Toward Visual Intelligence of Service Robot

Ming XIE
School of Mechanical & Production Engineering
Nanyang Technological University
Singapore 639798
http://www.ntu.edu.sg/home/mmxie/me/me.htm

Abstract

This paper reports some achievements made under the ARC1/94 project funded by the Ministry of Education (Singapore). These include: (1) The development of a theoretical framework for robotic hand-eye coordination with uncalibrated cameras; (2) The development of a deterministic solution for robotic head-eye coordination; (3) The development of a simple and practical solution for ground plane obstacle detection which is an indispensable function that any autonomous mobile robot should have, and (4) The development of a quasi closed form solution to solve feature correspondence in rotational motion stereo.

1 Introduction

It is commonly believed now that robotics research reaches the stage of looking into the aspect of intelligence which will be the key issue to make industrial & service robots more user-friendly and hence more useful than current cases. The challenge here is how to define machine intelligence that we want to duplicate on robots. The understanding of intelligence by human beings is quite subjective. It may be difficult to objectively state the definition of machine intelligence. Then, the question is whether it is still possible to build intelligent robot without knowing precisely what we mean by robot's intelligence ? (it is easy to say that an intelligent system must be reactive, sensitive, and have certain abilities to understand and to communicate, etc). We believe that one possible way of developing intelligent robot is to make robot imitate nature or human beings. For the case of understanding the role that a machine vision system can play inside a service robot, one strategy is to think of the role being played by human vision in connection with human head, human arms/hands, and human legs/body. Following this way of thinking, it is easy for us to list out some intelligent human behaviors guided by human vision, for example:

1. Head-Eye Coordination (Human eyes can easily guide human head's motion in searching for or looking at objects of interest).
2. Hand-Eye Coordination (Human eyes can effortlessly guide human arms/hands in doing very complicated actions or operations such as grasping objects, doing painting, playing sports, etc).
3. Leg-Eye Coordination (Human beings can easily move around an unknown environment without damaging our body by avoiding collision with other objects. This is greatly due to our visual perception capability).
4. Description of Any 3D Unknown Scene (Human eyes are able to describe, with the help of human brain, the content of an unknown scene without too much difficulty in general).

The next section describes a viewpoint of interaction between vision system and robot control in the context of service robot. The subsequent sections report our achievements made in this field of vision-guided execution of robot tasks. The last section concludes this paper.

2 Vision and Execution Control of Robot Task

Generally speaking, service robot refers to robot that can provide useful services/works to both human beings and machines themselves. In order to build such a robot, one strategy is to integrate the following major systems into a common platform (called "service robot"):

- Wireless communication system (for robot to dialogue with remote systems or human beings).
- Speech recognition system (for robot to understand voice message).
- Speech synthesis system (for robot to talk to human beings).
- Mobile base system (for robot to move around).

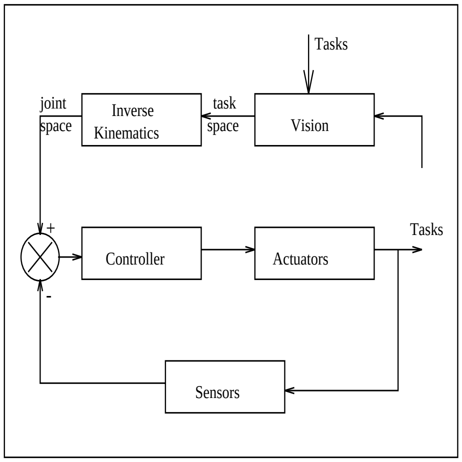

Figure 1: Vision and execution control of robot task.

- Manipulation/hand system (for robot to do actions/operations).
- Vision system (for robot to see and communicate image data).

If we look at the progress made in the computer science engineering and its industry, it is easy to notice that it is no more a dream to develop a computer (a kind of robot's head or brain) that can communicate, talk, understand voice message and think. In robotics, it is no more a difficult thing to develop mechanisms that are able to drive the motions of robot's body (legs or wheels) and robot's head/arm/hands. But, one important limiting factor for the intelligence/flexibility of current robot is certainly in robot's capability of performing and integrating visual perception with the control of robot's motions (legs, wheels, head, or arms/hands, etc). A lot of expectations and efforts have been put on the development of general purpose 3D vision system (eg., stereovision) because of the common believe that the availability of accurate 3D information will be the only way to enable the effective guidance of robot's motions in a 3D space.

However, an important question that can be raised here is: What are the possible roles that a vision system can play inside a robot system ? Most people believe that a vision system must be considered as a sensor that are used to measure 2D or 3D geometric information related to a 3D space. Restricted by this viewpoint, a lot of research works have been dedicated to the topic of visual servoing where a vision system directly involves in the feedback control loop to generate feedback information [2]. There are two inconveniences with this strategy of development:

1. How to analyse the convergence and stability of the motion control system that includes visual feedback control loop ?
2. How to manage the complexity of robot's motion control if the closed control loop includes a vision system [5] ?

