

Active Stereo Tracking of $N \leq 3$ Targets Using Line Scan Cameras

Joao P. Barreto, *Member, IEEE*, Luis Perdigoto, *Student Member, IEEE*, Rui Caseiro, and Helder Araujo

Abstract—This paper presents a general approach for the simultaneous tracking of multiple moving targets using a generic active stereo setup. The problem is formulated on the plane, where cameras are modeled as “line scan cameras,” and targets are described as points with unconstrained motion. We propose to control the active system parameters in such a manner that the images of the targets in the two views are related by a homography. This homography is specified during the design stage and, thus, can be used to implicitly encode the desired tracking behavior. Such formulation leads to an elegant geometric framework that enables a systematic and thorough analysis of the problem at hand. The benefits of the approach are illustrated by applying the framework to two distinct stereo configurations. In the first case, we assume two pan-tilt-zoom cameras, with rotation and zoom control, which are arbitrarily placed in the working environment. It is proved that such a stereo setup can track up to $N = 3$ free-moving targets, while assuring that the image location of each target is the same for both views. The second example considers a robot head with neck pan motion and independent eye rotation. For this case, it is shown that it is not possible to track more than $N = 2$ targets because of the lack of zoom. The theoretical framework is used to derive the control equations, and the implementation of the tracking behavior is described in detail. The correctness of the results is confirmed through simulations and real tracking experiments.

Index Terms—Active vision, computer vision, stereo, visual servoing, visual tracking.

I. INTRODUCTION

ACTIVE tracking is a part of the *active vision* paradigm [1], [2], where visual systems adapt themselves to the observed

Manuscript received July 6, 2009; revised January 17, 2010 and March 25, 2010; accepted March 25, 2010. Date of publication April 26, 2010; date of current version June 9, 2010. This paper was recommended for publication by Associate Editor D. Kragic and Editor G. Oriolo upon evaluation of the reviewers' comments. The work of J. P. Barreto, H. Araujo, and L. Perdigoto was supported by the Portuguese Science Foundation through Grants PTDC/EEA-ACR/68887/2006 and SFRH/BD/50281/2009. This work was also supported by the project Perception on Purpose (POP), funded by the European Union under the reference FP6-IST-2004-027268. Parts of the work were presented at the IEEE International Conference on Computer Vision and Pattern Recognition, Miami, FL, June 2009.

J. P. Barreto, R. Caseiro, and H. Araujo are with the Institute for Systems and Robotics and the Department of Electrical and Computer Engineering, Faculty of Science and Technology, University of Coimbra, 3030 Coimbra, Portugal (e-mail: jpbarto@isr.uc.pt; ruicaseiro@isr.uc.pt; helder@isr.uc.pt).

L. Perdigoto is with the Institute for Systems and Robotics, University of Coimbra, 3030 Coimbra, Portugal, and also with the Department of Electrical Engineering, Polytechnic Institute of Leiria, 2400 Leiria, Portugal (e-mail: perdigoto@estg.ipleiria.pt).

This paper has supplementary downloadable material available at <http://ieeexplore.ieee.org>, provided by the author. The material includes three multimedia AVI format movie clips that show experimental results in active tracking.

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TRO.2010.2047300

environment, either to obtain extra information or to perform a task more efficiently. Active tracking consists of controlling the degrees of freedom (DOF) of robotized cameras such that specific scene objects are imaged in a certain manner. An example of active tracking is fixation, where camera control assures that gaze is kept on the same object over time.

Fixation can be performed with either one camera (monocular fixation) or two cameras (binocular fixation). The former typically employs a pan-tilt-zoom (PTZ) camera such that the point of interest is aligned with the optical axis and projected at the image center (fovea) [3], [4]. The latter usually considers a stereo head [5], with the point of interest being foveated by intersecting the optical axes of both cameras at the exact target location (*the vergence/fixation point*) [5]–[7]. Since the fixation point lies in the horopter [8], many binocular systems use target disparity between retinas as feedback-control signal [9]. In general, and at the low image level, fixation can be formulated as a regulation-control problem that does not require explicit target identification or expensive image processing [10]. The ability to fixate can be helpful to simplify a broad range of high-level vision tasks, which include object recognition [11], 3-D reconstruction [12], robot navigation [13], monocular depth inference [14], and robot docking [15].

While fixation concerns tracking a single object, this paper addresses the problem of using an active stereo setup to simultaneously track $N > 1$ free-moving points of interest. Sommerland and Reid proposed an information theoretical framework for tracking multiple targets with multiple PTZs [16]. However, they addressed problems, e.g., sensor-target assignment, camera hand-off, and zoom control with no missing new objects, whereas the focus of this paper is toward extending the classical binocular fixation framework for the case multiple points of interest. We aim to push the single focus of the attention paradigm, which is typical of binocular fixation, toward a more general multifocal attention framework [17].

In this paper, we show that it is possible to control the cameras' parameters such that the two views of the N targets are related by a homography. This homography H —henceforth, called the *configuration homography*—is specified in advance and maps points of interest in one image into corresponding points of interest in the other image. A suitable choice of H can either ensure that the N objects are simultaneously visible in both images and/or enforce a particular relation between views that can simplify certain high-level visual tasks. This formulation leads to a geometric framework that, for a particular stereo configuration S , desired homography H , and number of targets N , enables the feasibility of the tracking task to be determined,

as well as the derivation of the relevant constraints on the control parameters.

This problem is fully formulated in the plane, with the cameras being modeled as “line scan cameras,” and the targets being described as 2-D points. Such simplified model is often used in binocular-fixation algorithms, where the *control in tilt* assures the alignment between the object and the plane defined by the cameras’ optical axes [6], allowing for the tracking in that plane. For the case of multiple target tracking, the alignment in tilt is often impossible to achieve (e.g., for $N > 3$, the points of interest are in general noncoplanar). However, and for practical purposes, we can always consider projecting the 3-D target motion onto the plane defined by the cameras’ optical axes. We will show that this solution is in particular effective for indoor applications, where trajectories are usually parallel to the ground plane.

A. Structure and Main Contributions

Our main contribution is the theoretical framework to control the active camera system such that the images of the targets in the two views are related by a homography. The formulation is general and enables a systematic and thorough analysis of the tracking by an arbitrary active stereo setup. The analysis includes devising possible strategies, deciding about the feasibility of the task, and deriving the control equations required for the practical implementation. The benefits of the framework are illustrated by applying it to two distinct stereo configurations. The first scenario assumes two PTZ cameras that have been arbitrarily placed in the work environment, while the second scenario focuses on active tracking using a custom made robot head with neck-pan motion and independent eye rotation.

The structure of the paper is as follows: Section II introduces the notation and relevant background concepts. Section III presents the projection model for “line scan cameras” and inherent geometry. The *homographic curve* associated to an arbitrary *configuration homography* is derived in Section IV, as a function of the system kinematics and camera’s intrinsic parameters. The tracking problem is cast as the control of the active system parameters such that the homographic curve goes through the free-moving targets. Section V applies the mathematical formulation to the case of tracking multiple targets using two PTZ cameras. The feasibility study shows that it is possible to track up to $N = 3$ free-moving targets, while assuring that the image location of each target is the same in both views. For $N = 2$, there is an infinite number of solutions for the control parameters that give rise to different possible tracking strategies. We implement two of these strategies using commercial PTZ units with low-bandwidth position control. For $N = 3$, the number of solutions is finite. Simulation results show that the tracking is possible in practice, but it requires cameras with a wide field-of-view (FOV). Section VI discusses active tracking using the robot head. It is shown that since the system has no zoom control, it is only possible to track a maximum of $N = 2$ targets. Experimental results of the active-tracking behavior are presented, with the theoretical framework being applied to derive control equations in position and velocity, respectively. Section VII

concludes with a discussion about the possible new applications of the framework.

This paper extends the Perdigoto *et al.* approach, where the mathematical framework of the *configuration homographies* was presented for the first time. Thus, there is a substantial overlap between [18] and the contents of Sections II–IV. This paper provides further details about the application of the framework to the two situations stated earlier. In particular, the space of solutions for the different tracking tasks are derived and discussed, simulation results for the tracking of $N = 3$ targets with the PTZs, as well as real experiments in tracking $N = 2$ targets, are presented, and a detailed description of the tracking implementation using the POPEYE robot head is provided [see Fig. 11(a)].

II. NOTATION AND BACKGROUND

We do not distinguish between a projective transformation and the matrix representing it. Matrices are represented by symbols in sans serif font, e.g., M , and vectors by bold symbols, e.g., \mathbf{Q} . Equality of matrices or vectors up to a scalar factor is written as \sim . Points, lines, and conic curves, unless stated otherwise, are represented in projective homogeneous coordinates. The following sections briefly introduce background concepts that will be used in the subsequent parts.

A. Vectorization of Matrix Equations

Let Y , A , X , and B be rectangular matrices such that

$$Y = A X B.$$

The aforementioned equality can be rewritten as

$$\text{vec}(Y) = (B^T \otimes A) \text{vec}(X)$$

where \otimes denotes the Kronecker product, while $\text{vec}(X)$ and $\text{vec}(Y)$ are the column-wise vectorizations of matrices X and Y (cf., [19, Ch. 4]). It is also convenient to keep in mind the following property of the Kronecker product:

$$(A B) \otimes (C D) = (A \otimes C) (B \otimes D).$$

B. Vector Representation of Conic Curves

Let us consider a point in the plane, with homogeneous coordinates

$$\mathbf{X} = (x \quad y \quad z)^T$$

and a conic curve represented by the symmetric matrix

$$\Omega \sim \begin{pmatrix} a & b/2 & d/2 \\ b/2 & c & e/2 \\ d/2 & e/2 & f \end{pmatrix}.$$

The point \mathbf{X} is on the conic curve *iff*

$$\mathbf{X}^T \Omega \mathbf{X} = 0.$$

This second-order polynomial can be rewritten in the following form:

$$\omega^T \hat{\mathbf{X}} = 0$$

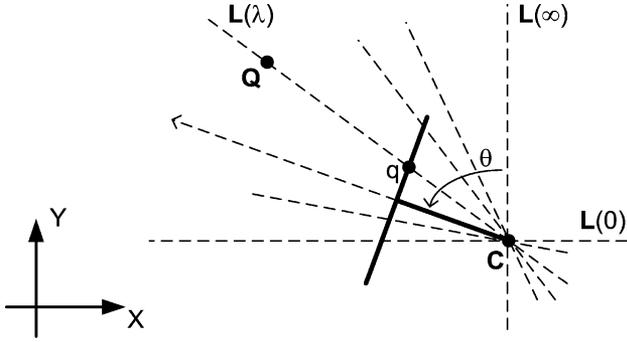


Fig. 1. Line-scan camera model, where OXY is the world system of coordinates. The camera is centered in C and rotated by an angle θ with respect to the Y -axis. The 2-D point Q is projected into q in the line image. The dashed lines $L(\lambda)$ represent the pencil of the back projection.

with \hat{X} being the *lifted point coordinates* of X

$$\hat{X} = (x^2 \quad xy \quad y^2 \quad xz \quad yz \quad z^2)^T$$

and ω a vector representation of the conic curve

$$\omega = (a \quad b \quad c \quad d \quad e \quad f)^T.$$

III. MODELING THE LINE-SCAN CAMERA

In this paper, the objects are described as free-moving points, and the cameras are modeled as “line-scan cameras” that can translate and rotate around an axis orthogonal to the plane of motion. The geometry of unidimensional cameras has already been studied under different contexts of application [20], [21]. This section introduces the projection and back-projection models that will be considered in the remaining part of the paper.

