

Active Stereo Tracking of Multiple Free-Moving Targets

Luis Perdigoto^{1,2}, Joao P. Barreto¹, Rui Caseiro¹, Helder Araujo¹ *

¹Institute for Systems and Robotics, Dept. of Electrical and Computer Engineering
University of Coimbra, 3030 Coimbra, Portugal

{jpbbar, ruicaseiro, helder}@isr.uc.pt

²ESTG, Polytechnic Institute of Leiria, 2411 Leiria, Portugal

perdigot@estg.ipleiria.pt

Abstract

This article presents a general approach for the active stereo tracking of multiple moving targets. 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 an homography. This homography is specified during the design stage and implicitly encodes the tracking behavior. It is shown that this formulation leads to an elegant geometric framework that enables to decide about the feasibility of a particular active tracking task. We apply it to prove that two cameras with rotation and zoom control, can track up to three moving targets, while assuring that the image location of each target is the same for both views. In addition, the framework is also useful for devising tracking strategies and deriving the required control equations. This feature is illustrated through a real experiment on tracking two independent targets using a binocular stereo head.

1. Introduction

Active tracking is a part of the *active vision* paradigm [15, 2], where visual systems adapt themselves to the observed environment in order to obtain extra information or perform a task more efficiently. Active tracking consists in 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 the gaze direction is maintained 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) [17]. The latter usually considers a stereo head [1], with the point of interest being foveated by intersecting the optical axes of both cameras at the exact target location (*the vergence/fixation point*). Since the fixation point lies in the horopter [8], many binocular systems use target disparity between retinas as feedback control signal [6]. 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 [3]. However, the ability to fixate can be helpful in simplifying a broad range of high-level vision tasks (e.g., object recognition [12], 3D reconstruction [16] and robot navigation [10]).

While fixation concerns tracking a single object or point of interest, our article addresses the problem of using an active stereo setup to simultaneously track $N > 1$ free moving targets. Sommerland and Reid have recently proposed an information theoretical framework for tracking multiple targets with multiple PTZs [14]. However, they address problems like sensor-target assignment, camera hand-off and zoom control without missing new objects, whereas the focus of our work is towards extending the classical binocular fixation framework for the case multiple points of interest. We aim to push the single focus of attention paradigm, typical of binocular fixation, towards a more general multi-focal attention framework [5].

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 an 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

*The authors acknowledge the support of project POP (Perception on Purpose), project FP6-IST-2004-027268, funded by the European Union. Luis Perdigoto acknowledges the support of the Portuguese Science Foundation through grant SFRH/BD/39096/2007. Joao Barreto is grateful to the Portuguese Science Foundation by generous funding through grant PTDC/EEA-ACR/68887/2006.

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, desired homography H , and number of targets N , enables to decide about the feasibility of the tracking task and derive the relevant constraints on the control parameters.

The problem is fully formulated in the plane, with the cameras being modeled as "line scan cameras", and the targets being described as 2D 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 camera optical axes [1], 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 non-coplanar). However, and for practical purposes, we can always consider projecting the 3D target motion onto the plane defined by the camera optical axes. We will show that this solution is particularly effective for indoor applications, where trajectories are usually parallel to the ground plane.

Although the focus of this paper is placed on active tracking, our framework has a wider scope for application. Our work can potentially be used in any system where collaborative inter-image processing is accomplished/facilitated by proper camera placement and orientation, as the same geometric principles apply.

2. 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 and lines, unless stated otherwise, are represented in projective homogeneous coordinates.

