An information criterion is developed for selecting the model structure. Based on the selected model, we use information entropy as the uncertainty measure of the model, and analyze the uncertainty to predict the information gain for new viewpoints to be taken. As a result, we can obtain the prediction of the information gain about the object.

The information gain is then mapped to the view space. The view that has the maximal information gain about the object is then selected as the *next best viewpoint*. The viewpoint planning procedure is illustrated in Figure 3. The implementation of the method is based on the same setup as shown in Figure 2. The perception process consists of a sequence of four repeated steps: acquiring data of the object surface from a viewpoint, registering the acquired data, integrating the new data with the partial model, and determining the next viewpoint. At each view, we choose a new viewpoint that has

maximum information gain. Then the robot moves to the new viewpoint and takes another measurement to update the object model until the terminating condition is met (determined by a criterion from the information gain) and a complete model is obtained.

Y. F. Li and H. Y. Chen
Department of Manufacturing Engineering and Engineering Management
City University of Hong Kong, China
E-mail: meyfli@cityu.edu.hk
http://personal.cityu.edu.hk/~meyfli/

References
1. Y. F. Li and S. Y. Chen, *Automatic Recalibration of an Active Structured Light Vision System,* **IEEE Trans. on Robotics and Automation 19** (2), pp. 259-268, April 2003.
2. Y. F. Li and R. S. Lu, *Uncalibrated Euclidean 3D Reconstruction Using an Active Vision System,* **IEEE Trans. on Robotics and Automation 20** (1), pp. 15-25, Feb. 2004.
3. S. Abrams, P. K. Allen, and K. Tarabanis, *Computing camera viewpoints in an active robot work cell,* **Int'l J. of Robotics Research 18** (3), pp. 267-288, March 1999.
4. S. Chen and Y. F. Li, *Automatic Sensor Placement for Model-Based Robot Vision,* **IEEE Trans. on System Man and Cybernetics, Part B, 34** (1), pp. 393-408, February 2004.

Virtual fixtures for telemanipulation: control and application

tual wall stiffness is given the simple function: $k \leq \min(2b/T, f_c/D)$.

Application to MIS

There are several important considerations for practical application of force feedback and virtual fixtures in robot-assisted minimally invasive surgery, all of which are subjects on ongoing work.

1. Degrees of freedom of force feedback
It is impractical to expect that robotic MIS tools can sense forces and torques in all degrees of freedom, especially when the tools are disposable.

2. Appropriate virtual-fixture geometry
The geometry can be selected by the surgeon, through computer vision recognition and/or modeling of tissue deformation.

3. Tuning of virtual fixture assistance
We are exploring methods for automatic tuning of virtual-fixture strength based on hidden-Markov-model recognition of operator motions.

4. Uncertainty in robot position relative to anatomical structures due to unmodeled dynamics
Robot design, modeling and control approaches are required to place accurate virtual fixtures.

Allison M. Okamura and Jake J. Abbott
Department of Mechanical Engineering
Engineering Research Center for Computer-Integrated Surgical Systems and Technologies
The Johns Hopkins University
Baltimore, USA
E-mail: aokamura@jhu.edu
http://haptics.me.juh.edu

References
1. B. Hannaford, *A Design Framework for Teleoperators with Kinesthetic Feedback,* **IEEE Trans. Robotics and Automation 5** (4), pp. 426-434, 1989.
2. A. Bettini, P. Marayong, S. Lang, A. M. Okamura, and G. D. Hager, *Vision Assisted Control for Manipulation Using Virtual Fixtures,* **IEEE Int'l Trans. Robotics and Automation,** 2004 (in press).
3. J. J. Abbott, G. D. Hager, and A. M. Okamura, *Steady-Hand Teleoperation with Virtual Fixtures,* **12th IEEE Int'l Workshop on Robot and Human Interactive Communication,** pp. 145-151, 2003.
4. J. E. Colgate and G. G. Schenkel, *Passivity of a Class of Sampled-Data Systems: Application to Haptic Interfaces,* **J. Robotic Systems, 14** (1), pp. 37-47, 1997.