A. Projection Matrix

Fig. 1 shows a line scan camera with projection center

$$C = (C_x \quad C_y \quad 1)^T$$

and matrix of intrinsic parameters

$$K \sim \begin{pmatrix} f & 0 \\ 0 & 1 \end{pmatrix}$$

with f standing for the focal length. With no loss of generality, it will be assumed that the origin of the image is coincident with the principal point.

Let Q be a generic point in the plane, and q is the 1-D projective representation of its image. The projection for the line scan camera can be carried as follows:

$$q \sim KR(1 \quad -C')Q \quad (1)$$

where 1 denotes the 2×2 identity matrix, C' is the nonhomogeneous representation of the projection center, and R encodes the rotation of the camera by an angle θ

$$R = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}.$$

The result of (1) is a 2×3 version of the standard-projection matrix for the case of 1-D cameras [8].

B. Back-Projection Pencil

Let us now consider the problem of computing the back projection of an image point q . We define the matrix U such that

$$U \sim \begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}.$$

Since U denotes a rotation by an angle of 90° , it is easy to verify that

$$q^T U q = 0 \quad \forall q \in \mathbb{P}^1.$$

By left multiplying both sides of (1) by $q^T U$, it follows that

$$\underbrace{q^T U K R}_{L^T} (1 \quad -C') Q = 0 \quad (2)$$

where L is a vector with length 3 that can be interpreted as the homogeneous representation of a line in the plane. Since L goes through Q and C , respectively, then it corresponds to the back projection of the image point q .

Let us define λ such that

$$\begin{pmatrix} \lambda \\ 1 \end{pmatrix} \sim R^T K^T U^T q. \quad (3)$$

For each image point q , there is a λ value that parametrizes the corresponding back-projection line according to the formula

$$L(\lambda) \sim (1 \quad -C')^T \begin{pmatrix} \lambda \\ 1 \end{pmatrix}$$

where $L(\lambda)$ is the pencil of lines going through the camera center C . It can be shown that for $\lambda = 0$, the line is parallel to the X -axis, while for $\lambda = \infty$, the line becomes parallel to the Y -axis (see Fig. 1).

IV. HOMOGRAPHIC CURVE

We propose using a 1-D configuration homography H to specify the desired tracking behavior. The idea is to control the active stereo system such that the two views of the N targets are mapped one into the other by the homography H . This section discusses the locus of points in the working plane whose stereo projections are related by a given homography. We show that this locus is in general a conic curve going through the two camera centers. The curve—henceforth called the *homographic curve*—depends both on the chosen H and on the configuration of the stereo setup. Thus, the active-tracking problem can be formulated as the manipulation of the cameras' parameters such that the homographic curve goes through the N free-moving targets. Note that the homographic curves generalize both the horopter [8] and the isodisparity curves proposed in [22]. The former is the homographic curve for $H \sim 1$, with points being projected at the same location in both images. The latter corresponds to the case of H being a 1-D translation that shifts image points by a constant amount.

A. Image Homography and Pencil Homography

Fig. 2 shows two cameras with centers C_1 and C_2 , rotation matrices R_1 and R_2 , and intrinsics K_1 and K_2 , respectively. Let

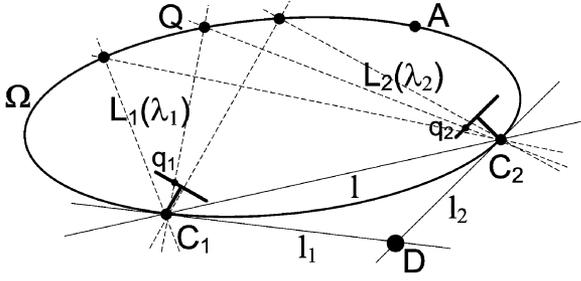


Fig. 2. Homography H induces an homographic relation between the back-projection pencils. It follows from Steiner's theorem that corresponding lines in two homographically related pencils intersect on a conic locus. Ω is called the homographic curve of H because the images of any point Q on Ω satisfy $q_2 \sim Hq_1$.

the desired image homography be

$$H \sim \begin{pmatrix} a & b \\ c & d \end{pmatrix} \quad (4)$$

where H maps points q_1 in the first view into points q_2 in the second view

$$q_2 \sim Hq_1.$$

Let us now consider the parameterization of the back-projection pencils discussed in Section III-B. Each image point q corresponds to a back-projection line that is parameterized by λ . Let λ_1 and λ_2 be the parameters associated with q_1 and q_2 . Inverting (3)

$$q \sim UK^{-T}R \begin{pmatrix} \lambda \\ 1 \end{pmatrix}$$

and replacing in the homography equation, we obtain that

$$\begin{pmatrix} \lambda_2 \\ 1 \end{pmatrix} \sim H_L \begin{pmatrix} \lambda_1 \\ 1 \end{pmatrix}$$

with

$$H_L \sim R_2^T K_2^T U^T H U K_1^{-T} R_1.$$

The image homography H defines a correspondence between back-projection lines. This correspondence is described by H_L that maps lines of the pencil going through C_1 into lines of the pencil going through C_2 . The equation relating H and H_L has been provided earlier and can be rewritten in a vectorized form (see Section II-A). It follows that

$$\text{vec}(H_L) \sim MF \text{vec}(H) \quad (5)$$

where M is a 4×4 matrix depending on the rotation angles θ_1 and θ_2

$$M \sim R_1^T \otimes R_2^T$$

$$\sim \begin{pmatrix} 1 & -\tan \theta_2 & -\tan \theta_1 & \tan \theta_1 \tan \theta_2 \\ \tan \theta_2 & 1 & -\tan \theta_1 \tan \theta_2 & -\tan \theta_1 \\ \tan \theta_1 & -\tan \theta_1 \tan \theta_2 & 1 & -\tan \theta_2 \\ \tan \theta_1 \tan \theta_2 & \tan \theta_1 & \tan \theta_2 & 1 \end{pmatrix}$$

and F is a matrix encoding the intrinsic parameters

$$F \sim (K_1^{-1} \otimes K_2^T)(U^T \otimes U^T) \\ \sim \begin{pmatrix} 0 & 0 & 0 & f_2 \\ 0 & 0 & -1 & 0 \\ 0 & -f_1 f_2 & 0 & 0 \\ f_1 & 0 & 0 & 0 \end{pmatrix}.$$

B. Equation of the Homographic Curve

The homography H transforms points q_1 into points q_2 , and establishes an implicit correspondence between the respective back-projection lines. This correspondence is parameterized by the 1-D pencil homography H_L that maps lines going through C_1 into the lines going through C_2 . Let Q be the intersection point of two corresponding back-projection lines. Since the projection of Q in the two views must satisfy the original image homography H , the homographic curve that we are looking for is the locus of all intersection points Q .

There is a well-known result from projective geometry—the *Steiner's theorem*—stating that locus of intersection of two homographically related pencils is a conic curve going through the centers of the pencils [8], [23]. Since our back-projection pencils are related by a homography H_L , the homographic curve is always a conic Ω going through the camera centers C_1 and C_2 (see Fig. 2). Given the 1-D pencil homography H_L and the position of centers C_1 and C_2 , it is possible to derive an explicit expression for the conic Ω . The procedure is described in detail in [23, cf., Chs. 5 and 6, respectively, for details and background] and briefly outlined as follows.

- 1) Let l be the line defined by C_1 and C_2 , respectively, as shown in Fig. 2. Since l belongs simultaneously to both back-projection pencils, we compute the corresponding parameters λ_1 and λ_2 such that

$$l \sim L_1(\lambda_1) \sim L_2(\lambda_2).$$

- 2) Let us assume l as a line belonging to the pencil L_1 . The 1-D homography H_L maps l into a line l_2 in the second pencil. We determine l_2 taking into account that

$$l_2 \sim L_2(\lambda'_2)$$

with

$$\begin{pmatrix} \lambda'_2 \\ 1 \end{pmatrix} \sim H_L \begin{pmatrix} \lambda_1 \\ 1 \end{pmatrix}.$$

In a similar manner, the inverse homography H_L^{-1} transforms l , as an element of pencil L_2 , into a line l_1 belonging to the first pencil. Let us repeat the aforementioned computation to determine l_1 .

- 3) Let us compute D as the point of intersection of l_1 and l_2 , respectively. It can be proved that l_1 and l_2 are the tangents to the curve in the points, where l intersects Ω [23]. Thus, D and l are always pole-polar with respect to the conic Ω satisfying

$$l \sim \Omega D.$$

- 4) Let us determine an additional point \mathbf{A} on Ω by intersecting a random pair of corresponding lines \mathbf{m}_1 and \mathbf{m}_2 , respectively. For this, we select an arbitrary value $\eta_1 \neq \lambda_1$, obtain $\mathbf{m}_1 \sim \mathbf{L}_1(\eta_1)$, and compute \mathbf{m}_2 in the second pencil by considering

$$\mathbf{m}_2 \sim \mathbf{L}_2(\eta_2)$$

such that

$$\begin{pmatrix} \eta_2 \\ 1 \end{pmatrix} \sim H_L \begin{pmatrix} \eta_1 \\ 1 \end{pmatrix}.$$

- 5) The 5 DOF of the conic curve Ω are fully constrained by the four points \mathbf{C}_1 , \mathbf{C}_2 , \mathbf{A} , and \mathbf{D} , respectively. Note that \mathbf{D} defines two independent constraints in the conic parameters because of the pole–polar relation with \mathbf{l} . Let the four points define a canonical projective basis in the plane [23]. It can be proved that in this case, the curve parametrization is always

$$\Omega' \sim \begin{pmatrix} 0 & 0 & -0.5 \\ 0 & 1 & 0 \\ -0.5 & 0 & 0 \end{pmatrix}.$$

- 6) The Euclidean parametrization of the conic curve can be determined by applying a change of coordinates \mathbf{S} that maps the basis points back to their Euclidean coordinates. It follows that

$$\Omega \sim \mathbf{S}^{-T} \Omega' \mathbf{S}^{-1}$$

with \mathbf{S} being a 3×3 matrix given by

$$\mathbf{S} \sim (\mathbf{C}_1 \ \mathbf{D} \ \mathbf{C}_2) \text{diag}((\mathbf{C}_1 \ \mathbf{D} \ \mathbf{C}_2)^{-1} \mathbf{A}).$$

Following the aforementioned steps, we derive Ω as a function of the two camera centers and the homographic relation H_L between pencils. After some tedious algebraic manipulations, the vectorized form ω of the conic curve can be written as follows:

$$\omega \sim \mathbf{N} \text{vec}(H_L)$$

with \mathbf{N} depending on the nonhomogeneous coordinates of \mathbf{C}_1 and \mathbf{C}_2 , respectively

$$\mathbf{N} \sim \begin{pmatrix} 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & -1 \\ 0 & 1 & 0 & 0 \\ -C_{1,y} & 0 & C_{1,x} + C_{2,x} & C_{2,y} \\ -C_{2,x} & -C_{1,y} - C_{2,y} & 0 & C_{1,x} \\ C_{1,y} C_{2,x} & C_{1,y} C_{2,y} & -C_{1,x} C_{2,x} & -C_{1,x} C_{2,y} \end{pmatrix}.$$

Replacing $\text{vec}(H_L)$ by the result of (5), we finally, obtain ω in terms of the original configuration homography \mathbf{H} defined in the image plane

$$\omega \sim \mathbf{NMF} \text{vec}(\mathbf{H}). \quad (6)$$

Equation (6) is nicely factorized in matrix \mathbf{N} that encodes the position of the centers (or alternatively the translational component of camera motion), matrix \mathbf{M} that depends on the

cameras' rotations, and matrix \mathbf{F} , which is a function of the optical parameters.