Vectorization of matrix equations: Let Y , A , X and B be rectangular matrices such that

$$Y = AXB.$$

The equality above can be re-written 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 (c.f. chapter 4 in [9]). It is also convenient to keep in mind the following property of the Kronecker product

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

Vector representation of conic curves: Consider a point in the plane, with homogeneous coordinates

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

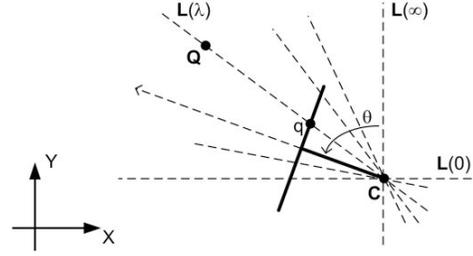


Figure 1. Line scan camera model where OXY is the world system of coordinates. The camera is centered in C and rotated by an angle θ with respect to the Y axis. The 2D point Q is projected into q in the line image. The dashed lines $L(\lambda)$ represent the pencil of back-projection.

and a conic curve represented by the symmetric matrix

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

Point \mathbf{X} is on the conic curve iff $\mathbf{X}^T \Omega \mathbf{X} = 0$. This second order polynomial can be re-written in the following form

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

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

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

and ω a vector representation of the conic curve

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

3. Modeling the Line Scan Camera

In this article 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 uni-dimensional cameras has already been studied under different contexts of application [7, 4]. This section introduces the projection and back-projection models that will be considered in the remaining of the paper.

3.1. Projection Matrix

Fig. 1 shows a line scan camera with projection center

$$\mathbf{C} = (C_x \ C_y \ 1)^T,$$

and matrix of intrinsic parameters

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

where f stands for the focal length. Without loss of generality, it will be assumed that the origin of the image is coincident with the principal point.

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

$$\mathbf{q} \sim \mathbf{K} \mathbf{R} (\mathbf{I} - \mathbf{C}') \mathbf{Q}. \quad (1)$$

\mathbf{I} denotes the 2×2 identity matrix, \mathbf{C}' is the non-homogeneous representation of the projection center and \mathbf{R} encodes the rotation of the camera by an angle θ :

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

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

3.2. Back-Projection Pencil

Let's now consider the problem of computing the back-projection of an image point \mathbf{q} . Define the matrix \mathbf{U} such that

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

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

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

By left multiplying both sides of equation 1 by $\mathbf{q}^T \mathbf{U}$, it follows that

$$\underbrace{\mathbf{q}^T \mathbf{U} \mathbf{K} \mathbf{R} (\mathbf{I} - \mathbf{C}')}_{\mathbf{L}^T} \mathbf{Q} = 0.$$

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

Define λ such that

$$\begin{pmatrix} \lambda \\ 1 \end{pmatrix} \sim \mathbf{R}^T \mathbf{K}^T \mathbf{U}^T \mathbf{q}. \quad (2)$$

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

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

$\mathbf{L}(\lambda)$ is the pencil of lines going through the camera center \mathbf{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 (Fig. 1).

4. The Homographic Curve

We propose using a 1-D configuration homography \mathbf{H} to specify the desired tracking behavior. The idea is to control the active stereo system such that the two views of the N

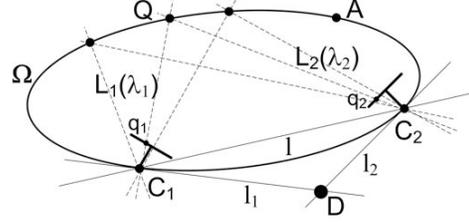


Figure 2. The homography \mathbf{H} induces an homographic relation between the back-projection pencils. It follows from Steiner's theorem that corresponding lines in two homographically related pencils intersect on a conic locus. Ω is called the homographic curve of \mathbf{H} because the images of any point \mathbf{Q} on Ω satisfy $\mathbf{q}_2 \sim \mathbf{H} \mathbf{q}_1$.

targets are mapped one into the other by the homography \mathbf{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 \mathbf{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. Remark that the homographic curves generalize both the horopter [8] and the iso-disparity curves proposed in [11]. The former is the homographic curve for $\mathbf{H} \sim \mathbf{I}$, with points being projected at the same location in both images. The latter corresponds to the case of \mathbf{H} being a 1-D translation that shifts image points by a constant amount.

