

A Study of the Role of the Observer In the Hand/eye Coordination Problem

Jean-Yves Hervé

Groupe de Recherche en Perception et Robotique
Département de génie électrique et de génie informatique
Ecole Polytechnique
B.P. 6079, succ. Centre-ville
Montréal (Québec)

Abstract

This paper approaches the hand/eye coordination problem as a skill learning as well as a control problem. We show—or rather, confirm analytically the intuitive concept—that all poses of the observer are not equally good in terms of the visual servoing to perform, and that a good pose can be chosen that facilitates both the local control problem and the learning of the visuo-motor coupling.

We represent the relation between visual input and movements of the actuators by a hyper-surface, whose shape we propose to learn, through the execution of simple hand displacement tasks. The shape of this hyper-surface depends on the geometry of the manipulator and on the position of the observer (camera) relative to it. If the observer is active, it can therefore, simply by changing its viewpoint, modify the shape of the hyper-surface, and sometimes even its topology.

We show that it is possible to establish a finite “family portrait,” indexed by the position of the observer relative to the manipulator, of all qualitatively different possible aspect of this hyper-surface. We also show how this information can be exploited to facilitate the skill learning process.

1 Introduction

1.1 Hand/eye Coordinations

The term “hand/eye coordination” is often used to represent several different concepts. With humans, the term often refers to the *current* degree of skill of an athlete, a craftsman, or an infant, but also to the process through which such sensorimotor skills are acquired. At the other end of the spectrum, in robotics, the term “hand/eye coordination” refers, almost exclusively, to the visual control of a robot manipulator in hand or tool positioning tasks.

This simplified form of hand/eye coordination is traditionally treated as an application of the cali-

bration problem: kinematic calibration of the manipulator, optical calibration of the visual system, and estimation of the pose of the “hand” and of the goal relative to the observer. Hand/eye coordination is then achieved when the visual module is capable of determining estimates of the current, initial state and of an expected final state that can be exploited by classical control processes.

Due in part to the complexity of pose estimation computations, the observer is generally a single camera, fixed with respect either to the base of the manipulator arm or with respect to the end-effector, as in the “eye in hand” configuration. In either case, it becomes possible to avoid one of the pose estimation problems and perform image-based control that directly exploits image measurements without computing any 3D information ([10], [3]). Still, such control rely heavily on precise estimates of a number of parameters of the optical and mechanical apparatus. Should the value of any of these parameters be changed (for example if the camera is displaced), the system would need to be re-calibrated, with no accumulation of experience from one time to the next, in clear contrast with the way humans seem to be dealing with the reaching problem.

1.2 Hand/Eye Coordination in Humans

We can only present here a rudimentary summary of the literature on the reaching problem. Hand/eye-coordination-as-a-learning-process (in infants) can be decomposed into three main phases. First is a phase of sensory accommodation, during which the child acquires focusing, object fixation, eye vergence control, *etc.* Next, the visuo-motor coupling is explored: the infant learns how to move one hand toward an object, with alternate glances from hand to object. At the end of the third phase, this simple visual control has been replaced by a more elaborate, anticipatory strategy, allowing the task to be initiated from a position outside the visual field.

The anticipatory aspect being even more promi-

ment with adults, it has been suggested that fast movements, such as the initial phase of precise hand movement, were *ballistic*, that is, programmed in advance and run without correction, and that only the slower homing-in phase is performed under visual control.

1.3 Application to Robotics

Without trying to precisely emulate the way human learn this visuo-motor coupling and accomplish hand positioning task, we can attempt to retain the same decomposition of hand movements into a fast, "ballistic" phase and a slow, precise phase that is performed under visual guidance. In many respects, we already know how to execute the second phase: visual servoing techniques have been successfully used to perform fine tool-positioning tasks in the case of an eye-in-hand system ([10], [3], [2]). As mentioned later in this paper, we have also implemented a simple image-based controller, for a 6-dof manipulator, in the case of the "eye looking at the hand" configuration [7]. The part of the problem that remains mostly unaddressed concerns therefore the generation of ballistic hand movements.

