

## VISION-BASED DETECTION OF AUTONOMOUS VEHICLES FOR PURSUIT-EVASION GAMES<sup>1</sup>

René Vidal      Shankar Sastry

*Department of Electrical Engineering and Computer Sciences  
University of California at Berkeley, Berkeley CA 94720, USA*

**Abstract:** We present a vision-based algorithm for the detection of multiple autonomous vehicles for a pursuit-evasion game scenario. Our algorithm computes estimates of the pose of multiple moving evaders from visual information collected by multiple moving pursuers, without previous knowledge of the segmentation of the image measurements or the number of moving evaders. We evaluate our algorithm in pursuit-evasion games with unmanned ground and aerial vehicles.

**Keywords:** Visual motion, multi-target motion estimation, autonomous vehicles.

### 1. INTRODUCTION

In this paper we consider a pursuit-evasion game scenario (Hespanha *et al.*, 1999; Vidal *et al.*, 2001; Kim *et al.*, 2001) in which a team of Unmanned Aerial Vehicles (UAVs) and Unmanned Ground Vehicles (UGVs) acting as *pursuers* tries to capture *evaders* within a bounded but unknown environment (see Fig. 1). Each pursuer is equipped with a monocular camera that is used to detect position, orientation and velocity of the evaders.

The problem of estimating the 3D pose of a moving camera observing a single static object is well studied in the computer vision community (Faugeras and Luong, 2001; Hartley and Zisserman, 2000; Soatto *et al.*, 1996; Tomasi and Kanade, 1992). However, the case in which multiple moving cameras observe multiple moving objects is very recent and only partially understood.

(Han and Kanade, 2000) proposed an algorithm for reconstructing a scene containing multiple moving points, some of them static and the others moving linearly with constant speed. The algorithm assumes a moving orthographic camera and does not require previous segmentation of the points. The case of a perspective camera was



Fig. 1. Berkeley test-bed for pursuit-evasion games.

studied in (Shashua and Levin, 2001), also under the assumption that points move linearly and with constant speed.

(Costeira and Kanade, 1995) proposed an algorithm for estimating the motion of multiple moving objects relative to a static orthographic camera, based on discrete image measurements for each object. They use a factorization method based on the fact that, under orthographic projection, discrete image measurements lie on a low-dimensional linear variety. Unfortunately, under full perspective projection such a variety is non-

<sup>1</sup> Research supported by ONR grant N00014-00-1-0621

linear (Torr, 1998), hence factorization methods cannot be used. However, (Irani, 1999) showed that infinitesimal image measurements do lie on a low-dimensional linear variety. She used the so-called *subspace constraints* to obtain a multi-frame algorithm for the estimation of the optical flow of a moving camera observing a static scene. She did not however use those constraints for motion estimation or 3D reconstruction.

In this paper, we propose a unified geometric representation of the perspective case that includes all the previous special cases. We use the subspace constraints to estimate the motion of multiple moving objects as observed by multiple moving cameras. We do not assume prior segmentation of the image points, nor do we restrict the motion of the objects to be linear or constant. Also, we do not assume previous knowledge of the number of moving objects. We show that the problem can be completely solved up to a certain parametric family that depends on the initial configuration of the objects and the cameras.

## 2. NOTATION AND PROBLEM STATEMENT

The **image**  $\mathbf{x} = [x, y, 1]^T$  of a point  $q = [q_1, q_2, q_3]^T$  (in the camera frame), is assumed to satisfy the *perspective projection* equation:

$$\mathbf{x} = q/Z, \quad (1)$$

where  $Z = q_3 > 0$  encodes the (unknown positive) **depth** of the point  $q$  with respect to its image  $\mathbf{x}$ .

The **optical flow**  $\mathbf{u}$  at point  $q$  is defined as the velocity of  $\mathbf{x}$  in the image plane, *i.e.*,

$$[\mathbf{u}^T, 0]^T = \dot{\mathbf{x}}.$$

The **motion** of pursuers and evaders is modeled as elements of  $SE(3) = \{(R, T) | R \in SO(3), T \in \mathbb{R}^3\}$  (discrete case) and  $se(3) = \{(\widehat{\omega}, v) | \widehat{\omega} \in so(3), v \in \mathbb{R}^3\}$  (differential case), where  $SO(3)$  and  $so(3)$  are the space of rotation and skew-symmetric matrices in  $\mathbb{R}^{3 \times 3}$ , respectively.

