Mosaicking Images

Panoramic Imaging for Miniature Robots

BY CHRISTIAN C. DOS SANTOS, SASCHA A. STOETER, PAUL E. RYBSKI, AND NIKOLAOS P. PAPANIKOLOPOULOS

Omnidirectional sensors hold a great promise for robot localization and navigation. However, the limited payload volume of miniature robots makes the use of omnidirectional vision sensors almost impossible. We overcame this problem by moving the robot around in order to create panoramic images. In addition to the general problems for building mosaics (i.e., computing the warping functions), this framework also has to take into account the noisy images delivered by the miniature robot. Two methods are presented. The solution for the general case allows for large rotations and a zoom factor. A special case, in which the homography is approximated by translations caused by minor changes in the optical system, is also considered, as computation time can be reduced significantly in this case.

Panoramic images require expensive lenses and custom-made hardware that, due to volume constraints, many robots (especially miniature ones) cannot accommodate. As a more cost effective alternative, a robot in front of a complex scene could restitute this landscape by taking several pictures and then trying to piece them together. This process, called "mosaicking images," tries to recreate a continuous picture from several overlapping images taken from the same scene. However, the result of this intuitive approach will be poor because of perspective distortions that appear when the camera is spun around its axes.

Hostage and disaster rescue missions, as well as toxic atmosphere surveillance, are two examples in which the use of robots can be beneficial to save human lives. We have developed a heterogeneous robot team consisting of two kinds of robots for these environments. The cylindrical Scout robot [1] is 11 cm long, 4 cm in diameter, and is equipped with a video camera (Figure 1). Locomotion is accomplished through a unique combination of rolling and jumping. The much larger Ranger robot is used to deploy Scouts in their area of operation and provides for computational resources.

Remote human rescue personnel desire to be provided with a complete and high-resolution view of a Scout's surrounding area. This is achieved by using the Scout's video camera as a visual sensor for mosaicking images. The small size of the Scout and its limited transmission power complicate the creation of mosaics because of noise in the images. Two video transmission frequencies are available. At the 900 MHz range, the signal penetrates objects or walls more easily than at the 2.4 GHz range. However, the higher frequency provides a clearer signal under the condition of line-of-sight.

The remainder of this article gives an overview of related work and introduces the proposed method for mosaicking images. Experimental results are presented, and the article closes with a conclusion and suggestions for future research directions.

1070-9932/04/$20.00©2004 IEEE

Related Work

Mosaicking of images has been in practice since long before the age of digital computers. Shortly after the photographic process was developed, photos were applied to topographical mapping [2]. Images acquired from hilltops or balloons were manually pieced together. After the development of airplanes, aerial photography became an exciting new field. The limited flying heights of the early airplanes and the need for large photo-maps forced imaging experts to construct mosaic images from overlapping photographs. This was initially done by manually mosaicking images that were acquired by calibrated equipment [3]. Starting in the second half of the 20th century, the need for mosaicking continued to increase as satellites began sending pictures from space.

With improvements in computer technology, computational techniques were developed to solve the mosaicking problem. The construction of mosaic images and the use of such images have been active areas of research in recent years. There has been a variety of new additions to the classic applications mentioned previously that aim primarily to enhance image resolution and the field of view. Image-based rendering [4], which combines the two complementary fields of computer vision and computer graphics [5], has become a major focus of attention. In computer graphics applications, images of the real world have traditionally been used as environment maps. In early applications, such environment maps were single images captured by fish-eye lenses or a sequence of images captured by wide-angle rectilinear lenses used as faces of a cube.

Mosaicking images on smooth surfaces allows an unlimited resolution and avoids discontinuities that can result from images acquired separately. Such immersive environments provide the users with an improved sense of presence in a virtual scene. A combination of such scenes used as nodes allows the users to navigate through a remote environment [6]. Computer vision methods can be used to generate intermediate views between the nodes.