4.1. Image Homography and Pencil Homography

Fig. 2 shows two cameras with centers \mathbf{C}_1 and \mathbf{C}_2 , rotation matrices \mathbf{R}_1 and \mathbf{R}_2 , and intrinsics \mathbf{K}_1 and \mathbf{K}_2 . Let the desired image homography be

$$\mathbf{H} \sim \begin{pmatrix} a & b \\ c & d \end{pmatrix}. \quad (3)$$

\mathbf{H} maps points \mathbf{q}_1 in the first view, into points \mathbf{q}_2 in the second view

$$\mathbf{q}_2 \sim \mathbf{H} \mathbf{q}_1.$$

Consider now the parametrization of the back-projection pencils discussed in section 3.2. Each image point \mathbf{q} is mapped into a λ value that defines the corresponding back projection line (equation 2). Let λ_1 and λ_2 be the parameters associated with \mathbf{q}_1 and \mathbf{q}_2 . By inverting equation 2 and replacing \mathbf{q}_1 and \mathbf{q}_2 in the above equation, it arises

$$\begin{pmatrix} \lambda_2 \\ 1 \end{pmatrix} \sim \underbrace{\mathbf{R}_2^T \mathbf{K}_2^T \mathbf{U}^T \mathbf{H} \mathbf{U} \mathbf{K}_1^{-T} \mathbf{R}_1}_{\mathbf{H}_L} \begin{pmatrix} \lambda_1 \\ 1 \end{pmatrix}.$$

The image homography \mathbf{H} defines a correspondence between back-projection lines. This correspondence is described by \mathbf{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 can be written in a vectorized form using the approach described in section 2. It follows that

$$\text{vec}(H_L) \sim M F \text{vec}(H), \quad (4)$$

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

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

and F is a matrix encoding the intrinsic parameters

$$F \sim (K_1^{-1} \otimes K_2^T) (U^{-1} \otimes U^T)$$

$$\sim \begin{pmatrix} 0 & 0 & 0 & f_2 \\ 0 & 0 & -1 & 0 \\ 0 & -f_1 & f_2 & 0 \\ f_1 & 0 & 0 & 0 \end{pmatrix}$$

4.2. The Equation of the Homographic Curve

The homography H transforms points \mathbf{q}_1 into points \mathbf{q}_2 establishing 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 \mathbf{Q} be the intersection point of two corresponding lines. It follows that the images of \mathbf{Q} in the two views must satisfy the original image homography H . This means that the homographic curve that we are aiming is the locus of the intersections points \mathbf{Q} . According to Steiner's theorem, two homographically related pencils intersect into a conic curve that goes through the centers of the pencils [13]. Thus, and as illustrated in Fig. 2, the homographic curve associated with H is always a conic Ω containing C_1 and C_2 . Given H_L and the camera centers, the conic Ω can be computed in a straightforward manner. The procedure is briefly outlined below (for further details c.f. chapters 5 and 6 of [13]).

- (i) Consider the line l going through the two centers. Since the line belongs to both pencils, determine the respective λ_1 and λ_2 values;
- (ii) Use H_L to map l into l_2 in the second pencil, and H_L^{-1} to map l into l_1 in the first pencil;
- (iii) Compute point \mathbf{D} where l_1 and l_2 intersect. It can be shown that \mathbf{D} and l are pole-polar with respect to Ω
- (iv) Find any two additional corresponding lines and determine the intersection point \mathbf{A} lying on the conic;
- (v) Let C_1 , \mathbf{D} , C_2 and \mathbf{A} define a canonical projective basis in the plane. For this particular parameterization the conic curve is

$$\Omega' \sim \begin{pmatrix} 0 & 0 & -1/2 \\ 0 & 1 & 0 \\ -1/2 & 0 & 0 \end{pmatrix}$$

