Localizing Un-Calibrated, Reactive Camera Motion in an Object Centered Coordinate System

Claus B. Madsen*and Henrik I. Christensen†
Laboratory of Image Analysis, Aalborg University, DK-9000 Aalborg Ø, Denmark

Abstract

Elements of theoretical analyses are presented that enable the localization of a moving monocular camera relative to an object centered coordinate system. Two reactive motion patterns are defined and a general trajectory model is presented. It is shown how qualitative characteristics of the changing appearance of the viewed object relate to the position of the camera, when engaged in either motion pattern.

1 Introduction

The work presented in this paper is motivated by a desire to use purposive viewpoint control to explore and measure 3D properties of objects without performing reconstruction. Active use of a possibility to change viewpoint is rapidly becoming an important research area, and some work has been done on range data, e.g., [3, 4, 6].

Other research has been directed towards moving a standard CCD camera to viewpoints, where the 3D interpretation of image data is especially simplified, e.g., [8, 7, 2]. In the absence of 3D data it is necessary to devise special means of establishing the relative position between the camera and the viewed object in order to interpret the image data. Single image aspects of this problem have been thoroughly investigated in [5, 1], only a unique solution does not exist without accurate a priori knowledge about object shape.

In [8] and [7] the position of the camera is guided by maximizing lengths of image lines, so as to align the image plane with edges on objects. The present paper proposes methods for obtaining known position of a camera relative to a local object centered coordinate system. The presented approach does not require positional feedback and is based on detecting extrema in angles between lines over an image sequence as the camera is moved. Image line direction is generally a more stable feature than length.

In section 2 we present some definitions and in section 3 we investigate a characteristic relationship between view point and angle between two image lines. Section 4 explores the effect of two different camera motion patterns in terms of resulting trajectories. The camera motion patterns are based on fixating on an object point and moving around the object in circles. Section 5 shows that when embarking on those trajectories, extrema in measured angles can be related to knowledge about the position of the camera.

The methods presented in this paper have been applied to the design of a view planning strategy, (a visual behaviour), aiming at measuring the true 3D angle between edges on polyhedral objects. Please refer to [2] for details;- the present paper is concerned with the theoretical aspects of locating the camera.

2 Basic Definitions

Figure 1 shows a two legged junction and a local coordinate frame defined by the junction. The paper is concerned with locating the camera relative to this coordinate frame.

Definition: *Junction:* A junction is formed by two lines in 3-D space, L_1 and L_2, intersecting at a point Q_F. These two lines span a plane.

Definition: *True angle, Ω:* The interior angle between the two space lines of a junction is called the true angle. $\Omega \in]0;\pi[$.

Definition: *Apparent angle, ω:* The apparent angle is defined as the interior angle between two image lines resulting from perspectively projecting two junction lines onto an image plane. $\omega \in]0;\pi[$.

Definition: *Orientation of bisecting line, α:* The virtual image line arising from bisecting the two projected junction lines will be denoted as the bisecting line. The angle this line makes with the camera coordinate x-axis defines the orientation, α, (of the bisecting

*Funded partly by the Danish Technical Research Council, partly by the ESPRIT Basic Research Action BR 7108, Vision As Process

†Funded by ESPRIT BR 3038/7108, Vision As Process

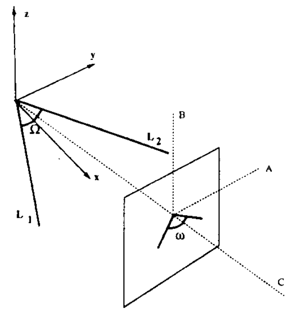

Figure 1: The Fixated Coordinate System (xyz) is defined relative to the junction formed by L_1 and L_2. A camera coordinate frame, $\vec{A}$, $\vec{B}$ and $\vec{C}$ determines the projection of the junction and the apparent angle, ω.

line). $\alpha \in [0; 2\pi[$.

Definition: *Fixated Coordinate System, (FCS):* The Fixated Coordinate System is defined as an orthonormal frame having origin at the junction point, Q_F. The x-axis of the FCS is located in the plane spanned by L_1 and L_2 so as to bisect the two lines; the z-axis is along the normal of the junction plane and the y-axis is defined by the cross product of the two other base vectors, (see figure 1).

