Monocular Model-Based 3D Tracking of Rigid Objects

Monocular Model-Based 3D Tracking of Rigid Objects

Vincent Lepetit

Ecole Polytechnique Fédérale de Lausanne (EPFL)
Computer Vision Laboratory
CH-1015 Lausanne, Switzerland
Vincent.Lepetit@epfl.ch

Pascal Fua

Ecole Polytechnique Fédérale de Lausanne (EPFL)
Computer Vision Laboratory
CH-1015 Lausanne, Switzerland
Pascal.Fua@epfl.ch

the essence of knowledge

Foundations and Trends® in Computer Graphics and Vision

Published, sold and distributed by:
now Publishers Inc.
PO Box 1024
Hanover, MA 02339
USA
Tel. +1 781 871 0245
www.nowpublishers.com
sales@nowpublishers.com

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

A Cataloging-in-Publication record is available from the Library of Congress

Printed on acid-free paper

ISBN: 1-933019-03-4; ISSNs: Paper version 1572-2740; Electronic version 1572-2759

Contents

1 Introduction **1**

1.1 Focus and Organization of the Survey 2
1.2 Different Approaches for Different Applications 2
1.3 Computer Vision-Based 3D Tracking 3

2 Mathematical Tools **5**

2.1 Camera Representation 5
2.2 Camera Pose Parameterization 11
2.3 Estimating the External Parameters Matrix 14
2.4 Least-Squares Minimization Techniques 19
2.5 Robust Estimation 22
2.6 Bayesian Tracking 26

3 Fiducial-Based Tracking **33**

3.1 Point Fiducials 33
3.2 Planar Fiducials 36

4 Natural Features, Model-Based Tracking **39**

4.1 Edge-Based Methods 40
4.2 Optical Flow-Based Methods 48

4.3 Template Matching 50
4.4 Interest Point-Based Methods 55
4.5 Tracking Without 3D Models 63
4.6 Natural Features Methods in Short 66

5 Tracking by Detection 67

5.1 Feature Point Extraction 68
5.2 Wide Baseline Matching 69
5.3 From Wide Baseline Matching to 3D Tracking 72
5.4 The End of Recursive Tracking? 73

6 Conclusion 75

6.1 Choosing the Appropriate Approach to 3D Tracking 75
6.2 Implementation Issues 79
6.3 The Future of 3D Tracking 80

References 83

1

Introduction

Tracking an object in a video sequence means continuously identifying its location when either the object or the camera are moving. There are a variety of approaches, depending on the type of object, the degrees of freedom of the object and the camera, and the target application.

2D tracking typically aims at following the image projection of objects or parts of objects whose 3D displacement results in a motion that can be modeled as a 2D transformation. An adaptive model is then required to handle appearance changes due to perspective effects or to deformation. It can provide the object's image position in terms of its centroid and scale or of an affine transformation [141, 26, 62]. Alternatively, more sophisticated models such as splines [16], deformable templates [142], 2D deformable meshes [112], or 2D articulated models [20] can be used. However, none of these methods involves recovering the actual position in space.

By contrast, 3D tracking aims at continuously recovering all six degrees of freedom that define the camera position and orientation relative to the scene, or, equivalently, the 3D displacement of an object relative to the camera.

1.1 Focus and Organization of the Survey

The 3D tracking literature is particularly massive both because there are almost as many approaches as target applications and because many different approaches to solving the same problem are possible. Here we focus on online model-based 3D tracking using a single camera. We will describe both marker-based techniques and marker-less natural features-based approaches for camera, scene, and object 3D tracking.

In particular, we will not consider batch methods to camera trajectory recovery: Because these methods can consider image sequence as a whole, they can rely on non-causal strategies that are not appropriate for online camera tracking. Furthermore, an excellent reference text for this topic already exists [54]. We will restrict ourselves to single-camera approaches because multi-camera systems require calibration of the stereo-rig and are therefore less popular. We will consider only rigid objects or scenes, as opposed to deformable [25, 89] or articulated objects such as human bodies [43, 121] that would take us too far afield.

