Photogrammetrische Woche 1991
Schriftenreihe, Heft 15, 1991
Institut für Photogrammetrie

# REAL-TIME PHOTOGRAMMETRY AND ROBOT VISION

H. Haggrén, Helsinki University of Technology

## 1. INTRODUCTION

A vision system analyzes images and produces descriptions of what is imaged whereas photogrammetry - and especially analytical photogrammetry - provides the coordinating means for spatial reconstruction during the reasoning process. If the physical reality is supposed to change, the vision systems are used for dynamic control of the environment and of the object geometries. For robots, these real-time scene updates are necessary in order to exactly coordinate within its environment and adapt the working procedures to the actual reality.

The vision systems are either 1) hand-held and relatively oriented camera set-ups or 2) fixed camera stations which are externally oriented in relation to the robot coordinate system. Accordingly, the scenes seen by the vision systems change either due to the robot's own movements or due to the actions in the environment. The advantages of real-time photogrammetry become essentially prominent in dynamic cases where fixed vision stations are used for analyzing and coordinating time-varying object geometries and movements.

Within this paper the definitional framework will be first handled, both for vision and for photogrammetry. Then the general tasks of world reconnaissance by robots are outlined and respective sensor solutions for reconstruction are overviewed. Even if the use of analytical photogrammetry for robot vision is so far rare, four different application examples are submitted in order to disclose the potential applicability of photogrammetry to robot vision. Finally, spectacles are provided for wider discussions concerning the future interdisciplinary research activities.

## 2. DEFINITIONAL FRAMEWORK

When examining photogrammetry and its use for robot vision it is necessary to define first the framework of the terminology. In a manner the utilization of photogrammetric techniques, either for monoscopic or stereoscopic robot vision, is already wide and manifold. This is largely not due to the active research carried out by the photogrammetric society, but is far more an immediate result of the straighforward problem solving performed by the robot vision society, when developing intelligent actuators. Nevertheless, all modern photogrammetry has the common origin in the works of perspective and projective geometry and dates back to the second half of 19th century.

Robot vision is a subdiscipline of computer vision. The extremely short reaction times, which are allotted to the entire actuation process including all vision computation, has practically limited the use of photogrammetric procedures to the simplified ones with reduced numbers of perspective parameters. In order to provide some wider relevances to the field of robot vision, we concentrate us later in this paper to some feasible solutions of modern analytical photogrammetry.

### 2.1 Computer vision, machine vision, robot vision

The purpose of computer vision is to make computers capable of understanding environments from visual information [/16/, SHIRAI, 1987]. What can be included to visual information? Visual information, as it relates to digital computation, is any kind of information which once has been digitized, then processed, analyzed, and finally visualized. For all computer vision processes we imply that the inputs are two-dimensional images, either the original ones or their derivatives.

In most cases of computer vision the entire reasoning chain - from imaging to scene description - is a result of several steps of intelligent image processing. The preparatory stage of this reasoning is referred to as image analysis, while the final understanding requires subsequent processing and is called scene analysis [See e.g. /9/, HORN, 1986]. The division is somewhat arbitrary, except insofar as image analysis starts with an image, while scene analysis begins with a sketch of extracted and meaningful symbols. All algorithmic vision research is generally included to the field of computer vision, without any futher division whether the systems are applied to some actual machine or robot related vision task or not [See e.g. /2/, FAUGERAS, 1990].

From engineering viewpoint, according to [/10/, HUANG, 1990], the general goal of vision is to build automated systems capable of analyzing and understanding three-dimensional and time-varying scenes. He further divides the tasks of computer vision into three levels: 1. Reconstruction for recovering of three-dimensional information, 2. Recognition for detecting and identifying objects, and 3. Understanding in order to finally figure out what is going on in the scene. Photogrammetry is primarily a technique of reconstruction.