- (vi) Determine the projective transformation S that maps the points back to their Euclidean coordinates. The final conic is

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

Ω depends both of the camera centers and the homographic relation H_L between the pencils. Considering the conic in its vectorized form ω , it follows that

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

with N being a function of the non-homogeneous coordinates of C_1 and C_2

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

The map of the image homography H into its corresponding homographic curve, is obtained by replacing $\text{vec}(H_L)$ for the result of equation 4.

$$\omega \sim N M F \text{vec}(H). \quad (5)$$

Remark that the equation is nicely factorized in matrix N , that encodes the position of the centers (or alternatively the translational component of camera motion), matrix M , that depends of the cameras rotations, and matrix F that is a function of the optical parameters.

4.3. Discussion

This section further analyzes equation 5, in order to gain insight about the homographic curve and its dependencies. The product of matrices N and M is a 6×4 matrix, where each column μ_i can be interpreted as the vectorized representation of a conic.

$$N M \sim (\mu_1 \quad \mu_2 \quad \mu_3 \quad \mu_4).$$

Consider the focal lengths in F and the scalar entries of H (equation 3). It follows from equation 5 that

$$\omega \sim d f_2 \mu_1 - b \mu_2 - c f_1 f_2 \mu_3 + a f_1 \mu_4. \quad (6)$$

Equation 6 denotes a linear system of conics with basis μ_i , with $i = 1, \dots, 4$ [13]. The equation shows that in general an homographic curve ω belongs to a 4D subspace in the space of all conics. This subspace is fully defined by the kinematic configuration of the stereo setup, because the conics μ_i in the basis only depend 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 H .

Let \mathbf{V} be the fixation point and \mathbf{T} the point in the plane that is projected at infinity in both views (Fig. 3(a)). If C_1 and C_2 are fixed points, then the coordinates of \mathbf{V} and \mathbf{T} only depend on the rotation angles θ_1 and θ_2 . 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

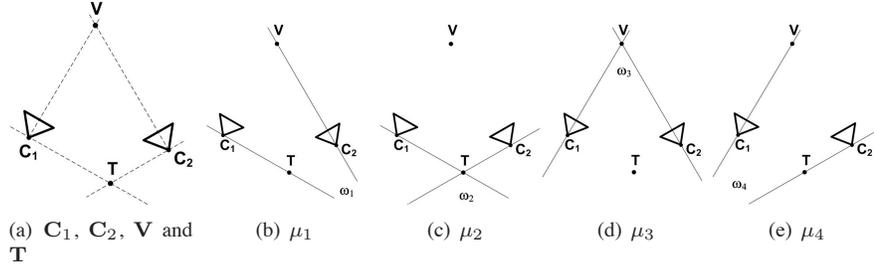


Figure 3. μ_i , $i = 1, \dots, 4$ are rank 2 degenerate conics and form the basis of the linear system of conics of equation 6. As shown in (b)-(e), the pair of lines composing each degenerate conic go through points C_1 , C_2 , V (the fixation point) and T (the point at infinity). These points encode the kinematic configuration of the active stereo system and implicitly define a 4D linear subspace in the space of all conics, containing every possible homographic curve ω .

