Active Sensing with VIRTUE:

A Virtual Trinocular Stereo-Head

J. Lang
Department of Computer Science,
York University, Canada M3J 1P3
e-mail: j_lang@cs.yorku.ca

M.R.M. Jenkin
Department of Computer Science,
York University, Canada M3J 1P3
e-mail: jenkin@cs.yorku.ca

Abstract

VIRTUE: A virtual trinocular Stereo-Head, is a trinocular head implemented using a single camera mounted on the end-effector of a robotic manipulator. VIRTUE's three virtual cameras lie on the vertices of an equilateral triangle and symmetrically fixate points which lie on the normal which projects to the triangle's center. This arrangement allows for converging virtual cameras over a range of fixation distances. This paper describes VIRTUE's design; its forward and inverse kinematics, as well as the details of a trinocular stereo algorithm. A calibration procedure and some results are also included.

1 Introduction

The active observer paradigm has given rise to more integrated systems for nearly all tasks in computer vision. An active vision system [1,3,4] is a system which aims to achieve a specific computer vision task not only through some sort of image processing, but also by controlling the image acquisition process. In order to perform active image acquisition, an image sensor needs be relocatable and redirectable. In the context of active stereo vision, the sensor has to provide stereo-images of an object to be modeled from variable positions in space. This may be achieved by a specially designed stereo-head (e.g. [14,16,9]), by a pan and tilt unit mounted on a mobile robot (e.g. [6,8]) or, with even greater flexibility, by an "eye-in-hand" set-up (e.g. [19]). The term "eye-in-hand" refers to configurations consisting of an image sensor (eye) mounted at the end of a manipulator (hand).

In an active stereo system, a primary task is to fixate different locations under computer control and to recover scene structure from this sequence of fixations. An individual fixation point is the point in space where the optical axes of all image sensors intersect. (If the optical axes are parallel, the fixation point is said to be at infinity). In general, the number of cameras used in a stereo-head may vary from 2 to n. However, using three cameras is especially advantageous. Every match between any token (e.g. point, line, feature etc.) in two of the images can be verified with the aid of the third image given the visibility of the token in all three images. More than three cameras provide more redundancy between images but also increase the processing load (see [2, pp.159–160] for a more detailed discussion).

Existing trinocular stereo rigs have been built around static arrangements of cameras. Ayache's ([2]) is typical. It consists of three cameras mounted in a "L-shaped" configuration with parallel optical axes. This arrangement reduces the overlap of the visual fields but is well suited for imaging reasonably distant objects in mobile robot applications. When targets are to be imagined over a range of distances, a head capable of converging over a range of distances is more appropriate.

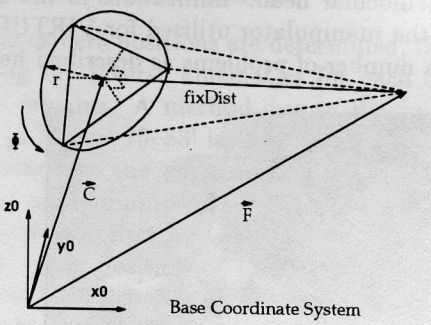

Figure 1: VIRTUE Concept

A completely controllable trinocular head has many potential degrees of freedom. In order to reduce these to a manageable number, VIRTUE was designed so that it was only capable of symmetric fixation. The three cameras which make up

VIRTUE are placed on a circle with radius r. The cameras are placed equidistantly around the circumference so that the nodal points of the cameras form an equilateral triangle. The cameras symmetrically fixate a point lying along the normal to the plane formed by the position of the three cameras, which projects to the center of the circle. This is sketched in figure 1.

Ignoring individual camera torsions, VIRTUE has eight degrees of freedom. The head is centered at $\vec{C}$, fixates a point $\vec{F}$, and has a variable baseline. VIRTUE is also free to rotate about the line $\vec{C} - \vec{F}$.

2 Implementing VIRTUE

A single eyed VIRtual TrinocUlar stEreo-head (VIRTUE) offers a number of advantages over the use of a specially designed stereo-head. The intrinsic camera parameters for each camera in the virtual trinocular head are identical (disregarding changes due to environmental conditions like temperature, pressure, etc.) since the same camera is used for each of the three sensors of the head. The underlying manipulator also has known kinematics and uses a standard controller. The disadvantages to the eye in hand approach are primarily related to simultaneity of imaging; and, the limited (five) degrees of freedom available with the arm. The lack of simultaneity of imaging is only a problem if VIRTUE is to imagine dynamic scenes or if the lighting changes due to the changing location of the robot arm. A more sophisticated manipulator capable of achieving arbitrary positions and orientations in its workspace could obviously achieve the same number of degrees of freedom as a specifically designed trinocular head. Limitations in the kinematics of the manipulator utilized for VIRTUE introduces a number of problems as described below.