**Problem Statement:** Let  $\mathbf{x}_j^i \in \mathbb{R}^3$  be the image of point  $q^i \in \mathbb{R}^3, i = 1, \dots, n$  in frame  $j = 0, \dots, m$ , with  $j = 0$  being the reference frame. Let  $\{\mathbf{u}_j^i\}$  be the optical flow of point  $\mathbf{x}_0^i$  between the reference frame and frame  $j = 1, \dots, m$ . Given the images  $\{\mathbf{x}_j^i\}$  and the flows  $\{\mathbf{u}_j^i\}$ , recover the number of moving objects, the object to which each point belongs to, the depth of the  $n$  points, the motion of each one of the objects and that of the observers.

To be consistent with the notation, we always use the superscript to enumerate the  $n$  different points and/or the object to which the point belongs to. We omit the superscript when we refer to a generic single point and/or object. The subscript is always used to enumerate the  $m$  different camera frames.

## 3. ESTIMATING THE RELATIVE POSE OF A SINGLE EVADER IN MULTIPLE VIEWS

Let us start with the simplest case of a single pursuer taking images of one evader. Let  $(R_e(t), T_e(t)) \in SE(3)$  and  $(R_p(t), T_p(t)) \in SE(3)$  be the poses of the evader and the pursuer at time  $t$  with respect to a fixed reference frame. Let  $Q$  be a point located on the evader with coordinates  $q \in \mathbb{R}^3$  relative to the evader frame. The coordinates of  $Q$  relative to the reference frame are:

$$q_e(t) = R_e(t)q + T_e(t)$$

and the coordinates of  $Q$  relative to the pursuer frame are:

$$q_{ep}(t) = R_p^T(t)R_e(t)q + R_p^T(t)(T_e(t) - T_p(t)). \quad (2)$$

### 3.1 Differential Case

Differentiating (2) yields:

$$\dot{q}_{ep} = (\dot{R}_p^T R_e + R_p^T \dot{R}_e)q + \dot{R}_p^T (T_e - T_p) + R_p^T (\dot{T}_e - \dot{T}_p). \quad (3)$$

Combining (2) and (3) gives:

$$\dot{q}_{ep} = (\dot{R}_p^T R_p + R_p^T \dot{R}_e R_e^T R_p)q_{ep} + R_p^T (\dot{T}_e - \dot{T}_p - \dot{R}_e R_e^T (T_e - T_p)). \quad (4)$$

Since  $\dot{R}R^T \in so(3)$ ,  $\widehat{R^T \dot{\omega}} = R^T \widehat{\dot{\omega}} R$  and  $\dot{R}^T R = -R^T \dot{R} R^T$  (Murray *et al.*, 1994), we may define the angular velocities  $\omega_p, \omega_e \in \mathbb{R}^3$  by:

$$\widehat{\omega}_e = \dot{R}_e R_e^T \quad \text{and} \quad \widehat{\omega}_p = \dot{R}_p R_p^T. \quad (5)$$

Combining (4) and (5) yields:

$$\begin{aligned} \dot{q}_{ep} &= [R_p^T (\omega_e - \omega_p)] \times q_{ep} + R_p^T (\dot{T}_e - \dot{T}_p - \widehat{\omega}_e (T_e - T_p)) \\ &= \widehat{\omega} q_{ep} + v, \end{aligned}$$

where  $\omega$  and  $v$  are the angular and translational velocities of the evader relative to the pursuer.

Under perspective projection, the optical flow  $\mathbf{u}$  of point  $Q$  is then given by:

$$\begin{aligned} \mathbf{u} &= \frac{d}{dt} \left( \frac{q_{ep}}{Z} \right) = \frac{1}{Z} \begin{bmatrix} 1 & 0 & -x \\ 0 & 1 & -y \end{bmatrix} \dot{q}_{ep} \\ &= \begin{bmatrix} -xy & 1+x^2 & -y & 1/Z & 0 & -x/Z \\ -(1+y^2) & xy & x & 0 & 1/Z & -y/Z \end{bmatrix} \begin{bmatrix} \omega \\ v \end{bmatrix} \end{aligned}$$

where  $q_{ep} = (X, Y, Z)^T$  and  $(x, y, 1)^T = q_{ep}/Z$ .