Although it is naturally possible to perform such gross motion planning if the kinematics of the manipulator and the position and orientation of the hand, the eye, and the object relative to each other are known, the performance of current visual reconstruction algorithms does not allow such direct computations. Furthermore, we believe that image-based qualitative methods are better adapted to problems in which the sources of uncertainty and inaccuracy are numerous. In particular, we would prefer to learn a global representation of the visuo-motor coupling between the visual feedback and the control to generate. For example, [8] showed that, given an adequate representation, learning can be performed by determining of a few parameters, through the direct observation of hand movements.

Our approach can be viewed as a geometric counterpart to that of [8]. We represent the visuo-motor coupling by a geometric object, and the goal of the learning process is to determine—qualitatively—the shape of this object. In order to determine what must be learned, we proceed by first analyzing the visuo-motor coupling. Next, we distinguish classes of patterns through which this coupling can be expressed. Finally, we propose the sketch of a learning strategy for the reaching problem.

2 The Perceptual Kinematic Map

This section presents a very brief overview of the mathematical concepts on which our study is based: the Perceptual Kinematic Map, or PKM, and its as-

sociated Perceptual Control Surface. Although the ideas and techniques of PKM-based control apply to the case of a full-fledged 6-dof manipulator ([7], [7]), the analysis of bifurcations presented in Section 3 was done specifically for a 2-dof manipulator. The notation in this section has been simplified to take this restriction into account.

2.1 Definition of the PKM

The PKM is a mapping between the joint space and the camera space [8] that maps the joint coordinates to a vector of image features:

$$\begin{aligned} \pi : \mathcal{J} &\longrightarrow \mathcal{C} \\ \mathbf{q} &\longmapsto \mathbf{c}. \end{aligned}$$

In the case of polyhedral objects, Weiss [10] proposed a list of easily extracted image features that could be used: areas of faces, lengths of edges, angles, centroids of faces (to which we can add corner points), and ratios of areas of faces, but more sophisticated information may be used as well [1].

2.2 The Kinematic Map

The type of manipulator we are considering in this paper is a 2-dof (2R) manipulator that has its rotation axes perpendicular to each other, which corresponds to the configuration presented by the first two joints J_0 and J_1 of most manipulators, as shown in Figure 1.

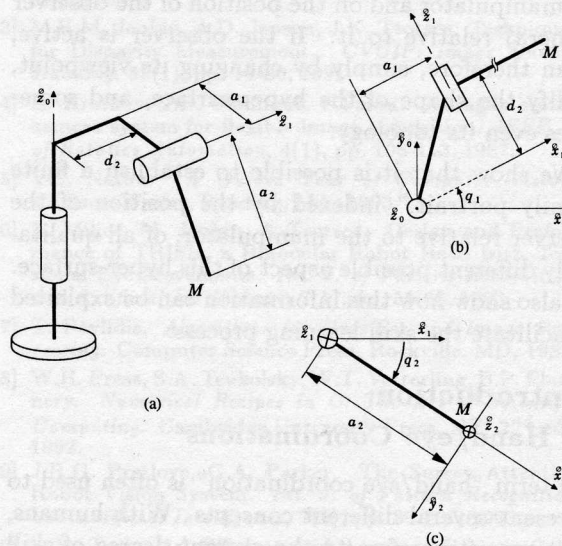


Figure 1 — Denavit-Hartenberg parameters for a 2-dof (2R) manipulator

Since we are only treating here the case of 2-dof manipulators, it is sufficient to track a point M that is attached to the terminal link of the arm (for example its tip, O_2). The kinematic map κ then simply maps the joint vector $\mathbf{q} = (q_1, q_2)'$ to the

Cartesian coordinates of M relative to \mathcal{R}_0 , \mathbf{M}_0 :

$$\begin{aligned} \kappa : \mathcal{J} &\longrightarrow \mathcal{K} \\ \mathbf{q} &\longmapsto \mathbf{M}_0 = (X_0, Y_0, Z_0)' \end{aligned}$$

The kinematic map's Jacobian matrix is then

$$\mathbf{K} = \begin{pmatrix} -a_1 S_1 - a_2 S_1 C_2 - d_2 C_1 & -a_2 C_1 S_2 \\ a_1 C_1 + a_2 C_1 C_2 - d_2 S_1 & -a_2 S_1 S_2 \\ 0 & -a_2 C_2 \end{pmatrix},$$

where we have used the shorthand notation $C_i = \cos q_i$ and $S_i = \sin q_i$.