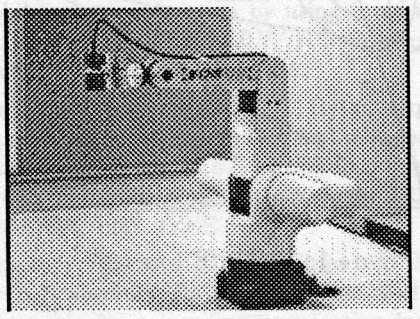

Figure 2: CRS-PLUS A150 Manipulator with Camera

At York University, VIRTUE is implemented using a CRS-Plus robotic manipulator. The CRS-Plus A150 (see fig. 2) manipulator is an all rotational joint robot (RRR). The manipulator possesses five degrees of freedom (DOF) which allows for arbitrary positioning of the camera within its workspace. It does not, however, allow for arbitrary control of the orientation of the end-effector and, hence, the image plane with respect to the optical axis of the camera cannot be independently controlled. This is due to the kinematics of the arm. The joints are arranged such that the arm is always in a plane containing the origin and z-axis of the base coordinate system.

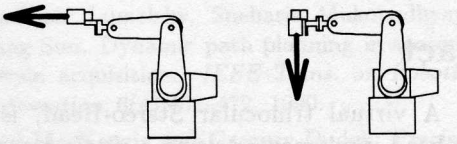

Figure 3: (a) Forward Mount, (b) Sideway Mount

In order to achieve a common fixation point for the three cameras of the head, the optical axis of the camera must be rotated with respected to the plane of the robot arm. The three virtual cameras have to be directable to points off the plane of the robot arm and, therefore, the camera mount holds the camera sideways rather than forwards. The sideway mount (see fig. 3(b)) allows the use of the wrist bend and roll to direct the camera in an arbitrary direction relative to the end of the lower arm of the manipulator. The camera will always lie on a circle around the end of the lower arm of the manipulator. The resulting single eye virtual trinocular stereo-head (VIRTUE) configuration is shown in fig. 4.

3 Controlling VIRTUE

3.1 Positioning the Head

The task of fixating with VIRTUE can be decomposed into the task of computing the poses associated with the virtual cameras that form VIRTUE and hence the appropriate arm configuration for the corresponding individual eye in hand configurations.

The center of the virtual camera-head is specified as a distance $fixDist$ form the the fixation point $\vec{F}$ of the head. The specification is completed by the azimuth ψ and elevation ζ of the head relative to the x-y-plane of the base-coordinate system. The position of the center of the head $\vec{C}$ is calculated as $\vec{C} = \vec{F} + fixDist \begin{bmatrix} \cos\psi\cos\zeta & \sin\psi\cos\zeta & \sin\zeta \end{bmatrix}^T$.

The cameras must be placed in a circle around the center of the camera-head. The plane of the

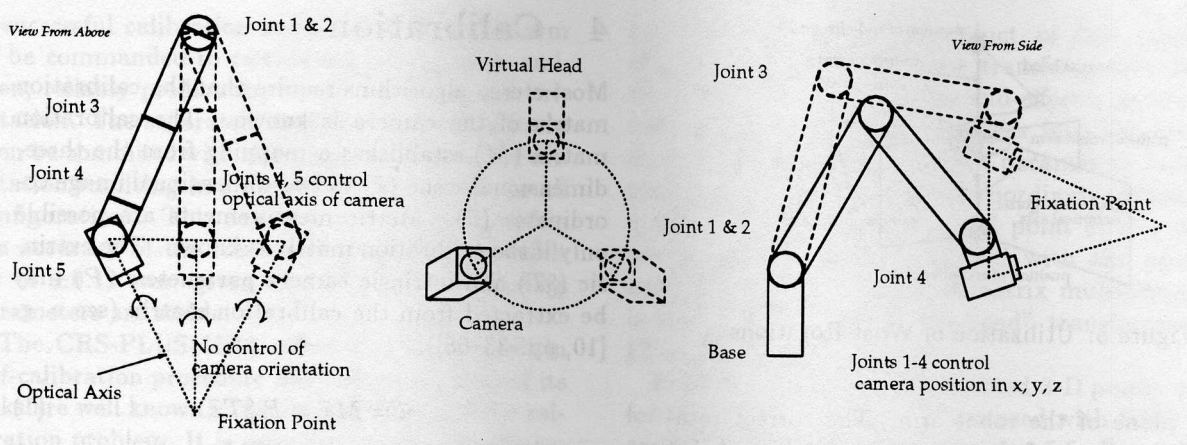

Figure 4: Single Eye Virtual Trinocular Stereo-Head (VIRTUE)