We will first introduce the key mathematical tools required for 3D tracking. We will then present marker-based techniques that use either point fiducials or planar markers to ease the tracking task. Next we will focus on techniques that rely on natural features. Finally, we will discuss recent advances that seek to increase tracking robustness to disappearance and reappearance of the target object by replacing frame-to-frame tracking by detection in each frame individually.

1.2 Different Approaches for Different Applications

3D tracking is a very useful tool in many different fields, and we briefly review some of them below.

1.2.1 Augmented Reality Applications

Many potential Augmented Reality (AR) applications have been explored, such as medical visualization, maintenance and repair, annotation, entertainment, aircraft navigation and targeting. They all involve superposing computer generated images on real scenes, which must be done at frame-rate in online systems. 3D real-time tracking is therefore a critical component of most AR applications. The objects in the real and virtual worlds must be properly aligned with respect to each other, and the system latency should also be low, lest the illusion that the two worlds coexist be compromised.

1.2.2 Visual Servoing

Visual servoing involves the use of one or more cameras and a Computer Vision system to control the position of a device such as a robotic arm relative to a part it has to manipulate, which requires detecting, tracking, servoing, and grasping. It therefore spans computer vision, robotics, kinematics, dynamics, control and real-time systems, and is used in a rich variety of applications such as lane tracking for cars, navigation for mobile platforms, and generic object manipulation.

The tracking information is required to measure the error between the current location of the robot and its reference or desired location from eye-in-hand cameras. As a consequence, the tracking algorithm must be robust, accurate, fast, and general.

1.2.3 Man–Machine Interfaces

3D tracking can be integrated into man–machine interfaces. For example, it could be used to continuously update the position of a hand-held object, which would then serve as a 3D pointer. This object would then become an instance of what is known as a *Tangible Interface*. Such interfaces aim at replacing traditional ones by allowing users to express their wishes by manipulating familiar objects and, thus, to take advantage of their everyday experience.

Eventually, this is expected to lead to more natural and intuitive interfaces. In this context, vision-based tracking is the appropriate technique for seamless interaction with physical objects.

1.3 Computer Vision-Based 3D Tracking

Many other technologies besides vision have been tried to achieve 3D tracking, but they all have their weaknesses: Mechanical trackers are accurate enough, although they tether the user to a limited working volume. Magnetic trackers are vulnerable to distortions by metal in the environment, which are a common occurrence, and also limit the range of displacements. Ultrasonic trackers suffer from noise and tend to be inaccurate at long ranges because of variations in the ambient temperature. Inertial trackers drift with time.

By contrast, vision has the potential to yield non-invasive, accurate and low-cost solutions to this problem, provided that one is willing to invest the effort required to develop sufficiently robust algorithms. In some cases, it is acceptable to add fiducials, such as LEDs or special markers, to the scene or target object to ease the registration task, as will be discussed in Section 3. Of course, this assumes that one or more fiducials are visible at all times. Otherwise, the registration falls apart. Moreover, it is not always possible to place fiducials. For example, Augmented Reality end-users do not like them because they are visible in the scene and it is not always possible to modify the environment before the application has to run.

It is therefore much more desirable to rely on naturally present features, such as edges, corners, or texture. Of course, this makes tracking far more difficult: Finding and following feature points or edges can be difficult because there are too few of them on many typical objects. Total, or even partial occlusion of the tracked objects typically results in tracking failure. The camera can easily move too fast so that the images are motion blurred; the lighting during a shot can change significantly; reflections and specularities may confuse the tracker. Even more importantly, an object may drastically change its aspect very quickly due to displacement. For example this happens when a camera films a building and goes around the corner, causing one wall to disappear and a new one to appear. In such cases, the features to be followed always change and the tracker must deal with features coming in and out of the picture. Sections 4 and 5 focus on solutions to these difficult problems.

References

[1] Y. Amit and D. Geman, "Shape quantization and recognition with randomized trees," *Neural Computation*, vol. 9, no. 7, pp. 1545–1588, 1997.

[2] M. Armstrong and A. Zisserman, "Robust object tracking," in *Proceedings of Asian Conference on Computer Vision*, pp. 58–62, 1995.

[3] ARToolKit http://www.hitl.washington.edu/artoolkit/.