2.3 Expression of the PKM

In the case of a 2-dof manipulator, the simplest camera space is the one generated by the image coordinates of m , the projection of M onto the image plane:

$$\begin{aligned} \pi : \mathcal{J} &\longrightarrow \mathcal{C} \\ \mathbf{q} &\longmapsto \mathbf{m} = (x, y)' \end{aligned}$$

Let \mathcal{R} be the coordinate system associated with the camera. The coordinate vector of M relative to \mathcal{R} is $\mathbf{M} = (X, Y, Z)'$:

$$\mathbf{M} = \mathbf{E}_{\psi, \theta, \phi} \cdot \mathbf{M}_0 + \mathbf{T}_0 \quad (1)$$

$$= \mathbf{E}_{\psi, \theta, \phi} \cdot \mathbf{M}_0 - \mathbf{E}_{\psi, \theta, \phi} \cdot \mathbf{C}, \quad (2)$$

where $\mathbf{E}_{\psi, \theta, \phi}$ is the 3×3 Euler rotation matrix of angles ψ, θ , and ϕ , \mathbf{T}_0 is the coordinate vector of O_0 relative to \mathcal{R} , and $\mathbf{C} = (X_c, Y_c, Z_c)'$ is the coordinate vector of the camera's optical center relative to \mathcal{R}_0 . $\mathbf{C} = -\mathbf{E} \cdot \mathbf{T}_0$.

If we model the camera as a pinhole of focal length f , then $\mathbf{m} = f/Z \cdot \mathbf{M}$ and, by differentiating \mathbf{m} with respect to the joint variables, we obtain

$$\mathbf{J} = \frac{\partial \mathbf{m}}{\partial \mathbf{q}} = \frac{\partial \mathbf{m}}{\partial \mathbf{M}} \cdot \frac{\partial \mathbf{M}}{\partial \mathbf{q}} = \frac{1}{Z} \cdot \mathbf{D} \cdot \mathbf{E}_{\psi, \theta, \phi} \cdot \mathbf{K} \cdot \dot{\mathbf{q}},$$

$$\text{where } \mathbf{D} = \begin{pmatrix} f & 0 & -x \\ 0 & f & -y \end{pmatrix}.$$

2.4 Simple PKM-based Control

In order to learn anything about the PKM, we need a basic image-based control scheme capable of moving the arm across the field of view. Specifically, we need a control technique capable of bringing the arm from its current position in the camera space, \mathbf{m}_i , corresponding to a (known) joint configuration \mathbf{q}_i , to a desired final position \mathbf{m}_f corresponding to an unknown joint configuration. We proposed in [7] to use a simple gradient descent algorithm to solve this problem:

$$\mathbf{q}(k+1) = \mathbf{q}(k) - \gamma \cdot \mathbf{J}' \cdot (\mathbf{m}(k) - \mathbf{m}_f).$$

The only unknown term in this algorithm is the Jacobian matrix of the PKM, \mathbf{J} , which can be estimated at the initial position and updated along the

trajectory by a Kalman filter [7]. The data about the PKM that we compute along such trajectories, however, is very local and short-lived: the value of \mathbf{J} at a particular point of the joint space. It is now time to look for more global information about the PKM.

2.5 The Perceptual Control Surface

In [6] we suggested studying the PKM geometrically, in the form of the hyper-surface that it implicitly defines, which we call the Perceptual Control Surface, or PCS. Figure 2(a) shows the very tube-like PCS of a typical 2D pendulum. Except in the case of the very degenerate configurations that are the subject of the next section, small perturbations of the parameters of the hand/eye system (due for example to vibrations or to displacements of the observer) will only result in small deformations of the PCS. The shape of the surface will be preserved (qualitatively), although the numerical values taken by the PKM in the camera space will all have changed.

We can go further towards a qualitative representation of the PKM if we notice that the shape of the PCS is *qualitatively* uniquely defined by the location and type of its singularities. Comparing Figures 2(a) and 2(b), we see that the singular curves form the "skeleton" around which the surface is fleshed. The object of our study will therefore from now on be the geometry of the singularities of the PKM and their behavior under perturbation of the hand/eye apparatus, in particular under displacements of the observer.