circle is orthogonal to the vector from the fixation point to the center of the camera head (see fig. 1). A plane not parallel to the z-plane may be described by parameters a, b, ρ such that $ax + by + z + \rho = 0$; the vector $\begin{bmatrix} a & b & 1 \end{bmatrix}^T$ is normal to the plane and the point $\begin{bmatrix} 0 & 0 & -\rho \end{bmatrix}^T$ lies in the plane. A similar parameterization is obtained if the plane is not orthogonal to z.

The parameters a, b, ρ need to be found based on the position of the fixation point $\vec{F}$ and the center of the head $\vec{C}$. Based on the description of the plane, the unit vectors $\hat{x}$, $\hat{y}$ and $\hat{z}$ of a coordinate system originating at the center of the head with the vector $\hat{z}$ pointing towards the fixation point are calculated as follows. $\hat{z} = \left(\vec{F} - \vec{C}\right) \Big/ \left|\vec{F} - \vec{C}\right|$, which is the unit vector from the center of the head to the fixation point. If, $\hat{z}'s$ z-component is scaled to 1, the parameters a, b of the plane can be calculated as $\begin{bmatrix} a & b & 1 \end{bmatrix}^T = \begin{bmatrix} \hat{z}_x/\hat{z}_z & \hat{z}_y/\hat{z}_z & 1 \end{bmatrix}^T$. Using the center of head as a point in the plane, the remaining parameter ρ can be computed as $\rho = -\begin{bmatrix} a & b & 1 \end{bmatrix}^T \cdot \vec{C}$.

Having established a representation of the plane of the stereo-head, and a unit vector $\hat{z}$ orthogonal to the plane, a unit-vector $\hat{x}$ in the plane is calculated. The intersection of the plane with the z-axis is given by $\begin{bmatrix} 0 & 0 & -\rho \end{bmatrix}^T$. Given that the center of the head does not lie on the z-axis, a unit vector $\hat{x}$ in the plane is calculated as $\hat{x} = \left(\vec{C} + \begin{bmatrix} 0 & 0 & -\rho \end{bmatrix}^T\right) \Big/ \left|\vec{C} + \begin{bmatrix} 0 & 0 & -\rho \end{bmatrix}^T\right|$. The third unit vector is then chosen to complete the coordinate system in a right-hand sense.

This result enables a representation of the orientation of the stereo-head and a positioning of VIRTUE's Cyclopean eye. Individual camera positions within VIRTUE are chosen to lie on a equilateral triangle centered at the midpoint of the head. An orientation of Φ with a base-line b of the stereo-head is understood to describe the following three camera positions (in the coordinate system of the head):

$$\frac{b}{\sqrt{3}} \begin{bmatrix} \cos(\Phi + n\pi) \\ \sin(\Phi + n\pi) \\ 0 \end{bmatrix}, n = -\frac{2}{3}, 0, \frac{2}{3}$$

The camera position in the (homogeneous) base-coordinate system ${}^0A$ can be obtained by coordinate transformation ${}^h_0T$ from the camera position in the coordinate system of the head ${}^hA$.

$${}^h_0T = \begin{bmatrix} \hat{x} & \hat{y} & \hat{z} & \vec{C} \\ 0 & 0 & 0 & 1 \end{bmatrix}$$

3.2 Inverse Kinematics

After the camera positions are determined, the task of solving the inverse kinematics for each camera position remains. A method due to Pieper as described in [7, pp. 75–83] is employed to map affine coordinates into the angle space of the manipulator. The method aims to establish the position of the end-effector first and, then solve for the orientation. The position of the end-effector of the CRS Plus A150 depends on the angle settings of θ_1 through θ_3. In the notation used for the arm, the 4^{th} and the 5^{th} frames are centered at the end effector. The angle θ_1 is the only angle defining the plane of the robot arm but has no influence on the radius r at which the end-effector is positioned nor on its height z.