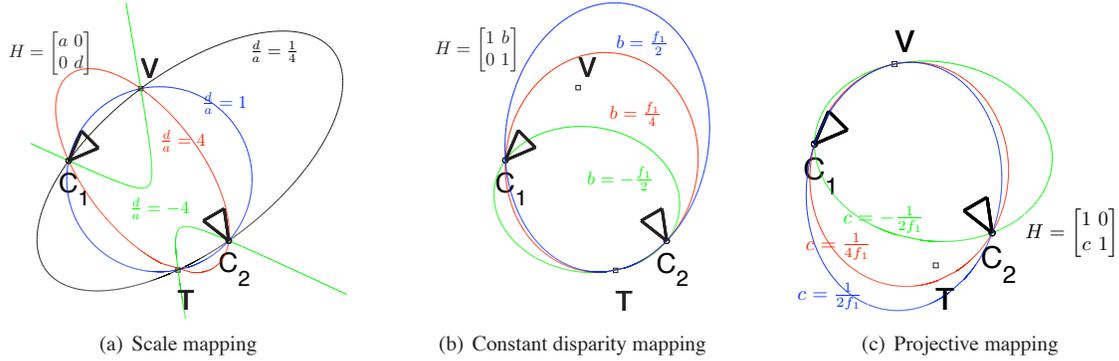


Figure 4. Dependence between the homographic curve and the selected homography H . The cameras are static and the focal lengths are constant and equal for both views (the linear subsystem of equation 6 is fixed). The entries of matrix H (equation 3) are changed in order to observe different shapes for the conic ω

shown in Fig. 3 (b)-(e), these lines can be found by knowing the locations of points C_1 , C_2 , V and T .

Fig. 4 shows the effect that the choice of the configuration homography H has in shaping the conic ω . It considers a particular kinematic configuration for the cameras such that C_1 , C_2 , V , T are fixed points that implicitly define a subspace in the space of all conics. Parameters f_1 and f_2 are assumed to be equal and constant. Fig. 4(a) concerns the case of 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 [13]. Since the pencil is defined by μ_1 and μ_4 that intersect at points C_1 , C_2 , V and T , then the homographic curve ω always goes through these points. 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 of T being 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 . Please note that the homographic cur-

ves of Fig. 4(b) are the *iso-disparity curves* discussed by Pollefeys et al. in the context of stereo reconstruction [11]. Finally, Fig. 4(c) shows the homographic curves for the case of H being a true projective transformation. Since b is zero, the linear system of equation 6 becomes a conic net [13], with point V being common to every member (the image centers are always mapped one into the other).

The remaining sections discuss the usage of the framework to study the feasibility of a tracking task and find the relevant control laws. The discussion is carried for two different examples: the tracking of N targets with PTZ cameras that undergo pure rotation motion (section 5); and the tracking of N objects using a binocular stereo head without zoom control (section 6). Henceforth, we will assume the configuration homography H to be 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 an useful feature for many real application scenarios. Please note that assuming $H \sim I$ does not imply a loss of generality. The framework can be equally employed for different choices of the configuration homographies, motivated by the need of meeting particular requirements of a certain tracking problem.

5. Active Tracking with Two PTZ Cameras

Consider two PTZ cameras with centers C_1 and C_2 . Both cameras undergo independent pan rotation and have zoom control that enables manipulating the ratio between the focal lengths

$$\rho = f_1/f_2.$$

The degrees of freedom (DOF) of the active stereo system are the ratio ρ and the rotation angles θ_1 and θ_2 . The goal 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 the configuration homography is $H \sim I$, it follows from equation 6 that the curve ω is always a member of the conic pencil

$$\omega \sim \mu_1 + \rho \mu_4. \quad (7)$$

In this case the homographic curve ω is the horopter of the stereo setup [8]. The curve contains points V and T , which depend on the cameras rotation angles, as well as the fixed projection centers C_1 and C_2 (Fig. 3(b) and (e)). Since ω is a function of ρ , θ_1 and θ_2 , the problem can be stated as controlling the system DOF such that the homographic curve goes through the locations of the N targets.

5.1. Tracking for the case of $N = 1$

Let the target have coordinates Q at a certain time instant. From the previous discussion it follows that the problem is feasible if there is an homographic curve ω such that:

$$\underbrace{(\hat{Q} \quad \hat{C}_1 \quad \hat{C}_2 \quad \hat{V} \quad \hat{T})^T}_A \omega = 0,$$

with $\hat{\cdot}$ denoting lifted point coordinates (c.f. section 2).