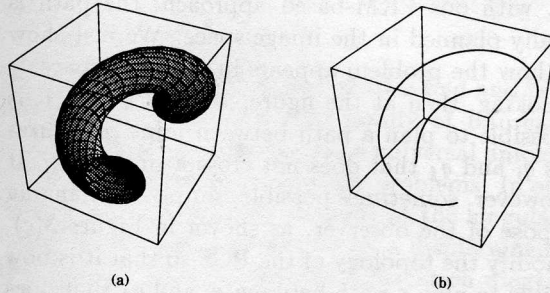


Figure 2 — (a) A Perceptual Control Surface (PCS); (b) Its singularities

3 Bifurcations

3.1 Presentation of the Problem

We suggested in [6] that, since displacements of the camera can change the topology of the PCS, the pose of the camera could (and therefore should) be chosen so as to result in "good" changes in the topology of the PCS. For example, the total number of singularities—or, more simply, the number of

singularities between the current position and the goal—could be reduced. The problem that we are studying here could therefore be considered as an example of *sensor planning* [9] where this planning is not done directly in order to improve the perception of the shape of a physical object, but in order to improve *the shape* of a topological one.

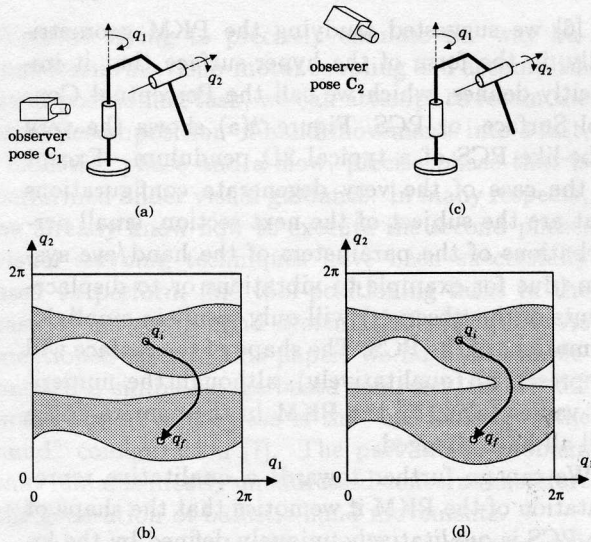


Figure 3 — Effect of a displacement of the observer on the geometry of singularities of the PKM

As an example, let us consider the hand/eye configuration represented in Figure 3(a) and the corresponding singularities, drawn in the joint space (Figure 3(b)). We should first emphasize the fact that, with our PKM-based approach, the path is actually planned in the image space. We just show here how the problem appears in the joint space.

Looking, then at the figure, we notice that it is impossible to plan a path between joint configurations q_i and q_f that does not cross a singularity. It is, however, *sometimes* possible, simply by changing the pose of the observer, as shown in Figure 3(c), to modify the topology of the PCS so that it is now possible to plan a path between q_i and q_f that does not cross a singularity, as shown in Figure 3(d). Obviously, only “perceptual” singularities can be so modified; displacements of the observer have no effect on kinematic singularities.

We will see that we can obtain a “family portrait” of the possible configurations of the singularities, that is, a finite set of 2D curves, indexed by the position of the camera relative to the manipulator. The learning of the PCS can therefore be guided by the prior recognition of a particular configuration of the family portrait, using measurements made by the low-level controller.

3.2 Bifurcations and Transition Sets

The underlying philosophy of our approach is that of the theory of singularities, applied to the study of bifurcation problems and in particular to the concept of an *organizing center*, defined as a configuration for which the system exhibits its most singular behavior [4]. A very interesting aspect of this theory is that quasi-global results may often be obtained by the application of local analysis near an appropriately chosen organizing center. Parameters of the system can therefore be treated as *perturbation* terms in the neighborhood of the organizing center.

Another important concept of the approach is that of the *transition set*, Σ , of a bifurcation problem, that is, of the set of n -tuples of perturbation parameters for which one of the basic elementary phenomena—bifurcation, hysteresis, and double limit point [4]—are observed. Generally, Σ defines a segmentation of the space of perturbation parameters, W , into connected components. Figure 4 shows the transition set (on the left) and the persistent forms of the bifurcation diagram for the pitchfork organizing center $x^3 + \lambda x + \alpha + \beta x^2$, where α and β are the perturbation terms.