The proceeding question would be: Where is the environment? Machine vision implies most likely a real scene during the image processing. This is understandable, as vision is now used either for obtaining dynamic information of the present processes, or, for controlling machines and processes with an immediate feedback. HORN characterizes machine vision systems simply according to their outputs. These must satisfy two criteria: 1. They must bear some relationship to what is being imaged, and, 2. They must contain all the information needed for the given task [/9/, HORN, 1986]. If the machines are manipulators, actuators, robots, unmanned vehicles and so forth, we definitely deal with robot vision systems as part of larger entity that interacts with the environment.

## 2.2 Real-time photogrammetry

In general, photogrammetry implies all means used for obtaining reliable geometric information about physical objects and the environment through processes of recording, measuring, and interpreting images [/17/, SLAMA, 1980]. The real-time use was first introduced in the early 70's by the National Research Council of Canada, and later reported in [/12/, KRATKY, 1978] and [/14/, PINKNEY, 1978]. The system, called Real-Time Photogrammetry System, is a monocular, single camera set-up on the end effector of the CANADARM manipulator. It is based on the concept of 1) tracking a square formed set of four target dots on the object and 2) resecting the position and orientation of the end effector according to the obsereved skewness of the target square.

How immediate is real-time? It is certainly relative, and the term 'real-time' in connection with photogrammetry refers solely to the maximum time available for the entire process of recording, measuring and interpreting the images. It may relate to the quick reaction cycles of robots, or to some looser but process depended time specifications, which are set for the application at hand, and is synonymous with the term 'on-line'. The principal requirement is, that the photogrammetric reconstruction should not cause any unexpected delays for the dynamic actions within the physical reality to be controlled. The real problems in keeping the photogrammetric procedure in due course should be faced more like engineering problems than theoretical ones.

In order to cope with the real-time requirements, the photogrammetric reconstruction task can be simplified. The two alternatives are: 1. The recording and measuring of some few object points in each scene only, which is still adequate for e.g. pose determination, or 2. The use of reduced sets of parameters in the collinearity model of perspective imaging, which still gives all necessary data for e.g. scene analysis. Respectively, the reconstructions result either 1) in accurate three-dimensional coordinates, although the environment model would remain incomplete, or 2) in adequately completed scene descriptions, but in a geometrically distorted and non-linear coordinate system.

In the case of autonomously moving vehicles or robots the complete scene descriptions are of primary interest. Therefore the perspective models are generally simplified and the imaging systems are calibrated according to the so called pinhole camera model. The set-ups are arranged for monocular imaging, or, like in the case of human perception, extended to binocular stereo [/9/, HORN, 1986; /16/, SHIRAI, 1987].

It is necessary to distinguish here between the 'real' photogrammetric and the pinhole-based vision collinearity in order to emphasize the unique features of modern analytical photogrammetry. It will be done without any further attitudes to what should be currently the most practical solution for some specific task or problem of robot vision, and in spite of knowing that the reconstruction procedures would in most cases run out only after the given time limits.

The basic concept in analytical photogrammety is that the set-up geometries cope with any number of camera stations and with any kind of camera orientations. Consequently, analytical reconstruction is founded on extended perspective formulation, which is capable of compensating both linear and nonlinear imaging distortions of the presumed collinearity. As it concerns the reliability aspects when determining the values of increasing number of perspective parameters, the coordinate transformations between the two-dimensional images and a common three-dimensional object space are normally overconstrained. The systems become thus redundant and are explicitely capable of continuos selfcalibrating [/7/, HEIKKILÄ, 1990].

In the case of multi-camera set-ups there are two principal ways of applying photogrammetry to robot vision: 1. The triangulation of some distinctive object points, and 2. The stereo restitution of the entire object and its environment. Both result in exact three-dimensional object space coordinates. Regarding the robot vision examples, triangulation only is used for real-time tasks, such as for object tracking [/1/, BEYER et.al., 1989] or pose determination [/4/, HAGGRÉN et.al., 1990]. As the stereo restitution is appropriate for more time-consuming on-line tasks, like for initial and comprehensive scene analysis, for three-dimensional object recognition, or, for geometric object verification, the available photogrammetric examples for robot vision are still partly illusory [/8/, HOBROUGH et.al., 1985].