A is a 5×6 matrix that is function of θ_1 and θ_2 (this dependency is due to points V and T). In general, A has an uni-dimensional null space $\mathcal{N}(A)$, that can be interpreted as a 6×1 vector representing a conic curve. This curve belongs to the conic pencil of equation 7, because it goes through the four intersections of μ_1 and μ_4 . By replacing ω by $\mathcal{N}(A)$ and solving with respect to ρ , it arises

$$\rho = \frac{(C_{2,x} - Q_x + (C_{2,y} - Q_1) \tan \theta_2)(Q_y - C_{1,y} + (C_{1,x} - Q_x) \tan \theta_1)}{(Q_y - C_{2,y} + (C_{2,x} - Q_x) \tan \theta_2)(C_{1,x} - Q_x + (C_{1,y} - Q_y) \tan \theta_1)}$$

The equation above is written in terms of the non-homogeneous coordinates of C_1 , C_2 and Q . Any triplet of values $(\rho, \theta_1, \theta_2)$ satisfying it, is an admissible solution for the active tracking problem. Remark that the tracking is still feasible for the situation of cameras without 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 .

5.2. Tracking for the case of $N = 2$

Let Q_1 and Q_2 be the two free-moving targets. Repeating the reasoning of the previous section, the tracking for

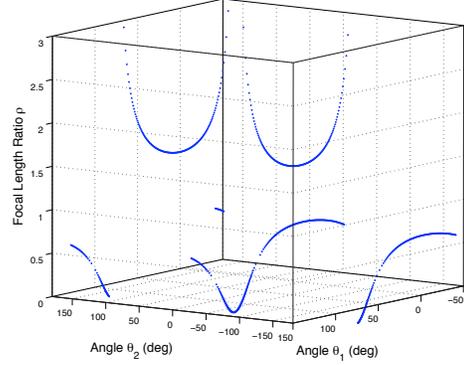


Figure 5. Solutions for the case of $N = 2$: The set of feasible solutions defines a 3D curve in the space of the control variables $(\rho, \theta_1, \theta_2)$. This curve depends on points Q_1, Q_2, C_1 and C_2 .

the case of $N = 2$ is feasible if there is a non-trivial solution for the following equation

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

Unfortunately B is in general a 6×6 non-singular matrix. However, and since the B is a function of the camera 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 non-trivial solution $\omega \sim \mathcal{N}(B)$, with $\mathcal{N}(B)$ denoting the 1-D null space of B . The solution $\omega \sim \mathcal{N}(B)$ must satisfy the equality of equation 7, which leads to an additional constraint involving the focal length ratio ρ .

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

Any solution $(\rho, \theta_1, \theta_2)$ of the system of equations is a feasible solution for the tracking problem. It assures that the homographic curve goes through Q_1 and Q_2 , and that targets are projected in the same location in both images. Each one of the above 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 Q_1, Q_2, C_1 and C_2 . Remark that for $N = 2$ the tracking problem still has an infinite number of solutions. However, there are ranges in ρ, θ_1 and θ_2 for which there are no feasible solutions. Moreover, and unlike what happens for $N = 1$, the accomplishment of the tracking objectives requires controlling the 3 DOF of the active system (e.g. if the system had no zoom control, then the tracking would not be in general possible).

5.3. Tracking for the case of $N = 3$

Repeating the approach, it follows that the solutions $(\rho, \theta_1, \theta_2)$ for the simultaneous tracking of Q_1, Q_2 and Q_3

must satisfy

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

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

$$C \sim (\hat{Q}_1 \quad \hat{Q}_2 \quad \hat{Q}_3 \quad \hat{C}_1 \quad \hat{C}_2 \quad \hat{V} \quad \hat{T})^T.$$

In general the system of equations has 8 solutions. However, some of the triplets $(\rho, \theta_1, \theta_2)$ are either redundant or have no physical meaning (e.g. $\rho < 0$). We will not pursue the discussion further but, and according to our simulations, the tracking for $N = 3$ is usually achievable in practice.