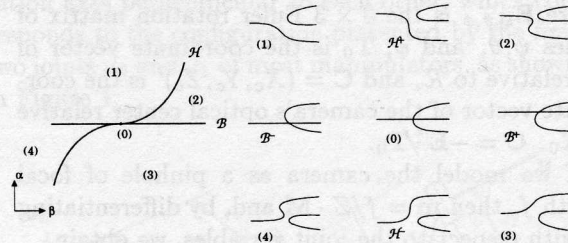


Figure 4 — Persistent perturbations of the pitchfork bifurcation: $x^3 - \lambda x + \alpha + \beta x^2 = 0$.

The main theorem of the approach states that, given $G : U \times L \times W \rightarrow \mathbb{R}$, a family of bifurcation problems that satisfy $G(x, \lambda; \alpha) \neq 0$ on $\partial U \times L \times W$, the persistent bifurcation diagrams in the unfolding G are enumerated by the connected components of $W \sim \Sigma$.

If we work within the framework that we briefly exposed, our immediate objective is then to determine the organizing centers of the PKM, their types, and their transition sets.

3.3 Notation

We are studying here the singularities of the PKM, that is, configurations for which

$$J = \det \mathbf{J} = 0, \quad (3)$$

which, once developed, turns out to be a rather complex expression. It is, however, possible to make numerous simplifications, for example if we notice,

first, that the orientation of the observer has no effect on solutions of (3), and, next, that our problem is cylindrically symmetrical with respect to the first joint axis, \bar{z}_0 . This means that we can simplify the problem by using the distance between the observer and the first joint, $R_c = (X_c^2 + Y_c^2)^{1/2}$, rather than coordinates X_c and Y_c , as long as we use a relative angle to represent the state of the the first joint.

Condition (3) then simplifies to

$$h(q_2, q_1) = 0, \quad (4)$$

where

$$h(q_2, q_1) = (a_1 + a_2 C_2) \cdot (k_3(q_2) - C_2 \cdot k_2(q_1)) + d_2 C_2 \cdot k_1(q_1)$$

and

- $k_1(q_1) = d_2 + R_c \cdot \sin q_1$,
- $k_2(q_1) = R_c \cdot \cos q_1$,
- $k_3(q_2) = a_2 + a_1 \cdot C_2 + Z_c \cdot S_2$.

The singularities of the PKM that were drawn in Figure 2 were the solutions of (4) obtained for one particular manipulator (that is, a triple of values for a_1 , a_2 , and d_2) and one particular pose of the observer (values for Z_c and R_c). We want now to take a more general view at the 5-parameter family of problems

$$H(q_2, q_1; a_1, a_2, d_2, Z_c, R_c) = 0, \quad (5)$$

in which three of the parameters are passive (the D-H parameters of the manipulator) and two are active: the pose parameters of the observer.

3.4 Identification of the Bifurcations

The first comment to make regarding the general bifurcation problem (5) is that it is remarkably complex. In fact, it cannot even be solved as a general problem and must instead be decomposed into several smaller problems, based on the geometrical configuration of the manipulator. This decomposition of the set of manipulator configurations into classes is defined by criteria based on D-H parameters d_2 and a_1 : $d_2 = 0$ or $d_2 \neq 0$; $a_1 = 0$, $0 < |a_1| < a_2$, $|a_1| = a_2$, or $|a_1| > a_2$.

For each of the eight manipulator geometry types so defined, we have identified all "organizing centers," that is, hand/eye configurations (q_2, q_1, Z_c, R_c) corresponding to the locally most singular forms of H . Some of the results of this analysis were very surprising. For example, it revealed that one of the worst possible manipulator geometry—in terms of the perceptual singularities it contributes to create—is used in most industrial manipulators, such as the PUMA 560: $d_2 \neq 0$ and $a_1 = 0$. For this type of manipulator, a bifurcation of infinite codimension can be observed, a mathematical object far too complex

to handle. Other types of manipulators give rise to better behaved PKMs.

We will treat the example of manipulators such that $d_2 \neq 0$ and $|a_1| > a_2$ (complete topological analysis corresponding to each of the eight types of geometries can be found in [5]). In this case, the following organizing centers can be observed:

- Three quartic fold bifurcations.
- One pitchfork bifurcation.
- One simple bifurcation, which is a "dual" of the pitchfork bifurcation, since it is observed for the same pose of the observer, and value of q_1 , but a different value of q_2 .
- In addition, two of the quartic fold bifurcations admit a "dual" bifurcation of codimension zero.