As a reverse problem, the three-dimensional (3-D) structure of scenes can be reconstructed from multiple nodes [7]. Among other major applications of image mosaicking in computer vision are image stabilization, resolution enhancement, and video processing. An overview of mosaicking can be found in [8].

There are many efforts that include the development and use of omnidirectional cameras (mounted often on larger mobile robots). Nayar and his team [9] studied various aspects of omnidirectional vision, including the computation of ego-motion using omnidirectional cameras. Geyer and Daniilidis [10] worked on calibration issues for catadioptric cameras. Menegatti and Pagello [11] studied the use of omnidirectional vision in the problem of mapping a multi-robot system. Suzuki et al. [12] discussed behavior learning for a robot that possessed omnidirectional sensors. An extensive overview of omnidirectional vision research efforts can be found in [13] and in the Proceedings of the IEEE Workshops on Omnidirectional Vision.

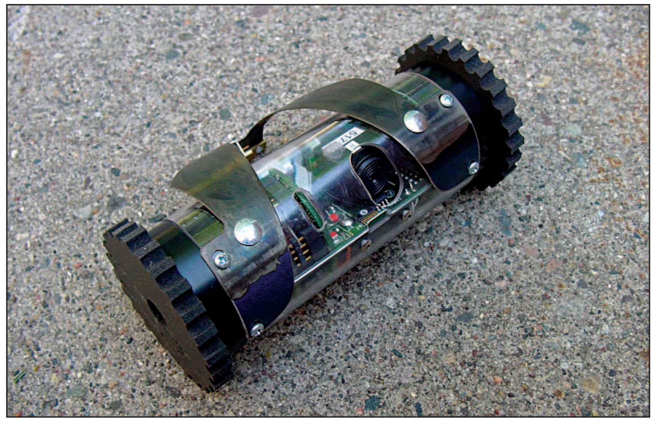

Figure 1. *The miniature Scout robot.*

Mosaicking Images

The fundamental input for all mosaicking methods is a set of images. Their positions with respect to each other must be determined in order to merge them into a single mosaic. In the following, it is assumed that two neighboring images share at least an empirically chosen 15% of their content, and that the images present enough features to compute their relative positions. The rotational and scaling factors should stay similar, and the optical distortion should be insignificant. Only grayscale images are used.

The images are preprocessed before any other steps to improve the accuracy of the results. First, a low-pass filter such as a Gaussian 3 × 3 filter is applied to reduce the noise and to smooth the images. An MDIF, a first-order derivative filter, is then applied, which is the combination of a low-pass filter and a derivative operator. Finally, the gradient is calculated, and the image is normalized.

General Method

In the general case, the parameters of the camera taking images from its surroundings can change from one image to the next. It is therefore important to take all the factors (i.e., shift, scaling, and rotation around the several axes of the camera) into account for an exact mosaic. The noise in the images should be relatively low.

Derivation of Relationships

Rotation, scaling, and shift factors between two images must be characterized. Let $\mathbf{p} = (x, y)^T$ be the coordinates of a point in the first image and $\mathbf{p}' = (x', y')^T$ the coordinates of the same point in the second one. A two-dimensional (2-D) affine transformation is described by $\mathbf{p}' = \Lambda\,\mathbf{p} + \mathbf{t}$, where $\mathbf{t}$ represents the translation vector and Λ contains the rotating factors and the scaling factor. With the introduction of two variables to correct the perspective, the complete transformation can be written as a homogeneous matrix $\mathbf{H}$:

$$\begin{pmatrix} X' \\ Y' \\ W \end{pmatrix} = \underbrace{\begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & 1 \end{pmatrix}}_{\mathbf{H}} \begin{pmatrix} x \\ y \\ 1 \end{pmatrix}. \qquad (1)$$

Finally, the link between X' and x' is given by $x' = X'/W$ and the one for Y' and y' is $y' = Y'/W$.