For $N > 3$ it is easy to verify that 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 5 points. Thus, and taking into account that ω must go through the two projection centers, the tracking for $N > 3$ is in general not feasible.

6. Tracking with an Active Stereo Head

This section discusses the tracking of multiple targets using an active stereo head. The two cameras are mounted on a robotic platform that rotates around an axis orthogonal to the working plane. The rotation angle α places the camera centers at different pairs of antipodal points on a circle with diameter B (the baseline). The two cameras have independent pan rotation (angles θ_1 and θ_2), equal focal lengths, and no zoom control.

We aim at tracking N targets assuming a configuration homography $H \sim I$. Since the focal lengths are equal, then $\rho = 1$ and equation 6 becomes

$$\omega \sim \mu_1 + \mu_4.$$

The horopter of two cameras with the same intrinsic parameters is the well known Vieth-Müller circle [8]. Thus, the above conic ω is a circle, containing points C_1, C_2, V and T , as well as the circular points I and J [13, 8]. While in section 5 the camera centers are fixed points, now points C_1 and C_2 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 . 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.

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 of the circle of diameter B . The three points Q, C_1 and C_2

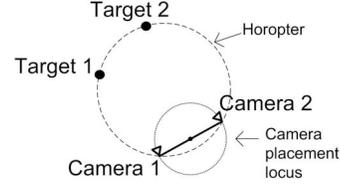


Figure 6. Simultaneous tracking of $N = 2$ targets. The platform rotation places C_1 and C_2 at antipodal positions on the circle with diameter B . The horopter (the curve ω for $H \sim I$) is the Vieth-Müller circle that goes through the targets and the camera centers.

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.

6.1. Tracking for the case of $N = 2$

Let Q_1 and Q_2 be target coordinates. The tracking problem is feasible *iff* there is a circle ω that simultaneously goes through the targets and the points C_1, C_2 and V . This means that the following equation must admit a non-trivial solution

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

The existence of a non-trivial solution requires matrix D to be rank deficient. It follows that

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

where the subscripts denote the matrix lines. The two equations are constraints on the controllable parameters $(\alpha, \theta_1, \theta_2)$. 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_1 and Q_2 (Fig. 6). 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 $(\theta_1$ and θ_2 must only assure that V lies on ω)

Fig. 7 shows a pair of stereo frames acquired during a real experiment in tracking two moving objects. The control was based on the discussion of the previous paragraph, with the fixation point placed in the middle of the targets in order to take full advantage of the cameras field of view. Unfortunately, and due to space limitations, we neither present details about control and singularities, nor discuss technicalities of the implementation. *However, the results are shown in the video provided as supplementary material.* To the best of our knowledge this is the first experiment in simultaneously tracking two free-moving targets with a stereo head.

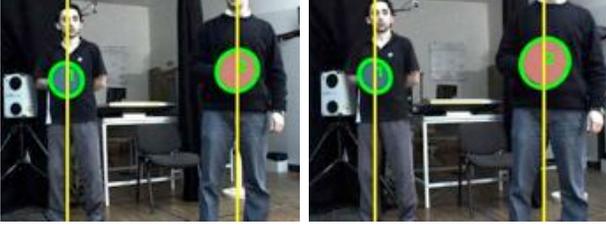


Figure 7. Simultaneous tracking of 2 moving objects using an active stereo head. The control of the stereo head tries to project the targets at the same horizontal location in both images ($H \sim I$). The workspace is 4×4 m, the FOV= 45° , and $B = 0.3$ m. The image processing is based on color which explains the markers. The head actuators are equipped with encoders enabling the reconstruction of the target positions by simple stereo triangulation.

6.2. 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

$$G \sim (\hat{Q}_1 \quad \hat{Q}_2 \quad \hat{Q}_3 \quad \hat{I} \quad \hat{J} \quad \hat{C}_1 \quad \hat{C}_2 \quad \hat{V})^T.$$

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 is not feasible.