## 3. RECONSTRUCTION FOR WORLD RECONNAISSANCE

The reconstruction tasks of robot vision can be divided into global and local scales. A descriptive example on global scale would be the one for three-dimensional scene analysis and navigation of autonomously moving vehicles when coping with an unknown environment. The local tasks relate primarily to already known environments, where only close-up changes or objects should be recognized. Typical examples of local tasks are the ones of a) guarding and obstacle avoidance within a robotic cell, or even closer, b) particle sorting or bin picking on a conveyor belt.

All reconstruction is based on imaging and image analysis. The techniques are either active or passive ones. The sensors are installed on the moving robot or manipulator itself, most likely close to the end effector. Usually the final touch will be controlled by other means, like with proximity and force sensors, and the reconstruction techniques are used for approaching the target only and until its near proximity. As the reconstruction accuracy is relative to the distance between the camera and the object and is highest in the near proximity, the disclosures caused by non-linear image distortions or by simplified perspective models will be diminished. Here we shortly summarize those techniques, mainly for close-up tasks, which are commonly presented in the literature [See e.g.: /9/, HORN, 1986; /10/, HUANG, 1990; /16/, SHIRAI, 1987].

The active imaging techniques are based on direct ranging or indirect ranging. In the case of direct ranging, the images are produced using scanning laser spot projections. The range values are derived directly from the time-of-flight or phase shift observations of the back reflected light signal, and the resulting range image is sampled and ordered according to the predetermined scan directions. In the case of indirect ranging, the range information is achieved first after image interpretation. The set-up is like the one of stereo photogrammetry, but instead of the second camera one uses structured light projectors. Presuming that the light projections - light spots, sections or squared grids - are of known internal geometry, the ranges may be triangulated from the corresponding image observations. In moiré-technique, two gratings are projected simultaneously onto the object surface, and the range information is interpreted directly from produced interference patterns.

The passive sensors are either stereoscopic or monocular. Stereoscopic range information is derived from the parallax observations of the image pair according to the known relative orientation of the two cameras. In the case of monocular sensors the range information is inferred from motion, optical flow, shading, texture, phocus and many others. The prerequisite of these so called 'shape-from-X'-techniques is, that the illumination and the surface properties are carefully known. As the use of passive sensors require much more computation than the use of active ones, their main references are more or less on the research side.

## 4. PHOTOGRAMMETRY FOR ROBOT VISION

The examples given here are chosen in order to disclose the potential applicability of analytical photogrammetric techniques to the reconstruction task for vision systems. Although not been all accomplished, it becomes clear, that most of the problems to be solved in this connection are really of engineering nature. With realistic task specifications, efficient use of analog processing and proper modification of image processing algorithms the real-time requirements can be fulfilled, and without loosing the profits of using rigorous photogrammetric formulation.

### 4.1 Space vision

Since 70ies the real-time photogrammetry systems of the National Research Council of Canada have been developed further, primarily as the Canadian contribution to the space station. One example of the industrial robotics applications is the adaptive orientation of a robot for unloading parts from monorail conveyors in automobile factories [/15/, PINKNEY et.al., 1986].

For the space station the primary aim is to build on-orbit, closed-loop controls for the manipulators as a part of a mobile servicing system, and to provide teleoperational techniques for the astronauts [/13/, MACLEAN et.al., 1990]. The servicing system is a collection of robotic elements configured to support the assembly, maintenance and servicing of the space station. The current vision system is of the same architecture of 1) tracking a set of targets of specific geometrical figure and 2) resecting for orientation and pose determination, like mentioned earlier. Here the reconstruction procedure may be additionally combined and supported with range data achieved by a laser scanner. The man-in-the-loop teleoperations will be supported with a stereographic stereovideo system. This should permit even dynamic superimposition of real-world objects, and will be driven on-line by data provided by either the vision system or the laser scanner. The tests of all these integrated operations are currently underway.

