Camera self orientation and docking maneuver using normal flow

P. Questa

Laboratory for Integrated Advanced Robotics (LIRA - Lab)
Department of Communication, Computer and Systems Science (DIST)
University of Genova
Via Opera Pia 13 - 16145 Genova, Italy
e-mail : questa@lira.dist.unige.it
Telephone (+39) 10.353.2946 - Fax (+39) 10.353.2948

E. Grossmann

Département de Mathématiques-Informatique
Université d'aix-Marseille II
163 Avenue de Luminy
13288 Marseille Cedex 9, France
e-mail : grossman@massilia.univ-mrs.fr

G. Sandini

Laboratory for Integrated Advanced Robotics (LIRA - Lab)
Department of Communication, Computer and Systems Science (DIST)
University of Genova
Via Opera Pia 13 - 16145 Genova, Italy
e-mail : giulio@lira.dist.unige.it
Telephone (+39) 10.353.2779 - Fax (+39) 10.353.2948

ABSTRACT

In this paper the possibility of driving egomotion in order to dock with a non-moving target using information on the first order approximation of the optic flow in the center of the image is presented; specifically it will be shown how a camera mounted on the end-effector of a six degrees of freedom robot arm can be driven toward a planar surface whose position and orientation in space are unknown, avoiding collisions with the target and orienting the camera with its viewing direction perpendicular to the surface. It will be described how it is possible to achieve this task combining the effect of three simple independent processes: braking, fixation and sideway translation.

Keywords: normal flow, affine flow, time-to-contact, egomotion, docking.

1 INTRODUCTION

Positioning a camera with respect to a target is an important issue in computer vision; it allows for simplification of other tasks such as inspection and object manipulation, docking maneuver or automatic power re-charge in case of mobile vehicle. Using the dynamic of the pattern movements in the optic array to actively control the position in space of a camera allows to perform the task in a robust way.[1,7,5] One of the simplest measures that it is possible to compute in the optic array is the normal flow, i.e. the optic flow in the image gradient direction[2]; normal flow computation is sufficient to retrieve a linear, *affine*, model of the motion of the light patterns in image sequences; this model is described by an image *translation* that specifies the position change in the optic array of the centroid of the pattern, a 2D rigid rotation, *curl*, an isotropic expansion (or contraction), *divergence*, that specifies a change in scale and a pure deformation, *shear*, that describe a distortion of the image pattern (expansion in a specific direction and a contraction in the direction perpendicular to the first one, in such a way that the area of the pattern results unchanged). If the movements of the pattern in the optic array are induced only by motion of the camera, i.e. the camera is facing a *still* world, then the affine parameters are strictly related to the kind of movement performed: a rotation of the camera around its viewing direction will affect only the curl parameter; a translation along the optical axis will affect only the divergence term; panning or tilting around the nodal point will result in a simple translation of the image while moving parallel to the image plane will result in an image translation and a pure shear. The divergence and shear parameters are particularly interesting because if information on the observer motion, i.e. egomotion, is available together with the estimate of the affine transform parameters, then it is possible to recover reliable information on time-to-contact and surface orientation.

In the following sections an application will be described; it will be shown with few very basic measures on the image plane how it is possible to perform a non trivial task such as docking; it will be shown how it is possible to control the position and orientation in space of a camera in order to dock with a non-moving target avoiding collisions and aligning the optical axis perpendicular to the target surface; section 2 will deal with the relationship between egomotion and affine flow parameters and how it is possible to estimate the value of the affine parameters from normal flow; in section 3 the docking maneuver control will be described in terms of strategies to be applied; in section 4 results of the experiments will be presented and open problems will be discussed.