Fortunately, these two problems can be overcome by adopting an alternative viewpoint on the role of a vision system inside a robot. Instead of thinking of using vision system as a sensor, one can think of the role of task execution control that a vision system can play during the course of conventional motion control of robot system. Fig.1 illustrates this idea. We can see in Fig.1 that the vision system is not included in the control loop of robot's actuators. More precisely, the vision system acts as an "on-line motion planner" or "task execution controller". It takes input from both the desired tasks and the current execution (path or trajectory) of tasks made by robot. The output from the vision system is just a new motion that the robot is going to execute in the next iteration. Under this strategy, the following interesting points hold:

1. The stability of the whole system depends on the initial stability condition of the robot's motion control system.
2. The convergence condition is ensured by the vision system and it is easy to prove the convergence in the framework of vision system.
3. The inclusion of vision system does not interfere with the original motion control loop of robot system. So, it is easy to manage the complexity of the whole system.

The works reported in the following three sections give practical examples that support the idea of considering a vision system as "on-line motion planner" or "task execution controller".

3 Head-Eye Coordination: A Closed Form Solution

The generic problem of robotic head-eye coordination can be stated as how to use data of vision system mounted on a robotic head to compute and drive the head's motion so as to satisfy one or both of the following two conditions:

- the image seen by the vision system equals to a reference image.
- the relative position of the head with respect to a static or moving target does not change.

Fig.head-eye schematically illustrates the robotic head-eye coordination problem. So far, all solutions for solving robotic head-eye coordination are iterative

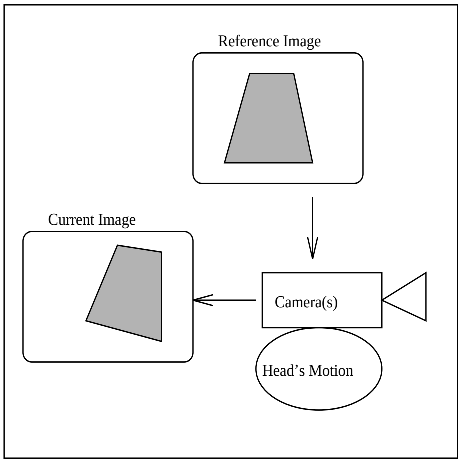

Figure 2: Vision and control of robot's head motion.

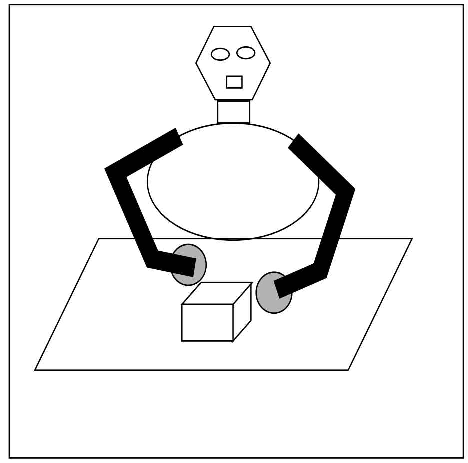

Figure 3: Vision and control of robot's hand motion.

in nature [5]. In [10], we have proposed a deterministic method. This method can be shortly summarized by the following proposition:

> **Proposition I**: If the location of a vision system in a 3D space at one time instance is defined as a view, then any third unknown view (ie., a reference view where a reference image has been taken) of the vision system can be determined from any pair of two known views. If a set of at least six non-collinear line segments are available, the computation of a third unknown view just requires the solution of a linear equation system.

One typical example of application of head-eye coordination is road-following or car-following in a motor highway. If the image (seen by human driver or a vision system) of a car in front of us does not change, we can say that the distance between this car and our car does not change. More examples of application can be found in electronics, manufacturing and defence industries [10].

4 Robotics Hand-Eye Coordination: The Simplest Solution

The capability of doing manipulations with robot relies on the motion control of robotic arm or hand. The basic case of motion control is the motion control between two points (eg., to drive a robot's tool tip from a starting point A to a target point B) because the motion control among a set of points is just a sequentially combined two point controls.

Human eyes can effortlessly guide human arms/hands to do operations. Now, the question is how to develop robotic hand-eye coordination system as illustrated by Fig.3.

Let's consider the ideal case of two point control in a 3D space without obstacles. An immediate and trivial solution is to use a 3D vision system to locate the positions of the starting point A and the destination point B. Then, one can guide the motion of robot's tool tip from A to B by following a path of straight line. This constitutes the optimal solution in terms of the shortest path. However, we should not ignore the following two situations:

1. The development of robust 3D vision is still under investigations of many researchers [7].
2. In most applications with robot, it is not absolutely necessary to guide robot by following path segments of straight line as long as a path segment is close to a straight line.