C. Discussion

This section further analyzes (6) in order to gain insight about the homographic curve and its dependencies. The product of matrices \mathbf{N} and \mathbf{M} is a 6×4 matrix, where each column μ_i can be interpreted as the vectorized representation of a conic

$$\mathbf{NM} \sim (\mu_1 \ \mu_2 \ \mu_3 \ \mu_4).$$

Let us consider the focal lengths in \mathbf{F} and the scalar entries of \mathbf{H} (4). It follows from (6) that

$$\omega \sim d f_2 \mu_1 - b \mu_2 - c f_1 f_2 \mu_3 + a f_1 \mu_4. \quad (7)$$

Equation (7) denotes a linear system of conics with basis μ_i , where $i = 1, \dots, 4$ [23]. It shows that in general, a homographic curve ω belongs to a 4-D subspace in the space of all conics. This subspace is fully defined by the kinematic configuration of the stereo setup because conics μ_i , in the basis, depend only on the rotation and translation of the cameras. The coordinates of ω in the linear system of conics are a function of the intrinsic parameters and the desired configuration homography \mathbf{H} , respectively.

Let \mathbf{V} be the fixation point and \mathbf{T} be the point in the plane that is projected at infinity in both views [see Fig. 3(a)]. If \mathbf{C}_1 and \mathbf{C}_2 are fixed points, then the coordinates of \mathbf{V} and \mathbf{T} depend only on the rotation angles θ_1 and θ_2 (8). It is curious to verify that conics μ_i are rank 2 degenerate conics corresponding to pairs of lines in the plane. Moreover, and as shown in Fig.3(b)–(e), these lines can be found by knowing the locations of points \mathbf{C}_1 , \mathbf{C}_2 , \mathbf{V} , and \mathbf{T} , respectively

$\mathbf{V} =$

$$\begin{pmatrix} -C_{2,x} \tan \theta_1 + C_{1,x} \tan \theta_2 + (C_{1,y} - C_{2,y}) \tan \theta_1 \tan \theta_2 \\ -C_{1,x} + C_{2,x} - C_{1,y} \tan \theta_1 + C_{2,y} \tan \theta_2 \\ \tan \theta_2 - \tan \theta_1 \end{pmatrix}$$

$\mathbf{T} =$

$$\begin{pmatrix} C_{1,y} - C_{2,y} - C_{1,x} \tan \theta_1 + C_{2,x} \tan \theta_2 \\ -C_{2,y} \tan \theta_1 + C_{1,y} \tan \theta_2 + (-C_{1,x} + C_{2,x}) \tan \theta_1 \tan \theta_2 \\ \tan \theta_2 - \tan \theta_1 \end{pmatrix}. \quad (8)$$

Fig. 4(a)–(c) shows the effect that the choice of the configuration homography \mathbf{H} has in shaping the conic ω . It considers a particular kinematic configuration for the cameras, such that \mathbf{C}_1 , \mathbf{C}_2 , \mathbf{V} , and \mathbf{T} are fixed points that implicitly define a subspace in the space of all conics. Parameters f_1 and f_2 , respectively, are assumed to be equal and constant.

Fig. 4(a) concerns the case of \mathbf{H} being a diagonal matrix. The different conics ω are generated by varying the ratio d/a . In this case, the configuration homography specifies a scaled mapping between images, and the linear system of conics becomes a conic pencil [23]. Since the pencil is defined by μ_1 and μ_4 that intersect at points \mathbf{C}_1 , \mathbf{C}_2 , \mathbf{V} , and \mathbf{T} , respectively, then the homographic curve ω always goes through these points.

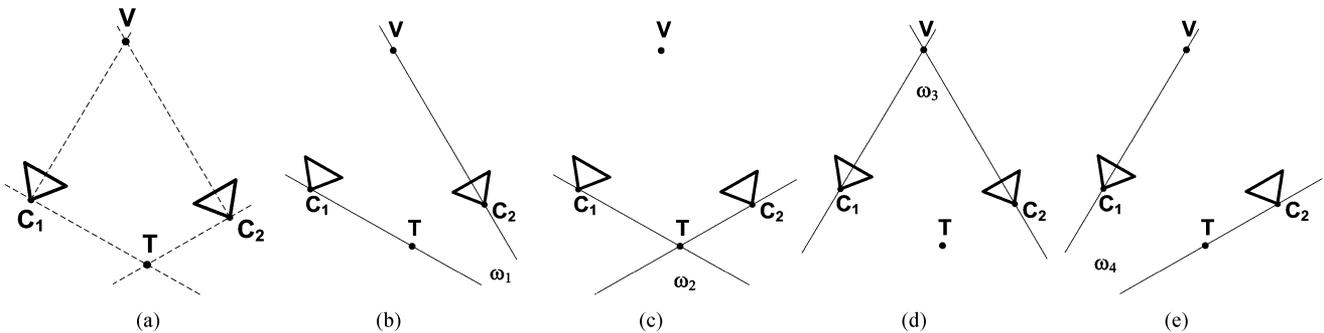


Fig. 3. Geometric interpretation of (7). V is the fixation point defined by the intersection of the optical axes. T denotes the point at infinity and can be understood as the intersection of the lines going through the camera centers which are parallel to the “line images.” μ_i , $i = 1, \dots, 4$ are rank 2 degenerate conics and form the basis of the linear system of conics described by (7). As shown in (b)–(e), the pair of lines composing each degenerate conic goes through points C_1 , C_2 , V , and T , respectively. These points encode the kinematic configuration of the active stereo system and implicitly define a 4-D linear subspace in the space of all conics, containing every possible homographic curve ω . (a) C_1 , C_2 , V , and T . (b) μ_1 . (c) μ_2 . (d) μ_3 . (e) μ_4 .

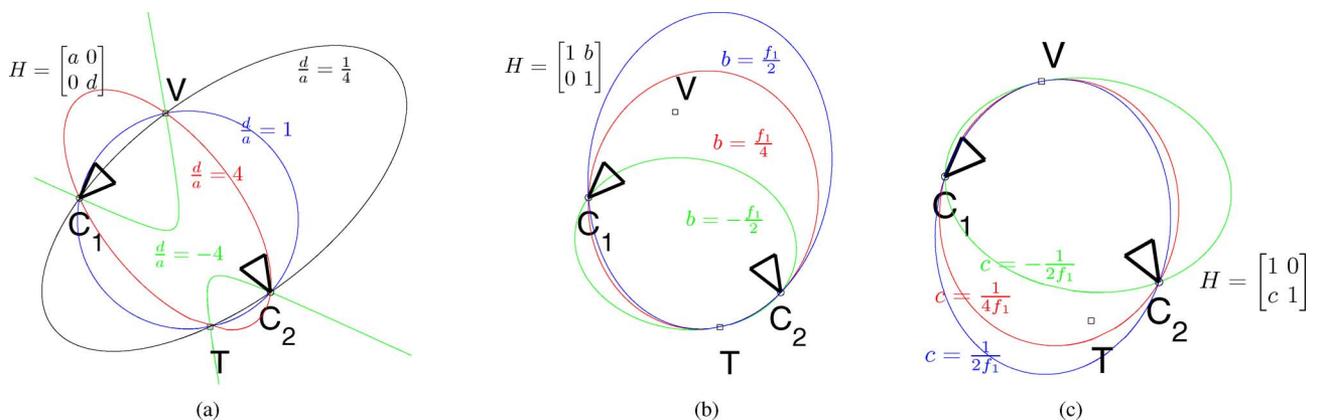


Fig. 4. Dependence between the homographic curve and the selected homography H . The cameras are static, and the focal lengths are constant and equal for both views (the linear subsystem of (7) is fixed). The entries of matrix H (4) are changed in order to observe different shapes for the conic ω . (a) Scale mapping. (b) Constant disparity mapping. (c) Projective mapping.

In Fig. 4(b), the configuration homography specifies a disparity of b pixels between the stereo pair of images. H is an Euclidean transformation that maps the point at infinity, in the first view, into the point at infinity in the second view. This explains the fact that T is always in the conic ω . On the other hand, and since H specifies a shift between images, the image centers are not mapped one into the other. This is in accordance with the observation that ω does not go through the fixation point V . The homographic curves in Fig. 4(b) are basically the *isodisparity curves* discussed by Pollefeys *et al.* in the context of stereo reconstruction [22].

Finally, Fig. 4(c) shows the homographic curves for the case of H being a nonaffine projective transformation. Since b is zero, the linear system of (7) becomes a conic net [23], with point V being common to every member (the image centers are always mapped one into the other).

V. ACTIVE TRACKING WITH TWO PAN-TILT-ZOOM CAMERAS

We derived the homographic curve and study its dependence with respect to the specified configuration homography H , camera-intrinsic parameters, and kinematic configuration of the stereo setup. This section shows how to apply the established framework to solve practical problems of active tracking. This

is illustrated by considering a pair of PTZ cameras with pure rotation motion and zoom control. The two units are arbitrarily placed in the working space at locations C_1 and C_2 , respectively. Therefore, the DOFs of the active stereo system are the pan angles of each camera θ_1 and θ_2 and the ratio ρ between the focal lengths that can be manipulated using the zoom control

$$\rho = \frac{f_1}{f_2}.$$

Henceforth, we will assume that the desired configuration homography H is the identity I . In this case, the active-tracking behavior assures that the N targets are imaged at the same location in both views, which might be a useful feature for many real application scenarios. Note that assuming $H \sim I$ does not imply a loss of generality. The framework can be similarly employed for different choices of the configuration homographies, motivated by the need to meet the particular requirements of a certain tracking problem.

Our objective in this paper is to track a set of N free-moving targets in such a manner that they are imaged at the same position in both views. Since $H \sim I$, it follows from (7) that the curve ω is always a member of the conic pencil

$$\omega \sim \mu_1 + \rho \mu_4. \quad (9)$$

For this particular case, the homographic curve ω is the horopter of the stereo setup [8]. The curve contains points \mathbf{V} and \mathbf{T} , which depend on the rotation angles of the cameras, as well as the fixed-projection centers \mathbf{C}_1 and \mathbf{C}_2 [see Fig. 3(b) and (e)]. Since ω is a function of ρ , θ_1 , and θ_2 , respectively, the problem can be stated as controlling the system's DOF such that the homographic curve goes through the locations of the N targets.

The remaining part of this section discusses the tracking for an increasing number N of free-moving targets. With the help of our geometric framework, we will be able to prove that under the described conditions, it is possible to track up to $N = 3$ objects.

A. Tracking for the Case of $N = 1$

Let the target have coordinates \mathbf{Q} at a certain time instant. From (9) and Fig. 3(b) and (d), respectively, it follows that the problem is feasible if there is an homographic curve ω such that

$$\underbrace{(\hat{\mathbf{Q}} \quad \hat{\mathbf{C}}_1 \quad \hat{\mathbf{C}}_2 \quad \hat{\mathbf{V}} \quad \hat{\mathbf{T}})^T}_{\mathbf{A}} \omega = 0$$

where $\hat{\cdot}$ is the lifted point coordinates (cf., Section II).

Here, \mathbf{A} is a 5×6 matrix that is a function of θ_1 and θ_2 [this dependency is because of points \mathbf{V} and \mathbf{T} , whose coordinates are provided in (8)]. In general, \mathbf{A} has an unidimensional null space $\mathcal{N}(\mathbf{A})$, which can be interpreted as a 6×1 vector representing a conic curve. This curve belongs to the conic pencil of (9), because it goes through the four intersections of μ_1 and μ_4 , respectively. By replacing ω by $\mathcal{N}(\mathbf{A})$ and solving with respect to ρ , it yields the result, shown at the bottom of the page.