To describe the motion between two images, the factors of $\mathbf{H}$ are solved for by using the following:

Figure 2. Mosaic from eight images with parametric intensity adjustment.

$$x' = \frac{a_{11}x + a_{12}y + a_{13}}{a_{31}x + a_{32}y + 1} \quad (2)$$

$$y' = \frac{a_{21}x + a_{22}y + a_{23}}{a_{31}x + a_{32}y + 1}. \quad (3)$$

The eight unknown parameters can be calculated without any 3-D information by using the correspondences between the points of the images as described in the following. Consequently, one of the two images has to be transformed by (2) and (3) into the base of the other in order to obtain an exact mosaic.

Detecting Image Correspondences

The relative positions of two images with respect to each other are detected by finding the best fit of points with strong curvature radii. A Harris detector is used to find these points that corresponds to perceived corners [14]. A corner is a point exhibiting a sturdy intensity change in several directions. Considering the four elementary directions, at least two of them must undergo a significant intensity change. If $I(x, y)$ is defined to be the intensity value of an image in point (x, y), then I_x is the derivative image with respect to x, and I_y is the derivative image with respect to y. With the $\hat{I}$'s denoting Gaussian operators, $\mathbf{M}$ is defined as:

$$\mathbf{M} = \begin{pmatrix} \widehat{I_x^2} & \widehat{I_x I_y} \\ \widehat{I_x I_y} & \widehat{I_y^2} \end{pmatrix}. \quad (4)$$

Note that the Gaussian operator must be applied to the image containing the squares of the derivatives, not to the raw image. The Prewitt masks are used to calculate the derivatives. Two small eigenvalues in $\mathbf{M}$ designate a constant intensity region, one large and one small eigenvalue designate an edge, and two large eigenvalues designate a corner. If both eigenvalues of $\mathbf{M}$ are large, it means that a small displacement in any direction will cause a significant change in the intensity level. This means that this point is a corner. The corner-responding function is defined as:

$$\mathbf{R}(x, y) = \left(k + \frac{1}{k}\right) |\det(\mathbf{M})| - |\mathrm{trace}(\mathbf{M})^2 - 2\det(\mathbf{M})|. \quad (5)$$

Thus, the sharper the corner, the higher the value of $\mathbf{R}(x, y)$. The best corners are selected with an appropriately chosen threshold. Harris recommends an empirically chosen $k = 0.05$.

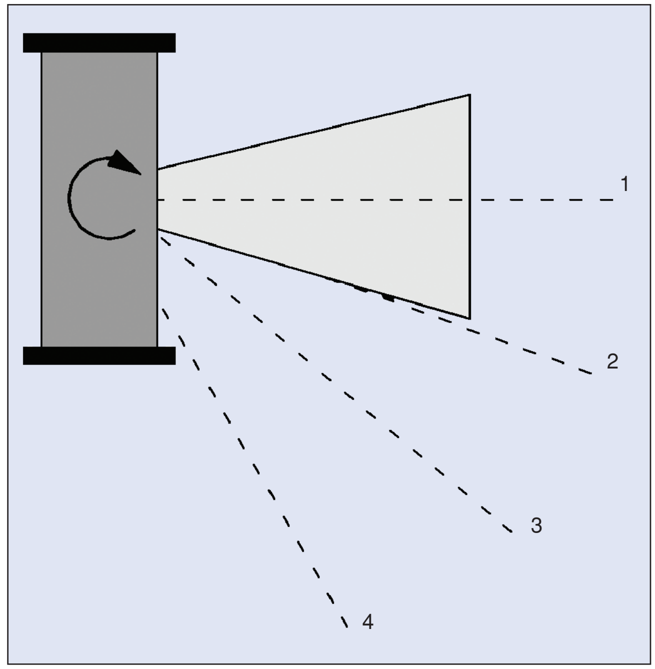

Figure 3. Experimental setup.

Finding correspondences between two pictures is therefore simplified to finding correspondences between the two sets of corners from the images. A corner in one image is selected and given a mark for each possible combination with the corners from the other image.