Inspired by the above situations, the following interesting question can be raised: can we use imprecise 3D output of a vision system to guide the motion of robot's tool tip in an iterative way and to guarantee the convergence of robot's motion control ? Fortunately, the answer is yes. So far, a lot of works have been done toward this direction in order to avoid the necessity of doing metric 3D reconstruction [3]. The basic idea is to directly use 2D image information to generate imprecise estimations of displacements of robot's tool tip as long as the accumulation of the estimated displacements converge toward the target point B.

In this area, our contribution is to have developed the simplest solutions that make use of constant mapping matrix J [11]. This work can be summarized by the following proposition:

> **Proposition II**: In the case of using 2D image information to implement robotic hand-

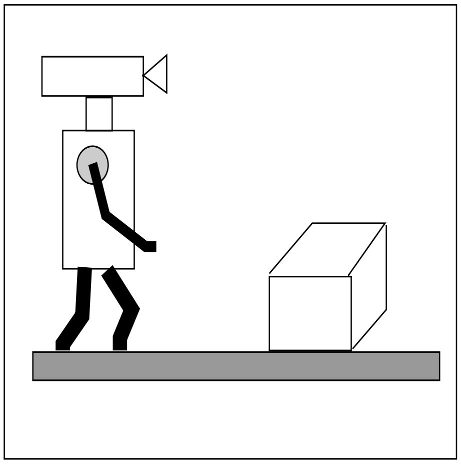

Figure 4: Vision and control of robot's leg motion.

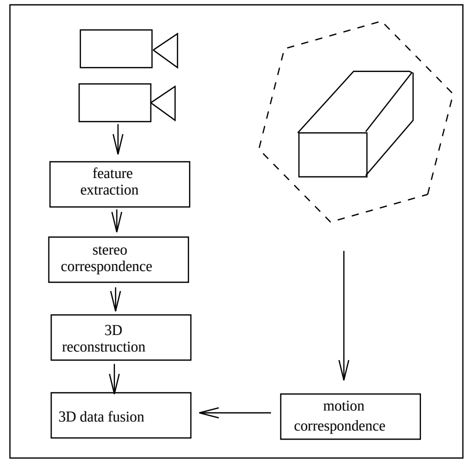

Figure 5: Stereovision for the perception of 3D scene.

eye coordination with uncalibrated stereo cameras, there exist more than one constant mapping matrix J that projects the variations $\triangle I$ of image feature (in 2D image space) to the variations $\triangle P$ of robot tool tip's position/orientation (in robot's 3D space), that is:

$$\triangle P = G \bullet J_{3\times 4} \bullet \triangle I. \quad (1)$$

so that robot tool tip's position P converges to a moving or static target's position B within a reasonable number of iterations, i.e.:

$$P + \sum \triangle P \Rightarrow B. \quad (2)$$

(G is control gain matrix).

Most importantly, the following facts hold:

1. One can effectively use constant mapping matrix J.
2. There are multiple solutions in practice to compute a constant mapping matrix J.
3. The convergence of robotic hand-eye coordination can be easily proved with the help of triangulation principle of stereovision.
4. There is absolutely no need to know any parameters of the vision system or cameras.

5 Leg-Eye Coordination: Use of 2D Projective Transformation

The generic problem of leg-eye coordination can be stated as how to determine free moving space (or alternatively space occupied by obstacles) in front of robot's body as illustrated by Fig.4.

Without any additional constraint, the only solution to this problem in a 3D space is to use a 3D vision system to obtain a 3D description of the scene seen by a robot. As mentioned earlier, the development of robust 3D vision system is still under active investigations. Now, the question is: is it possible to avoid the use of 3D vision for some constrained environment ? The answer is affirmative for the case of detecting free moving space on a ground plane (eg., planar road or planar floor).

It is well known that any two image planes are related to each other by a 2D projective transformation constrained by a 2D plane [1]. In other words, if one uses a stereo pair of cameras and the primitives of the two images are the projection of (2D) objects inside a 2D plane in front of the stereovision, the two image planes of the stereovision can be related to each other by a 2D projective transformation. This provides a necessary (but not sufficient) condition for determining whether an image primitive belongs to a ground plane or an obstacle. Based on this idea, we have developed an algorithm of obstacle detection on a ground plane [9]. Our on-going work in this area is to incorporate matching constraint so as to come out a necessary and sufficient condition for the determination of obstacles on a ground plane.

6 3D Perception: A quasi Closed Form Solution For Feature Correspondence in Rotational Motion Stereo

The correspondence problem in motion stereo can be shortly stated as how to identify, from two consecutive images, a pair of geometrical primitives (including image points) which are the projection of a same object primitive. Fig.5 illustrates the roles of stereo and motion correspondence inside the 3D perception system based on the use of stereovision.