2 EGOMOTION AND AFFINE FLOW PARAMETERS

The relationship between the motion parameter of the optic array and the affine flow parameters is very simple and it has been exhaustively described.[3,4]

It is well known that any arbitrary movement of a camera can be decomposed in a unique manner into a translation and a rotation around an axis going through a given point. Let the translational velocity be denoted $\mathbf{V}$ and the rotational velocity $\mathbf{R}$; let $\mathbf{r}$ be the direction of the viewing direction; then we can decompose $\mathbf{V}$ and $\mathbf{R}$ into radial components (V_r, R_r), which are the components in the direction of $\mathbf{r}$, and into transverse components $(\mathbf{V}_t, \mathbf{R}_t)$, (see figure 1) which are components perpendicular to $\mathbf{r}$. In order to eliminate all the measures of distance from the beginning, let us introduce the distance d from the observed patch and define the *specific radial velocity*

$$A_r = \frac{V_r}{d} \tag{1}$$

and a *specific transverse velocity*

$$A_t = \frac{\mathbf{V}_t}{d}. \tag{2}$$

These quantity have an intuitive content:

R_r is a rotation around the line of sight;

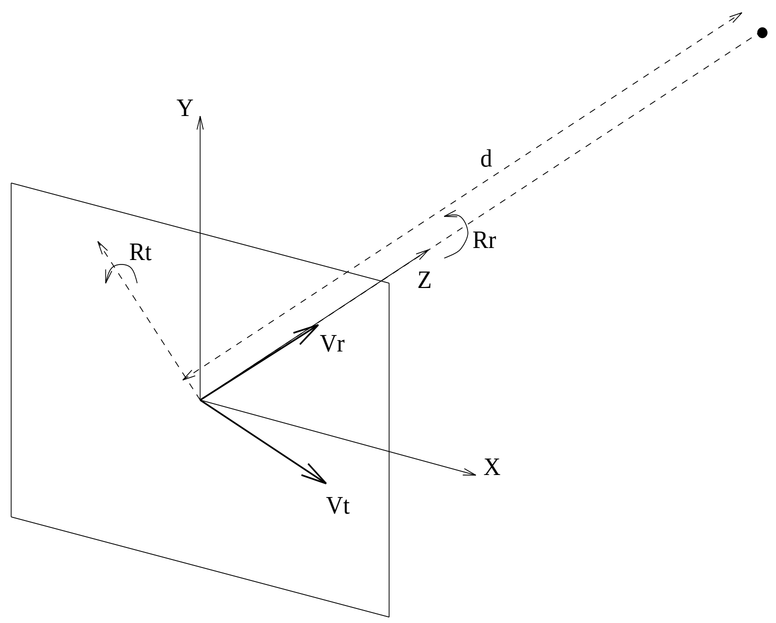

Figure 1:

$\mathbf{R}_t$ describes a pan and/or tilt of the camera;

A_r is the reciprocal of the time necessary to reach the patch at distance d proceeding with radial velocity $\mathbf{V}_r$ (i.e. the inverse of the time-to-contact);

$\mathbf{A}_t$ is a translation parallel to the image plane; is the measurement of the instantaneous rate of turn in the optic array due to translation.

The orientation of a patch has a vectorial expression: it has a magnitude (or *slant*) and a direction (or *tilt*); the slant could be, for example, represented by the angle between the viewing direction and the normal to the patch, while the tilt could be defined as the direction of the increasing distance to different points on the patch. If we then define the orientation of the surface as $\mathbf{F}$, it is possible to express its value as

$$\mathbf{F} = \frac{\mathbf{grad}\ d}{d} \tag{3}$$

Let the translational affine parameter be (u_0, v_0), the curl be R, the divergence D and the shear S; introduce for simplicity

$$\begin{cases} S_1 = S\ cos2\mu \\ S_2 = S\ sin2\mu \end{cases} \tag{4}$$

where μ is the axis of expansion of the shear term; The relationship between the motion and the affine flow parameters is then given by:

$$\begin{bmatrix} u_0 \\ v_0 \end{bmatrix} = -\mathbf{A}_t - \mathbf{R}_t \tag{5}$$

$$D = \mathbf{F} \cdot \mathbf{A_t} + \mathbf{2A_r} \tag{6}$$

$$R = |\mathbf{F} \times \mathbf{A}_t| - 2R_r \tag{7}$$

$$S = \sqrt[2]{S_1^2 + S_2^2} = |\mathbf{F}|\ |\mathbf{A}_t| \tag{8}$$

It is worth noting that the divergence D and the shear S are unaffected by arbitrary rotation of the camera.

