Vision-Based Navigation – A Survey

Henrik I Christensen & Jan-Olof Eklundh
Centre for Autonomous Systems
Numerical Analysis and Computing Science
Royal Institute of Technology
S-100 44 Stockholm, Sweden

{hic,joe}@nada.kth.se

Abstract

A brief review of the functionalities needed for vision based navigation is provided together with a discussion of the basic characteristics of the available algorithms. From the discussion a set of issues related to current research is outlined. It is finally discussed how available methods may be integrated into operational systems for applications like service robotics.

1 Introduction

Vision for navigation has been an active area of research for more than two decades. Gradually more advanced techniques for vision based control have been developed. Initial work was focussed on mapping of structured environments so as to allow obstacle detection and navigation in a well controlled setting. Much of this work adopted a geometric approach to interpretation of the environment. In parallel an effort was devoted to identification of structures in the environment to allow estimation of ego-position. This information is useful for initialization of the system and it also provides valuable information for position based servoing. Again much of the work has been approached from a geometric point of view, with little consideration to the intended application. Recently significant interest has been directed towards methods that integrate vision and control into closely coupled behaviours that allow for direct control of robots to navigation in complex environments. In such work the interaction between vision and other components of a system is considered explicitly. Vision is here used in a control loop to servo on structures in the environment. It is, however, characteristic that few of these methods have been integrated into fully operational systems.

In this paper different techniques used for visual navigation are surveyed and the characteristics of the techniques are outlined. For the discussion the techniques are grouped into three major categories according to the outline above, i.e.:

- Mapping of the environment
- Landmark recognition
- Visual servoing

Based on the survey of techniques it is suggested how a combination of these techniques can be used for indoor navigation in the context of a service robot. It should be noted that this presentation will only consider navigation for mobile platforms, as the techniques needed for vision based manipulation can be considered as a subset of these.

2 Mapping of the Environment

One of the earliest reports of vision based navigation was the Stanford Cart developed by Moravec [38, 39]. In this work a binocular set of cameras is used for recovery of structures in an indoor environment. The structures (cones on a floor) were approximated as shallow objects. Once the position and the size of the objects had been determined they were mapped onto the floor plane as a 2-D region representing an obstacle. The floor plane was thus segmented into regions of free-space and obstacles, which

allowed for determination of a path from the current position to a goal point. Stereo reconstruction is here used only to delimit regions sticking out of the plane, as such objects are considered obstacles. Many more advanced approaches to stereo reconstruction have been reported in literature, see [12] for a review. One example of more advanced used of stereo reconstruction for mobile navigation is the work by Faugeras et al. at INRIA [3, 4]. In this work a "complete" model of the environment is maintained through fusion of multiple stereo maps using Kalman filtering. The model of the environment is here used for explicit navigation to given locations, i.e. the reconstructed model of the environment is matched to an a-priori given model, to allow identification of obstacles and identification of known structures.

One problem with stereo reconstruction is the need for accurate calibration of the stereo-rig to allow computation of 3D information from the disparity map. To partly resolve this problem it has recently been demonstrated how structure can be recovered from uncalibrated stereo cameras [18]. From an uncalibrated stereo set it is possible to recover structure modulo a projective transformation. Even with the uncalibrated setup it is, however, possible to determine qualitative depth which allow determination of a depth ordering among objects. From this it is possible to focus attention on the closest objects.

For navigation in an indoor environment (i.e., on a planar surface), it is often adequate to know the location of objects on the floor place, as initially described by Moravec [39]. An efficient method to perform object-floor segmentation is through use of the inverse-perspective transformation introduced by Mallot et al. [35]. In this method images from two cameras are mapped back onto the floor plane. And the mapped images are then subtracted. Any structure not lying in the floor place will violate this inverse transformation and show up as a difference region, corresponding to an objects above or below the floor plane. This method allow for easy computation of local floor maps. Again it does, however require good calibration to be useful.