The aforementioned equation is written in terms of the non-homogeneous coordinates of \mathbf{C}_1 , \mathbf{C}_2 , and \mathbf{Q} . Any triplet of values $(\rho, \theta_1, \theta_2)$ satisfying it is an admissible solution for the active-tracking problem. It should be noted that the tracking is still feasible for the situation of cameras with no zoom control. In this case, the ratio ρ is a constant, and the equation expresses a constraint over the rotation angles θ_1 and θ_2 , respectively.

Former works in active fixation have already proved the feasibility of tracking a single target whenever the focal length is equal in both cameras [5]–[7], [24]. For $\rho = 1$, the homographic curve ω of (9) becomes a circle (the so-called *Vieth–Müller circle* [8]), and any pair (θ_1, θ_2) that places \mathbf{V} in the circle defined by \mathbf{C}_1 , \mathbf{C}_2 , and \mathbf{Q} assures that the target projection is the same in both views. The constraint derived earlier is a generalization of this previous result for the case of $\rho \neq 1$.

B. Tracking for the Case of $N = 2$

Let \mathbf{Q}_a and \mathbf{Q}_b be two free-moving targets. Repeating the reasoning of the previous section, the tracking for the case of $N = 2$ is feasible if there is a nontrivial solution for the follow-

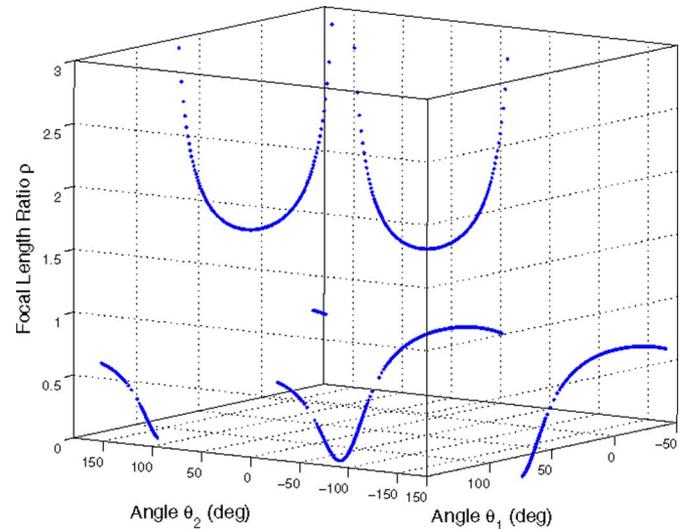


Fig. 5. Solutions for the case of $N = 2$. The set of feasible solutions defines a 3-D curve in the space of the control variables $(\rho, \theta_1, \theta_2)$. This curve depends on points \mathbf{Q}_a , \mathbf{Q}_b , \mathbf{C}_1 , and \mathbf{C}_2 , respectively.

ing equation:

$$\underbrace{(\hat{\mathbf{Q}}_a \quad \hat{\mathbf{Q}}_b \quad \hat{\mathbf{C}}_1 \quad \hat{\mathbf{C}}_2 \quad \hat{\mathbf{V}} \quad \hat{\mathbf{T}})^T}_{\mathbf{B}} \omega = 0.$$

Unfortunately, \mathbf{B} is, in general, a nonsingular 6×6 matrix. However, and since \mathbf{B} is a function of the camera's rotation angles, the values of θ_1 and θ_2 can be chosen such that the matrix becomes rank deficient. In this case, the equation admits a nontrivial solution $\omega \sim \mathcal{N}(\mathbf{B})$, with $\mathcal{N}(\mathbf{B})$ denoting the 1-D null space of \mathbf{B} . The solution $\omega \sim \mathcal{N}(\mathbf{B})$ must satisfy the equality of (9), which leads to an additional constraint involving the ratio of the focal length ρ

$$\begin{cases} \det(\mathbf{B}) = 0 \\ \mathcal{N}(\mathbf{B}) \sim \mu_1 + \rho \mu_4. \end{cases} \quad (10)$$

1) *Space of Solutions:* Any solution $(\rho, \theta_1, \theta_2)$ of the aforementioned system of equations is a feasible solution for the tracking problem. It assures that the homographic curve goes through \mathbf{Q}_a and \mathbf{Q}_b and that the targets are projected in the same location in both images. Each one of the aforementioned equations is a constraint on the control variables, defining a surface in the space of parameters $(\rho, \theta_1, \theta_2)$. The feasible solutions are points lying in the locus of intersection of these two surfaces. Fig. 5 plots an example of this locus for particular values of \mathbf{Q}_a , \mathbf{Q}_b , \mathbf{C}_1 , and \mathbf{C}_2 , respectively.

Thus, this particular tracking problem has an infinite number of solutions for the control parameters. In general, the topology of the solution space is difficult to characterize because of the dependence on the position of the free-moving targets and camera centers. However, and according to our simulations, the feasible solutions seem to define a highly nonlinear 3-D curve

$$\rho = \frac{(C_{2,x} - Q_x + (C_{2,y} - Q_y) \tan \theta_2)(Q_y - C_{1,y} + (C_{1,x} - Q_x) \tan \theta_1)}{(Q_y - C_{2,y} + (C_{2,x} - Q_x) \tan \theta_2)(C_{1,x} - Q_x + (C_{1,y} - Q_y) \tan \theta_1)}.$$

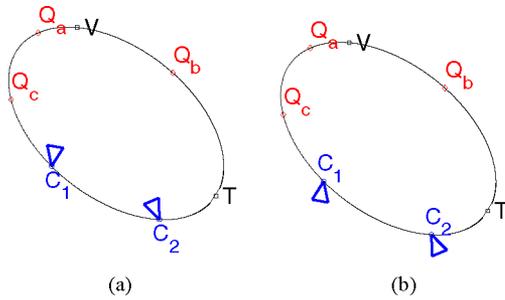


Fig. 6. Redundant solutions while tracking $N = 3$ targets using two PTZ cameras. The system of (11) has eight distinct solutions that can be grouped in two sets of four elements with the same focal length ratio ρ . Parts (a) and (b) correspond to two solutions in the same group with the rotation angles differing by 180° . These solutions are redundant in the sense that they do not change the positions, where targets are imaged. (a) $\rho = 0.6$. (b) $\rho = 0.6$.

in the space of the parameters. As shown in Fig. 5, there is typically a range of ρ values for which the curve is not defined. This means that the ratio of focal lengths cannot arbitrarily fixated; otherwise, it might not exist a feasible solution for the tracking problem. Thus, and unlike to what happens for $N = 1$, the active zoom control is mandatory to accomplish the tracking of $N = 2$ targets, which undergo free motion. This space of solutions is further discussed in Section V-E, which reports some real experiments in the simultaneous tracking of two targets.

C. Tracking for the Case of $N = 3$

Repeating the previous approach, it follows that the solutions $(\rho, \theta_1, \theta_2)$ for the simultaneous tracking of $Q_a, Q_b,$ and Q_c must satisfy

$$\begin{cases} \det(C_{1\dots 6}) = 0 \\ \det(C_{1\dots 5,7}) = 0 \\ \mathcal{N}(C) \sim \mu_1 + \rho \mu_4 \end{cases} \quad (11)$$

where the numbers in subscript denote lines in the 7×6 matrix C

$$C \sim (\hat{Q}_a \quad \hat{Q}_b \quad \hat{Q}_c \quad \hat{C}_1 \quad \hat{C}_2 \quad \hat{V} \quad \hat{T})^T.$$

1) *Space of Solutions:* In general, the system of (11) has eight distinct solution triplets $(\rho, \theta_1, \theta_2)$. These solutions can be clustered in two groups of four with all the elements having the same focal length ratio ρ . The solutions within each group differ only by angles of 180° in the parameters θ_1 and/or θ_2 , which means that the targets are projected in the same image position, regardless of the selected triplet [see Fig. 6(a) and (b)]. Thus, each group has a unique solution with practical significance, i.e., the one for which both cameras are forward looking the targets. The two significant solutions arising from the two groups have inverse values for the focal length ratio ρ and a difference of 90° in the cameras' orientations [see Fig. 7(a) and (b)]. Since, for a particular configuration of the free-moving targets, there are only two effective feasible solutions, it is easy to conclude that the tracking for $N = 3$ can only be accomplished by simultaneously controlling the orientation of both cameras and the focal length ratio as well.

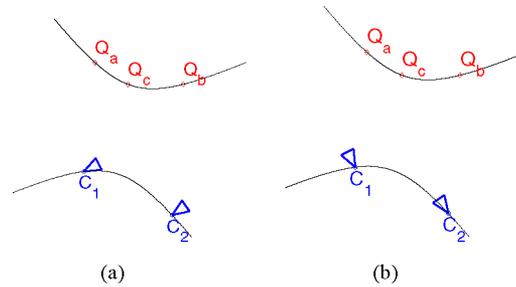


Fig. 7. Tracking $N = 3$ targets using two PTZ cameras. Two distinct solutions with practical significance are shown. The cameras' orientation differ by an angle of 90° , and there is an inversion in the focal length ratios. (a) $\rho = 5$. (b) $\rho = 0.2$.

2) *Simulation Results:* Fig. 8(a)–(c) presents simulation results in simultaneously tracking $N = 3$ free-moving targets using two PTZs.¹ The simulation workspace is depicted in Fig. 8(a), where the two cameras undergo independent pan motion. The left camera has a constant FOV of 45° , and the right camera has variable zoom to enable the control of the focal length ratio ρ . The targets $Q_a, Q_b,$ and Q_c move in front of the cameras according to the arbitrarily defined trajectories. The tracking objective is to keep the targets at the same image positions in both views, which is achieved by solving (11). As discussed earlier, for each time instant, there are two distinct solutions for the control variables. We select the one that is closer in an Euclidean sense to the current system configuration.

Fig. 8(b) plots the targets' image positions along time. The three targets are projected in the same locations in the two views, which proves the feasibility of the tracking task and the correctness of the derived control equations. The active system configuration is shown in Fig. 8(c). The system's DOF vary smoothly along time, which suggests a continuous space of solutions. This is an important requirement for the practical feasibility of the tracking task. However, and since the first- and second-order derivatives of the control parameters tend to take high values, the tracking must be carried by PTZ units with good dynamic performance. The simulation does not take into account the physical limitations such as the camera's FOV or the variation range of the tunable parameters. From Fig. 8(b), it follows that the assumed FOV of 45° is clearly insufficient for assuring the success of the tracking in an hypothetical real experiment. Although the objects move smoothly in front of the cameras, the specified active system would quickly lose track because it would not be able to keep the three targets simultaneously visible. This fact is not changed by using the alternative solution space arising from the second group of solutions of (11).

The aforementioned observations were confirmed in repeated simulations, assuming a broad range of target trajectories. It seems safe to conclude that the tracking of $N = 3$ targets is feasible in practice, but it requires the use of PTZ units with wide FOV lenses (e.g., fish-eye lenses) and mechanical DOF with high-performance-dynamic response. Unfortunately, and to the best of our knowledge, there are no commercial systems

¹A video of the simulation is provided as supplementary material (Clip 2).