### 4.2 High-speed object tracking

At the Swiss Federal Institute of Technology real-time photogrammetric techniques were used for positioning the ball-in-flight for a robot playing table tennis [/1/, BEYER et.al., 1989]. The ball trajectories were tracked in 20 msec time intervals (50 Hz rate) in order to provide the robot with sufficient data for deriving all parameters for returning the ball. The fixed camera set-up consisted of two shuttered CCD-cameras, one placed on the left side and one above the robot. All computation for locating the ball in the images was performed using analog processing. The rest, consisting the calculation of image coordinates, the compensation of systematic imaging errors, and the intersection of three-dimensional coordinates, was processed by software.

The camera set-up was geometrically calibrated using a target plate of 30 balls, which was moved along in four positions. The entire calibration volume covered the expected playing volume and the positions of the targets within this volume were controllrd externally. The verified accuracy of the photogrammetric intersection after the calibration was of 1/10th of the pixel spacing and corresponded to less than one millimeter in all three coordinates within the playing volume. The improvement of the accuracy was in the order of eight compared to the case without compensation of systematic imaging errors.

### 4.3 Adaptive orientation

The application is of car body orientation for robotic manufacturing, which was performed in a joint project of the Helsinki University of Technology and the company Mapvision Ltd [/4/, HAGGRÉN et.al., 1990]. The basis of the photogrammetric station here is the real-time photogrammetric machine vision system called Mapvision. The station consists of a fixed set-up of four cameras and is used in order to provide a sealant robot with actual orientation data of each body on the line. The real-time requirement is of five seconds. The procedure consists of 1) imaging, 2) two-dimensional locating of four natural target features in each of four images, 3) intersection of the three-dimensional coordinates, 4) coordinate transformation between the actual body and the robot coordinate system, and 5) feeding the orientation data for the robot. All photogrammetric processing is performed by software.

The Mapvision systems are calibrated on-site using scale references only, and are capable of continuous self-calibration in order to control automatically all later instabilities of the set-up fixtures of the cameras. The mathematical model compensates all systematical imaging errors. Within this application both the external control points and body orientation points are searched in the images using stored templates and the two-dimensional search is performed by least-squares matching technique. The repeatability of the image coordinates varies between 1/10th to 1/20th of the pixel spacing, and corresponds to an RMS-value after space coordinate transformations of 0.3 to 0.4 mm. The distances between the cameras and the orientation points on the car body vary from 1.5 to 2.5 meters.

### 4.4 Real-time stereopsis

The utmost goal in developing the Stereopsis System has been to build a modular robot vision system, which is analogous with the human vision system, but with exceeding performance as it concerns the metric coordination and speed. The three modules are: 1. The camera head, 2. The three-dimensional converter, and 3. The application package. These notes are derived from a writing, which has been published in 1985, and are no doubt outdated but relevant to be presented in this connection [/8/, HOBROUGH et.al., 1985].

The set-up configuration of the camera head is the one of the normal case of stereo photogrammetry, where two solid state cameras are oriented relative to each other and the image surfaces are controlled for their exact co-planarity. The basic functions for ranging will be thus minimized, and consist of 1) synchronized scanning along homologous image lines, which are parallel to the camera base, 2) finding out homologous features on the left and right side by analog correlation, 3) determining the parallax by the time difference between the scanning of homologous image points, and 4) calculating the range using parallax equations only. The object features of primary interest are here the edges of the objects. The first Stereopsis System was designed to provide a range value for every pixel in the field of view every 17 milliseconds, which is somewhat quicker than the equivalent human functions.