[4] K. Arun, T. Huang, and S. Blostein, "Least-squares fitting of two 3-D points sets," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 9, no. 5, pp. 698–700, 1987.

[5] A. Azarbayejani and A. P. Pentland, "Recursive estimation of motion, structure and focal length," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 6, pp. 562–575, 1995.

[6] R. Azuma, "A survey of augmented reality," in *Computer Graphics, SIGGRAPH Proceedings*, pp. 1–8, August 1995.

[7] S. Baker and I. Matthews, "Lucas-Kanade 20 years on: A unifying framework," *International Journal of Computer Vision*, pp. 221–255, March 2004.

[8] S. Balcisoy, M. Kallmann, R. Torre, P. Fua, and D. Thalmann, "Interaction techniques with virtual humans in mixed environment," in *International Symposium on Mixed Reality* (Yokohama, Japan), March 2001.

[9] Y. Bar-Shalom and T. Fortmann, *Tracking and Data Association.* Academic Press, 1988.

[10] S. Basu, I. Essa, and A. Pentland, "Motion regularization for model-based head tracking," in *International Conference on Pattern Recognition* (Vienna, Austria), 1996.

[11] A. Baumberg, "Reliable feature matching across widely separated views," in *Conference on Computer Vision and Pattern Recognition*, pp. 774–781, 2000.

[12] P. A. Beardsley, A. Zisserman, and D. W. Murray, "Sequential update of projective and affine structure from motion," *International Journal of Computer Vision*, vol. 23, no. 3, pp. 235–259, 1997.

[13] J. Beis and D. Lowe, "Shape indexing using approximate nearest-neighbour search in high-dimensional spaces," in *Conference on Computer Vision and Pattern Recognition* (Puerto Rico), pp. 1000–1006, 1997.

[14] S. Birchfield, "KLT: An Implementation of the Kanade-Lucas-Tomasi Feature Tracker," http://www.ces.clemson.edu/˜stb/klt/.

[15] M. J. Black and Y. Yacoob, "Recognizing facial expressions in image sequences using local parameterized models of image motion," *International Journal of Computer Vision*, pp. 23–48, 1997.

[16] A. Blake and M. Isard, *Active Contours*. Springer-Verlag, 1998.

[17] D. Brown, "Close range camera calibration," *Photogrammetric Engineering*, vol. 37, no. 8, pp. 855–866, 1971.

[18] R. Brown and P. Hwang, *Introduction to Random Signals and Applied Kalman Filtering, Second Edition*. John Wiley & Sons, Inc., 1992.

[19] M. Cascia, S. Sclaroff, and V. Athitsos, "Fast, reliable head tracking under varying illumination: An approach based on registration of texture-mapped 3D models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, April 2000.

[20] T.-J. Cham and J. Rehg, "A multiple hypothesis approach to figure tracking," in *Conference on Computer Vision and Pattern Recognition* (Ft. Collins, CO), June 1999.

[21] K. Chia, A. Cheok, and S. Prince, "Online 6 DOF augmented reality registration from natural features," in *International Symposium on Mixed and Augmented Reality*, 2002.

[22] A. Chiuso, P. Favaro, H. Jin, and S. Soatto, "Structure from motion causally integrated over time," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 4, pp. 523–535, 2002.

[23] Y. Cho, W. Lee, and U. Neumann, "A multi-ring color fiducial system and intensity-invariant detection method for scalable fiducial-tracking augmented reality," in *International Workshop on Augmented Reality*, 1998.

[24] D. Claus and A. Fitzgibbon, "Reliable fiducial detection in natural scenes," in *European Conference on Computer Vision*, pp. 469–480, Springer-Verlag, May 2004.

[25] L. Cohen and I. Cohen, "Finite-element methods for active contour models and balloons for 2-D and 3-D images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, pp. 1131–1147, November 1993.

[26] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, pp. 564–575, 2003.

[27] A. Comport, E. Marchand, and F. Chaumette, "A real-time tracker for markerless augmented reality," in *International Symposium on Mixed and Augmented Reality* (Tokyo, Japan), September 2003.