Another method for mapping of the environment is through use of motion information. If there are independently moving objects within the field of view this will give rise to a motion field that allow segmentation of the object from the background. When the platform is moving other structures in the environment will also introduce a motion parallax. This motion information can be used to locate other moving agents and for mapping/reconstruction of the environment [36]. Let the robot move with a translation velocity of $\vec{\mathbf{V}} = (V_x, V_y, V_z)^T$ and angular velocity $\vec{\mathbf{\Omega}} = (\Omega_x, \Omega_y, \Omega_z)^T$. If $\vec{\mathbf{X}}$ then denote a point on an object, with respect to the camera coordinate system, and $\vec{\mathbf{x}} = (x, y, 1)^T$ is the corresponding point on the image place at $Z = 1$, then the motion of the point is [25]:

$$\dot{\mathbf{x}} = \frac{\hat{\mathbf{z}}}{\hat{\mathbf{z}}^T \vec{\mathbf{X}}} \times (\vec{\mathbf{X}} \times \vec{\mathbf{x}}) + \hat{\mathbf{z}} \times (\vec{\mathbf{x}} \times (\vec{\mathbf{x}} \times \vec{\mathbf{\Omega}}))$$

where $\hat{\mathbf{z}}$ is a unit vector in the direction of the optical axis. Knowing the velocity of the platform and being able to estimate the image motion it becomes possible to estimate the structure of objects in the environment [34]. Using for example optical flow [26, 25] it is possible to generate a flow field along image contours. The physical interpretation of the motion field has been widely discussed in the literature, good examples of this are [29, 28]. In addition to establishing structure from motion, it is also possible to find invariance motion patterns that contains critical information about the motion of the observer and other objects [28]. One example of this is the divergence of the motion field due to motion towards or away from objects, normally termed the Focus of Expansion (FOE). From the motion field it is further possible to compute the time-to-collision (TTC) (i.e. when the images motion goes to infinity). By minimizing the TTC the robot will basically traverse the safest possible path in the environment. These techniques have been used extensively for early navigation, see for example [25, 27, 1, 37, 40, 45].

As for stereo reconstruction it is possible to recover a simplified flow field through projection of the flow onto the ground plane, which allow for extraction of obstacles that does not conform to the motion estimated for the background/ground plane. An example of this can for example be found in [9].

A recurring problem with motion based techniques is the computation of reliable motion information. In most cases it is difficult to compute the correct optical flow and the computations are at the same time noise sensitive. This has first of all led to studies of the noise sensitivity [11]. In addition it has led to studies

of how uncalibrated vision can be used for obstacle detection and navigation [43]. More importantly it has, however, also resulted in new methods that allow construction of qualitative maps that allow determination of motion invariant and qualitative/ordinal depth maps [20]. Using such an approach several of the traditional noise and motion constraints can be removed at the methods become more tractable for navigation in real environment.

One example of qualitative based navigation using processing dedicated to the specific problem at hand can be found in [21], where qualitative vision maps are extracted to allow control of a mobile platform or a static manipulator.

3 Landmark Recognition

While navigation among obstacles is an important task, it must be accompanied by methods for estimating ego-position in order to allow operation in real settings. To achieve this it is necessary to have access to a map of the environment, which specifies the layout and identifies structures that may be used for computing the position of the robot. Initial work on ego-position estimation relied on explicit floor maps of the environment in terms of a 2D model of the environment and in some cases even a 3-D (CAD) model [33, 19]. Using a 3-D model of the environment, it is possible to extract features like linear structures and subsequently match them again the model. In practical terms a pose estimation is performed in order to allow computation of ego-position with respect to the model [17]. To enable handling of outliers, noisy features etc. it is convenient to use robust matching estimation techniques as described in [33]. Recently such an approach has also been used for navigation in complex environments [5, 24].

An alternative to full CAD models is to use models for particular structures (landmarks) in the environment, like doorways, windows, or corridors for local estimation of the position. I.e., in most cases a rough estimate of the position can be derived from odometric feedback and it is then only necessary to correct the position estimate. Examples of such an approach can be found in [32, 10].