7. Conclusions

This article extends the active fixation framework for the case of $N > 1$ points of interest. The tracking behavior is specified by selecting a configuration homography that defines how the stereo images of the targets should relate. We show that the locus of points whose stereo projections are consistent with an homography is a plane conic, and that the tracking problem can be casted 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 by showing that two PTZ cameras can track up to $N = 3$ targets while keeping them in the horopter. In addition, and to the best of our knowledge, we also presented the first experiment in simultaneously tracking two targets with an active stereo head.

As future work, we intend to investigate the possibility of using configuration mappings other than the homography (e.g. second order functions), in an attempt to extend the approach for the case of $N > 3$ targets. We also believe that the application domain goes beyond robotics and surveillance. The framework provides a powerful and insightful geometric analysis that is also relevant for problems in camera placement, camera networks, and stereo reconstruc-

tion (remark that the homographic curves generalize the iso-disparity curves presented in [11]).

References

- [1] C. S. Andersen and H. I. Christensen. Using multiple cues for controlling an agile camera head. In *In IAPR Workshop on Visual Behaviors*, pages 97–101. IAPR, IEEE Press, 1994.
- [2] R. Bajcsy. Active perception. In *Proceeding of IEEE*, 1988.
- [3] J. P. Barreto, J. Batista, and H. Araujo. Model predictive control to improve visual control of motion: Applications in active tracking of moving targets. In *IEEE Int. Conf. on Pattern Recognition*, Barcelona, September 2000.
- [4] R. Bolles, H. Baker, and D. H. Marimont. Epipolar-plane image analysis: An approach to determining structure from motion. *International Journal of Computer Vision*, 1, 1987.
- [5] P. Cavanagh and G. A. Alvarez. Tracking multiple targets with multifocal attention. *Trends in Cognitive Sciences*, 9(7):349–354, July 2005.
- [6] D. Coombs and C. Brown. Real time binocular smooth pursuit. *Int. Journal on Computer Vision*, 1993.
- [7] R. Gupta and R. Hartley. Linear pushbroom cameras. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 19(9):963–975, 1997.
- [8] R. Hartley and A. Zisserman. *Multiple View Geometry in Computer Vision. 2nd edn.* Cambridge University Press, 2004.
- [9] R. Horn and C. Johnson. *Topics in Matrix Analysis.* Cambridge University Press, 1991.
- [10] D. W. Murray, I. D. Reid, and A. J. Davison. Steering without representation using active fixation. *Perception*, 26:1519–1528, 1997.
- [11] M. Pollefeys and S. Sinha. Iso-disparity surfaces for general stereo configurations. In *ECCV'2004 - Proc. of European Conf. in Computer Vision*, 2004.
- [12] S. D. Roy, S. Chaudhury, and S. Banerjee. Active recognition through next view planning: a survey. *Pattern Recognition*, 37(3), March 2004.
- [13] J. Semple and G. Kneebone. *Algebraic Projective Geometry.* Clarendon Press, 1998.
- [14] E. Sommerlade and I. Reid. Cooperative surveillance of multiple targets using mutual information. In *Proceedings of the ECCV Workshop on Multi-camera and Multi-modal Sensor Fusion Algorithms and Applications (M2SFA2)*, October 2008.
- [15] I. Y. Aloimonos and A. Bandyopadhyay. Active vision. In *Proc. of IEEE Int. Conf. in Computer Vision*, 1987.
- [16] T. O. Y. Satoh and K. Deguchi. Binocular motion tracking by gaze fixation control and 3d shape reconstruction. *Advanced Robotics*, 17(16):1057–1072, 2003.
- [17] A. Yeung and N. Barnes. Efficient active monocular fixation using the log-polar sensor. *Int. J. Intell. Syst. Technol. Appl.*, 1(1/2):157–173, 2005.