[28] A. Davison, "Real-time simultaneous localisation and mapping with a single camera," in *Proceedings of International Conference on Computer Vision*, pp. 1403–1410, October 2003.

[29] D. DeCarlo and D. Metaxas, "Optical flow constraints on deformable models with applications to face tracking," *International Journal of Computer Vision*, vol. 38, pp. 99–127, 2000.

[30] F. Dellaert and R. Collins, "Fast image-based tracking by selective pixel integration," in *ICCV Workshop of Frame-Rate Vision*, pp. 1–22, 1999.

[31] D. DeMenthon and L. S. Davis, "Model-based object pose in 25 lines of code," *International Journal of Computer Vision*, vol. 15, pp. 123–141, 1995.

[32] R. Deriche and O. Faugeras, "Tracking line segments," *Image and Vision Computing*, vol. 8, pp. 271–270, 1990.

[33] R. Deriche and G. Giraudon, "A Computational approach for corner and vertex detection," *International Journal of Computer Vision*, vol. 10, no. 2, pp. 101–124, 1993.

[34] A. Doucet, N. de Freitas, and N. Gordon, eds., *Sequential Monte Carlo Methods in Practice*. Springer-Verlag, 2001.

[35] T. Drummond and R. Cipolla, "Real-time visual tracking of complex structures," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, pp. 932–946, July 2002.

[36] R. Evans, "Filtering of pose estimates genrated by the RAPiD tracker in applications," in *British Machine Vision Conference* (Oxford), pp. 79–84, 1990.

[37] O. Faugeras, *Three-Dimensional Computer Vision: A Geometric Viewpoint*. MIT Press, 1993.

[38] M. Fischler and R. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography," *Communications ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[39] A. Fitzgibbon and A. Zisserman, "Automatic camera recovery for closed or open image sequences," in *European Conference on Computer Vision* (Freiburg, Germany), pp. 311–326, June 1998.

[40] W. Förstner, "A feature-based correspondence algorithm for image matching," *International Archives of Photogrammetry and Remote Sensing*, vol. 26, no. 3, pp. 150–166, 1986.

[41] E. Foxlin and L. Naimark, "Miniaturization, claibration and accuracy evaluation of a hybrid self-tracker," in *International Symposium on Mixed and Augmented Reality* (Tokyo, Japan), 2003.

[42] J. Fryer and D. Goodin, "In-flight aerial camera calibration from photography of linear features," *Photogrammetric Engineering and Remote Sensing*, vol. 55, no. 12, pp. 1751–1754, 1989.

[43] D. Gavrila, "The visual analysis of human movement: A survey," *Computer Vision and Image Understanding*, vol. 73, January 1999.

[44] Y. Genc, S. Riedel, F. Souvannavong, and N. Navab, "Marker-less tracking for augmented reality: A learning-based approach," in *International Symposium on Mixed and Augmented Reality*, 2002.

[45] D. Gennery, "Visual tracking of known three-dimensional objects," *International Journal of Computer Vision*, vol. 7, no. 1, pp. 243–270, 1992.

[46] C. Geyer and K. Daniilidis, "Omnidirectional video," *The Visual Computer*, vol. 19, pp. 405–416, October 2003.

[47] F. Grassia, "Practical parameterization of rotations using the exponential map," *Journal of Graphics Tools*, vol. 3, no. 3, pp. 29–48, 1998.

[48] A. Gruen and T. Huang, eds., *Calibration and Orientation of Cameras in Computer Vision*. Vol. 34 of *Springer Series in Information Sciences*, Springer-Verlag, 2001.

[49] M. Haag and H.-M. Nagel, "Combination of edge element and optical flow estimates for 3-D model-based vehicle tracking intraffic image sequences," *International Journal of Computer Vision*, vol. 35, no. 3, pp. 295–319, 1999.

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

[51] R. Haralick, D. Lee, K. Ottenburg, and M. Nolle, "Analysis and solutions of the three point perspective pose estimation problem," in *Conference on Computer Vision and Pattern Recognition*, pp. 592–598, June 1991.

[52] C. Harris, *Tracking with Rigid Objects*. MIT Press, 1992.