Calendar

2004

Eighth International Conference on Cognitive and Neural Systems
19-22 May
Boston, Massachusetts USA
http://www.cns.bu.edu/meetings/

International Conference on Design Computing and Cognition (DCC'04).
19-21 July
Cambridge, MA USA
http://www.arch.usyd.edu.au/kcdc/conferences/dcc04/

ICINCO 2004—International Conference on Informatics in Control, Automation, and Robotics.
25-28 August
Setobal, Portugal
http://www.icinco.org

Brain Inspired Cognitive Systems— BICS2004
29 August-1 September
Scotland, UK
Including:
- First International ICSC Symposium on Cognitive Neuro Science (CNS 2004)
- Second International ICSC Symposium on Biologically Inspired Systems (BIS 2004)
- Third International ICSC Symposium on Neural Computation (NC'2004)

http://www.icsc-naiso.org/conferences/bics2004/program.html
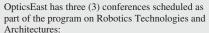
IEEE SMC 2004 International Conference on Systems, Man, and Cybernetics
10-13 October
The Hague, The Netherlands
http://www.ieeesmc2004.tudelft.nl/

OpticsEast
25-28 October
Philadelphia, PA
OpticsEast has three (3) conferences scheduled as part of the program on Robotics Technologies and Architectures:
- Intelligent Robots and Computer Vision XXII: Algorithms, Techniques, and Active Vision
- Mobile Robots XVII
- Sensor Fusion and Decentralized Control in Robotic Systems V

http://www.spie.org/conferences/calls/o4/oe/

For More Information Contact
SPIE • PO Box 10, Bellingham, WA 98227-0010
Tel: +1 360 676 3290 • +1 360 647 1445
Email: spie@spie.org • Web: www.spie.org

Position sensors for robotic-hand grasping

using appropriate pre-selection methods[4] that filter out infeasible combinations of polygons—because their mutual distance does not fit the measurements, for example—the search can be performed quickly. Inherently, the position of the object makes available the contact polygons on the object and hence the location of the contact from the perspective of the object (see Figure 2).

Steffen Haidacher and Gerd Hirzinger
Institute of Robotics and Mechatronics
German Aerospace Center
Wessling, Germany
E-mail: Steffen.Haidacher@dlr.de

References
1. M. H. Lee and H. R. Nicholls, *Tactile Sensing for Mechatronics—a State of the Art Survey,* **Mechatronics 9,** pp. 1-31, 1999.
2. Steffen Haidacher and Gerd Hirzinger, *Contact Point Identification in Multi-Fingered Grasps Exploiting Kinematic Constraints,* **Proc. Int'l Conf. Robotics and Automation,** Washington, DC, USA, May 2002.
3. David J. Montana, *The Kinematics of Multi-Fingered Manipulation,* **IEEE Transactions of Robotics and Automation 11** (4), pp. 491-503, August 1995.
4. Steffen Haidacher and Gerd Hirzinger, *Estimating Finger Contact Location and Object Pose from Contact Measurements in 3-D Grasping,* **Proc. Int'l Conf. Robotics and Automation,** Taipei, Taiwan, September 2003.

Contributing to the Robotics and Machine Perception Newsletter

News, events, and articles

If there are things you'd like to see us cover in the newsletter, please let our Technical Editor, Sunny Bains, know by the deadline indicated on the right. She can be reached at sunny@spie.org. Before submitting an article, please check out our full submission guidelines at:

`http://www.sunnybains.com/newslet.html`

Special issues

Proposals for topical issues are welcomed and should include:

- A brief summary of the proposed topic;
- The reason why it is of current interest;
- A brief resume{acute e!} of the proposed guest editor.

Special issue proposals should be submitted (by the deadline) to Sunny Bains and will be reviewed by the Technical Group Chair.

Upcoming deadlines

30 April 2004: *Special issue proposals.*

7 May 2004: *Ideas for articles you'd like to write (or read).*

July 9 2004: *Calendar items for the 12 months starting September 2004.*

Join the Technical Group

...and receive this newsletter

Membership Application

Please Print   ☐ Prof.   ☐ Dr.   ☐ Mr.   ☐ Miss   ☐ Mrs.   ☐ Ms.

First Name, Middle Initial, Last Name _____

Position _____ SPIE Member Number _____

Business Affiliation _____

Dept./Bldg./Mail Stop/etc. _____

Street Address or P.O. Box _____

City/State _____ Zip/Postal Code _____ Country _____

Telephone _____ Telefax _____

E-mail Address _____

Technical Group Membership fee is $30/year, or $15/year for full SPIE members.

❏   Robotics & Machine Perception
     Total amount enclosed for Technical Group membership $ _____

❏   **Check enclosed.** Payment in U.S. dollars (by draft on a U.S. bank, or international money order) is required. Do not send currency. Transfers from banks must include a copy of the transfer order.

❏   **Charge to my:**   ❏ VISA   ❏ MasterCard   ❏ American Express   ❏ Diners Club   ❏ Discover

Account # _____ Expiration date _____

Signature _____
          (required for credit card orders)

This newsletter is printed as a benefit of the **Robotics & Machine Perception Technical Group.** Membership allows you to communicate and network with colleagues worldwide.

As well as a semi-annual copy of the *Robotics & Machine Perception* newsletter, benefits include SPIE's monthly publication, **oe***magazine,* and a membership directory.