The projection of a two-legged junction onto an image plane produces two image lines that meet at a point. Basically, three types of information are present in such an image: 1) the apparent angle, 2) the orientation of the bisecting line and 3) the lengths of the two lines. The presented approach is based on extrema in apparent angle and signs of change in orientation, which are computable from directions only. Distortions from image processing make it very hard to robustly determine lengths of image lines.

3 Apparent Angles

To study the appearance of the junction under view variation we construct a general perspective projection transformation, $M_{F \to C}$, mapping points from the FCS to camera coordinates. Let the general position of the origin of the camera coordinate frame be given by the spherical coordinates (d, θ, ϕ), (d is the distance from the focal point to the origin of the FCS). If $\vec{A}$, $\vec{B}$ and $\vec{C}$ denote the three base vectors of the camera coordinate system, we have from the fixation assumption:

$$\vec{C} = \begin{bmatrix} \sin(\theta)\cos(\phi) & \sin(\theta)\sin(\phi) & \cos(\theta) \end{bmatrix}^T \quad (1)$$

The first base vector of the camera coordinate system can be constructed by the normalized vector product of the z-axis of the FCS, $\begin{bmatrix} 0 & 0 & 1 \end{bmatrix}^T$, and $\vec{C}$, eq. (2). $\vec{B}$ is determined by the vector product of $\vec{C}$ and $\vec{A}$:

$$\vec{A} = \begin{bmatrix} -\sin(\phi) & \cos(\phi) & 0 \end{bmatrix}^T \quad (2)$$

$$\vec{B} = \begin{bmatrix} -\cos(\phi)\cos(\theta) & -\sin(\phi)\cos(\theta) & \sin(\theta) \end{bmatrix}^T \quad (3)$$

The general FCS to camera coordinate transform can be constructed directly from the three vectors $\vec{A}$, $\vec{B}$ and $\vec{C}$, as done in eq. (4), (using f to denote focal length):

$$M_{F \to C} = \begin{bmatrix} \vec{A} & \vec{B} & \vec{C} & 1/f \cdot \vec{C} \\ 0 & 0 & -d & 1 + \frac{d}{f} \end{bmatrix} \quad (4)$$

Two points, P_F and R_F, on L_1 and L_2 respectively, can be expressed in homogeneous FCS coordinates as:

$$P_F = \begin{bmatrix} \cos(\Omega/2) & -\sin(\Omega/2) & 0 & 1 \end{bmatrix} \quad (5)$$

$$R_F = \begin{bmatrix} \cos(\Omega/2) & \sin(\Omega/2) & 0 & 1 \end{bmatrix} \quad (6)$$

The apparent angle is the angle between the image lines that L_1 and L_2 project onto. The fixation constraint provides us with the knowledge, that the junction point always projects to image coordinates $\begin{bmatrix} 0 & 0 \end{bmatrix}$. Thus, a vector $\vec{l_1}$ along the image of L_1 can be expressed by the the projection of P_F, (obtained by multiplying P_F by $M_{F \to C}$); likewise, the projection of R_F specify a vector $\vec{l_2}$ along the image of L_2. Some manipulation results in the following expressions for image direction vectors of the projected junction lines:

$$\vec{l_1} = \begin{bmatrix} -\sin(\Omega/2 + \phi) \\ -\cos(\theta)\cos(\Omega/2 + \phi) \end{bmatrix} \quad (7)$$

$$\vec{l_2} = \begin{bmatrix} \sin(\Omega/2 - \phi) \\ -\cos(\theta)\cos(\Omega/2 - \phi) \end{bmatrix} \quad (8)$$

Let the angles that $\vec{l_1}$ and $\vec{l_2}$ make with $\vec{A}$, (the x-axis of the camera coordinate system), be denoted by β_1 and β_2 respectively. The apparent angle, ω, is the difference between the two angles, β_1 and β_2. Using the addition formula for tangent:

$$\tan(\omega) = \frac{2\cos(\theta)\sin(\Omega)}{(\cos^2(\theta) + 1)\cos(\Omega) - \sin^2(\theta)\cos(2\phi)} \quad (9)$$

Note, that the apparent angle, ω, is a function of true angle and the view point on the unit view sphere only. Parameters such as focal length and distance to junction do not influence ω. Figure 2 shows the apparent

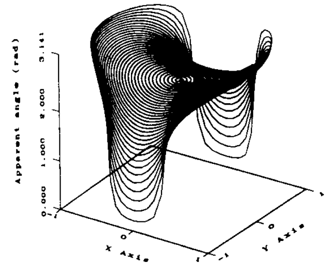

Figure 2: A plot of apparent angle over the northern view semi-sphere. The true angle in this case is $3\pi/4$. Each point on the surface represents a three-tuple $< x, y, \omega >$, where x and y specify a view point in FCS coordinates according to the sphere equation, and ω is the apparent angle.