Assume that the pursuer has measurements for the optical flow  $\mathbf{u}_j^i = (u_j^i, v_j^i)^T$  of point  $\mathbf{x}_0^i, i = 1 \dots n$  between frames  $j = 0$  and  $j = 1 \dots m$ , and define the matrix of *rotational flows*  $\Psi$  and the matrix of *translational flows*  $\Phi$  as:

$$\begin{aligned} \Psi &= \begin{bmatrix} -\{xy\} & \{1+x^2\} & -\{y\} \\ -\{1+y^2\} & \{xy\} & \{x\} \end{bmatrix} \in \mathbb{R}^{2n \times 3}, \\ \Phi &= \begin{bmatrix} \{1/Z\} & 0 & -\{x/Z\} \\ 0 & \{1/Z\} & -\{y/Z\} \end{bmatrix} \in \mathbb{R}^{2n \times 3}. \end{aligned}$$

where (for example)  $\{xy\}^T = [x^1 y^1, \dots, x^n y^n]$ .

Also let

$$U = \begin{bmatrix} \mathbf{u}_1^1 & \cdots & \mathbf{u}_m^1 \\ \vdots & & \vdots \\ \mathbf{u}_1^n & \cdots & \mathbf{u}_m^n \end{bmatrix} \quad \text{and} \quad V = \begin{bmatrix} \mathbf{v}_1^1 & \cdots & \mathbf{v}_m^1 \\ \vdots & & \vdots \\ \mathbf{v}_1^n & \cdots & \mathbf{v}_m^n \end{bmatrix}.$$

Then, the *optical flow* matrix  $W \in \mathbb{R}^{2n \times m}$  satisfies:

$$W = \begin{bmatrix} U \\ V \end{bmatrix} = [\Psi \ \Phi]_{2n \times 6} \begin{bmatrix} \omega_1 & \cdots & \omega_m \\ v_1 & \cdots & v_m \end{bmatrix}_{6 \times m} = SM^T$$

where  $\omega_j$  and  $v_j$  are the velocities of the evader relative to the pursuer in the  $j^{\text{th}}$  frame.

We call  $S \in \mathbb{R}^{2n \times 6}$  the *structure* matrix and  $M \in \mathbb{R}^{m \times 6}$  the *motion* matrix. We conclude that, for general translation and rotation, the optical flow matrix  $W$  has rank 6.

We would like to estimate the relative velocities  $(\omega_j, v_j)$  and depth  $Z^i$  from image points  $\mathbf{x}_0^i$  and optical flows  $\mathbf{u}_j^i$ . We can do so by factorizing  $W$  into its motion and structure components. For, consider the SVD of  $W = U\Sigma V^T$  and let  $\tilde{S} = U$  and  $\tilde{M} = V\Sigma$ . Then we have  $S = \tilde{S}A$  and  $M = \tilde{M}A^{-T}$  for some  $A \in \mathbb{R}^{6 \times 6}$ . Let  $A_k$  be the  $k$ -th column of  $A$ . Then the columns of  $A$  must satisfy:

$$\tilde{S}A_{1-3} = \Psi \quad \text{and} \quad \tilde{S}A_{4-6} = \Phi.$$

Since  $\Psi$  is known,  $A_{1-3}$  can be immediately computed. The remaining columns of  $A$  and the vector of the depths  $\{1/Z\}$  can be obtained up to scale from:

$$\begin{bmatrix} -I & \tilde{S}_u & 0 & 0 \\ -I & 0 & \tilde{S}_v & 0 \\ \text{diag}(\{x\}) & 0 & 0 & \tilde{S}_u \\ \text{diag}(\{y\}) & 0 & 0 & \tilde{S}_v \\ 0 & \tilde{S}_v & 0 & 0 \\ 0 & 0 & \tilde{S}_u & 0 \end{bmatrix} \begin{bmatrix} \{1/Z\} \\ A_4 \\ A_5 \\ A_6 \end{bmatrix} = 0.$$

where  $\tilde{S}_u \in \mathbb{R}^{n \times 6}$  and  $\tilde{S}_v \in \mathbb{R}^{n \times 6}$  are the upper and lower part of  $\tilde{S}$ , respectively. To summarize, given the image points of at least 4 points on one evader, and the optical flow of those points in at least 6 frames, one can estimate depth  $Z$  of each point and the angular and translational velocities  $\omega$  and  $v$  of the evader, relative to the pursuer.