Recently there has also been significant interest in image based techniques for position / landmark recognition. The idea is here to acquire a rich set of images of the environment. During operation the robot will then perform a matching between the image acquired from its current position and the images acquired during a learning phase. Based on an estimate of the pose of the robot with respect to the learned images, it is simple to compute the current position as the position of the learned images is stored in the environmental images. One example of such an approach can be found in [2]. In general the storage of many images of the environment is expensive in terms of memory. To circumvent this problem various approaches like Principal Component Analysis have been applied [41, 44]. One problem with the use of landmark based navigation using raw images is the need for matching of images with a large baseline distance. This problem has recently been addressed by several researchers but so far no satisfactory/robust solution has been reported.

4 Visual Servoing

The approaches outlined sections 2 and 3 all have their outset in basic computer vision problems. When trying to construct integrated systems it becomes obvious that there is a need to integrate the computer vision techniques with other methods in order to allow for efficient processing. A tightly coupled perception-action pair is often referred to as a behaviour. Such behaviour based approaches to robotics are often attributed to Brooks [6, 7, 8].

One of the earliest to adopt an integrated approach to visual navigation or visual servoing was the group of Dickmanns [16, 14, 15, 13]. The application is here outdoor navigation for a car driving at high speed on the high-way. In this approach the structure of a lane (with markings etc.) is used for dedicated visual servoing to allow the car to stay in a particular lane. Gradually the developed system has been extended to its state today, when fully autonomous driving on regular road is possible, see [13] for a description of the latest system. It is characteristic for this system that the outset is the overall system design and the requirement for control of the car, which has motivated the visual processing and led to efficient vision algorithms.

Another example of tight integration between vision and control is the work at Yale by

Hager [22, 23]. In this approach the relation between image features and changes in robot position is modeled by the image Jacobian. Using this transformation is becomes possible to perform servoing in the image plane and relate changes direct to the navigation of the overall system. This approach has the advantage that it does not require a projection back into the world as the error term, usually used in control, has been projected onto the image plane. An example of how this method may be used for navigation of an indoor robot can be found in [30, 31]. A problem in this approach is that the image Jacobian is dependent on the position of the camera (i.e. it assumed a rigid camera transformation that is well-known).

Another approach to visual servoing is though use of an image database of the environment as outlined in section 3. An image is here acquired at the current position and compared to an a-priori image. Through computation of the location of the epi-pole for the prior image it becomes possible to servo on the position of the epi-pole. When the robot arrives as the epi-pole, i.e. when the images are the same, the robot has arrived that the position corresponding to where the prior image was recorded. Such an approach has for example been reported by [42]. One problem with this approach is that the estimation of the position of the epi-pole often is noisy and as the distance between the position of the robot and the location where the prior image was recorded is reduced the sensitivity to changes becomes small. There is thus a need for careful selection of good prior images and there is a need for a rather dense sampling of the environment.

5 Building Systems

For the design of operational systems it is not obvious that any one of the methods outlined above are adequate. For a systems like an intelligent service robot that is to carry out navigation and manipulation in a domestic setting these methods must be combined into operational systems.

For navigation between locations it is possible to exploit landmark based methods based on either raw image models (provided the large baseline problem can be solved) or prototypical models of structures like doorways, furniture, and domestic appliances like refrigerators can be used for local feature based servoing.

For detection of obstacles methods like inverse perspective mapping or optical flow can be used. Here a dominating problem is the figure ground problem, that must be solved in a robust manner to ensure efficient operation.

Finally, for doing manipulation there is a need for object recognition and servoing on recognised structures to allow pick-up of objects. Here the problem of fusion with other sensory modalities becomes apparent. To perform robust pick-up of objects it is necessary for shift control between visual servoing and force/torque and/or tactile control as contact with the object is encountered. This problem is yet to be studied in its full complexity.

While an abundance of methods already are available for performing a rich set of these tasks there is still a need for basic research in combination of visual methods so as to allow interaction with the rich environment found for instance in domestic applications. It is here key that a *systems point of view* is adopted from the outset.

Acknowledgement

This research has been sponsored by the Swedish Strategic Research Foundation, this support is gratefully acknowledged.

References

[1] Y. Aloimonos, editor. *Visual Navigation.* Computer Vision. Lawrence Erlbaum Associates, Mahwah, NJ, 1997.