It is possible to estimate the values of the affine parameters from measures of the normal flow; consider a point $\mathbf{P}$ in the scene, whose position is given by the triplet (X, Y, Z) and whose projection onto the image plane is $\mathbf{p}$ with coordinate (x, y). For any movement of the camera, $\mathbf{p}$ will move on the image with a given velocity $\dot{\mathbf{p}} = (u, v)$. We can write[1,8] the first order optic flow approximation of (u, v) as

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} u_0 \\ v_0 \end{bmatrix} + \begin{bmatrix} D+S_1 & S_2-R \\ R+S_2 & D-S_1 \end{bmatrix} \cdot \begin{bmatrix} x \\ y \end{bmatrix} \tag{9}$$

It is possible to express (9) in the unknown terms u_0, v_0, D, R, S_1, S_2 as

$$\begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} 1 & 0 & x & -y & x & y \\ 0 & 1 & y & x & -y & x \end{bmatrix} \cdot \begin{bmatrix} u_0 \\ v_0 \\ D \\ R \\ S_1 \\ S_2 \end{bmatrix} \tag{10}$$

It is possible to further develop the first term of equation (10) assuming the brightness constancy[2] (if $E(x, y)$ is the gray level associated with $\mathbf{p}$ at time t and if $\mathbf{p}$ moves, it still will have the same gray level value at time $t+1$); then

$$\frac{dE}{dt} = \frac{\partial E}{\partial x}u + \frac{\partial E}{\partial y}v + \frac{\partial E}{\partial t} = 0 \tag{11}$$

or

$$\dot{E} = -\begin{bmatrix} \frac{\partial E}{\partial x} & \frac{\partial E}{\partial y} \end{bmatrix} \cdot \begin{bmatrix} u \\ v \end{bmatrix} \tag{12}$$

Substituting (12) into (10) gives

$$-\dot{E} = \begin{bmatrix} E_x & E_y & xE_x + yE_y & -yE_x + xE_y & xE_x - yE_y & yE_x + xE_y \end{bmatrix} \begin{bmatrix} u_0 \\ v_0 \\ D \\ R \\ S_1 \\ S_2 \end{bmatrix} \tag{13}$$

where $E_x = \frac{\partial E}{\partial x}$ and $E_y = \frac{\partial E}{\partial y}$. The system in (13) expresses six unknowns in one equation; it is obvious that only an approximate solution can be found if the values of at least six flow vectors are known.

3 THE DOCKING MANEUVER

As already mentioned, the goal at this stage is to perform a *docking* maneuver; suppose having a camera with six degrees of freedom, i.e. freely moving in space. The camera has to move as near as possible to a planar surface that has a different orientation in space with respect to the camera frame; it has to avoid collision with the target and align its viewing direction with the normal of the surface and maintain fixation on the same target point. These two processes are independent since in the first case the velocity to be controlled is A_r while in the second case fixation is determined by $\mathbf{A}_t$ and $\mathbf{R}_t$. From now on, the first process will be referred to as the *braking* process and the second as the *aligning* process.