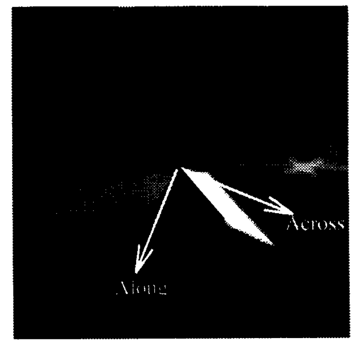

Figure 3: The two black lines mark the projection of two intersecting lines onto an image plane. The two white arrows define unique image directions forming the basis for reactive camera motion patterns, (along refers to 'along' the bisecting line; across is perpendicular to that). The arrow heads indicate what will be denoted as **positive speed. negative speed** means motion in the opposite direction.

angle as a function of view point over the entire northern hemisphere of the view sphere. It is seen that the topology is that of a saddle surface; this is the case for all true angle values.

The principal axes of the saddle surfaces are coincident with the x and y axes of the FCS. Thus, if camera motion patterns can be designed so that the view point will pass either principal axis, then extrema in apparent angle, (zero-crossings in change in apparent angle), can be related to a passing through the xz-plane or the yz-plane. The following section presents two such motion patterns. Also, it turns out that the sign of the change in orientation can be utilized for qualitative location in terms of which octant of the FCS, the camera is in.

4 Reactive Motion

A basic motion pattern has been investigated, where some *direction* in the image uniquely determines the parameters for a circle in space. First this section briefly looks at the case of a general direction;- later two specific cases will be addressed, that relate to the viewed junction.

Definition: *Motion along < direction >:* Let a view point be determined by spherical coordinates (d_0, θ_0, ϕ_0) in the FCS, valid at some instant in time, t_0. If an image line passes through the image center, then a plane in space, Π, is spanned by the origin of the FCS and two points on that image line, transformed to FCS coordinates. Motion along the direction of the image line is defined as a circle in Π of diameter d_0, centered at the origin of the FCS.

Implementation of such trajectories without known, absolute motion, follows immediately from performing fixation. Each time fixation is performed, the image plane will be tangent to a sphere centered at the fixation point, (Q_F). By performing sequences of translation in the image plane, (in the chosen direction), followed by fixation, the resulting trajectory will be a piecewise linear approximation to a planar circle in space.

If the translation between different frames is not very small compared to the distance to the object, the resulting trajectory will not be a true circle, but rather an elliptic trajectory. Section 3 proved, though, that the distance to the object does not influence the development of the apparent angle. The important fact is, that the the optical axis intersects the view sphere in a circle. This section develops a general model of this intersection, yet it will be referred to as the camera trajectory.

Let S_C and T_C be two points on the image line defining the direction of motion, where T_C is the image of the origin, Q_F, of the FCS:

$$S_C = [\ s_{c1} \quad s_{c2} \quad 0 \quad 1\ ] \quad T_C = [\ 0 \quad 0 \quad 0 \quad 1\ ] \quad (10)$$

When transforming these two points to FCS coordinates and supplementing with Q_F itself we have three points in the desired plane, Π. The inverse affine mapping of $M_{F \to C}$ transforms camera coordinates to FCS coordinates:

$$M_{C\rightarrow F} = \begin{bmatrix} \vec{A}^T & 0 \\ \vec{B}^T & 0 \\ \vec{C}^T & 0 \\ d_0 \cdot \vec{C}^T & 0 \end{bmatrix} \tag{11}$$