[2] C. S. Andersen, S. Jones, and J.L. Crowley. Apperance based processes for visual navigation. In C. Bautigam H.I. Christensen and C. Ridderström, editors, *5th Symposium on Intelligent Robotic Systems*, pages 227–236, Stockholm, July 1997.

[3] N. Ayache and O. Faugeras. Maintaning representations of the environment of a mobile robot. *IEEE Transactions on Robotics and Automation*, 5(6):804 – 819, December 1989.

[4] N. Ayache and F. Lustman. Trinocular stereo vision for robotics. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(1):73–84, January 1991.

[5] M. Beetz, W. Burgard, D. Fox, and A. B. Cremers. Active localisation for service robot applications. In C. Bautigam H.I. Christensen and C. Ridderström, editors, *5th Symposium on Intelligent Robotic Systems*, pages 175–187, Stockholm, July 1997.

[6] R. Brooks and L. A. Stein. Building brains for bodies. *Autonomous Robots*, 1(1), 1994.

[7] R. A. Brooks. A robust layered control system for a mobile robot. *IEEE Journal of Robotics and Automation*, RA - 2.(1):14 – 23, March 1986.

[8] R. A. Brooks. Intelligence without representation. *A.I. Memo*, 1263, 1990.

[9] S. Carlsson and J.-O. Eklundh. Object detection using model based prediction and motion parallax. In O. Faugeras, editor, *Proc. 1st European Conference on Computer Vision*, volume vol. 427 of Lecture Notes in Computer Science, pages 297–306, Berlin, April 1990. Springer Verlag.

[10] H. I. Christensen, N.O.S. Kirkeby, S. Kristensen, L. F. Knudsen, and E. Granum. Model-driven vision for in-door navigation. *Robotics and Autonomous Systems*, 12:199–207, 1994.

[11] K. Daniilidis and M. E. Spetsakis. *Visual Navigation*, chapter Understanding Noise Sensitivity in Structure from Motion, pages 60–88. Computer Vision. Lawrence Erlbaum Associates, Mahwah, NJ, 1997.

[12] U.R. Dhond and J. Aggarwal. Structure from stereo - a review. *Systems, Man and Cybernetics*, 19(6):1489–1511, Nov./Dec. 1989.

[13] E. Dickmanns. Autonomous vehicle control. In *Proc. IJCAI 1997*, Osaka, August 1997. IJCAI Inc. Invited Talk.

[14] E. D. Dickmanns and V. Graefe. Dynamic monocular machine vision. *Machine Vision and Applications*, 1:223–240, 1988.

[15] E. D. Dickmanns, B. Mysliwetz, and T. Christians. An integrated spatio-temporal approach to automatic visual guidance of autonomous vehicles. *IEEE Transactions on Systems, Man and Cybernetics*, 37(6):1273–1284, November/December 1990.

[16] E. D. Dickmanns and A. Zapp. A curvature-based scheme for improving road vehicle guidance by computer vision. In William J. Wolfe and Nelson Marquina, editors, *Mobile Robots*, pages 161–168. SPIE, Bellingham, 1987. vol. 727.

[17] R. M. Dolezal, T. N. Mudge, J. L. Turney, and R. A. Volz. Determining the pose of an object. In O. D. Faugeras and Robert Kelley, editors, *Computer Vision for Robots*, pages 68–71, 1986. Proc. SPIE 595.

[18] O. Faugeras. *Three-dimensional computer vision: A geometric Viewpoint.* MIT Press, Cambridge, MA, 1993.

[19] C. Fennema, A. Hanson, E. Riseman, J.R. Beveridge, and R. Kumar. Model-directed mobile robot navigation. Technical Report COINS TR 90-42, University of Massachusetts, Computer and Information Science, June 1990.

[20] C. Fermüller and Y. Aloimonos. Qualitative egomotion. *Intl. Jour. of Computer Vision*, 15:7–29, 1995.

[21] V. Grafe. perception and situation assessment for behaviour based robot control. In *Intelligent Autonomous Systems*, page (To appear), Sapporo, June 1998.