[53] C. Harris and M. Stephens, "A combined corner and edge detector," in *Fourth Alvey Vision Conference*, Manchester, 1988.

[54] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge University Press, 2000.

[55] J. Heikkila and O. Silven, "A four-step camera calibration procedure with implicit image correction," in *Conference on Computer Vision and Pattern Recognition*, pp. 1106–1112, 1997.

[56] W. A. Hoff, K. Nguyen, and T. Lyon, "Computer vision-based registration techniques for augmented reality," in *Proceedings of Intelligent Robots and Control Systems XV, Intelligent Control Systems and Advanced Manufacturing*, pp. 538–548, November 1996.

[57] B. Horn, "Closed-form solution of absolute orientation using unit quaternions," *Journal of the Optical Society of America*, vol. 4, pp. 629–642, April 1987.

[58] B. Horn, H. Hilden, and S. Negahdaripour, "Closed-form solution of absolute orientation using orthonormal matrices," *Journal of the Optical Society of America*, vol. 5, no. 7, pp. 1127–1135, 1987.

[59] P. Huber, *Robust Statistics*. New York: Wiley, 1981.

[60] Intel, "Open Source Computer Vision Library," http://www.intel.com/research/mrl/research/opencv/.

[61] M. Isard and A. Blake, "A smoothing filter for Condensation," in *European Conference on Computer Vision*, pp. 767–781, 1998.

[62] A. Jepson, D. J. Fleet, and T. El-Maraghi, "Robust on-line appearance models for vision tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 10, pp. 1296–1311, 2003.

[63] H. Jin, P. Favaro, and S. Soatto, "A semi-direct approach to structure from motion," *The Visual Computer*, vol. 19, pp. 1–18, 2003.

[64] F. Jurie, "Tracking objects with a recognition algorithm," *Pattern Recognition Letters*, vol. 3–4, no. 19, pp. 331–340, 1998.

[65] F. Jurie and M. Dhome, "A simple and efficient template matching algorithm," in *International Conference on Computer Vision* (Vancouver, Canada), July 2001.

[66] F. Jurie and M. Dhome, "Hyperplane approximation for template matching," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, pp. 996–100, July 2002.

[67] H. Kato and M. Billinghurst, "Marker racking and HMD Calibration for a video-based augmented reality conferencing system," in *IEEE and ACM International Workshop on Augmented Reality*, October 1999.

[68] H. Kato, M. Billinghurst, I. Poupyrev, K. Imamoto, and K. Tachibana, "Virtual object manipulation on a table-top AR environment," in *International Symposium on Augmented Reality*, pp. 111–119, 2000.

[69] G. Klein and T. Drummond, "Robust visual tracking for non-instrumented augmented reality," in *International Symposium on Mixed and Augmented Reality*, pp. 36–45, October 2003.

[70] D. Koller, K. Daniilidis, and H.-H. Nagel, "Model-based object tracking in monocular image sequences of road traffic scenes," *International Journal of Computer Vision*, vol. 10, pp. 257–281, July 1993.

[71] D. Koller, G. Klinker, E. Rose, D. Breen, R. Whitaker, and M. Tuceryan, "Real-time vision-based camera tracking for augmented reality applications," in *ACM Symposium on Virtual Reality Software and Technology* (Lausanne, Switzerland), pp. 87–94, September 1997.

[72] H. Kollnig and H.-H. Nagel, "3D pose estimation by directly matching polyhedral models to gray value gradients," *International Journal of Computer Vision*, vol. 23, pp. 283–302, July 1997.

[73] A. Kosaka and G. Nakazawa, "Vision-based motion tracking of rigid objects using prediction of uncertainties," in *International Conference on Robotics and Automation* (Nagoya, Japan), pp. 2637–2644, 1995.

[74] V. Lepetit, P. Lagger, and P. Fua, "Randomized trees for real-time keypoint recognition," in *Conference on Computer Vision and Pattern Recognition* (San Diego, CA), June 2005.

[75] H. Li, P. Roivainen, and R. Forchheimer, "3-D motion estimation in model-based facial image coding," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, pp. 545–555, June 1993.