The application package provides the robot with the data about the workspace at global scale and about the objects in it at local scale. The reconstruction tasks are for complete scene descriptions, including the reckognizing and monitoring of all objects to be tracked. The primary applications, which were mentioned, are the ones of reconnaissance, robotic assembly and navigation. As the non-linear imaging errors will be compensated during photogrammetric processing, the system can be used for accurate spatial reconstruction as well, like for geometric inspection. In order to inspect large objects, the reconstruction is progressing, section by section, until the entire object geometry is measured, and then the actual shape and size are compared to what was designed.

## 5. FURTHER OUTLOOK

For autonomous actuators, the photogrammetric vision systems are just a single type of sensors among many others, which all provide actual information about the real world. The main problem for an actuator is to interact within that world while performing its tasks, using all perceptual, cognitive and motor components. The information assimilation is efficient only if these components are controlled by a proper environmental model which should be the heart of the entire system [/11/, JAIN et.al., 1991].

The key issue is, that the ways of representation of the environment model remains reliable. The ideal three-dimensional representation consists of elementary objects, all of which are known by their individual location, orientation, geometry, shape, face, texture, etc. and functionally reasoned by their mutual relationships. The predominant role of photogrammetry is to provide the primary coordination in that three-dimensional space. The natural consequence of this evolution is, that the photogrammetric vision techniques are used not only for adequate self control of the robot themselves, nor for updating and verifying the environmental model, but also for controlling the external processes like in modern manufacturing [/5/, HAGGRÉN et.al., 1990].

The efficient utilization of such environmental model facilitates the real-time photogrammetric reconstruction procedures as well. After the scenes are ones described, the reconstruction task will be reduced to reckognizing and locating the actual changes only. Now we come to a third alternative of simplifying the reconstraction procedure: Instead of using rectangular euclidean geometry for perspective modelling, we will take use of projective geometry and perform the reconstructions through projective transformations [/3/, FUCHS, 1989].

Using proper problem formulation the projective transformations may be performed along two-dimensional planes without loosing the rigorous linearity [/6/, HAGGRÉN et.al., 1990]. They are easier of their solutions, as the orthogonality constraints are not included. However, all reconstruction results in rectangular three-dimensional coordinates, if some global features, which remain stable, are simultaneously used as projective references. This is due to the fact, that the rectangular euclidean geometry of the global space, according to which the environmental model is coordinated, induces a subordinate euclidean geometry in every actual plane chosen for the projective transformations.

## 6. CONCLUSION

It is evident that the rigorous analytical formulation of photogrammetry provides for robot vision a coordinating base, which is accurate and reliable throughout the entire working space. The real-time photogrammetric vision systems have been primarily developed during the last decade, first for space applications, but henceforth increasingly for wider industrial use. The few examples, which were presented here, indicate that the main progress will be on photogrammetric stations of fixed camera set-ups. There the dynamic reconstruction tasks can be solved by triangulation of some distinct points only, like in the cases of pose determination or dynamic object tracking.

The current trend in designing integrated autonomous systems will definitely be the proper basis for wider development of photogrammetric vision techniques, too. Within these systems the photogrammetric reconstruction procedure will be supported and controlled by the system itself, providing combined information, which comes from multiple sensors, from action control, and from the high-level reasoning process. The perhaps most interesting observation during this writing was, that there still are some potentials in reserve even within the photogrammetric theory. These potentials come up from the projective geometry, and should be thoroughly restudied without any prejudices.

## REFERENCES

/1/ BEYER, H. A., FÄSSLER, H. P. and WEN, J.: Real-Time Photogrammetry in High-Speed Robotics, Optical 3-D Measurements Techniques, A. Gruen, H. Kahmen, Editors, Wichmann, 1989, pp. 271-280.

/2/ FAUGERAS, O. (ed.): Computer Vision - ECCV 90, Springer-Verlag, 1990.

/3/ FUCHS, H.: Real-Time Algorithms for the Orientation of Robot Camera Systems, Optical 3-D Measurements Techniques, A. Gruen, H. Kahmen, Editors, Wichmann, 1989, pp. 456-469.