### 3.2 Discrete Case

We consider equation (2) at two time instants,  $t$  and  $t_0$  and eliminate  $q$  to obtain:

$$\begin{aligned} q_{ep}(t) &= R_p(t)^T R_e(t) R_e(t_0)^T R_p(t_0) q_{ep}(t_0) + \\ &\quad R_p(t)^T (T_e(t) - T_p(t)) - \\ &\quad R_p(t)^T R_e(t) R_e(t_0)^T (T_e(t_0) - T_p(t_0)) \\ &= R(t, t_0) q_{ep}(t_0) + T(t, t_0) \end{aligned}$$

where  $(R(t, t_0), T(t, t_0))$  can be interpreted as the change in the relative pose of the evader with respect to the pursuer between times  $t_0$  and  $t$ .

There are a number of methods to estimate  $(R, T)$  from image measurements. Here we choose a simple linear method based on the rank deficiency of the multiple view matrix (Ma *et al.*, 2001), because it exploits the fact that the depth vector is known.

Assume that we take measurements at discrete time instants  $t = t_1 \dots t_m$ , and let  $R_j = R(t_j, t_0)$ ,  $T_j = T(t_j, t_0)$  and  $q_j = q_{ep}(t_j)$ . Then we have:

$$\begin{aligned} q_j^i &= R_j q_0^i + T_j \\ Z_j^i \mathbf{x}_j^i &= Z^i R_j \mathbf{x}_0^i + T_j \\ 0 &= Z^i \widehat{\mathbf{x}}_j^i R_j \mathbf{x}_0^i + \widehat{\mathbf{x}}_j^i T_j. \end{aligned}$$

Solving for  $(R_j, T_j)$  is equivalent to finding vectors  $\mathbf{R}_j = [r_{11}, r_{12}, r_{13}, r_{21}, r_{22}, r_{23}, r_{31}, r_{32}, r_{33}]^T \in \mathbb{R}^9$  and  $\mathbf{T}_j = T_j \in \mathbb{R}^3$ ,  $j = 1, \dots, m$ , such that:

$$P_j \begin{bmatrix} \mathbf{R}_j \\ \mathbf{T}_j \end{bmatrix} = \begin{bmatrix} Z^1 \widehat{\mathbf{x}}_j^1 * \mathbf{x}_0^{1T} & \widehat{\mathbf{x}}_j^1 \\ Z^2 \widehat{\mathbf{x}}_j^2 * \mathbf{x}_0^{2T} & \widehat{\mathbf{x}}_j^2 \\ \vdots & \vdots \\ Z^n \widehat{\mathbf{x}}_j^n * \mathbf{x}_0^{nT} & \widehat{\mathbf{x}}_j^n \end{bmatrix} \begin{bmatrix} \mathbf{R}_j \\ \mathbf{T}_j \end{bmatrix} = 0 \in \mathbb{R}^{3n}, \quad (6)$$

where  $A * B$  is the *Kronecker product* of  $A$  and  $B$ .

It can be shown that  $P_j$  is of rank 11 if more than  $n \geq 6$  points in general position are given. In that case, the kernel of  $P_j$  is unique, and so is  $(R_j, T_j)$ . However, in the presence of noise,  $R_j$  may not be an element of  $SO(3)$ . In order to obtain an element of  $SO(3)$  we proceed as follows: Let  $\tilde{R}_j \in \mathbb{R}^{3 \times 3}$  and  $\tilde{T}_j \in \mathbb{R}^3$  be the (unique) solution of (6). Such a solution is obtained as the eigenvector of  $P_j$  associated to the smallest eigenvalue. Let  $\tilde{R}_j = U_j \Sigma_j V_j^T$  be the SVD of  $\tilde{R}_j$ . Then the solution of (6) in  $SO(3) \times \mathbb{R}^3$  is:

$$R_j = \text{sign}(\det(U_j V_j^T)) U_j V_j^T \in SO(3) \quad (7)$$

$$T_j = \frac{\text{sign}(\det(U_j V_j^T))}{\sqrt[3]{\det(\Sigma_j)}} \tilde{T}_j \in \mathbb{R}^3. \quad (8)$$