[76] J. Li, R. Laganiere, and G. Roth, "Online estimation of trifocal tensors for augmenting live video," in *International Symposium on Mixed and Augmented Reality*, pp. 182–190, 2004.

[77] T. Lindeberg, "Scale-space theory: A basic tool for analysing structures at different scales," *Journal of Applied Statistics*, vol. 21, no. 2, pp. 224–270, 1994.

[78] D. G. Lowe, "Fitting parameterized three-dimensional models to images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, pp. 441–450, June 1991.

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

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

[81] D. Lowe, "Local feature view clustering for 3D object recognition," in *Conference on Computer Vision and Pattern Recognition*, pp. 682–688, 2001.

[82] D. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 20, no. 2, pp. 91–110, 2004.

[83] B. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," in *International Joint Conference on Artificial Intelligence*, pp. 674–679, 1981.

[84] E. Marchand, P. Bouthemy, and F. Chaumette, "A 2D-3D model-based approach to real-time visual tracking," *Image and Vision Computing*, vol. 19, no. 13, pp. 941–955, 2001.

[85] E. Marchand, P. Bouthemy, F. Chaumette, and V. Moreau, "Robust real-time visual tracking using a 2D-3D model-based approach," in *International Conference on Computer Vision* (Corfu, Greece), pp. 262–268, September 1999.

[86] E. Marchand and F. Chaumette, "Virtual visual servoing: A framework for real-time augmented reality," in *Computer Graphics Forum, Eurographics* (Sarrebruck, Allemagne), pp. 289–298, September 2002.

[87] J. Matas, O. Chum, U. Martin, and T. Pajdla, "Robust wide baseline stereo from maximally stable extremal regions," in *British Machine Vision Conference* (London, UK), pp. 384–393, September 2002.

[88] C. McGlove, E. Mikhail, and J. Bethel, eds., *Manual of Photogrametry*. American Society for Photogrammetry and Remote Sensing, Fifth ed., 2004.

[89] D. Metaxas and D. Terzopoulos, "Shape and nonrigid motion estimation through physics-based synthesis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 15, no. 6, pp. 580–591, 1991.

[90] K. Mikolajczik and C. Schmid, "A performance evaluation of local descriptors," in *Conference on Computer Vision and Pattern Recognition*, pp. 257–263, June 2003.

[91] K. Mikolajczyk and C. Schmid, "An affine invariant interest point detector," in *European Conference on Computer Vision*, pp. 128–142, Springer, 2002. Copenhagen.

[92] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. V. Gool, "A comparison of affine region detectors," *Accepted to International Journal of Computer Vision*, 2005.

[93] F. Mindru, T. Moons, and L. VanGool, "Recognizing color patterns irrespective of viewpoint and illumination," in *Conference on Computer Vision and Pattern Recognition*, pp. 368–373, 1999.

[94] M. Moegring, C. Lessig, and O. Bimber, "Video see-through AR on consumer cell phones," in *International Symposium in Mixed and Augmented Reality*, pp. 252–253, 2004.

[95] N. Molton, A. Davison, and I. Reid, "Locally planar patch features for real-time structure from motion," in *British Machine Vision Conference*, September 2004.

[96] H. Moravec, *Robot Rover Visual Navigation*. Ann Arbor, Michigan: UMI Research Press, 1981.

[97] H. Moravec, "Towards automatic visual obstacle avoidance," in *International Joint Conference on Artificial Intelligence* (MIT, Cambridge, Mass.), p. 584, August 1977.

[98] L. Naimark and E. Foxlin, "Circular data matrix fiducial system and robust image processing for a wearable vision-interial self-tracker," in *International Symposium on Mixed and Augmented Reality* (Darmstadt, Germany), 2002.

[99] S. K. Nayar, S. A. Nene, and H. Murase, "Real-time 100 object recognition system," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 18, no. 12, pp. 1186–1198, 1996.

[100] D. Nister, O. Naroditsky, and J. Bergen, "Visual odometry," in *Conference on Computer Vision and Pattern Recognition*, pp. 652–659, June 2004.