/4/ HAGGRÉN, H. and HAAJANEN, L.: Target Search Using Template Images, Close-Range Photogrammetry Meets Machine Vision, A. Gruen, E. Baltsavias, Editors, Proc. SPIE 1395, 1990, pp. 572-578.

/5/ HAGGRÉN, H. and HEIKKILÄ, J.: Photogrammetry for Statistical Process Control, Vision '90 Conference Proceedings, K. W. White, Editor, SME, 1990, pp. 4:29-37.

/6/ HAGGRÉN, H. and NIINI, I.: Relative Orientation Using 2-D Projective Transformations, International Archives of Photogrammetry and Remote Sensing, Vol. 28, Part 3/2, Wuhan, 1990, pp. 234-245.

/7/ HEIKKILÄ, J.: Update Calibration of a Photogrammetric Station, Close-Range Photogrammetry Meets Machine Vision, A. Gruen, E. Baltsavias, Editors, Proc. SPIE 1395, 1990, pp. 1234-1241.

/8/ HOBROUGH, G. L. and HOBROUGH, T. B.: A Future for Realtime Photogrammetry, Mensuration, Photogrammétrie, Génie rural, 9/85, pp. 312-315.

/9/ HORN, B. K. P.: Robot Vision, MIT Electrical Engineering and Computer Science Series, The MIT Press & McGraw-Hill, 1986.

/10/ HUANG, Th. S.: Computer Vision and Dynamic Scene Analysis, tutorial print, ETH-Zürich, 1990.

/11/ JAIN, R. and ROTH, Y.: Towards Integrated Autonomous Systems, Applications of Artificial Intelligence IX, M. M. Trivedi, Editor, Proc. SPIE 1468, 1991, 99. 188-201.

/12/ KRATKY, V.: Analytical Study of Photogrammetric Solution for Real-Time Three-Dimensional Control, International Archives of Photogrammetry, Vol. 22, Part V.2, 1978.

/13/ MACLEAN, S. G., RIOUX, M., BLAIS, F., GRODSKI, J., MILGRAM, P., PINKNEY, H. F. L. and AIKENHEAD, B. A.: Vision System Development in a Space Simulation Laboratory, Close-Range Photogrammetry Meets Machine Vision, A. Gruen, E. Baltsavias, Editors, Proc. SPIE 1395, 1990, pp. 8-15.

/14/ PINKNEY, H. F. L.: Theory and Development of an On-Line 30 Hz Video Photogrammetry System for Real-Time 3-Dimensional Control, International Archives of Photogrammetry, Vol. 22, Part V.2, 1978.

/15/ PINKNEY, H. F. L. and PERRATT, C. I.: A Flexible Machine Vision Guidance System for 3-Dimensional Control Tasks, International Archives of Photogrammetry and Remote Sensing, Vol. 26, Part V, 1986, pp. 414-423.

/16/ SHIRAI, Y.: Three-Dimensional Computer Vision, Springer-Verlag, 1987.

/17/ SLAMA, C. C. (ed.): Manual of Photogrammetry, Fourth Edition, American Society of Photogrammetry, Falls Church, Virginia, 1980.

## ABSTRACT

The ideal goal for all robot vision research and development is to omit the need of human intervention during automated operations. Real-time photogrammetry is the adequate technique which provides the necessary space coordination for all vision based scene descriptions. Due to the high real-time requirements, the photogrammetric vision models are usually simplified, which causes disclosure problems regarding the environmental model in global scale. When using extended analytical photogrammetric models, this problem is avoided, which is exemplified in four different cases. Finally, some ideas are presented about the future trends when approaching the era of integrated autonomous systems.

Henrik Haggrén
Helsinki University of Technology
Institute of Photogrammetry and Remote Sensing
Otakaari 1 F
SF-02150 Espoo, Finland