It is easy to show how the braking maneuver can be driven without any need for recovery of 3D measures, i.e. the distance to the observed surface. Suppose for the moment the camera only moves along its optical axis; equation (6) thus becomes

$$D = 2A_r = 2\frac{V_r}{d} \tag{14}$$

and time-to-contact τ is directly linked to divergence:

$$\tau = \frac{2}{D} \tag{15}$$

Time-to-collision τ decreases as the distance to the surface decreases; but if it is possible to control V_r in such a way that τ remains constant in time then the camera will never collide and braking will be automatically achieved. Consider two successive estimates of the time-to-contact τ_i and τ_{i+1} at time i and $i+1$; suppose for simplicity that the time interval between i and $i+1$ has unitary duration; then

$$d_{i+1} = d_i - V_{r_i} \tag{16}$$

substituting (14) into (16) and superimposing $\tau_{i+1} = \tau_i$, we obtain

$$V_{r_{i+1}} = V_{r_i}(1 - \frac{1}{\tau_i}) \tag{17}$$

This simple rule allows to achieve braking in case of motion along the viewing direction only (or in case of alignment between the optical axis and the normal to the surface); but for the docking maneuver the alignment also has to be accomplished; in this case $\mathbf{A}_t$ is non-null and the different orientation between the optical axis and the normal to the surface gives a contribution to (6) in such a way that we can only retrieve boundary condition for the time-to-contact[6]; it is possible to show that within a large set of boundaries conditions on the initial position of the camera and its transverse velocity, the alignment maneuver will be accomplished before the braking will become critical (see appendix A); in addition to this it is worth noting that as the alignment process proceeds the term $\mathbf{F} \cdot \mathbf{A}_t$ will become smaller and will affect the estimate of τ less and less.

Alignment between the optical axis and the normal to the surface must be accomplished; it is possible to show how this task can be achieved with a combination of two subprocesses: sideway translation and fixation. In fact moving the camera in the direction of the increasing distance from the surface (tilt of the surface with respect to the camera) and always maintaining fixation on the same point will automatically accomplish the alignment task. It is possible to retrieve an estimate of the tilt[1]; in fact since the axis of maximum expansion μ of the shear bisects $\mathbf{F}$ and $\mathbf{A}_t$

$$\mu = \frac{\angle\mathbf{F} + \angle\mathbf{A}_t}{2} \tag{18}$$

and the direction of $\mathbf{A}_t$ is known, from (4) it is possible to estimate $\angle\mathbf{F}$ and then the new direction of the transverse velocity at time $i+1$ will be

$$\angle\mathbf{A}_{t_{i+1}} = \arctan\frac{\mathrm{S}_2}{\mathrm{S}_1} - \angle\mathbf{A}_{\mathrm{t_i}} \tag{19}$$

If fixation is maintained during the sideways translation then the camera will align the optical axis with the normal to the surface; fixation can be performed controlling only $\mathbf{R}_t$; since the camera is translating parallel to the image plane along the tilt of the surface, in order to fixate a rotation in the opposite direction of the translation is necessary in order to keep the fixated point in the center of the image. Thus, when fixating, the zero-order flow (u_0, v_0) of the first order approximation of the optic flow is null

$$\mathbf{A}_t + \mathbf{R}_t = \frac{\mathbf{V}_t}{d} + \mathbf{R}_t = \mathbf{0} \tag{20}$$

and then from (5) it is possible to use the estimate of the translation parameter (u_0, v_0) as an error measure. The direction of the rotation is determined by the knowledge of the direction of the sideway translation (19) and the amplitude will be determined by the amplitude of (u_0, v_0), whose measures can be fed back into a PID control block. As the camera approaches the surface, the zero order flow becomes larger in module and then a larger rotation is provided.

Otherwise, if an estimate of the distance d to the surface is available, it is possible to use it directly into (20). The easiest way to retrieve a measure of the distance from the observed surface is to express the value of d from (14)

$$d = 2\frac{V_r}{D} \tag{21}$$

Then, since $\mathbf{V}_t$ is known, $\mathbf{R}_t$ is defined.

Summarizing, it is possible to perform the docking maneuver splitting it into two independent processes: braking and aligning. These two processes can be controlled starting from two measures on the image plane which are the divergence and the shear of the first order approximation of the optic flow and with information on the egomotion. Docking is accomplished as a result of combining (17), (19) and (20).