The camera and the end of the camera mount are part of the robot arm structure and, therefore, lie

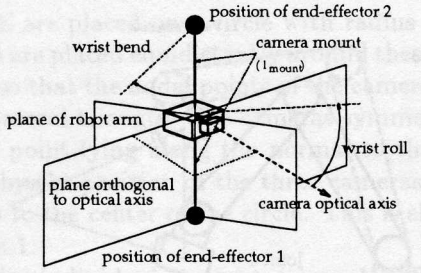

Figure 5: Utilization of Wrist Rotations

in the plane of the robot arm. The correct rotation of the plane of the robot arm is achieved if it intersects with the plane orthogonal to the optical axis of the individual camera at the camera position (see fig. 5). This rotation is in the waist of the arm, i.e. around the z-axis of the base coordinate system. The rotational axis in the wrist of the arm has to be used to direct the camera towards the fixation point. In order to achieve the correct orientation of wrist's rotational axis, the camera mount must lie on the line of intersection between the robot arm plane and the plane orthogonal to the camera's optical axis. The length of the camera mount l_{mount} determines two points on that line of intersection. In the case both planes are congruent, the length of the camera mount l_{mount} determines a complete circle in the plane of the robot arm. Singular configurations are therefore easily detected and can be dealt with.

The algorithm used for solving the inverse kinematics of the manipulator for a single camera position can be summarized:

- Use the camera position to determine the rotation of the waist.
- Determine the plane orthogonal to the optical axis of the camera containing the camera position.
- Derive a description for the plane of the robot arm.
- Determine the intersection of the plane of the robot arm with the plane orthogonal to the optical axis of the camera.
- Calculate the positions of the end of the lower arm (two points on the line of intersection, or, in the singular case the complete circle).
- Use each position of the end of the lower arm to solve for the shoulder and elbow rotation.
- Solve for the corresponding wrist bend and roll to reach the correct camera position and direction.

The inverse kinematics for the trinocular stereohead adds a preprocessing step to the above described solution for a single camera position.

4 Calibration

Most stereo algorithms require that the calibration matrix of the camera is known. The calibration matrix (M) establishes a mapping from the three-dimensional scene ($\vec{x}$) to two-dimensional image coordinates ($\vec{u}$). Metric measurements are possible only if the calibration matrix is known. The extrinsic (${}^c_0T$) and intrinsic camera parameters (P) may be extracted from the calibration matrix (see e. g. [10, pp. 33–68]).

$$\vec{u} = M\vec{x} = P\,{}^c_0T\vec{x} \qquad (1)$$

A CCD camera can be successfully modeled by a pinhole camera. The pinhole camera model has six extrinsic parameters for position and orientation of the camera in space (see e. g. [10, pp. 33–68] and [20]). The extrinsic camera parameters can be expressed as the four by four transformation matrix ${}^c_0T$ containing six independent parameters; three for translation and another three for rotation (see [7, pp. 43–56]).

A simple model for the intrinsic parameters consists of five quantities: The origin of the image coordinates relative to the optical center (u_0, v_0), the angle between the image coordinates (θ) and the effective focal length in pixels vertically and horizontally (α_u, α_v).

$$P = \begin{bmatrix} \alpha_u & -\alpha_u cot(\theta) & u_0 & 0 \\ 0 & \frac{\alpha_v}{sin(\theta)} & v_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

Most camera calibration procedures require a precisely located object of known dimensions. Tsai ([20]) specifies the necessary precision as at least an order of magnitude higher than the desired accuracy of the calibration matrix. The calibration project has to consist of more than six precisely known points in general position (see [10, pp. 58-61]). An object with two orthogonal or near orthogonal planes with a number of known points well above six satisfies. In VIRTUE two planes of the robot's base are used.

Camera calibration in the case of VIRTUE has to include not only the camera but also the robot arm (including the camera mount). The configuration of the robot arm determines the extrinsic parameters of the camera. The goal of the camera calibration has to be two-fold: establish a relationship between the angular settings of the joints of the manipulator and the extrinsic camera parameters; and, extract the intrinsic camera parameters themselves.

A successful calibration would allow the robot arm to be commanded to certain angular settings and, then, reliably predict the camera position and orientation. The separated intrinsic parameters could then be combined with this camera position and orientation resulting in a known camera model for all configurations. This can be expected work within the accuracy of the robot arm with its repeatability of ± 0.13 *mm* given that the intrinsic camera parameters are stable.