So far, there is no analytical solution having been found [12]. All the existing algorithms for correspondence problem can only be qualified to be algorithmic solutions because they make use of constraints to perform the search for matches. The commonly used constraints are: (a) feature similarity (including image intensity), (b) epipolar geometry, and (c) figural continuity, etc.

Now, the question is whether we can find a closed form or quasi closed form solution to motion stereo correspondence problem ? In other words, we are interested in finding a matching function MF(.) which outputs the match (u_2, v_2) in one image, given an input (u_1, v_1) in its preceding image, that is:

$$(u_2, v_2) = MF(u_1, v_1). \tag{3}$$

The answer to the above question is affirmative if the motion is a rotation or dominated by rotation with weak translation. The following proposition summarizes our solution:

> **Proposition III**: In motion stereo, the match (u_2, v_2) in one image of a point (u_1, v_1) in its preceding image is determined by a 2D projective transformation as follows:
>
> $$\begin{pmatrix} u_2 \\ v_2 \end{pmatrix} = MF(u_1, v_1) = \begin{pmatrix} \frac{m_{11} \bullet u_1 + m_{12} \bullet v_1 + m_{13}}{m_{31} \bullet u_1 + m_{32} \bullet v_1 + m_{33}} \\ \frac{m_{21} \bullet u_1 + m_{22} \bullet v_1 + m_{23}}{m_{31} \bullet u_1 + m_{32} \bullet v_1 + m_{33}} \end{pmatrix}. \tag{4}$$
>
> The coefficients $\{m_{ij}, i, j = 1, 2, 3\}$ are constants if the motion is a pure rotation. If the motion is not a pure rotation, only (m_{13}, m_{23}, m_{33}) are not constants and depend uniquely on the depth (ie., Z coordinate of the corresponding 3D point).

(NOTE: In Eq.4, (u_1, v_1) and (u_2, v_2) are index coordinates of image points. So, no intrinsic parameters of camera(s) are needed to support the matching. This constitutes an advantage).

7 Conclusions

The degree of intelligence of service robot largely depends on its capability of doing visual perception and integrating vision with its motion control in an effective way. The interaction between a vision system and the motion control system of a robot can not be a simple matter of treating vision as a feedback sensor. In fact, a vision system can actually fulfill the role of controlling the motion execution being performed by a service robot. Some typical examples of such a case are: (a) robotic head-eye coordination, (b) robotic hand-eye coordination and (c) robotic leg-eye coordination. Progress made by us and other researchers toward this direction is quite promising. Certainly, more works need to be done to address the issues such as: (a) effective integration of vision and robot's motion control (eg., use of task function approach [6]), (b) real-time performance of vision system, and (c) adaptive acquisition of images, etc.

References

[1] O. Faugeras, "Three-Dimensional Computer Vision: A Geometric Viewpoint", *MIT Press*, Cambridge, p33-69 & p165-40, 1993.

[2] K. Hashimoto (editor), Visual Servoing, World Scientific, Singapore, 1994.

[3] K. Hosoda, and M. Asada, "Versatile Visual Servoing without Knowledge of True Jacobian", IEEE/RSJ International Conference on Intelligent Robots and Systems, Germany, p186-193, 1994.

[4] I. Masaki, Vision-Based Vehicle Guidance, Springer, Berlin, 1992.

[5] B. J. Nelson and P. K. Khosla, "An Extendable Framework for Expectation-Based Visual Servoing Using Environment Models", Proc. IEEE Int. Conf. Robotics and Automation, Japan, p184-189, 1995.

[6] C. Samson, M. L. Borgne and B. Espiau, "Robot control: The task function approach", *Clarendon Press*, 1990.

[7] T. Vieville, E. Clergue, R. Enciso, and H. Mathieu, "Experimenting with 3D vision on a robotic head", *Robotics and Autonomous Systems 14*, p1-27, 1995.

[8] M. Xie, "A cooperative strategy for the matching of multi-level edge primitives", *Image and Vision Computing*, Vol.13, No.2, p89-99, March, 1995.

[9] M. Xie, "Matching Free Stereovision for Detecting Obstacles on a Ground Plane", *Machine Vision and Applications Vol.9*, p9-13, 1996.

[10] M. Xie, "Head-Eye Coordination: A Closed Form Solution", 2nd World Automation Congress, Montpellier/France, May 27-30, 1996. (recipient of MANSOUR best paper award from WAC96)

[11] M. Xie, "Robotic Hand-Eye Coordination: Solutions with uncalibrated stereo cameras", *to appear in IES Journal*, Jan/Feb, 1997.

[12] Z. Zhang, R. Deriche, O. Faugeras, and Q. T. Luong, "A robust technique for matching two uncalibrated images through the recovery of the unknown epipolar geometry", *Artificial Intelligence*, Vol.78, p87-119, October 1995.