[22] G. D. Hager. Calibration-free visual control using projective invariance. In *ICCV*, pages 1009–1016, 1995.

[23] G. D. Hager. A modular system for robust positioning using feedback from stereo vision. Technical Report YALEU/DCS/RR-1074, Yale University, May 1995.

[24] U. Hanebeck, C. Fischer, and G. Schmidt. Roman: A mobile robotic assistant for in-door service applications. In *IROS-97*, volume vol. 2, pages 518–525, Grenoble, June 1997. IEEE, CS Press.

[25] B. Horn. *Robot Vision.* The MIT Press, Cambridge, Massachusetts, 1986.

[26] B.K.P. Horn and B. Schunk. Determining optical flow. *Artificial Intelligence*, 17:185–203, 1981.

[27] T. S. Huang, editor. *Image Sequence Analysis.* Information Sciences. Springer Verlag, Berlin, 1981.

[28] J. J. Koenderink. Optic flow. *Vision Research*, 26(1):161–179, 1986.

[29] J. J. Koenderink and A. J. van Doorn. Invariant properties of the motion parallax field due to the movement of rigid bodies relative to an observer. *Opt. Acta 22*, pages 773–791, 1975.

[30] J. Košecká. *Supervisory Control of Autonomous Mobile Agents.* PhD thesis, GRASP Laboratory, Departments of Computer Science, University of Pennsylvania, May 1997.

[31] J. Kosecka, H. I. Christensen, and R. Bajcsy. Experiments in behaviour composition. *Journal on Robotics and Autonomous Systems*, 19(3–4):287–318, March 1997.

[32] D. Kriegman, E. Triendl, and T. Binford. Stereo vision and navigation in buildings for mobile robots. *IEEE Transactions on Robotics and Automantion*, 5(6):792 – 802, December 1989.

[33] R. Kumar. *Model Dependent Inference of 3D Information From a Sequence of 2D Images.* PhD thesis, University of Massachusetts at Amherst, February 1992.

[34] H.C. Longuet-Higgins and K. Prazdny. The interpretation of moving retinal images. *Proceedings of the Royal Society, London B*, 208:385–397, 1980.

[35] H.A. Mallot, H.H. Bulthoff, J.J. Little, and S. Bohrer. Inverse perspective mapping simpilfies optical flow computation and obstacle detection. *Biological Cybernetics*, 64:177–185, 1991.

[36] D. C. Marr. *Vision.* W.H. Freeman and Co., San Francisco, CA, 1982.

[37] W. Martin and J. Aggarwal. *Motion Understanding: Robot and Human Vision.* Kluwer Academic Publishers, Boston, 1988.

[38] H. P. Moravec. Towards automatic visual obstacle avoidance. In *Proceedings of Int. Joint. Conf. on Artificial Intelligence in Cambridge,MA*, page page 584, 1977.

[39] H. P. Moravec. The Stanford cart and the CMU rover. *Proceedings IEEE*, 71:872 – 884, 1983.

[40] H. H. Nagel. On the estimation of optical flow: Relations between different approaches and some new results. *Artificial Intelligence*, pages 299–324, 1987.

[41] A. Pentland. Content-based indexing of images and video. *Philosophical Transactions of the Royal Society (Biological Sciences)*, 352(1358):1283–1290, August 1997.

[42] E. Rivlin, S. Dickinson, and R. Basri. Surfing the epi-poles. In *6th ICCV*, page (To appear). IEEE, CS Press, January 1998.

[43] L. Robert, C. Zeller, O. Faugeras, and M. Hebert. *Visual Navigation*, chapter Application of nonmetric vision to some visually guided robotics tasks, pages 89–134. Computer Vision. Lawrence Erlbaum Associates, Mahwah, NJ, 1997.

[44] B. Schiele. *Object Recognition using Multi-Dimensional Receptive Field Histograms.* PhD thesis, Institute Nationale Polytechnique de Grenoble, Grenoble, July 1997.

[45] A. M. Waxman, T. R. Kushner, E. Liang, and T. Siddalingariah. A visual navigation system for autonomous land vehicles. *IEEE Transactions on Robotics and Automation*, 3(2):124–141, April 1987.