The CRS-PLUS 150A robot arm has a built-in self-calibration procedure and the dimensions of its links are well known. This is used to reduce the calibration problem. It is enough to establish a transformation from the camera to the end of the lower arm of the robot, as well as find the intrinsic camera parameters.

4.1 Calibration Procedure

In order to recover the calibration matrix, correspondence of precisely known points in the scene and in the image is established. An approach described in [12, pp. 312-315] and in [10, pp. 55-58]) expresses the elements of the calibration matrix as a function of a number of scene, and their corresponding image, points. Employing equation 1, a linear system of equations is obtained for the 11 unique parameters of M. Given more than 5.5 matches a least squares solution can be used to solve for M.

A more stable result than that obtained with the linear least squares technique is reported using non-linear constrained optimization techniques [10, pp. 63-66]). Two non-linear constraints are suggested: one ensures unit length of the third row of the extrinsic rotation parameter and the other imposes orthogonality between the image coordinates. The orthogonality constraint eliminates the intrinsic parameter θ representing the angle between the image coordinate axes. VIRTUE uses this second approach. A non-linear least squares technique due to Levenberg-Marquardt performs the following minimization (where $\lambda = 1$).

$$\sum_i \left| u_i - \frac{M_1 \cdot \vec{x}}{M_3 \cdot \vec{x}} - \lambda(1 - |M_3|)^2 \right|^2 + \left| v_i - \frac{M_2 \cdot \vec{x}}{M_3 \cdot \vec{x}} - \lambda(1 - |M_3|)^2 \right|^2$$

The coordinates of the calibration object are specified in the robot base frame. Therefore, the resulting calibration matrix is a mapping from the general robot base coordinate system to image coordinates. The calibration of the "hand-eye" configuration is thus a by-product of the camera-calibration. The coordinate transformation from the robot base frame to the end-effector is known from the mapping of joint angles into base coordinates. The external camera parameters form a transformation from the base coordinate system to a system centered at the nodal point of the camera where the z-axis is aligned with the optical axis of the camera. Simple matrix multiplication is used to retrieve the "eye-hand" transformation ${}^c_5T = {}^c_0T\,{}^0_5T$.

Figure 7 shows the reconstructed 3-D points used for calibration. The viewpoints used with VIRTUE are the same as specified in section 5.5. For a view of the calibration object see figure 6.

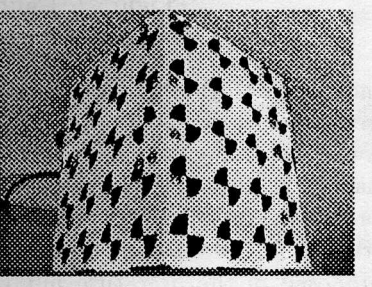

Figure 6: Image of Calibration Object

The main source of error in the calibration process is due to imprecise knowledge of the coordinates of the calibration object in the robot's base frame. A second source of error is due to uncertainty in the rigid transformation between the coordinate system of the calibration object and the robot base frame. The calibration object itself has a rather large error which is assumed to be around 2.5 *mm*.

Another major source of errors in the calibration matrix is due to the barrel distortion of the lens. This distortion is not negligible for a lens with a short focal length ($f = 8.5$ *mm*). Modeling the distortion is undesirable because of the nonlinearity of such a model. A non-linear model would greatly complicate the correspondence problem for line segments. Another approach is simply to use only the center region of the image. An improvement of 100% of the calibration matrix when only the inner 90% of the image are used is reported in [20].

5 Matching Three Views

Stereo-vision is a common approach for recovering the third spatial dimension of a scene. Given the projection of a feature into two or more cam-