A correlation is computed by multiplication, and the result is normalized to the range of $[-1, 1]$ for comparing matched corners. The best score defines the corresponding corners. Scores near 1 mean that corners are similar, while those close to -1 are considered different. Once the corresponding corners are found, only the best pairs are kept.

Resolution of Levenberg-Marquardt

The homography of two pictures taken from the same scene and the same point can be determined by (2) and (3). Given the matched corners of these images, the energy function $\mathbf{E} = \sum_i ((x'_i, y'_i) - f(x_i, y_i))^2$ which is the sum of the squared distances between the corners of one image and the projection of the corresponding corners in the same basis must be minimized. The homography f, minimizing this value, presents the best possible solution (i.e., the eight parameters a_{ij} that minimize this sum must be found). The minimization of this quadratic function with eight parameters is realized with the method of Levenberg-Marquardt. While an empirically derived matrix can be used as an initial estimate of the homography, a better initial estimate can be obtained from the optimized method presented in the following section.

Optimized Method

While universally applicable, the presented method has a significant drawback: It takes unacceptably long to compute the parameters. Reviewing the initial problem, it can be found that the robot used to obtain the images only moves to a small extent, and that the sequence of images is known. The difference between the resulting images is roughly a shift with a small distortion, due to the modification of the rotational and scaling factors. Thus, only translation must be accounted for. The shift can be determined by means of a simple correlation. A template is extracted from one image, and its correspondence calculated in the other image through convolution. The area of interest can be restricted, as the motion of the Scout is known to a certain extent. Vertical displacement is expected to be less than 10%, while the horizontal displacement is mostly a function of the heading change.

The match can be found either via subtraction or multiplication. First, the correlation is obtained by the computationally less demanding subtraction. If the result of the sum is higher than a certain level of doubt, the correlation is established by multiplication. The fluctuation between two images (i.e., the Gaussian intensity differences) is accounted for by a similarity measure 8σ, where a typical value for σ is 2.5.

After calculating the shifts, the images can be merged into a single panoramic mosaic. The intensity values in the overlapping regions are adjusted with a parametric function. A typical result of this algorithm is shown in Figure 2.

In order to make the routine completely automated, the program should be able to detect when the Scout has completed a full circle. A separate template serves as the termination criterion. A backup termination criterion is available to increase the robustness of the algorithm. This criterion stops the computation after the mosaic has reached a certain size. The resulting mosaic covers more than 360° in this case.

Experimental Results

The methods were embedded into an existing distributed software architecture [15], and experiments were set up to verify the presented method. As shown in Figure 3, a Scout was instructed to rotate in the horizontal plane. Images were taken at fixed intervals when the robot was fully stopped, in order to prevent interference from the motors. The video was transmitted wirelessly at both 900 MHz and 2.4 GHz. At the lower frequency, a significant amount of noise was visible in the images for a typical office environment as the one used for the experiments. In addition, a high quality Panasonic video camera, and a Sony digital still camera from the two-megapixels class with virtually noiseless images were tested.

Only perfect mosaics are counted as a success. A failure is registered if just a single correspondence could not be established. The mosaic is still usable in these cases, as the other images were fit correctly.

The mosaics were classified by hand. The Scouts operate in vastly different environments, which prevents facile assumptions that could help in automatically determining failures in order to adjust the matching parameters.

The limited payload volume of miniature robots makes the use of omnidirectional vision sensors almost impossible.

Results of the General Method

The general method was evaluated by merging three images into a mosaic. The small number of images is a direct result from the high computational complexity of this method. One image is chosen as the basis into which the other images are transformed.

Figure 4 shows a mosaic obtained from this method using the clearer 2.4 GHz frequency. The success rate of the method is 60% when executed automatically. With human assistance in selecting an appropriate number of corners to base the correlation on, the success rate increased to 89%. The operation took four minutes on a Pentium II 450 MHz Linux workstation, with most of this time spent on finding the correspondences.