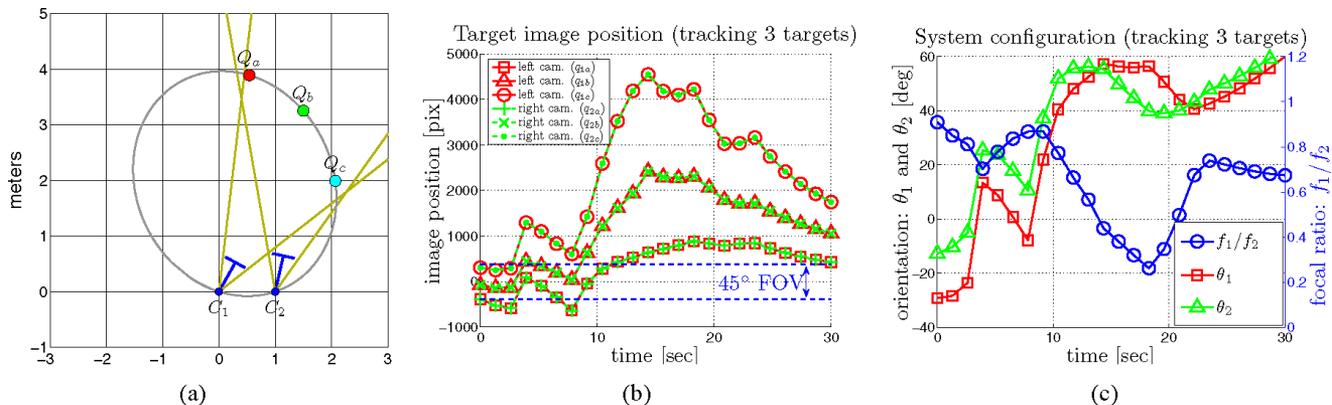


Fig. 8. Simulation results in tracking $N = 3$ targets using two PTZ cameras (the video of the simulation is available as supplementary material). (a) Instant of the simulation. The targets move along arbitrarily defined trajectories. Both cameras undergo independent pan motion, and the right camera has zoom control. The yellow lines that intersect at each camera represent the FOV that is constant for the left camera and variable for the right camera because of the zoom DOF. The gray closed curve denotes the homographic conic. (b) and (c) Plots of targets' image positions and the evolution of the system configuration along time are shown, respectively. The tracking objective of keeping the targets at the same position in the stereo views is met. The system parameters vary smoothly, which suggests that the space of solutions is continuous. The tracking is feasible in practice for cameras with an FOV $\geq 145^\circ$.

with such characteristics. A real implementation of the described tracking behavior would require the development of customized PTZ units.

D. Tracking for the Case of $N > 3$

Following the same line of reasoning used to study the tracking feasibility for $N = 1, 2$, and 3 , respectively, it is easy to verify that for $N > 3$, the constraints outnumber the DOF of the active system. This means that in general there is no solution for the problem. Such conclusion is not surprising because the homographic curve is a conic defined by a maximum of five points. Thus, and taking into account that ω must also go through the two projection centers, the tracking for $N > 3$ is in general not feasible.

E. Experiments in Tracking $N = 2$ Targets

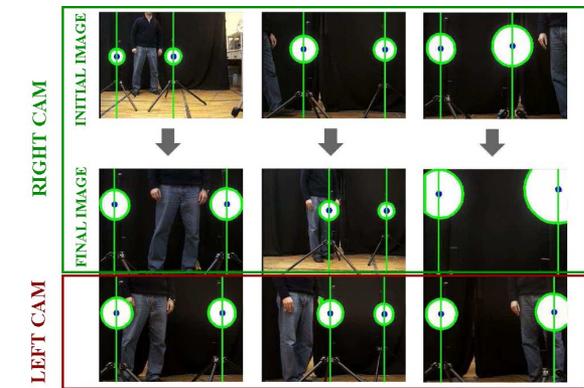
This section reports real tracking results using a stereo pair of PTZ camera units.² The experimental setup consists of two calibrated Sony EVI-D31 PTZ charge-coupled device color cameras. Being primarily intended for applications, e.g., surveillance or teleconferencing, these commercial units do not feature the high communication rate required for the real-time simultaneous control of its DOF, thus, hindering a smooth-tracking implementation. The tracking is, thus, achieved in a discontinuous three-step process: First, the targets are detected in the stereo images, and their position in the working space is estimated by triangulation; second, the solution for the control parameters is determined by solving the appropriate system of equations; and third, an alignment saccade is performed by sending the commands to the PTZ units through a serial channel. The baseline between the cameras is approximately 1.5 m, and we use white markers as targets that are easily distinguished from the background, thus keeping image detection simple and robust.

²The experiment video is available as supplementary material (Clip 1).

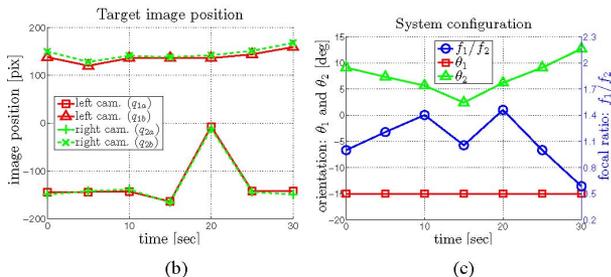
The tracking of a single object has been implemented in the past and, therefore, will not be discussed here. In Section V-C2, we verified through simulation that the tracking for $N = 3$ targets can only be accomplished by PTZ units with specific features, namely, wide FOV lenses and high-performance-dynamic behavior. Unfortunately, our Sony cameras do not match these requirements. The FOV ranges from 5° to 47.5° , respectively, and the mechanical performance is relatively limited. Thus, and despite our efforts, it was impossible to obtain a successful real implementation because of hardware limitations. The remaining part of this section reports two experiments in aligning the PTZs with respect to $N = 2$ targets, such that they are projected in the same locations in both views.

1) *Experiment 1*: Since the problem of tracking $N = 2$ targets admits multiple possible solutions (see Section V-B), we need additional constraints to select a unique active system configuration at each time instant. In this experiment, we keep the left PTZ stationary, while the targets undergo arbitrary motion within the camera FOV. The DOF to be controlled are the rotation angle θ_2 and the focal length ratio ρ , which depends on the zoom of the right camera. After replacing θ_1 by a constant k in (10), we obtain four distinct solutions for (θ_2, ρ) . The first equation becomes a one variable equation that admits four solutions for θ_2 . The second equation on ρ is always possible. Since the conic $\omega = \mathcal{N}(B)$ goes through points C_1, C_2, V , and T , then it is always an element of the pencil defined by μ_1 and μ_4 (see Fig. 3 and [23], respectively). This is consistent with the result of Fig. 5, where a generic vertical plane $\theta_1 = k$ cuts the curve of the solution space in four distinct points. Two of these solutions are discarded because they orient the camera toward a backward-looking direction. For each tracking instant, we choose the effective solution that is closer to the current system configuration in order to assure smooth camera motion. Fig. 9(a)–(c) shows the experimental results.

2) *Experiment 2*: In this second experiment, the two targets are kept static, while the left camera sweeps a predefined



(a)



(b)

(c)

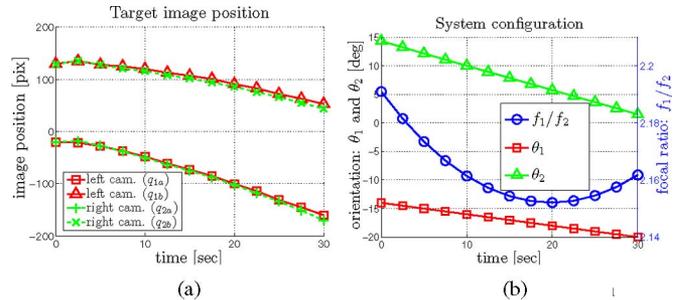
Fig. 9. Experiment 1 in tracking two targets with two PTZ cameras. The targets were moved a total of seven times. The left camera is kept stationary, while the right camera performs a saccadic motion to align the targets' image position with the ones observed in the left view. (a) Right images before and after the alignment saccade. (b) and (c) Plots of the image positions and the system configuration along time. The tracking objective is fully accomplished with the targets being projected at the same locations in both retinas. (a) Tracking instants corresponding to three of the seven motions. (b) Targets' positions in both views. (c) System configuration.

set of pan and zoom positions. The objective is to control the configuration of the right camera such that the two targets are imaged in the same positions as in the left view. The controlled DOF are θ_2 and ρ , and the mathematical formulation is equivalent to experiment 1. The differences are in the application, where the left camera acts as a “master” following an arbitrary trajectory in pan and zoom, and the right camera acts as a “slave” that is automatically controlled to maintain the targets aligned. Fig. 10(a) and (b) plots results of this experiment that can be watched in the video provided as supplementary material (Clip 1).

VI. TRACKING WITH AN ACTIVE STEREO HEAD

This section discusses the tracking of $N > 1$ targets using the POPEYE active stereo head shown in Fig. 11(a). We start by using the theoretical framework for studying the feasibility of the tracking task. It is shown that the tracking for $N = 3$ cannot be accomplished because of the lack of zoom control. The tracking for the case of $N = 2$ is implemented, and results of real experiments are presented.

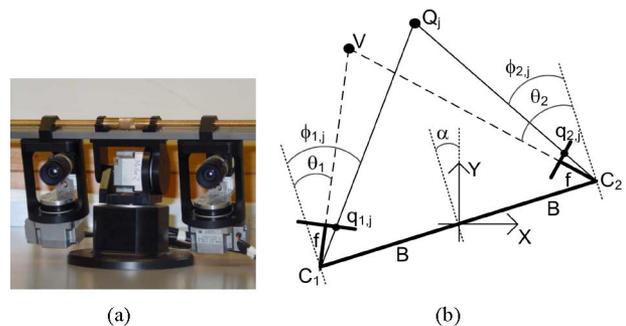
While the commercial PTZ cameras have hardware limitations that prevent achieving a smooth active-tracking behavior, the POPEYE head is a custom-made system that enables a high-performance-visual control of motion [10]. The implementation of a suitable closed-control loop requires mapping image infor-



(a)

(b)

Fig. 10. Experiment 2 in tracking two targets with PTZ cameras. The left camera sweeps a predefined set of positions in pan and zoom, while the right camera motion is controlled such that the targets' image are the same as for the left view. (a) and (b) Plots of the target's image position and the system configuration after each alignment saccade, respectively, which show that the tracking objective can be successfully accomplished. This master-slave configuration can be helpful in surveillance scenarios, where an operator controls the master camera for screening certain regions of the working space, and the automatic control of the slave camera assures a secondary view of those regions. (a) Targets' image position in both views. (b) System configuration.



(a)

(b)

Fig. 11. POPEYE head has a total of 4 DOF and follows the Helmholtz configuration, with both eyes having a common tilt rotation (neck tilt). In our experiments, the common neck tilt is kept constant such that the cameras' optical axes are always coplanar. In (b), a scheme of the assumed kinematic configuration is shown. The world reference frame is placed at the system's rotation center. The angles α , θ_1 , and θ_2 represent the neck, left camera, and right camera pan rotations, as seen by the system's encoders. Target Q_j is projected into points $q_{1,j}$ and $q_{2,j}$ in left and right images, respectively. $\phi_{1,j}$ and $\phi_{2,j}$ denote the angles between the line of sight and the direction orthogonal to the rotating baseline. V is the fixation point, and the length of the baseline is $2B$. (a) POPEYE head. (b) Kinematic configuration.

mation into position and velocity commands for the system actuators. A significant part of this section is devoted to applying our mathematical formulation to derive these control equations.

A. POPEYE Active Stereo Head

Fig. 11(a) shows the POPEYE active stereo head used in our experiments. The robotic platform has four rotational DOFs—neck pan, neck tilt, and individual eye pan—and follows the Helmholtz configuration [25]. The actuators are equipped with optical encoders in the back-shaft that enable accurate knowledge of the position of the DOFs in real time. The system has two similar firewire color cameras, with no zoom and equal focal length. In our experiments, we will control the pan rotations of the neck, left camera, and right camera (the neck tilt is not used). The system is initially aligned so that all three pan rotation axes are parallel relative to each other and orthogonal to the plane that

contains the optical axes of the two cameras. In the follow-up of the discussed geometric framework, the cameras are modeled as line scan cameras. In terms of implementation, the acquired 2-D images are processed to compute targets motion, but only the horizontal components of position and velocity are used for the subsequent control steps.