4 EXPERIMENTAL RESULTS

An experimental set-up has been prepared (see figure 2): a camera is mounted on the end effector of a six degrees of freedom robot arm (PUMA 260) in such a way that the rotation between the camera and the end effector coordinates systems is negligible; the camera is facing a planar surface and the robot is commanded to move the camera toward the surface.

Figure 2: Setup for the docking experiment

Computation of the normal flow and estimate of the affine parameters is performed on a HP 747 while the robot is controlled from a Sparc IPC running RCCL. Docking will be considered successfully accomplished if the camera will end the maneuver in front of the surface[1] with the optical axis aligned with the normal to the surface. In order to show the correctness of the approach the estimated distance from time-to-contact and the angle between normal to the surface and viewing direction are reported below as a result of repeated trials.

[1] when the camera is near to the surface, since during the maneuver (17) is applied, its velocity is near to zero, thus computation of the optic flow is affected by large errors; it has been decided to stop the braking process whenever a fixed velocity has been reached.

Quantitative analysis of data will not be treated; in order to do it correctly, a calibrated system is necessary: this would require the knowledge of intrinsic parameters of the camera and the exact relationship between the camera and the robot frames; thus, a qualitative interpretation of the results will be discussed.

In figure 3 the real and the estimated distance from the observed surface during the braking process are reported as mean values over 10 different trials, all having the same initial conditions in terms of initial camera position and orientation; the real distance is retrieved from the robot encoders while the estimated is computed from the value of divergence and from the knowledge of the camera velocity (14). As can be seen the error tends to decrease as the maneuver proceeds; this fact can be explained considering that, as already mentioned, the estimated distance is computed without taking into account the first term of equation (6), but as the alignment process proceeds the error induced by this term becomes smaller.

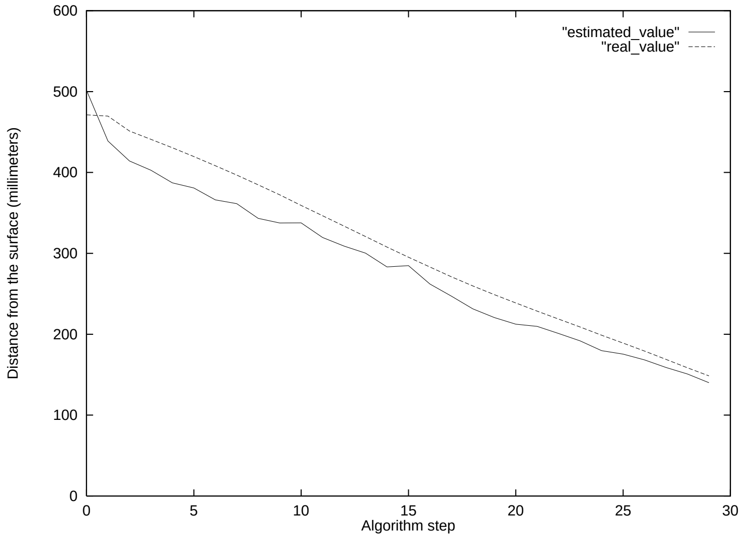

Figure 3: Real and estimated distance from the surface during the docking experiments.

The value of the angle between the normal to the surface and the camera optical axis during the maneuver is reported in figure 4; the plot is the mean angle over 20 different trials having all the same initial conditions, while the second one is the standard deviation from the mean values of the 20 samples. After an initial phase, where the error angle decreases linearly, the alignment process tend to stabilize on a constant error of 8 degrees; in reality the process tends to be very unstable, oscillating in a range between 4 and 12 degrees, as can be seen from the standard deviation in figure 5. The fact that these oscillations are not, as one could have expected, around the zero degrees value can be explained with few considerations: since the camera is mounted by hand on the end effector of the robot and since no calibration is performed to retrieve the rotation matrix between the camera and the robot frames, there could be, in fact, a rotation of some degrees between the two coordinate frames; the camera itself could have a sensor not properly mounted in such a way that optical axis and ccd are not perpendicular.