Results of the Optimized Method

The shift was determined correctly in 95% of the tests runs for less noisy images. For noisy images, such as the ones produced by the 900 MHz transmitter, the success rate of properly merging all images of a mosaic was 20%. To increase the success rate, a function that determines the noise level was added. If too much noise was detected, the image is discarded and another one taken. This way, the error rate dropped to 10%.

Run-time for the examples presented in Figure 5 was about three minutes on the described computer. This is a dramatic increase in speed from the general method in which only three images were merged.

Summary and Conclusions

Two methods of mosaicking images were developed, one for the simple case of a shift between two following images and another one for the general case. For the simple mosaicking, the error rate is acceptable for both cases of noisy or less noisy images. Owing to the fact that future generations of Scouts will be provided with the 2.4 GHz transmission frequency, panoramic mosaics can be created with a high level of success. The operation is completely automated, and the execution time is close to three minutes for each mosaic.

Figure 4. *An example of the general method.*

For the case of mosaicking images in general, the major problem is that the quality of the images given by the Scout does not allow for choosing a small number of points for the Harris detector. This leads to long execution times. Furthermore, it is extremely hard to realize a panoramic vision by such a method because it is still unreliable without the intervention of an operator. Building 360° mosaics is not practical. Nevertheless, the general method can be used for creating precise mosaics of three consecutive images, thus providing wide angle images. Images resulting from this operation represent 100° mosaics. Alas, the time of calculation is still a considerable four minutes.

Future Work

For the optimized method, computation speed is a minor reason for concern. Execution time could be reduced by decreasing the size of the search area while realizing a closed-loop control on the motion of the Scout. Feedback on the rotation of the Scout's wheels would allow for a reduced search area. This problem could be solved by a faster computer.

For the general method, execution time should be reduced and the success rate of the operation increased. A more reliable way of matching corners would address both problems. The use of parametric corners promises to yield improved results. Nevertheless, the presented theory is still correct and can be applied to the case of an elevated Scout. Research is currently underway for a grappling hook attachment to the Scout, allowing it to elevate itself onto a roof. A set of images could be taken during the elevation process to create a mosaic.

Given the new functionality of a 360° view, exciting new opportunities arise. As the Scout has increased awareness of its environment, it can act in a more educated way. Additional functions could also be implemented easily (e.g., determining the horizon).

Acknowledgment

We would like to thank the anonymous reviewers for their valuable comments. This material is based on work supported by the National Science Foundation through awards CNS-0224363 and CNS-0324864, Microsoft Corporation, and the Defense Advanced Research Projects Agency, Microsystems Technology Office (Distributed Robotics), ARPA Order G155, Program Code 8H20, issued by DARPA/CMD under Contract MDA972-98-C-0008.


(a) Panasonic Camera, Nine Images


(b) Wireless 2.4-GHz Transmission, Nine Images

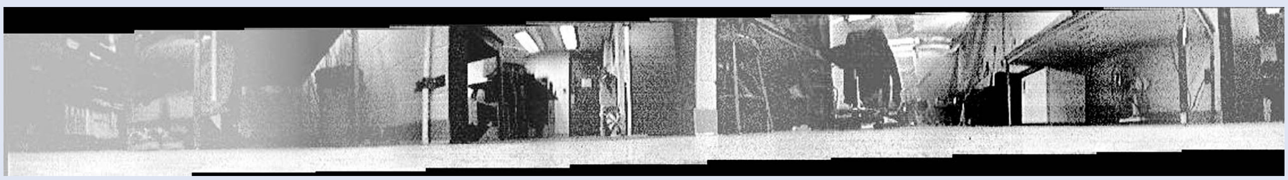
(c) Wireless 900-MHz Transmission, Ten Images

Figure 5. *Examples of simple mosaics generated by the optimized method.*

Keywords
Panoramic imaging, mosaicking, miniature robots, omnidirectional camera systems.