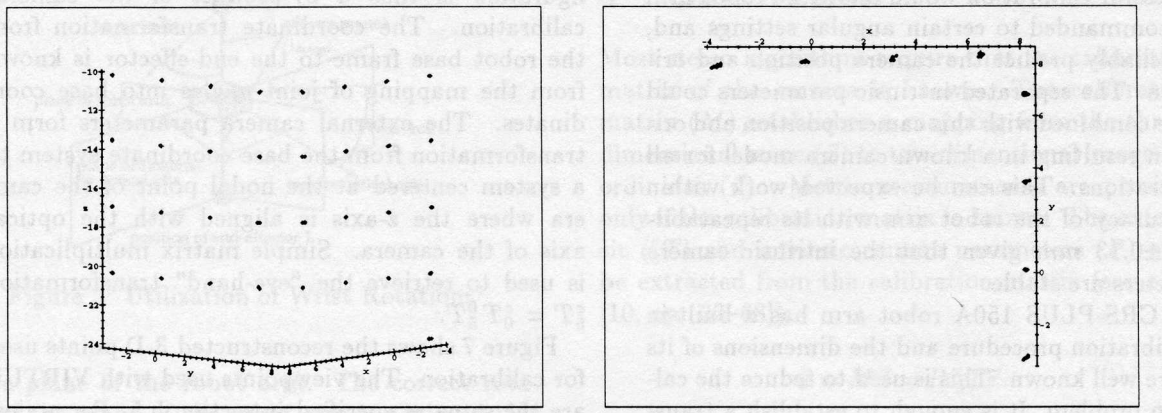

Figure 7: Calibration Points: Sideview - Topview

eras, and knowledge of the camera imaging process, the three-dimensional location of the feature can be recovered by triangulation. Actual stereo-algorithms have to address one major problem: establishing correspondences between image features. Several algorithms for establishing correspondence have been proposed in the literature but no optimal algorithm exists for all image scenes (for an overview, see [13]). In order to reduce the complexity of the matching process, different constraints can be applied. Perhaps the most powerful constraint in stereo matching is the epipolar constraint. The epipolar constraint refers to the fact that the line that passes through an image point and the optical center of one camera maps to a line in the image plane of another camera and that searches for the image point are thus constrained to a line in the other image. Other commonly used constraints are the uniqueness and the ordering constraints. The uniqueness constraint states that every token in one image has at most one corresponding token in the other image. The ordering constraint assumes that the order of tokens on the same epipolar line is the same in different images. While the epipolar constraint is true in all cases (given that the pinhole camera model holds), the uniqueness and order constraint may be violated in real image scenes (for a more detailed discussion, see [15, pp. 226–240]).

One matching approach which obtains good results in man-made environments like offices is due to Ayache [2]. The approach is based on matching intensity edges approximated by straight line segments. Ayache has formulated algorithms for binocular and trinocular stereo vision. The trinocular stereo algorithm uses only the epipolar constraint and some heuristics to match line segments. In particular, it does not depend on the commonly applied but not universally true uniqueness and ordering constraints.

VIRTUE uses Ayache's approach to recover object structure. Ayache's algorithm can be broken down into the following steps; obtaining the images, applying an edge detector, approximating the intensity edges by straight lines, rectifying the images, finding matching segments and, finally, reconstructing the 3-D lines.

5.1 Edge Detection and Line Approximation

The Canny edge detector [5] is used to detect local line segments. The current implementation uses between two and six filter directions and provides independent control of the 2-D Canny filter.

Peaks of the filter response are declared edges. The peaks are found using spline fitting in the direction of the edge detector (see [11, pp. 504-507]). The detected peaks are mapped back to pixel coordinates and the responses for all directions of edge detection are combined. The results are edge contours with a width of one or two pixels (disregarding the distortion at contour corners).

The edge contours are thinned (see [17, pp. 199-201] for a description of the algorithm) and approximated as straight lines. The contours are represented as chains. Each chain is initially approximated as a straight line from its endpoints. A collinearity test due to Pavlidis ([17, pp. 281-288]) is performed breaking the chains, such that a maximum error in pixel displacement is not violated. These line segments are then improved using a robust line estimator (see [18, pp. 703-705]). The re-

sulting segments are joined if their ends are close and the collinearity test succeeds (which is employed again).

5.2 Image Rectification

The coordinates of the line approximations are transfomed in a rectification operation (for a definition see [2, pp. 30-36]) in order to speed up the matching process. Rectification ensures that the image points of identical scene points (u_i, v_i) have the properties that; $v_1 = v_2$, $u_3 = u_1$ and $-u_2 = v_3$. In addition, the nodal points of the camera before and after rectification are the same and the focal plane of all three cameras are identical. This is equivalent to placing the cameras so that the epipolar lines are horizontal between views one and two, and vertical between one and three. The image rectification applied here is due to Ayache [2, pp. 30-36].

5.3 Matching

Stereo-matching establishes the correspondence between 2-D line segments. The approach of Ayache [2, pp. 129-143] is to reduce the matching problem by the epipolar constraint for the ends of line segments. In the trinocular case, for each line segment in one image a corresponding line segment in one of the other images is searched for. A match is only accepted if a match between line segments in two images can be verified by the third image. In VIRTUE the matching algorithm is applied three times, each time with a different image of a trinocular view as the third verification image.