By multiplying with $M_{C\rightarrow F}$ the points S_C and T_C be transformed to their FCS counterparts, S_F and T_F respectively. The normal of the sought plane can then be expressed as:

$$\begin{aligned} \vec{N} &= \overline{T_F S_F} \times \overline{Q_F T_F} \\ &= \begin{bmatrix} s_{c1}\cos(\phi_0)\cos(\theta_0) - s_{c2}\sin(\phi_0) \\ s_{c1}\sin(\phi_0)\cos(\theta_0) + s_{c2}\cos(\phi_0) \\ -s_{c1}\sin(\theta_0) \end{bmatrix} \end{aligned} \tag{12}$$

Eq. (12) is the plane normal corresponding to an arbitrary image line direction given by s_{c1} and s_{c2}. Two specific directions have been investigated in detail, (figure 3): **motion along bisector**, using the direction of virtual image line which bisects the viewed junction, and **motion across bisector** using a direction perpendicular to the bisecting line. Subsequently, **motion along bisector** will be modelled as a vector function.

To get a point on the bisecting line the images of the legs of the junction can be used. Eqs. (7) and (8) give coordinates of the direction vectors of those two image lines. By normalizing to unit length and performing vector addition we get a vector along the bisecting line. Inserting the image coordinates of the point on the bisecting line into eq. (12) yields the three coordinates of the normal of the plane, (n_1, n_2 and n_3):

$$\cos(\theta_0)\sin(\Omega/2)\left(1/|\vec{l}_1| - 1/|\vec{l}_2|\right) \tag{13}$$

$$-\cos(\theta_0)\cos(\Omega/2)\left(1/|\vec{l}_1| + 1/|\vec{l}_2|\right) \tag{14}$$

$$\sin(\theta_0)\left(\sin(\Omega/2+\phi_0)/|\vec{l}_1| + \sin(\Omega/2-\phi_0)/|\vec{l}_2|\right) \tag{15}$$

By defining two angles from the coordinates of the normal as $\zeta = \arctan(n_2/n_1)$ and $\xi = \arctan(n_1/n_3)$ respectively, we can take a simple vector function for a circle in the xy-plane and transform it to have $\vec{N}$ as normal vector. The resulting general view point trajectory vector function in FCS coordinates for **motion along bisector** becomes:

$$\vec{T}_F = \begin{bmatrix} \cos(\nu t)\cos(\xi)\cos(\zeta) - \sin(\nu t)\sin(\zeta) \\ \cos(\nu t)\cos(\xi)\sin(\zeta) + \sin(\nu t)\cos(\zeta) \\ -\cos(\nu t)\sin(\xi) \end{bmatrix} \tag{16}$$

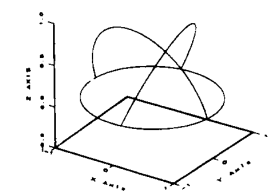

Figure 4: Perspective view of trajectories of both **motion along bisector** and **motion across bisector** given an initial view point in the fourth octant, (where the trajectories cross). The plot is in FCS coordinates, i.e., the viewed junction lies in the plane of the unit circle with legs symmetrically placed on both sides of the x-axis, (one in octant 1, the other in octant 4).

The vector function in eq. (16) is meant for subsequent qualitative time analysis of change in apparent angle and orientation. The frequency ν and time t just serve as a vehicle for representing the fact that the camera is moving. We have assumed un-calibrated motion, i.e., ν is unknown and it might even be time-varying.

5 View point location

Our main concern is to define robustly detectable features: features that are not influenced by the, possibly changing, speed of the camera. We will thus concentrate on event type characteristics, i.e., sign changes in the time derivatives.

To analyze the effect of moving the camera along the previously presented trajectories we first construct the time varying frame of the camera coordinate system. Again, **motion along bisector** will be used as example. Due to limited space, the derivations can only be sketched.

We can arbitrarily chose to orient the ordinate axis of the camera base frame along the normalized velocity vector of the trajectory:

$$\vec{B}(t) = 1/\nu \cdot \delta\vec{T}_F/\delta t \tag{17}$$

Because of fixation, the third base vector, $\vec{C}(t)$, is oriented along the direction vector of the camera frame origin, $\vec{T}_F(t)$:

$$\vec{C}(t) = \vec{T}_F/\delta t \tag{18}$$

$\delta\alpha(t)/\delta t$	$-$	$y < 0$
	0	$y = 0$
	$+$	$y > 0$
	$-+$	
	$+-$	

$\delta\omega(t)/\delta t$	$-$	$x < 0$
	0	
	$+$	$x > 0$
	$-+$	$x = 0$
	$+-$	$x > 0$

Table 1: Localization from events for **motion along bisector, positive speed**. Reverse all signs for interpretation during **negative speed**.