References

[1] P.E. Rybski, N.P. Papanikolopoulos, S.A. Stoeter, D.G. Krantz, K.B. Yesin, M. Gini, R. Voyles, D.F. Hougen, B. Nelson, and M.D. Erickson, "Enlisting rangers and scouts for reconnaissance and surveillance," *IEEE Robot. Automat. Mag.*, vol. 7, no. 4, pp. 14–24, 2000.

[2] P.R. Wolf, *Elements of Photogrammetry*. New York: McGraw-Hill, 1983.

[3] P. Kolonia, "When more is better," *Popular Photography*, vol. 58, no. 1, pp. 30–34, Jan. 1994.

[4] S.B. Kang, "Survey of image-based rendering techniques," in *SPIE Videometrics VI*, vol. 3641, 1999, pp. 2–16.

[5] J. Lengyel, "The convergence of graphics and vision," *Computer*, vol. 31, no. 7, pp. 46–53, 1998.

[6] S.E. Chen, "QuickTime VR—an image-based approach to virtual environment navigation," in *Proc. of SIGGRAPH*, Los Angeles, CA, 1995, pp. 29–38.

[7] H.-Y. Shum, M. Han, and R. Szeliski, "Interactive construction of 3d models from panoramic mosaics," In *Proc. IEEE Computer Soc. Conf. Computer Vision and Pattern Recognition*, Santa Barbara, CA, 1998, pp. 427–433.

[8] S. Gumustekin (July 1999). An introduction to image mosaicing [Online]. Available: http://likya.iyte.edu.tr/eee/sevgum/research/mosaicing99/

[9] J. Gluckman and S. Nayar, "Ego-motion and omnidirectional cameras," in *Proc. ICCV*, 1998, pp. 999–1005.

[10] C. Geyer and K. Daniilidis, "Catadioptric camera calibration," in *Proc. ICCV*, 1999, pp. 398–404.

[11] E. Menegatti and E. Pagello, "Omnidirectional distributed vision for multi-robot mapping," in *Proc. Int. Symp. on Distributed Autonomous Robotic Systems (DARS02)*, Fukuoka, Japan, 2002.

[12] S. Suzuki, T. Kato, M. Asada, and K. Hosoda, "Behavior learning for a mobile robot with omnidirectional vision enhanced by an active zoom mechanism," in *Conf. Rec. IEEE/IAS Annu. Meeting,*, 1998 pp. 244–249.

[13] R. Benosman and S.B. Kang, *Panoramic Vision: Sensors, Theory and Applications*. New York: Springer Verlag, 2001.

[14] C. Harris and M. Stephens, "A combined corner and edge detector," in *Proc. 4th Alvey Vision Conf.*, 1988, pp. 147–151.

[15] S.A. Stoeter, P.E. Rybski, M.D. Erickson, M. Gini, D.F. Hougen, D.G. Krantz, N. Papanikolopoulos, and M. Wyman, "A robot team for exploration and surveillance: Design and architecture," in *Proc. Int. Conf. Intelligent Autonomous Systems*, Venice, Italy, 2000, pp. 767–774.

Christian C. Dos Santos received a diploma of engineer in robotics and computer science in 2001 from "Ecole Nationale Superieure de Physique" in Strasbourg, France. He also obtained the French equivalent of an M.S. in computer science and cybernetics the same year in the same institution. He is working on the development of an integrated validation station for the first totally interactive biochemistry analyzer.

Sascha A. Stoeter obtained his M.S. and Ph.D. degrees in computer and information sciences in 1997 and 2003, respectively, from the University of Minnesota. Before entering the Ph.D. program in Minnesota, he was a research assistant at the Institute for Robotics and Process Control in Braunschweig, Germany. He is a member of the IEEE and Computer Professionals for Social Responsibility.