The matcher is conservative and produces only a small number of false matches. However, it relies heavily on the assumption that a three-dimensional line segment maps to one line segment in all three images. This assumption is rather optimistic, especially if the line approximation produces a fragile segmentation or if there exists considerable occlusion in the images. Line approximation will always be fragile for continuously curved edges, e.g. from spheres.

5.4 3-D Line Reconstruction

The last step in the stereo-algorithm is to reconstruct a 3-D line from matching 2-D image lines. Reconstruction based on the endpoints of the 2-D image line segments could be used if these endpoints in the image would be the exact mappings of the endpoints of the corresponding 3-D line segment. However, segmentation errors make this approach inappropriate. A better approach is to calculate the intersection of the planes defined by the line segments in the image; see Ayache [2, pp. 238-240] for more details.

5.5 Preliminary Results

The result of the stereo-algorithm, i.e. the 3-D lines, is shown in figure 9. The result is obtained as follows: Three camera positions are selected by VIRTUE given a fixation point of $[\ 9.5 \quad 8.0 \quad -15.5\ ]$. The azimuth angle was set to $\frac{\pi}{4}$, the elevation angle to 0 and the viewing distance to 19.0 *cm*. The base length for VIRTUE was set to 14.0 *cm*. The camera positions determined by VIRTUE are $[\ 18.0 \quad 26.3 \quad -19.6\ ]$, $[\ 27.9 \quad 16.5 \quad -19.5\ ]$ and $[\ 22.9 \quad 21.5 \quad -7.42\ ]$. Two images are recorded from each viewpoint, the first one of the calibration object, the second of the object to be modeled. Both views from the third viewpoint are shown in figures 6 and 8. Figure 9 shows the result of the 3-D line reconstruction using the automated edge detection and line approximation of section 5.1. The underlying grid shows a ground "truth" but with high error since the object is only a pattern printed on paper.

6 Conclusion

This paper describes the development of VIRTUE: A single eye virtual trinocular stereo-head. VIRTUE is a trinocular head with individual cameras positioned on an equilateral triangle centered at the midpoint of the head. VIRTUE's three virtual cameras symmetrically fixate points which lie on the normal to the equilateral triangle which projects to the triangle's center.

An implementation using a CRS-Plus robotic manipulator is described. The CRS-Plus manipulator was shown to be sufficient to allow for arbitrary po-

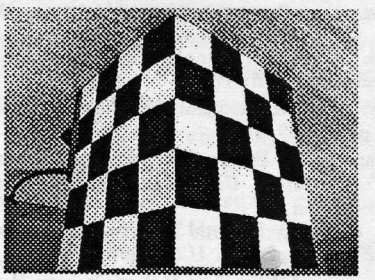

Figure 8: Image of Modeling Object

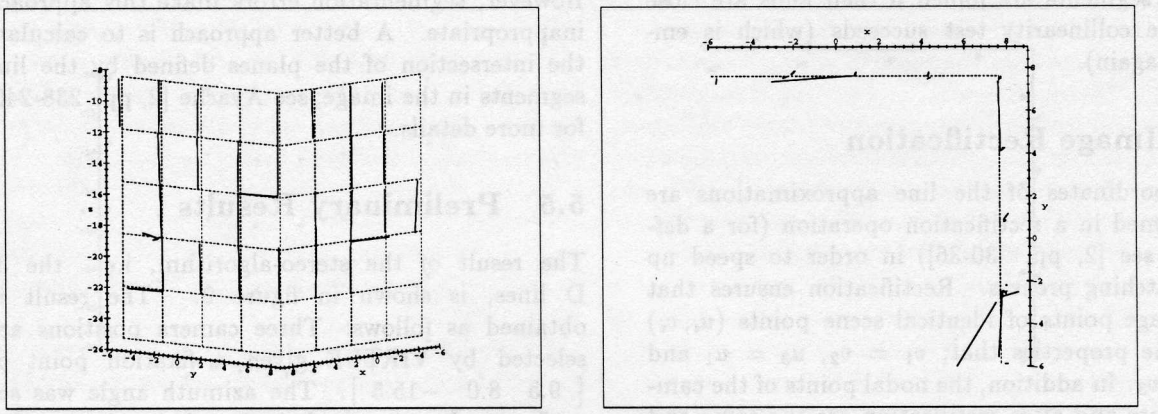

Figure 9: Reconstructed 3-D Lines: Sideview - Topview (Automatic Line Segmentation)

sitioning of the head within its workspace but not, however, for arbitrary control of the orientation of the image plane with respect to the optical axis of each virtual camera.