$\delta\alpha(t)/\delta t$	$-$	$x > 0$
	0	$x = 0$
	$+$	$x < 0$
	$-+$	
	$+-$	

$\delta\omega(t)/\delta t$	$-$	$y > 0$
	0	
	$+$	$y < 0$
	$-+$	$y > 0$
	$+-$	$y = 0$

Table 2: Localization from events for **motion across bisector, positive speed**.

$\vec{A}(t)$ may then be found as the vector product, $\vec{B}(t) \times \vec{C}(t)$, producing an expression for the abscissa axis, (this vector is not time dependent, since it is actually parallel to the normal of the plane).

From these three base vectors we can construct a time varying FCS to camera coordinate transformation matrix, $M_{F\rightarrow C}(t)$, using eq. (4).

From here on we can repeat the approach taken in section 3, i.e, transform the two points P_F and R_F to image coordinates and use them directly to form vectors along the projected junction legs.

$$\vec{l}_1(t) = (P_F \cdot M_{F\rightarrow C}(t))^T \quad (19)$$

$$\vec{l}_2(t) = (R_F \cdot M_{F\rightarrow C}(t))^T \quad (20)$$

If $\beta_1(t)$ and $\beta_2(t)$ are the time dependent counterparts of the image line orientations defined in section 3, we then have for the apparent angle and the orientation of bisecting line respectively:

$$\tan(\omega(t)) = \tan(\beta_1(t) - \beta_2(t)) \quad (21)$$

$$\tan(2\alpha(t)) = \tan(\beta_1(t) + \beta_2(t)) \quad (22)$$

Here we can use the addition formula for tangent and insert the expression for $\tan(\beta_1(t))$ and $\tan(\beta_2(t))$, to arrive at formulas describing time functions for the apparent angle and the orientation. Differentiation of these functions with respect to the time parameter t has been done, and the signs and zero-crossings have been analyzed. Tables 1 and 2 present the positional knowledge attainable from signs and sign transitions in change in apparent angle and orientation.

It is seen from the tables that 1) any combination of signs in change on apparent angle and orientation *uniquely* determines the octant in which the camera is, and 2) zero-crossings in change in apparent angle is an indicator of crossing through the xz-plane or the yz-plane.

6 Summary

The paper presents two schemes for moving a camera around an object, emphasizing how the position of the camera can be related to a local object centered coordinate system. Robust, event type characteristics, e.g., extrema in angles between image lines, are shown to be coincident with moving the camera through well-defined planes in the local coordinate system.

The results in the present paper have been applied to the design of a view planning strategy enabling the determination of angles in polyhedral scenes. This work can be found in [2].

References

[1] K. Kanatani. *Group-Theoretical Methods in Image Understanding.* Springer Series in Information Sciences. Springer-Verlag, Berlin, Germany, 1990.

[2] C. B. Madsen and H. I Christensen. Reactive view planning for quantification of local geometry. In *Proceedings: IEEE Conference on Computer Vision and Pattern Recognition, Seattle, Washington*, June 1994.

[3] J. Maver and R. Bajcsy. How to decide from the first view where to look next. In *Proceedings: DARPA Image Understanding Workshop*, pages 482 – 496, Pittsburgh, PA, September 1990.

[4] J. Maver and R. Bajcsy. Occlusions as a guide for planning the next view. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(5):417 – 433, May 1993.

[5] T. Shakunaga and H. Kaneko. Perspective angle transform: Principle of shape from angles. *International Journal of Computer Vision*, (3):239 – 254, 1989.

[6] P. Whaite and F. P. Ferrie. From uncertainty to visual exploration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(10):1038 – 1049, October 1991.

[7] D. Wilkes and J. K. Tsotsos. Acitve object recognition. In *Proceedings: IEEE Conference on Computer Vision and Pattern Recognition, Champaign, Illinois*, pages 136–141, June 1992.

[8] J. Y. Zheng, F. Kishino, Q. Chen, and S. Tsuji. Active camera controlling for manipulation. In *Proceedings: IEEE Conference on Computer Vision and Pattern Recognition, Maui, Hawaii*, pages 413 – 418, 1991.