5 CONCLUSIONS

A method for docking in 3D has benn presented. This method relies on differential measures of the dynamic of the grey level patterns on the optic array and on the knowledge of egomotion. It has been described how this measures can be used to drive the docking maneuver. Results of experiments has been presented and discussed.

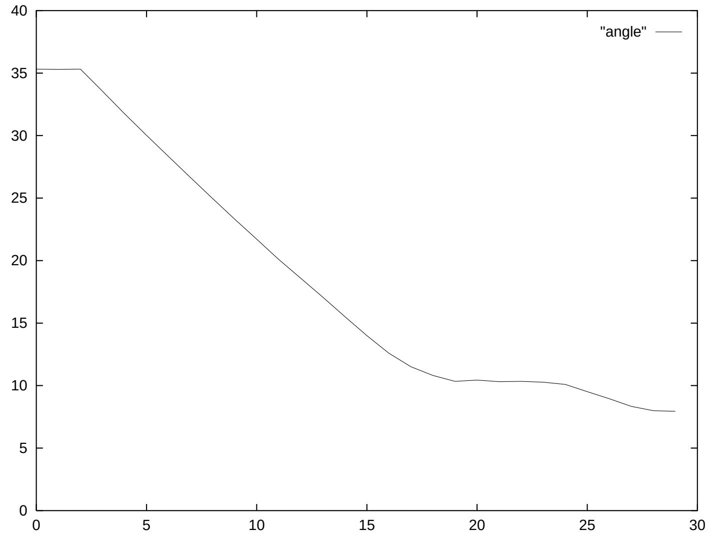

Figure 4: Angle between the optical axis and the normal to the surface during the docking experiments (in reality the zero error angle should be 180 degrees).

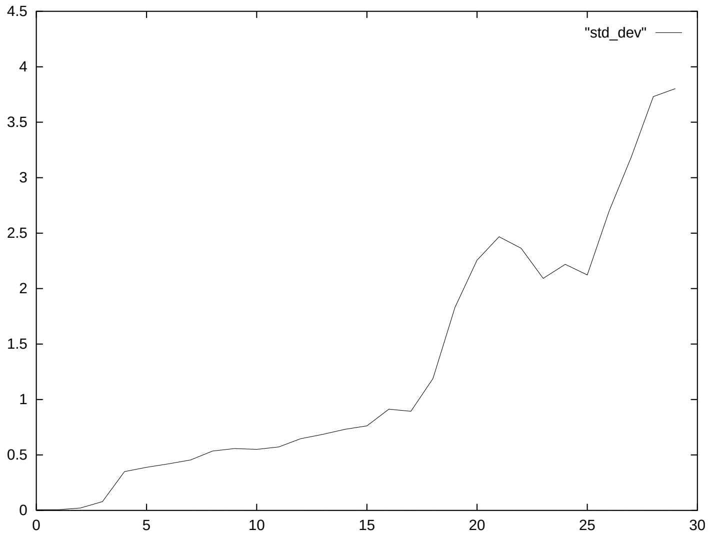

Figure 5: Standard deviation

6 ACKNOWLEDGMENTS

The research described in this paper has been supported by the Esprit project VAP-II (BR 3038/BR 7108).