A trinocular stereo-algorithm has been implemented using VIRTUE. The trinocular stereo-algorithm is based on algorithm by Ayache [2]. It has been shown to produce reasonable results in the case of a simple object. An active object modelling system based on VIRTUE is currently being developed.

VIRTUE's calibration method is currently being refined in order to obtain a more accurate estimate of the eye-hand relation. Also, as mentioned in section 4.1, introducing a separate reference frame for the calibration object and the manipulator is desirable. Without this information it is not possible to compose a calibration matrix for an arbitrary viewpoint with accuracy. Improvements may also be achieved by using a multiple view calibration procedure and a more precisely machined calibration object.

References

[1] J.Y. Aloimonos, I. Weiss, A. Bandyopadhyay. Active Vision. *Int. J. of Computer Vision*, Vol.1, pp. 333-356, 1987.

[2] N. Ayache. *Artificial Vision for Mobile Robots.* MIT Press, Cambridge, MA, 1991.

[3] R. Bajcsy Active Perception Vs. Passive Perception. *WCVRC*, pp. 55-59, 1985.

[4] D.H. Ballard, C.M. Brown. Principles of Animate Vision. *CVGIP: Image Understanding*, 56(1), pp 3-21, 1992.

[5] J. Canny. A Computational Approach to Edge Detection. *IEEE Trans. PAMI*, 1(2), 1986.

[6] H.I.Christensen. A Low-Cost Robot Camera Head. *Int. J. of Pattern Recognition and Artificial Intelligence*, 7(1), pp. 69-87, 1993.

[7] J.J. Craig. *Introduction to Robotics: Mechanics and Control.* Addison-Wesley, Reading, MA, 2nd ed., 1989.

[8] J.L. Crowley, P. Bobet, M. Mesrabi. Layered Control of a Binocular Camera Head. *Int. J. of Pattern Recognition and Artificial Intelligence*, 7(1), pp. 109-122, 1993.

[9] N.J. Ferrier, J.J. Clark. The Harvard Binocular Head. *Int. J. of Pattern Recognition and Artificial Intelligence*, 7(1), pp. 9-31, 1993.

[10] O. Faugeras. Three-Dimensional Computer Vision. MIT Press, Cambridge, MA, 1993.

[11] J.D. Foley, A. van Dam, S.K. Feiner, J.F. Hughes. *Computer Graphics: Principles and Practice.* Addison-Wesley, Reading, MA, 2nd ed., 1990.

[12] B.K.P. Horn. *Robot Vision.* MIT Press, Cambridge, MA, 1986.

[13] M.R.M. Jenkin, A.D. Jepson, J.K. Tsotsos. Techniques for Disparity Measurement. *CVGIP: Image Understanding*, 53(1), pp. 14-30, 1991.

[14] E. Krotkov, J.F. Summers, F.Fuma. An agile stereo camera system for flexible image acquisition. *IEEE J. of Robotics Automation*, 4(1), pp. 108-113, 1987.

[15] V.S. Nalwa. *A Guided Tour of Computer Vision.* Addison-Wesley, Reading, MA, 1993.

[16] E. Milios, M. Jenkin, J. Tsotsos. Design and Performance of TRISH, a Binocular Robot Head with Torsional Eye Movements. *Int. J. of Pattern Recognition and Artificial Intelligence*, 7(1), pp. 51-68, 1993.

[17] T. Pavlidis. *Algorithms for Graphics and Image Processing.* Computer Science Press, Rockville, MD, 1982.

[18] W.H. Press, S.A. Teukolsky, W.T. Vetterling, B.P. Flannery. *Numerical Recipes in C: The Art of Scientific Computing.* Cambridge University Press, NY, 2nd ed., 1992.

[19] J.R.G. Pretlove, G.A. Parker. The Surrey Attentive Robot Vision System. *Int. J. of Pattern Recognition and Artificial Intelligence*, 7(1), pp. 89-107, 1993.

[20] R.Y. Tsai. An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision. *Proc. of the IEEE Computer Soc. Conf. on Computer Vision and Pattern Recognition*, pp. 364-374, Miami Beach, Fl, 1986.

Acknowledgements

The authors greatfully acknowledge the support of the Federal Networks of Centres of Excellence IRIS project, and NSERC Canada.