[101] D. Oberkampf, D. DeMenthon, and L. Davis, "Iterative pose estimation using coplanar feature points," *Computer Vision and Image Understanding*, vol. 63, pp. 495–511, 1996.

[102] J. Pilet, V. Lepetit, and P. Fua, "Real-time non-rigid surface detection," in *Conference on Computer Vision and Pattern Recognition* (San Diego, CA), June 2005.

[103] M. Pollefeys, R. Koch, and L. VanGool, "Self-calibration and metric reconstruction in spite of varying and unknown internal camera parameters," in *International Conference on Computer Vision*, 1998.

[104] L. Quan and Z. Lan, "Linear n-point camera pose determination," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 21, pp. 774–780, July 1999.

[105] A. Rahimi and T. Darrell, "Location estimation with a differential update network," in *Neural Information Processing Systems*, pp. 1049–1056, 2002.

[106] A. Rahimi, L.-P. Morency, and T. Darrell, "Reducing drift in parametric motion tracking," in *International Conference on Computer Vision*, pp. 315–322, 2001.

[107] S. Ravela, B. Draper, J. Lim, and R. Weiss, "Adaptive tracking and model registration across distinct aspects," in *International Conference on Intelligent Robots and Systems*, pp. 174–180, 1995.

[108] J. Rekimoto, "Matrix: A realtime object identification and registration method for augmented reality," in *Asia Pacific Computer Human Interaction*, 1998.

[109] A. Ruf, M. Tonko, R. Horaud, and H.-H. Nagel, "Visual tracking by adaptive kinematic prediction," in *Proceedings of International Conference on Intelligent Robots and Systems*, pp. 893–898, September 1997.

[110] F. Schaffalitzky and A. Zisserman, "Multi-view matching for unordered image sets, or "How do I organize my holiday snaps?"," in *Proceedings of European Conference on Computer Vision*, pp. 414–431, 2002.

[111] C. Schmid and R. Mohr, "Local grayvalue invariants for image retrieval," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, pp. 530–534, May 1997.

[112] S. Sclaroff and J. Isidoro, "Active blobs," in *International Conference on Computer Vision*, pp. 1146–1153, 1998.

[113] S. Se, D. G. Lowe, and J. Little, "Mobile robot localization and mapping with uncertainty using scale-invariant visual landmarks," *International Journal of Robotics Research*, vol. 22, no. 8, pp. 735–758, 2002.

[114] Y. Shan, Z. Liu, and Z. Zhang, "Model-based bundle adjustment with application to face modeling," in *International Conference on Computer Vision* (Vancouver, Canada), July 2001.

[115] J. Sherrah and S. Gong, "Fusion of 2D face alignment and 3D head pose estimation for robust and real-time performance," in *Proceedings of IEEE International Workshop on Recognition, Analysis and Tracking of Faces and Gestures in Real-Time Systems*, September 1999.

[116] J. Shi and C. Tomasi, "Good features to track," in *Conference on Computer Vision and Pattern Recognition* (Seattle), June 1994.

[117] G. Simon and M.-O. Berger, "A two-stage robust statistical method for temporal registration from features of various type," in *International Conference on Computer Vision* (Bombay, India), pp. 261–266, January 1998.

[118] G. Simon and M.-O. Berger, "Pose estimation from planar structures," *Computer Graphics and Applications*, vol. 22, pp. 46–53, November 2002.

[119] G. Simon, A. Fitzgibbon, and A. Zisserman, "Markerless tracking using planar structures in the scene," in *International Symposium on Mixed and Augmented Reality*, pp. 120–128, October 2000.

[120] I. Skrypnyk and D. G. Lowe, "Scene modelling, recognition and tracking with invariant image features," in *International Symposium on Mixed and Augmented Reality* (Arlington, VA), pp. 110–119, November 2004.

[121] C. Sminchisescu and B. Triggs, "Estimating articulated human motion with covariance scaled sampling," *International Journal of Robotics Research*, vol. 22, no. 6, pp. 371–393, 2003.

[122] S. M. Smith and J. M. Brady, "SUSAN – A new approach to low level image processing," Tech. Rep. TR95SMS1c, Oxford University, Chertsey, Surrey, UK, 1995.