1) *Kinematic Configuration*: As discussed in [3], it is highly convenient to model the head kinematics using a system of coordinates aligned with the controllable DOF and feedback variables. Fig. 11(b) shows a scheme of the system's geometry and the parameters considered for describing it. The world reference frame is placed in the center of the platform with the Z -axis being aligned with the neck pan rotation axis. The neck rotation α controls the position of the camera centers, which can be placed in antipodal points of a circle with diameter $2B$ (the baseline). Their coordinates in the world reference frame are given by

$$\mathbf{C}_{1/2} \sim \begin{pmatrix} \mp B \cos(\alpha) \\ \mp B \sin(\alpha) \\ 1 \end{pmatrix}. \quad (12)$$

The pan rotation angles for each camera are θ_1 and θ_2 , which are measured with respect to the line orthogonal to the baseline. All the angles are assumed to be counterclockwise, and for $\alpha = 0$, $\theta_1 = 0$, $\theta_2 = 0$, the baseline is aligned with the X -axis, and both optical axes are parallel to the Y -axis (initial frontal-parallel configuration).

We would like to emphasize that there are two important differences between the current kinematic description and the one that is considered in Section V. First, the cameras' locations \mathbf{C}_1 and \mathbf{C}_2 are no longer independent from each other but are simultaneously controlled by α . Second, the pan rotation angles are θ_1 and θ_2 are no longer measured with respect to the world reference frame but with respect to a line orthogonal to the moving platform.

2) *Projection*: Henceforth, and to improve clarity, we will use the subscript i to indicate the camera number ($i = 1, 2$) and the subscript j to refer to the target ($j = a, b$). Let \mathbf{Q}_j be the homogeneous coordinates of a target j in the world reference frame. The target is projected into $\mathbf{q}_{1,j}$ in the first camera and into $\mathbf{q}_{2,j}$ in the second camera. From the result of (1), concerning the projection into line scan cameras, it follows that

$$\mathbf{q}_{i,j} \sim \mathbf{K} \mathbf{R}_i (1 - \mathbf{C}'_i) \mathbf{Q}_j$$

where \mathbf{K}_i denotes the matrix of intrinsic parameters, \mathbf{C}'_i is the nonhomogeneous representation of the camera center (12), and \mathbf{R}_i is the camera rotation with respect to the world reference frame

$$\mathbf{R}_i = \begin{pmatrix} \cos(\theta_i + \alpha) & \sin(\theta_i + \alpha) \\ -\sin(\theta_i + \alpha) & \cos(\theta_i + \alpha) \end{pmatrix}.$$

3) *Stereo Reconstruction of \mathbf{Q}_j* : The target position \mathbf{Q}_j can be recovered from the pair of stereo images ($\mathbf{q}_{1,j}$, $\mathbf{q}_{2,j}$) and from the pose of the robot head (α , θ_1 , θ_2).

Let the homogeneous coordinates of the image point be

$$\mathbf{q}_{i,j} \sim \begin{pmatrix} q_{i,j} \\ 1 \end{pmatrix}$$

and let $\phi_{i,j}$ be the angle between the line of back projection and the normal to the baseline [see Fig. 11(b)]. It follows that

$$\phi_{i,j} = \theta_i + \arctan\left(\frac{q_{i,j}}{f}\right) \quad (13)$$

with f denoting the camera focal length.

From (2), we know that the back-projection line for image point $\mathbf{q}_{i,j}$ is given by

$$\mathbf{L}_{i,j} \sim (1 - \mathbf{C}'_i)^T \mathbf{R}_i^T \mathbf{K}^T \mathbf{U}^T \mathbf{q}_{i,j}.$$

Considering the back-projection lines from the two cameras, and using triangulation to recover the target position, it follows that

$$\mathbf{Q}_j \sim \mathbf{L}_{1,j} \times \mathbf{L}_{2,j}.$$

After some algebraic manipulations, and taking into account the result of (13), the target position can be rewritten as

$$\mathbf{Q}_j \sim \begin{pmatrix} Q_{j,x} \\ Q_{j,y} \\ 1 \end{pmatrix} = \begin{pmatrix} B \frac{\cos \alpha (\tan \phi_{1,j} + \tan \phi_{2,j}) + 2 \sin \alpha}{\tan \phi_{1,j} - \tan \phi_{2,j}} \\ B \frac{\sin \alpha (\tan \phi_{1,j} + \tan \phi_{2,j}) - 2 \cos \alpha}{\tan \phi_{1,j} - \tan \phi_{2,j}} \\ 1 \end{pmatrix}. \quad (14)$$

Equation (14) provides the target world coordinates \mathbf{Q}_j directly as a function of the encoder readings— α , θ_1 , θ_2 —and the stereo image information— $q_{1,j}$, $q_{2,j}$. The fixation point \mathbf{V} can be computed in a similar manner by simply replacing $q_{1,j}$ and $q_{2,j}$ by zero

$$\mathbf{V} \sim \begin{pmatrix} B \frac{\cos \alpha (\tan \theta_1 + \tan \theta_2) + 2 \sin \alpha}{\tan \theta_1 - \tan \theta_2} \\ B \frac{\sin \alpha (\tan \theta_1 + \tan \theta_2) - 2 \cos \alpha}{\tan \theta_1 - \tan \theta_2} \\ 1 \end{pmatrix}. \quad (15)$$

B. Feasibility Analysis

We aim at tracking N targets with the POPEYE stereo head, assuming a configuration homography $\mathbf{H} \sim \mathbf{I}$. This section presents the feasibility analysis for the problem at hand. The analysis is similar in spirit to the one carried in Section V for the case of two PTZ cameras.

Since the focal lengths of the cameras are equal, then $\rho = 1$, and thus, (7) becomes

$$\omega \sim \boldsymbol{\mu}_1 + \boldsymbol{\mu}_4.$$

The horopter of two cameras with the same intrinsic parameters is the well-known Vieth–Müller circle [8]. Thus, the aforementioned conic ω is a circle, containing points \mathbf{C}_1 , \mathbf{C}_2 , \mathbf{V} , and \mathbf{T} , as well as the circular points \mathbf{I} and \mathbf{J} [8], [23]. While in Section V, the camera centers are fixed points, the points \mathbf{C}_1

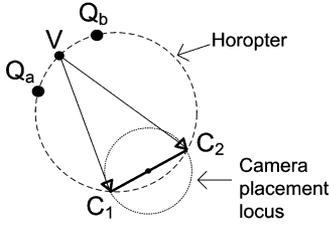


Fig. 12. Simultaneous tracking of $N = 2$ targets. The platform rotation places C_1 and C_2 at antipodal positions on the circle with diameter $2B$. The horopter (the curve ω for $H \sim I$) is the Vieth–Müller circle that goes through the targets and the camera centers.

and C_2 now depend on the rotation angle α . This means that μ_1 and μ_4 are a function of the DOF of the system α , θ_1 , and θ_2 , respectively. The fact that ω is a circle assures that point V is aligned with the curve *iff* point T is also aligned. Henceforth, we will ignore T because it adds no information to the problem.

1) *Tracking for the Case of $N = 1$:* The tracking of a single target Q is trivial. Let α take a particular value such that C_1 and C_2 are antipodal points on the circle of diameter $2B$. The three points Q , C_1 , and C_2 define a circle in a unique manner. Any choice of angles (θ_1, θ_2) that aligns the fixation point V with this circle is a feasible solution for the tracking problem.

2) *Tracking for the Case of $N = 2$:* Let Q_a and Q_b be the target coordinates. The tracking problem is feasible *iff* there is a circle ω that simultaneously goes through Q_a , Q_b and points C_1 , C_2 , and V . This means that the following equation must admit a nontrivial solution:

$$\underbrace{(\hat{Q}_a \quad \hat{Q}_b \quad \hat{I} \quad \hat{J} \quad \hat{C}_1 \quad \hat{C}_2 \quad \hat{V})^T}_{D} \omega = 0.$$

The existence of a nontrivial solution requires the 7×6 matrix D to be rank deficient. It follows that

$$\begin{cases} \det(D_{1,\dots,6}) = 0 \\ \det(D_{1,\dots,5,7}) = 0 \end{cases}$$

where the subscripts denote the matrix lines. The two equations are constraints on the controllable parameters α , θ_1 , and θ_2 , respectively. The first is a constraint only on α , and provides a single solution of practical significance. It implies that the platform rotation is uniquely defined by Q_a and Q_b (see Fig. 12). The second equation is a condition that is satisfied by any pair (θ_1, θ_2) that places the fixation point on the circle defined by the targets and the camera centers. Therefore, the simultaneous tracking of $N = 2$ targets is a feasible problem, with a unique solution for the platform rotation, and multiple solutions for the pan angles (θ_1 and θ_2 must only assure that V lies on ω)

3) *Tracking for the Case of $N = 3$:* For the case of $N = 3$, the matrix D in the previous section gives place to the 8×6 matrix G

$$G \sim (\hat{Q}_a \quad \hat{Q}_b \quad \hat{Q}_c \quad \hat{I} \quad \hat{J} \quad \hat{C}_1 \quad \hat{C}_2 \quad \hat{V})^T.$$

Enforcing the rank deficiency would lead to two independent constraints on the angle α ($\det(G_{1,\dots,6}) = 0 \wedge \det(G_{1,\dots,5,7}) = 0$) that are either impossible or do not have a common solution. Thus, for $N > 2$, the tracking problem

is in general not feasible, as opposed to the case analyzed in Section V-C. This results from the fact that we have considered (in the case of the stereo head) the ratio of focal lengths constant and equal to one, i.e., $\rho = 1$. In Section V-C, it was shown that tracking $N = 3$ targets requires the control of ρ .

C. Tracking Strategy/Constraints in the DOF

Section VI-B2 shows that the stereo head is able to keep $N = 2$ targets in the horopter *iff* the following conditions hold:

$$\begin{cases} \det(D_{1,\dots,6}) = 0 \\ \det(D_{1,\dots,5,7}) = 0 \end{cases}$$

with

$$D \sim (\hat{Q}_a \quad \hat{Q}_b \quad \hat{I} \quad \hat{J} \quad \hat{C}_1 \quad \hat{C}_2 \quad \hat{V})^T.$$

Let us take into account the kinematics of our robot head and replace C_1 , C_2 , and V with the result given in (12) and (15). Note that matrix D becomes a function of the target world coordinates and of the system DOF (α , θ_1 , and θ_2).

As stated in Section VI-B2, the first constraint $\det(D_{1,\dots,6}) = 0$ does not involve parameters θ_1 and θ_2 . By solving the equation with respect to the neck pan α we obtain

$$\alpha = \arctan \frac{(Q_{b,x}^2 + Q_{b,y}^2 - B^2)Q_{a,y} - (Q_{a,x}^2 + Q_{a,y}^2 - B^2)Q_{b,y}}{(Q_{b,x}^2 + Q_{b,y}^2 - B^2)Q_{a,x} - (Q_{a,x}^2 + Q_{a,y}^2 - B^2)Q_{b,x}}. \quad (16)$$

Angle α is uniquely defined by the target locations Q_a and Q_b , respectively. The correct α value places the cameras in such a manner that points Q_a , Q_b , C_1 , and C_2 lie on the circular horopter (see Fig. 12).