SPIE members are invited to join for the reduced fee of $15. If you are not a member of SPIE, the annual membership fee of $30 will cover all technical group membership services. For complete information and a full application form, contact SPIE.

Send this form (or photocopy) to:
SPIE • P.O. Box 10
Bellingham, WA 98227-0010 USA
Tel: +1 360 676 3290
Fax: +1 360 647 1445
E-mail: spie@spie.org

http://www.spie.org/info/robotics

Please send me

☐ Information about full SPIE membership
☐ Information about other SPIE technical groups
☐ FREE technical publications catalog

Reference Code: 3995

ROBOTICSONLINE

Robotics Web
Discussion Forum

You are invited to participate in SPIE's online discussion forum on Robotics & Machine Perception. To post a message, log in to create a user account. For options see "subscribe to this forum."

You'll find our forums well-designed and easy to use, with many helpful features such as automated email notifications, easy-to-follow 'threads,' and searchability. There is a full FAQ for more details on how to use the forums.

Main link to the new Robotics & Machine Perception forum: http://spie.org/app/forums/tech/

Related questions or suggestions can be sent to forums@spie.org.

SPIE The International Society for Optical Engineering

Virtual fixtures for telemanipulation: control and application in robot-assisted minimally-invasive surgery

Telemanipulation is the direct human control of a robotic manipulator, where the operator and the manipulator are at different locations. Robot-assisted minimally invasive surgery (MIS), in which robotic tools are inserted into the patient via small ports, allows a surgeon increased accuracy, dexterity, and visualization over traditional MIS. In bilateral telemanipulation systems, the human operator manipulates a haptic (force-feedback) interface that sends position commands to a remote robot, and interaction forces between the robot and the environment are displayed to the operator. The performance of a telemanipulation system is typically judged by three criteria: stability, tracking, and transparency.[1] In addition to traditional telemanipulation, we are interested in the application of *virtual fixtures* for operator assistance in MIS tasks. A virtual fixture is a constraint, implemented in software, that attempts to force a robot's movement along desired paths or prevent a robot from moving into forbidden regions.[2] The potential benefit of virtual fixtures is safer and faster operation: they attempt to capitalize on the accuracy of robotic systems, while maintaining a degree of operator control.

Types of virtual fixtures
The goals of traditional telemanipulator design all revolve around giving the user the highest possible control over the slave. In contrast, the goal of a virtual fixture is to remove some control from the user. Because these goals generally conflict with one another, it is not obvious how to best implement virtual fixtures on a telemanipulation system. We have considered two types: guidance and forbidden-region virtual fixtures.

The guidance variety are designed to work with admittance control systems, where the
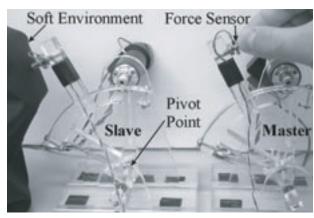

Figure 1. A one-degree-of-freedom telemanipulation testbed, used to compare virtual fixture algorithms implemented on various underlying telemanipulation control laws.
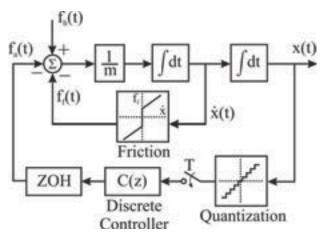

Figure 2. System model for a virtual-wall controller, including discrete and continuous elements.

velocity of the manipulator is proportional to the force applied by the human operator. A guidance virtual fixture can be implemented by simply rejecting the force inputs in certain directions.[2] Since typical bilateral-telemanipu-

lation systems are of the impedance type, a pseudo-admittance control system was designed to allow this type of virtual fixture without the use of a force sensor.[3] Forbidden-region virtual fixtures can be implemented with virtual springs, or through motion scaling. Predicting the passivity of the virtual spring method is described in the following section. The motion-scaling type attenuates or rejects inputs from the haptic interface at the remote robot. We have found experimentally that such operator assistance is most efficient when the operator receives some haptic feedback regarding the location and strength of the virtual fixture.

Virtual fixture passivity
For impedance-controlled telemanipulation systems, stability problems arise from the combination of discrete and continuous system elements.[4] This is a well-known problem in the design of virtual environments, where "virtual walls" constructed from spring models are used to create haptic objects. We have developed a sufficient condition for virtual wall passivity that accounts for quantization effects resulting from measuring position with optical encoders, and also assumes nothing about the bandwidth of the human operator. We consider a one-degree-of-freedom haptic interface, modeled as a mass with coulomb (f_c) and viscous (b) friction, controlled as shown in Figure 2. The virtual wall is implemented on a digital computer with a fixed and known sampling rate, T, and the position sensor has resolution D. Using energy conservation arguments, as well as techniques from optimal control, an explicit upper bound on vir-

Change Service Requested

DATED MATERIAL