## 4. ESTIMATING THE RELATIVE POSE OF MULTIPLE EVADERS IN MULTIPLE VIEWS

So far, we have assumed that the scene contains a single moving evader. Now, we consider the case in which a single pursuer observes  $n_e$  evaders. The new optical flow matrix  $W$  will contain additional rows corresponding to measurements from the different evaders. However, we cannot directly apply the equations in the previous section to solve for the relative motion of each evader, because we do not know which measurements in  $W$  correspond to which evader. We therefore need to consider the segmentation problem first, *i.e.*, the problem of separating all the measurements into  $n_e$  classes:

$$\mathcal{I}^k = \{i \in \{1 \dots n\} \mid \forall j \in \{1 \dots m\} \mathbf{x}_j^i \in \text{evader } k\}.$$

Furthermore, we assume that  $n_e$  itself is unknown.

Assume that each pursuer tracks  $n^k$  image points for evader  $k$  and let  $n = \sum n^k$  be the total number of points tracked. If the segmentation of these points were known, then the multi-body optical flow matrix could be written as:

$$\begin{aligned} W &= \begin{bmatrix} U_1 \\ \vdots \\ U_{n_e} \\ V_1 \\ \vdots \\ V_{n_e} \end{bmatrix} = \begin{bmatrix} \tilde{S}_{u1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \tilde{S}_{un_e} \\ \tilde{S}_{v1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \tilde{S}_{vn_e} \end{bmatrix} \begin{bmatrix} \tilde{M}_1^T \\ \vdots \\ \tilde{M}_{n_e}^T \end{bmatrix} = \tilde{S} \tilde{M}^T \\ &= \tilde{S} \begin{bmatrix} A_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A_{n_e} \end{bmatrix} \begin{bmatrix} A_1^{-1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & A_{n_e}^{-1} \end{bmatrix} \tilde{M}^T \\ &= \tilde{S} A A^{-1} \tilde{M}^T = S M^T. \end{aligned}$$

where  $\tilde{S}_{uk}$  and  $\tilde{S}_{vk} \in \mathbb{R}^{n^k \times 6}$ ,  $k = 1 \dots n_e$ ,  $\tilde{S}$  and  $S \in \mathbb{R}^{2n \times 6n_e}$ ,  $A \in \mathbb{R}^{6n_e \times 6n_e}$  and  $\tilde{M}, M \in \mathbb{R}^{m \times 6n_e}$ . In reality, the segmentation is unknown and the rows of  $W$  may be in a different order. However, such a permutation does not affect the rank of  $W$ . We conclude that

$$\boxed{n_e = \text{rank}(W)/6.} \quad (9)$$

Furthermore, the permutation will affect the rows of  $\tilde{S}$  hence those of  $S$ , but  $A$ ,  $\tilde{M}$  and  $M$  are unaffected. Therefore, from the SVD of  $W = \mathcal{U}\Sigma V$  we have  $\mathcal{U}\mathcal{U}^T =$

$$\begin{bmatrix} P \begin{bmatrix} \tilde{S}_{u1} \tilde{S}_{u1}^T & 0 \\ \vdots & \vdots \\ 0 & \tilde{S}_{un_e} \tilde{S}_{un_e}^T \end{bmatrix} P^T & P \begin{bmatrix} \tilde{S}_{u1} \tilde{S}_{v1}^T & 0 \\ \vdots & \vdots \\ 0 & \tilde{S}_{un_e} \tilde{S}_{vn_e}^T \end{bmatrix} P^T \\ P \begin{bmatrix} \tilde{S}_{v1} \tilde{S}_{u1}^T & 0 \\ \vdots & \vdots \\ 0 & \tilde{S}_{vn_e} \tilde{S}_{un_e}^T \end{bmatrix} P^T & P \begin{bmatrix} \tilde{S}_{v1} \tilde{S}_{v1}^T & 0 \\ \vdots & \vdots \\ 0 & \tilde{S}_{vn_e} \tilde{S}_{vn_e}^T \end{bmatrix} P^T \end{bmatrix},$$

where  $P \in \mathbb{R}^{n \times n}$  is the unknown permutation.