A APPENDIX

In order to compute time-to-contact from equation (6), a simplification has been introduced to obtain (14). It is possible to show that for a large set of initial position of the camera with respect to the target surface and for a large set of values of the ratio between V_r and $\mathbf{V_t}$ (or $\mathbf{A_t}$ and A_r), the alignment process will end before the camera will reach the target and then the braking becomes critical; in order to better understand this, assume the camera is a point in space; let the coordinate frame be centered on the surface and let one of the axes be the normal to the surface; the direction of V_r at any instant of time is, due to the fixation process, along the line between the camera and the origin, while the direction of $\mathbf{V}_t$ is perpendicular to the first one and passes through the normal to the surface; let the initial position of the camera be any position on a plane parallel to the surface at a fixed distance. The alignment process will end whenever the camera crosses the normal to the surface; if the crossing happens before the camera has reached the origin, then for those initial conditions in terms of position and velocities the alignment is performed before the end of the braking process, in such a way that the conditions that have been set in order to write (14) hold. In figure 6, the crossing points are plotted for a set of initial conditions, i.e. different values of the ratio $\frac{|V_r|}{|\mathbf{V}_t|}$ and different initial position of the camera on a plane at fixed distance from the surface. Squared points in the plot are points where the assumption fails. As can be seen for values of the ratio between radial and transverse velocity that are below 7 and for any initial position of the camera, the alignment process will end before collision. Thus error introduced by using (14) during the maneuver will not affect the achievement of the task.

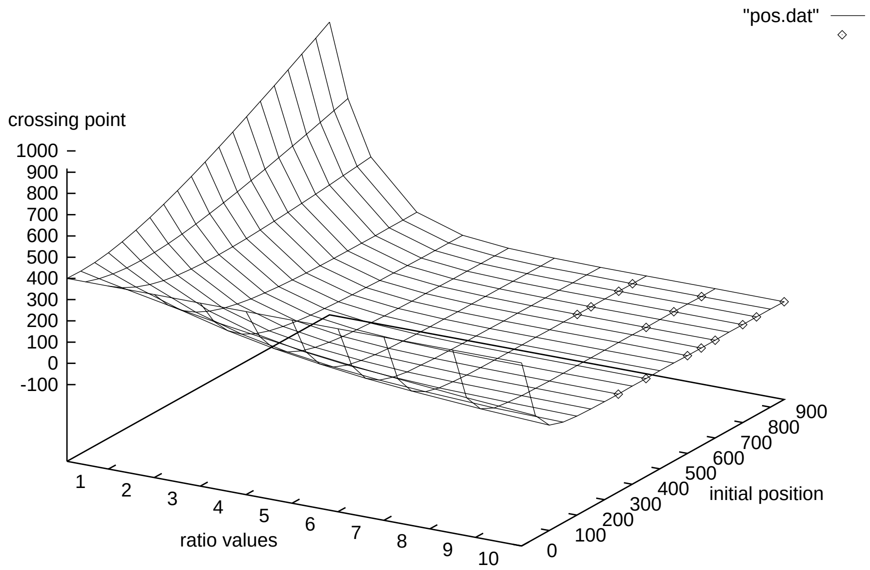

Figure 6:

B REFERENCES

[1] R. Cipolla and A. Blake. Surface Orientation and Time to Contact from Image Divergence and Deformation. *Proc. of the Second European Conference on Computer Vision* Santa Margherita Ligure, Italy, May 1992.

[2] B.K.P. Horn and B.G. Shunk. Determining Optical Flow. *Artificial Intelligence*,Vol. 17 : 185-203, 1981.

[3] 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. *Optica Acta*, 22 (9):773–791, 1975.

[4] J.J. Koenderink and A.J. Van Doorn. Optic flow. *Vision Research*, 26 (1):161–179, 1986.

[5] J.A. Santos Victor and G. Sandini Visual Behaviors for Docking. Technical Report LIRA-TR 2/94; University of Genova, Italy, June 1994.

[6] M. Subbarao. Bounds on Time to collision and Rotational Component from First Order Derivatives of Image Flow. *Computer Vision, Graphics and Image Processing* Vol. 50:329-341, 1986

[7] V. Sundareswaran, P. Bouthemy, F. Chaumette. Active Camera Self-Orientation using Dynamic Image Parameters *Proc. of the Third European Conference on Computer Vision* Stockholm, Sweden, May 1994.

[8] H. Tunley and D. Young. First Order Optic Flow from Log-Polar Sampled Images. *Proc. of the Third European Conference on Computer Vision* Stockholm, Sweden, May 1994.