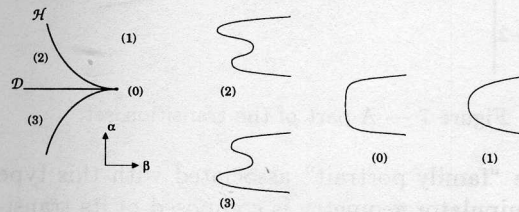


Figure 6 — Persistent perturbations of the quartic fold bifurcation $x^4 - \lambda + \alpha x + \beta x^2 = 0$.

We proved in [5] that at the quartic fold and pitchfork organizing centers, the family of bifurcation problems $H(q_2, q_1; Z_c, R_c)$ was a universal unfolding of the corresponding elementary problems. In other words, all possible perturbations of the singularity conditions can be represented by a displacement of the observer, and all possible bifurcation diagrams for the PKM of this type of manipulator can be associated with one connected component of the space (Z_c, R_c) .

The complete bifurcation diagram of the PKM is therefore a complex composition of the bifurcation diagrams of elementary bifurcations such as the pitchfork bifurcation (Figure 4), the simple bifurcation (Figure 5), and the quartic fold (Figure 6). In either diagram, we see that the perturbation space is decomposed into a finite number of connected components and that a (qualitatively) different bifurcation diagram corresponds to each of these connected components.

Figure 7 shows a part of the transition set associated with the type of manipulator geometry that we are considering here. Any displacement of the observer across one of the curves composing the transition set results in the occurrence of one of three basic topological phenomena: a bifurcation; the appearance or disappearance of a hysteresis point; or the appearance or disappearance of a double limit point. Figure 8 gives examples of bifurcation diagrams observed when the camera is displaced in the neighborhood of the pitchfork organizing center.

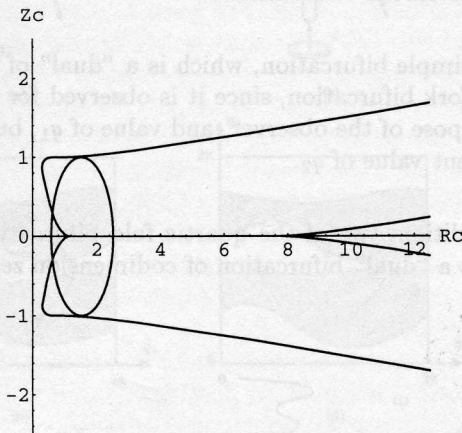


Figure 7 — A part of the transition set.

The “family portrait” associated with this type of manipulator geometry is composed of its transition set and of the (finite) catalog of all classes of bifurcation diagrams that are allowed.

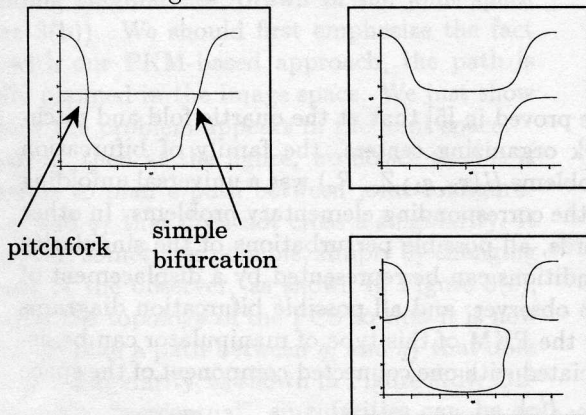


Figure 8 — Examples of bifurcation diagrams observed in the neighborhood of the pitchfork organizing center.

4 Recognition-Guided Learning

The bifurcation diagrams shown in Figure 8 represent locations of singularities of the PKM in the joint space for different poses of the observer. Any displacement of the observer across the transition set results in a change in the topology of the PCS.

If the complete transition set has been defined, then this topological change can be chosen *a priori* so that beneficial effects can be observed on the PCS.

There are therefore several aspects of learning that must be pursued. The simplest kind is the recognition of the type of bifurcation diagram currently imposed by the pose of the observer. This task is simple enough, since simple point-to-point trajectories generated by the basic control sketched in Subsection 2.4 routinely intersect singularities and hence inform us about their locations and orientations. We see in Figure 9(a) an example of such reports of “singularity crossings” collected along six consecutive trajectories of the arm. After a while, enough information about the singular curves should be available to decide to which kind of diagram in the “family” portrait the one being intersected belongs to—for example, here, the one shown in Figure 9(b). Since the model of the current diagram has been recognized, its complete recovery is then greatly facilitated. For example, arm movements can now be planned to verify if singularities can actually be found where they are expected.