Consider the second constraint $\det(D_{1,\dots,5,7}) = 0$. Replacing α by the expression derived earlier yields an equation on the angles θ_1 and θ_2 , respectively. The condition is satisfied by any duplet of values (θ_1, θ_2) that places the fixation point V anywhere on the circle defined by the targets and camera centers. The tracking problem is undetermined in the sense that there exist an infinite number of solutions for the control parameters. To further restrict the problem, we decided to impose an additional constraint, which requires point V to be kept in the middle of the arc defined by Q_1 and Q_2 at all times (see Fig. 12). This is accomplished by rotating each camera so that its optical axis halves the angle defined by the two targets back-projection lines or, in other words, by keeping the two target images symmetric with respect to image center. Angles θ_1 and θ_2 can be computed in a straightforward manner as a function of targets world coordinates and neck pan angle

$$\theta_{1/2} = \frac{1}{2} \arctan \frac{-Q_{a,x} \mp B \cos \alpha}{Q_{a,y} \pm B \sin \alpha} + \frac{1}{2} \arctan \frac{-Q_{b,x} \mp B \cos \alpha}{Q_{b,y} \pm B \sin \alpha} - \alpha. \quad (17)$$

Note that the constraint in the position of the fixation point is beneficial in terms of tracking. By forcing V to be in the middle of Q_a and Q_b , we take full advantage of the cameras FOV in the situation of the targets being moving apart [see Fig. 14(b)].

TABLE I
SUMMARY OF THE CONTROL EQUATIONS FOR SIMULTANEOUSLY TRACKING
TWO TARGETS USING THE POPEYE STEREO HEAD

Control Equations
$\dot{\alpha}^{cmd} = \frac{4 \cdot \Upsilon}{\sin(2(\phi_{2b} - \phi_{1b})) \cdot \Psi} + \dot{\alpha}^r + G \arctan\left(-2 \frac{\Gamma}{\Lambda}\right)$
$\dot{\theta}_i^{cmd} = \frac{\dot{\phi}_{ia}^r + \dot{\phi}_{ib}^r}{2} + G \left(\frac{\phi_{ia}^r + \phi_{ib}^r}{2} - \theta_i^r \right)$
Auxiliary expressions
$\phi_{i,j}^r = \theta_i^r + \arctan\left(\frac{q_{i,j}}{f}\right)$
$\dot{\phi}_{i,j}^r = \dot{\theta}_i^r + \dot{q}_{ij} \frac{\cos^2(\phi_{i,j}^r - \theta_i^r)}{f}$
$\Upsilon = \dot{\phi}_{2b}^r - \dot{\phi}_{2a}^r + \dot{\phi}_{1a}^r - \dot{\phi}_{1b}^r$
$\Psi = \tan \phi_{1a}^r - \tan \phi_{1b}^r - \tan \phi_{2b}^r + \tan \phi_{2a}^r$
$\Gamma = \tan(\phi_{2b}^r - \phi_{1b}^r) - \tan(\phi_{2a}^r - \phi_{1a}^r)$
$\Lambda = (\tan \phi_{2b}^r + \tan \phi_{1b}^r) \tan(\phi_{2b}^r - \phi_{1b}^r) - (\tan \phi_{2a}^r + \tan \phi_{1a}^r) \tan(\phi_{2a}^r - \phi_{1a}^r)$

Equations (16) and (17) provide the values for α , θ_1 , θ_2 such that the two free-moving targets are projected in the same position in both retinas, and their images are symmetric with respect to the center. Since the target locations \mathbf{Q}_a and \mathbf{Q}_b can be determined from stereo triangulation (14), the correct angles α , θ_1 , and θ_2 can be directly computed from target image coordinates $q_{i,j}$ and current kinematic configuration of the robot head.

Henceforth, we will use the superscript d to denote the *desired* (or reference) values for the system DOF (α^d , θ_1^d , θ_2^d) and the superscript r to represent the *real* (or actual) angles measured by the system encoders (α^r , θ_1^r , and θ_2^r). The former are the angular positions for the active system to accomplish the defined tracking objectives, while the latter describe the current kinematic configuration of the robot head. Replacing $\mathbf{Q}_{i,j}$ in (16) by the result of (14), and taking into account the new notation, yields

$$\alpha^d = \alpha^r + \arctan\left(-2 \frac{\Gamma}{\Lambda}\right) \quad (18)$$

with Γ and Λ being auxiliary expressions given in Table I. Repeating the procedure for the cameras pan angles of (17), we obtain

$$\theta_i^d = \frac{\phi_{ia}^r + \phi_{ib}^r}{2}. \quad (19)$$

The aforementioned two equations provide suitable servo-control references for the DOFs of the robot head. The set point angles α^d and θ_i^d are conveniently expressed as a function of image measures $q_{i,j}$ and the encoder-feedback information α^r , θ_i^r . The following section describes the system-control architecture and final implementation.

D. Control Architecture

Each system axis α , θ_1 , and θ_2 follows a similar control scheme as shown in Fig. 13. This section presents a brief overview of the architecture that addresses the active visual

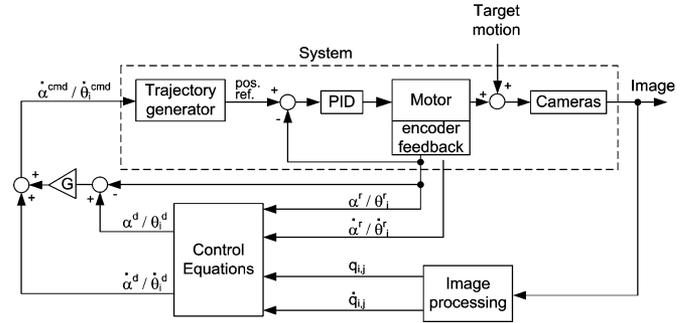


Fig. 13. Control scheme of the DOFs of the POPEYE robot head. Each DOF, i.e., α , θ_1 , and θ_2 , has a similar control architecture. The dashed box represents the inner most servo-loop, which runs at 8 kHz and controls the velocity of the brushless ac motor. The outer most loop runs at 15 Hz and uses both visual and mechanical feedback to compute the set point of the inner-most loop. Table I shows how to compute the reference for the axis controller, given the image measurements and the encoder information. The final servo command combines angular velocity with the position error weighted by a constant gain G . This paper does not discuss the tuning of the PIDs, trajectory generators, and gains. For details, see [10].

tracking as a regulation-control problem in the image plane (for a more detailed explanation, see [10]).

Although the axis controller implements a position-control loop, it can be configured to receive velocity commands. In such operation mode, the velocity reference is integrated by a trajectory generator that, in turn, updates the reference of the proportional integral derivative (PID) position control loop. In [10], it is shown that operating in velocity mode improves both the stability and the responsiveness of the active visual tracking. Therefore, the system axes will be controlled using this mode, which means that we must compute a velocity reference/command.

Since (16) and (17) provide the correct reference in angular position, we can differentiate them with respect to time in order to obtain angular velocities. It follows that

$$\dot{\alpha}^d = \frac{4 \Upsilon}{\sin(2(\phi_{2b} - \phi_{1b})) \Psi} + \dot{\alpha}^r \quad (20)$$

and

$$\dot{\theta}_i^d = \frac{\dot{\phi}_{i,a}^r + \dot{\phi}_{i,b}^r}{2}. \quad (21)$$

with Υ , Ψ , and $\phi_{i,j}^r$ shown in Table I.

Here $\dot{\alpha}^d$, $\dot{\theta}_1^d$, and $\dot{\theta}_2^d$ are the angular velocities of the axes that keep target image positions symmetric and the same in both retinas. Unfortunately, pure velocity commands do not encode position information. As a consequence, the system will not be able to compensate accumulated position errors, and the tracking will tend to drift. The problem is solved by adding the angular position error multiplied by a suitable constant gain G . The final commands (or set points) sent to the axes controllers are

$$\dot{\alpha}^{cmd} = \dot{\alpha}^d + G(\alpha^d - \alpha^r)$$

for the neck pan and

$$\dot{\theta}_i^{cmd} = \dot{\theta}_i^d + G(\theta_i^d - \theta_i^r)$$

for the eye pan. Note that, while the position term assures step-disturbance rejection, the velocity term works as a derivative

component that improves the system-transient response [10]. Table I summarizes the final control equations used in the experiment described in the following section.

E. Experimental Results

This section reports the tracking of two moving targets by the POPEYE robot head using the strategy established earlier.³ The objective of the experiment is to prove the concept, and not to necessarily provide a fully functional application. Therefore, the image processing was simplified by assuming as targets two color markers that are easily detected and tracked using the OpenCV implementation of the CAMSHIFT algorithm [26]. The markers are moved around by two persons, and the POPEYE head tries to project them in the same position in both cameras, while keeping the images symmetric with respect to center.

Fig. 14(e) shows the image positions of the targets during 2 min of tracking with no interruptions, while Fig. 14(a)–(d) concerns four particular tracking instants. The experiment proves that the adopted tracking scheme succeeds in projecting the targets in the same position in the two views. In addition, for each camera, the target image positions are symmetric with respect to the center. Such tracking behavior assures that the chances of target mutual occlusion are minimized and that the stereo reconstruction of the trajectories is usually possible [see third row in Fig. 14 (a)–(d)].

Mutual occlusion can only occur when the two targets and the middle point of the baseline become collinear. This is a singular configuration, which corresponds to a discontinuity in solution for the platform orientation α and causes the cameras to be placed in such a manner that both targets are projected in the image center. Fig. 14(d) shows a moment when this configuration occurs. The mechanical limits for the cameras' rotation are reached, and the tracking momentarily fails, which results in the peaks in the plots of Fig. 14(e). Handling this problem is, for now, beyond the scope of the paper. The current implementation is able to gracefully recover the tracking behavior.

Fig. 14(a)–(d) shows two other situations that deserve a particular remark. In Fig. 14(b), the targets move away from each other, but the tracking proves to be effective in maintaining them in the FOV of both cameras. The ability of taking full advantage of the available FOV results from the strategy of keeping the fixation point between the targets. The misalignment in image positions observed in Fig. 14(c) is caused by fast target motion toward different directions. The system, despite the natural difficulties in maintaining zero-tracking error, shows a stable behavior.

VII. DISCUSSIONS AND CONCLUSION

This paper has extended the active-fixation framework for the case of $N > 1$ points of interest. The tracking behavior has been specified by selecting a configuration homography that defines how the stereo images of the targets should relate. We have shown that the locus of points whose stereo projections are consistent with an homography is a plane conic and that

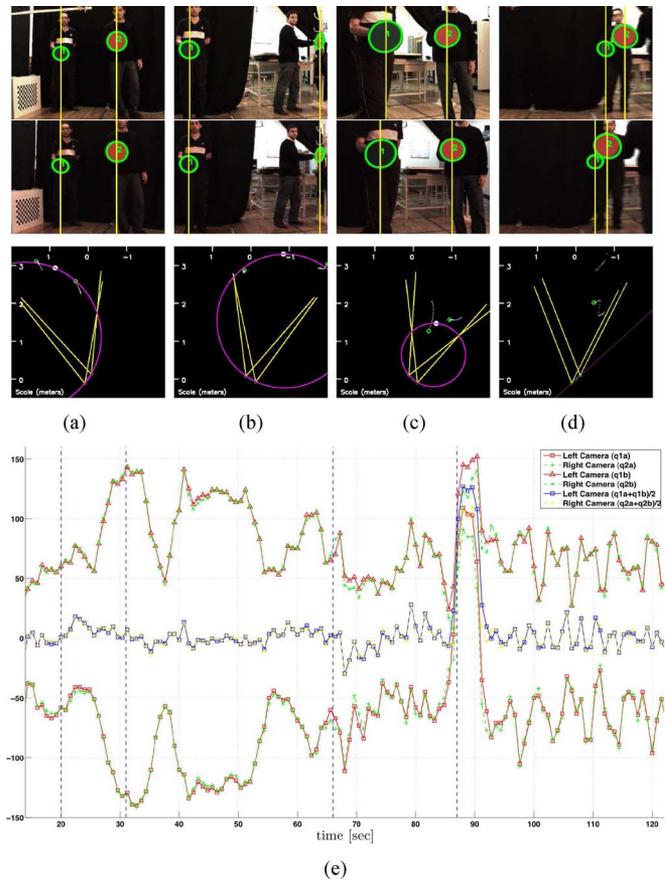


Fig. 14. Simultaneous tracking of two free-moving targets with the POPEYE head (the video of the experiment is available as supplementary material). In (e), plots of the image positions of the targets in the two views along almost 2 min of tracking with no interruptions are shown. The top and bottom red lines correspond, respectively, to the horizontal coordinates of Q_a and Q_b in the left-hand side image. The symmetry with respect to the center is rather obvious. The dashed green lines concern the horizontal target image coordinates in the second view (right-hand side camera). The red and green lines are coincident most of the time, which proves that the tracking objective has been successfully accomplished. (a)–(d) Four distinct tracking instants, with the two top images being the stereo views and the bottom image being a top view of the working plane. The green dots are the recovered target positions, the white dot is the fixation point V , and the magenta circle is an overlay of the homographic curve. Note that in (b), the targets move apart, but the system manages to keep them in both views taking full advantage of the cameras FOV. Instant (d) is a singularity that arises whenever the two objects are aligned with the cyclopean eye. In this case, one target occludes the other, and the tracking fails. However, the system is able to gracefully recover from this situation. (a) 20 s. (b) 31 s. (c) 66 s. (d) 87 s.