[123] A. State, G. Hirota, D. Chen, W. Garett, and M. Livingston, "Superior augmented reality registration by integrating landmark tracking and magnetic tracking," *Computer Graphics, SIGGRAPH Proceedings,*, pp. 429–438, July 1996.

[124] P. Sturm and S. Maybank, "On plane-based camera calibration: A general algorithm, singularities, applications," in *Conference on Computer Vision and Pattern Recognition*, pp. 432–437, June 1999.

[125] I. Sutherland, "Sketchpad: A Man Machine Graphical Communications System," Technical Report 296, MIT Lincoln Laboratories, 1963.

[126] R. Swaminathan and S. K. Nayar, "A perspective on distortions," in *Conference on Computer Vision and Pattern Recognition*, June 2003.

[127] R. Szeliski and S. Kang, "Recovering 3-D shape and motion from image streams using non linear least squares," *Journal of Visual Communication and Image Representation*, vol. 5, no. 1, pp. 10–28, 1994.

[128] S. Thrun, D. Fox, W. Burgard, and F. Dellaert, "Robust monte carlo localization for mobile robots," *Artificial Intelligence Journal*, vol. 128, no. 1–2, pp. 99–141, 2000.

[129] C. Tomasi and T. Kanade, "Detection and Tracking of Point Features," Technical Report CMU-CS-91-132, Carnegie Mellon University, April 1991.

[130] C. Tomasi and T. Kanade, "Shape and motion from image streams under orthography: A factorization method," *International Journal of Computer Vision*, vol. 9, no. 2, pp. 137–154, 1992.

[131] B. Tordoff, W. Mayol, T. de Campos, and D. Murray, "Head pose estimation for wearable robot control," in *British Machine Vision Conference*, pp. 807–816, 2002.

[132] B. Triggs, P. McLauchlan, R. Hartley, and A. Fitzgibbon, "Bundle adjustment – a modern synthesis," in *Vision Algorithms: Theory and Practice*, (B. Triggs, A. Zisserman, and R. Szeliski, eds.), pp. 298–372, Springer-Verlag, 2000.

[133] R. Tsai, "A versatile cameras calibration technique for high accuracy 3D machine vision mtrology using off-the-shelf TV cameras and lenses," *IEEE Journal of Robotics and Automation*, vol. 3, no. 4, pp. 323–344, 1987.

[134] T. Tuytelaars and L. VanGool, "Wide baseline stereo matching based on local, affinely invariant regions," in *British Machine Vision Conference*, pp. 412–422, 2000.

[135] M. Uenohara and T. Kanade, "Vision-based object registration for real-time image overlay," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[136] L. Vacchetti, V. Lepetit, and P. Fua, "Combining edge and texture information for real-time accurate 3D camera tracking," in *International Symposium on Mixed and Augmented Reality* (Arlington, VA), November 2004.

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

[138] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Conference on Computer Vision and Pattern Recognition*, pp. 511–518, 2001.

[139] D. Wagner and I. Schmalsteig, "First steps towards handheld augmented reality," in *Proceedings of the 7th International Conference on Wearable Computers* (New York, USA), 2003.

[140] G. Welch and G. Bishop, "An introduction to Kalman Filter," Technical Report TR95-041, Department of Computer Science, University of North Carolina, 1995. http://www.cs.unc.edu/˜welch/kalman/.

[141] C. Wren, A. Azarbayejani, T. Darrell, and A. Pentland, "Pfinder: Real-time tracking of the human body," in *Photonics East, SPIE*, 1995.

[142] A. Yuille, P. Hallinan, and D. Cohen, "Feature extraction from faces using deformable templates," *International Journal of Computer Vision*, vol. 8, no. 2, pp. 99–111, 1992.

[143] Z. Zhang, "A Flexible New Technique for Camera Calibration," http://research.microsoft.com/ zhang/Calib.

[144] Z. Zhang, "A flexible new technique for camera calibration.," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 1330–1334, 2000.

[145] Z. Zhang, R. Deriche, O. Faugeras, and Q. Luong, "A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry," *Artificial Intelligence*, vol. 78, pp. 87–119, 1995.