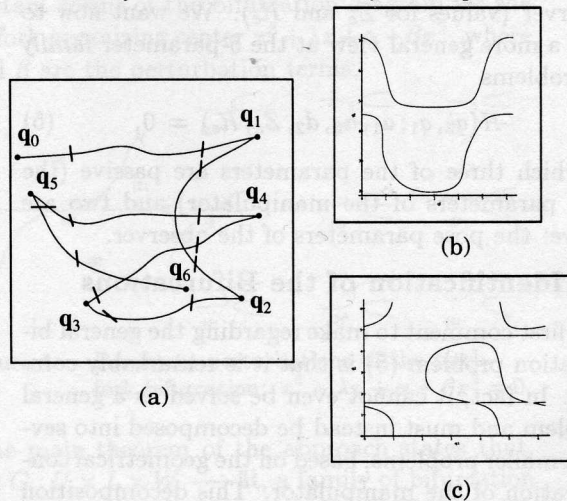


Figure 9 — Recognition of the bifurcation diagram in the family portrait

In general, it is sufficient to recognize the current bifurcation diagram to identify the type of the manipulator’s geometry; at most, only two candidates remain to choose from. To each type of geometry is associated a unique transition set, and a catalog of bifurcation diagrams. It now becomes possible to predict what bifurcation diagram—what shape of the PCS—can be expected if the observer moves in a certain way.

5 Conclusion

We have presented ongoing work on the role of the observer in the hand/eye coordination problem,

based on the concepts of Perceptual Kinematic Map and Perceptual Control Surface. We showed that the position of the observer can be chosen so that the problem becomes simpler, by removing some singularities or re-arranging them in a way more convenient for the controller. Finally, we proposed to use the displacements of the observer to guide the learning of the PKM through the recognition of some configurations of its singularities.

References

- [1] P. Cucka, Center for Automation Research, Technical Report in preparation.
- [2] J.T. Feddema, C.S.G. Lee, and O.R. Mitchell, "Weighted selection of image features for resolved rate visual feedback control," *IEEE Transactions on Robotics and Automation*, 7(1): 31-47, 1991.
- [3] J.T. Feddema and O.R. Mitchell, "Vision-guided servoing with feature-based trajectory generation," *IEEE Transactions on Robotics and Automation*, 5(5): 691-700, 1989.
- [4] M. Golubitsky and D.G. Schaeffer, *Singularities and Groups in Bifurcation Theory*, Vol. I, Springer-Verlag, New York, 1985.
- [5] J-Y. Hervé, "Role of the observer in the hand/eye coordination problem. Part I," rapport technique GRPR-9502, Groupe de Recherche en Perception et Robotique, Ecole Polytechnique de Montréal, Montréal, Québec, July 1995..
- [6] J-Y. Hervé, R. Sharma, and P. Cucka, "The geometry of visual coordination," in *Proceedings AAAI National Conference on Artificial Intelligence*, pp. 732-737, Anaheim, CA, 1991.
- [7] J-Y. Hervé, R. Sharma, and P. Cucka, "Toward robust vision-based control: Hand/eye coordination without calibration," in *Proceedings IEEE International Symposium on Intelligent Control*, pp. 457-462, Arlington, VA, 1991.
- [8] S.B. Skaar, W.H. Brockman, and R. Hanson, "Three-dimensional camera-space manipulation", *International Journal of Robotics Research*, 9(4): 22-39, 1990.
- [9] K.A. Tarabanis, P.K. Allen, and R.Y. Tsai, "A survey of sensor planning in computer vision," *IEEE Transactions on Robotics and Automation*, 11(1): 86-104, 1995.
- [10] L.E. Weiss, "Dynamic Visual Servo Control of Robots: An Adaptive Image-based Approach," Ph.D. Thesis, Dept. of Electrical and Computer Engineering, Carnegie-Mellon University, Pittsburgh, PA, 1984.
- [11] A.T. Welford, *Fundamentals of Skill*, Methuen, London, UK, 1968.