the tracking problem can be cast as the alignment of this conic with the moving targets. This formulation is quite convenient because it enables the systematic analysis of the feasibility of a given tracking task and the straightforward derivation of the relevant constraints and control laws. These features were illustrated through two particular application examples: active tracking using two PTZ cameras arbitrarily placed in the working environment (see Section V) and active tracking using the POPEYE robot head (see Section VI). In the former case, it is possible to track up to $N = 3$ free-moving targets, while in the latter case, the maximum number of targets is $N = 2$ because of the lack of zoom control. Simulation results showed that

³The experiment video is available as supplementary material (Clip 3).

tracking $N = 3$ targets can only be accomplished in practice by PTZ cameras equipped with wide angle lenses.

Our main contribution is the theoretical formulation using *configuration homographies*. The framework provides a geometric insight that can be helpful for applications other than the ones discussed. A nonexhaustive list of problems that might benefit from the proposed approach include the following.

- 1) *Active tracking using configuration matrices $H \neq I$* : The reported experiments in active tracking always assume a constant configuration homography $H = I$. However, we foresee that other useful tracking schemes can be accomplished by dynamically changing H . Let us consider a surveillance scenario where a static wide-angle camera is combined with an active PTZ unit. The former provides permanent visual coverage of the space, while the latter aims at obtaining visual detail of points of interest. In the simultaneous presence of $N = 2$ moving targets, the PTZ must zoom in as much as possible with no loss of the visibility of objects. Thus, the tracking objective can be formulated as controlling the PTZ such that the two targets are kept on opposite sides in the image. Since the mapping of the two targets in the wide-angle image into the two lateral limits of the PTZ view is always an affine 1-D transformation, the proposed framework can be applied to implement the described active-tracking system. The tracking task seems to be also feasible for $N = 3$; however, care must be taken because the mapping of three points defines a general projective transformation that can change the order in which targets are projected.
- 2) *Camera placement in surveillance and robotics*: The formulation can also be used for camera placement. In this case, the geometric framework is employed for shaping the homographic curve such that the stereo images of certain scene locations are related by a predefined homography H . A possible application is indoor surveillance, with cameras being placed and oriented in order for certain 3-D locations (e.g., entrances) to be imaged according to certain conditions (e.g., same position in the two views) that might simplify event detection (e.g., image difference for detecting that someone crossed the entrance) and promote collaborative image processing. Another possibility is to use a similar strategy to shape an homographic curve around a mobile robot equipped with stereo line scan cameras. This curve will work as a “bumper” to enable obstacle detection using very simple visual routines (e.g., image difference).
- 3) *Robot placement and formation control*: Another practical application might be the positioning of one or multiple robots relative to three reference points in the scene. In the former case, we would be interested in guaranteeing that the landmarks and the robot with the camera occupy predefined relative positions, which are encoded via a suitable configuration homography. Therefore, the goal of the control would be achieved by changing the position and orientation of the robot so that the image projections of the targets correspond to that specific homography. The approach can be easily extended to problems of robot for-

mation, where the different robots converge to relative poses encoded by different homographies.

- 4) *3-D stereo reconstruction*: Since the homographic curves are a generalization of the isodisparity curves used in stereo reconstruction [22], our results might also be relevant in this context.

A significant limitation of this framework is that it applies exclusively to targets undergoing a planar motion and cannot be easily extended to the case of unconstrained 3-D motion. The difficulty in the generalization is that while a 1-D homography between line scan images gives rise to a conic curve in the working plane, a 2-D homography between perspective images usually corresponds to a planar surface in 3-D. This is rather limitative and does not sound promising in terms of active-vision applications.

REFERENCES

- [1] I. Y. Aloimonos and A. Bandyopadhyay, “Active vision,” in *Proc. IEEE Int. Conf. Comput. Vis.*, 1987, pp. 35–54.
- [2] R. Bajcsy, “Active perception,” in *Proc. IEEE*, 1988, vol. 76, pp. 996–1005.
- [3] J. P. Barreto and H. Araujo, “A general framework for selecting world coordinate systems in perspective and catadioptric imaging applications,” *Int. J. Comput. Vis.*, vol. 57, no. 1, pp. 23–47, 2004.
- [4] A. Yeung and N. Barnes, “Efficient active monocular fixation using the log-polar sensor,” *Int. J. Intell. Syst. Technol. Appl.*, vol. 1, no. 1–2, pp. 157–173, 2005.
- [5] J. Batista, P. Peixoto, and H. Araujo, “The ISR multi-degrees-of-freedom active vision robot head,” in *Proc. Int. Conf. Mechatron. Mach. Vis. Practice*, 1995, pp. 1–7.
- [6] C. S. Andersen and H. I. Christensen, “Using multiple cues for controlling an agile camera head,” in *Proc. IAPR/IEEE Workshop Vis. Behav.*, 1994, pp. 97–101.
- [7] H. Christensen and J. O. Eklund, *Active Vision from Multiple Cue*. (Lecture Notes in Computer Science Series, vol. 1811). New York: Springer-Verlag, 2000.
- [8] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge, U.K.: Cambridge Univ. Press, 2004.
- [9] D. Coombs and C. Brown, “Real time binocular smooth pursuit,” *Int. J. Comput. Vis.*, vol. 11, no. 2, pp. 147–164, 1993.
- [10] J. P. Barreto, J. Batista, and H. Araujo, “Model predictive control to improve visual control of motion: Applications in active tracking of moving targets,” presented at the IEEE Int. Conf. Pattern Recognit, Barcelona, Spain, Sep. 2000.
- [11] S. D. Roy, S. Chaudhury, and S. Banerjee, “Active recognition through next view planning: A survey,” *Pattern Recognit.*, vol. 37, no. 3, pp. 429–446, Mar. 2004.
- [12] T. O. Y. Satoh and K. Deguchi, “Binocular motion tracking by gaze fixation control and 3D shape reconstruction,” *Adv. Robot.*, vol. 17, no. 16, pp. 1057–1072, 2003.
- [13] D. W. Murray, I. D. Reid, and A. J. Davison, “Steering without representation using active fixation,” *Perception*, vol. 26, pp. 1519–1528, 1997.
- [14] G. Sandini and M. Tistarelli, “Active tracking strategy for monocular depth inference over multiple frames,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 12, no. 1, pp. 13–27, Jan. 1990.
- [15] N. Barnes and G. Sandini, “Direction control for an active docking behaviour based on the rotational component of log-polar optic flow,” in *Proc. 6th Eur. Conf. Computer Vision-Part II*. London, U.K.: Springer-Verlag, 2000, pp. 167–181.
- [16] E. Sommerlade and I. Reid, “Cooperative surveillance of multiple targets using mutual information,” in *Proc. ECCV Workshop Multi-Camera Multi-Modal Sens. Fusion Algorithms Appl.*, Oct. 2008, pp. 1–12.
- [17] P. Cavanagh and G. A. Alvarez, “Tracking multiple targets with multifocal attention,” *Trends Cogn. Sci.*, vol. 9, no. 7, pp. 349–354, Jul. 2005.
- [18] L. Perdigoto, J. P. Barreto, R. Caseiro, and H. Araujo, “Active stereo tracking of multiple free moving targets,” in *Proc. IEEE Int. Conf. Pattern Recognit.*, Miami, FL, Jun. 2009, pp. 1–8.
- [19] R. Horn and C. Johnson, *Topics in Matrix Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 1991.

- [20] R. Gupta and R. Hartley, "Linear pushbroom cameras," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 19, no. 9, pp. 963–975, Sep. 1997.
- [21] R. Bolles, H. Baker, and D. H. Marimont, "Epipolar-plane image analysis: An approach to determining structure from motion," *Int. J. Comput. Vis.*, vol. 1, pp. 7–55, 1987.
- [22] M. Pollefeys and S. Sinha, "Iso-disparity surfaces for general stereo configurations," in *Proc. Eur. Conf. Comput. Vis.*, 2004, pp. 1–12.
- [23] J. Semple and G. Kneebone, *Algebraic Projective Geometry*. Gloucestershire, U.K.: Clarendon, 1998.
- [24] I. Reid and D. Murray, "Active tracking of foveated feature clusters using affine structure," *Int. J. Comput. Vis.*, vol. 18, pp. 41–60, 1996.
- [25] H. Helmholtz, *Treatise on Physiological Optics*. New York: Dover, 1925.
- [26] G. Bradski, "Computer vision face tracking for use in a perceptual user interface," *Intel Technol. J.*, vol. Q2, 1998.



and active vision.

Rui Caseiro received the B.Sc. degree in electrical engineering (specialization in automation) from the University of Coimbra, Coimbra, Portugal, in 2005.

Since 2007, he has been involved in several research projects, which include the European project "Perception on Purpose" and the "Brisa-Highway Project." He is currently a Researcher with the Institute of Systems and Robotics and the Department of Electrical and Computer Engineering, Faculty of Science and Technology, University of Coimbra. His research interests include computer vision, robotics,



Joao P. Barreto (M'99) received the "Licenciatura" and Ph.D. degrees from the University of Coimbra, Coimbra, Portugal, in 1997 and 2004, respectively.

From 2003 to 2004, he was a Postdoctoral Researcher with the University of Pennsylvania, Philadelphia. He has been an Assistant Professor with the Department of Electrical and Computer Engineering, Faculty of Science and Technology, University of Coimbra, since 2004, where he is also a Researcher with the Institute for Systems and Robotics. His current research interests include different topics in computer vision, with particular emphasis on camera models and 3-D reconstruction.

He is the author of more than 30 peer-reviewed publications and is also a reviewer for numerous conferences and journals proceedings.



Helder Araujo received the M.Sc. and Ph.D. degrees in electrical and computer engineering from the University of Coimbra, Coimbra, Portugal, in 1981 and 1989, respectively.

He is a Professor with the Department of Electrical and Computer Engineering, University of Coimbra. He is also a Researcher with the Institute for Systems and Robotics, Coimbra. His research interests include computer and robot vision, robot navigation and localization, and sensor modeling for robot navigation and localization.



Luis Perdigoto (S'09) received the M.S. degree in electrical and computer engineering from the University of Coimbra, Coimbra, Portugal, in 2007. He is currently working toward the Ph.D. degree with the Institute for Systems and Robotics, University of Coimbra.

He is also a Teaching Assistant with the Escola Superior de Tecnologia e Gestao de Leiria, Department of Electrical Engineering, Polytechnic Institute of Leiria, Leiria, Portugal.