Paul E. Rybski received an interdisciplinary B.A. in math/computer science, with an emphasis in cognitive science, in 1995 from Lawrence University in Appleton, Wisconsin. He received the M.S. in computer and information sciences in 2000 at the University of Minnesota and his Ph.D. in computer science, with a minor in cognitive science at the

same institution. His research interests include distributed robot control strategies, probabilistic algorithms for automated navigation/localization, visual feature-detection algorithms for servoing applications, human/robot interfaces, and development of real-time algorithms for embedded control systems. He is a Member of the IEEE, ACM, and AAAI.

Nikolaos P. Papanikolopoulos received the Diploma degree in electrical and computer engineering from the National Technical University of Athens, Athens, Greece, in 1987, the M.S.E.E. in electrical engineering from Carnegie Mellon University (CMU), Pittsburgh, Pennsylvania, in 1988, and the Ph.D. in electrical and computer engineering from Carnegie Mellon University, Pittsburgh, Pennsylvania, in 1992. Currently, he is a professor in the department of computer science at the University of Minnesota and director of the Center for Distributed Robotics. His research interests include robotics, sensors for transportation applications, control, and computer vision. He has authored or coauthored more than 170 journal and conference papers in these areas (41 refereed journal papers). He was finalist for the Anton Philips Award for Best Student Paper in the 1991 IEEE Int. Conf. on Robotics and Automation and recipient of the best Video Award in the 2000 IEEE Int. Conf. on Robotics and Automation. He was recipient of the Kritski fellowship in 1986 and 1987. He was a McKnight Land-Grant Professor at the University of Minnesota for the period 1995-1997, and has received the NSF Research Initiation and Early Career Development Awards. He was also awarded the Faculty Creativity Award from the University of Minnesota. One of his papers (co-authored by O. Masoud) was awarded the IEEE VTS 2001 Best Land Transportation Paper Award. He has also received grants from DARPA; Sandia National Laboratories; NSF; Microsoft; INEEL; USDOT, DHS, and Air Force; MN/DOT; Honeywell; and 3M.

Address for Correspondence: Nikos Papanikolopoulos, Department of Computer Science and Engineering, University of Minnesota, Minneapolis, MN 55455 USA. E-mail: npapas@cs.umn.edu.

Biomimetic Centering Behavior

BY ANTONIS A. ARGYROS, DIMITRIS P. TSAKIRIS, AND CÉDRIC GROYER

(continued from page 30)

Science, University of Crete, Greece, in visual motion analysis. He has been a postdoctoral fellow at the Royal Institute of Technology in Stockholm, Sweden, where he worked on vision-based, reactive robot navigation. In 1999 he joined the Computational Vision and Robotics Laboratory of ICS-FORTH, where he has been involved in many RTD projects in image analysis, computational vision and robotics. His current research interests include computational vision and robotics and particularly the visual perception of motion and 3-D structure, the development of robot behaviors based on visual information and alternative visual sensors.

Dimitris P. Tsakiris is a research scientist at the Institute of Computer Science of FORTH and a visiting professor at the University of Crete. He received his B.S. degree from the Department of Electrical Engineering of the National Technical University of Athens and his M.S. and Ph.D. degrees from the Department of Electrical Engineering of the University of Maryland at College Park. Prior to his current position, he was a Marie Curie postdoctoral fellow in INRIA, Sophia-Antipolis. His research interests lie in the areas of sensor-based robotics, nonlinear control, undulatory locomotion, geometric mechanics, and computational vision. He is a principal investigator and coinvestigator of several European IST projects and research networks related to these research areas.

Cédric Groyer is a research and development engineer at the Computational Vision and Robotics Laboratory of ICS/FORTH. He received his B.Sc. degree in École Nationale Supérieure de Physique de Marseille, with major in Signal Processing. His interests include image processing and computer vision, with emphasis on their application to mobile robotics.

Address for Correspondence: Dimitris P. Tsakiris, Institute of Computer Science-FORTH, Vassilika Vouton, P.O. Box 1385, GR-71110 Heraklion, Greece. Tel.: +30 2810 391708. Fax: +30 2810 391601. E-mail: tsakiris@ics.forth.gr. URL: http://www.ics.forth.gr/~tsakiris.