We define the *segmentation matrix*  $\Sigma$  as the sum of the diagonal blocks of  $\mathcal{U}\mathcal{U}^T$ , i.e.,

$$\Sigma = P \begin{bmatrix} \tilde{S}_{u1} \tilde{S}_{u1}^T + \tilde{S}_{v1} \tilde{S}_{v1}^T & 0 \\ \vdots & \vdots \\ 0 & \tilde{S}_{un_e} \tilde{S}_{un_e}^T + \tilde{S}_{vn_e} \tilde{S}_{vn_e}^T \end{bmatrix} P^T.$$

Then,  $\Sigma_{ij} > 0$  if and only if image points  $i$  and  $j$  belong to the same evader. Therefore the matrix  $\Sigma$  can be used to determine the class to which each image point belongs to.

Once the segmentation problem has been solved, one can apply the algorithms in Section 3 to estimate the motion of each object separately.

## 5. ABSOLUTE POSE ESTIMATION

So far, we have shown how to obtain the relative position, orientation and velocities of multiple moving evaders with respect to a single moving pursuer. However, for the pursuit-evasion scenario that we are considering (Vidal *et al.*, 2001), absolute position and orientation is needed. In this section, we show that it is not possible to solve the problem from image measurements only, unless some additional constraints are imposed. We consider additional constraints such as known position for the pursuers and concentration of the image points corresponding to each evader. We also consider the case in which multiple pursuers share their measurements.

### 5.1 Pose and velocity of the pursuers

In practice, not all the image measurements will correspond to the moving evaders. Some image points will correspond to 3D points in the background, which are actually not moving. Therefore, matrix  $W$  will be segmented into  $n_e + 1$  classes, one of them corresponding to static points. While points on each evader are concentrated in a specific region of the image, points on the background are distributed all over the image. Therefore, given the segmentation of the  $n_e + 1$  classes, the class of static points can be identified as the one with the largest spatial standard deviation in all frames.

Let the class of static points be the zero<sup>th</sup> class. Also, let  $(\omega_j^k, v_j^k)$  and  $(R_j^k, T_j^k)$  be the estimates of relative motion for class  $k$  in frame  $j$  as obtained by the algorithms in Sections 3 and 4. The zero<sup>th</sup> class contains information about the motion of the pursuer only. More explicitly, we have:

$$\omega_j^0 = -R_{pj}^T \omega_{pj} \quad \lambda^0 v_j^0 = -R_{pj}^T \dot{T}_{pj} \quad (10)$$

$$R_j^0 = R_{pj}^T R_{p0} \quad \lambda^0 T_j^0 = -R_{pj}^T (T_{pj} - T_{p0}) \quad (11)$$

where  $\lambda^0$  is the unknown scale lost under perspective projection. We are now interested in recovering the absolute motion  $(R_{p0}, T_{p0})$  and  $(\omega_{pj}, v_{pj})$ ,  $(R_{pj}, T_{pj})$ ,  $j = 1 \dots m$ . We can see from equations (10) and (11) that this cannot be done, because there are  $12m + 7$  unknowns and  $12m$  equations. Therefore, the absolute motion of the pursuer can be estimated up to a 7-parameter family, given (for example) by the initial rotation and translation of the pursuer and the scale lost under perspective projection.

For the case of a single pursuer, this ambiguity is not relevant, since it is equivalent to choosing the reference frame, which can be chosen to coincide with the initial location of the pursuer, i.e.,  $(R_{p0}, T_{p0}) = (I, 0)$ . For the case of multiple pursuers, we resolve this ambiguity by using additional information on the motion of the pursuers.

For example, we assume that the position of the pursuers  $T_{pj}, j = 0 \dots m$  is known from GPS measurements. Given that information, one can solve linearly for  $\lambda^0$  and  $R_{pj}^T, j = 0 \dots m$  from (11). Then, solving for  $(\omega_{pj}, \dot{T}_{pj}), j = 1 \dots m$  is trivial from (10).

## 5.2 Pose and velocities of the evaders

Given  $\lambda^0, (\omega_{pj}, \dot{T}_{pj}), j = 1 \dots m$  and  $(R_{pj}, T_{pj}), j = 0 \dots m$  we would like to solve for  $\lambda^k, (\omega_{ej}^k, \dot{T}_{ej}^k), j = 1 \dots m$  and  $(R_{ej}^k, T_{e0}^k), j = 0 \dots m, k = 1 \dots n_e$  from:

$$w_j^k = R_{pj}^T (\omega_{ej}^k - \omega_{pj}) \quad (12)$$

$$\lambda^k v_j^k = R_{pj}^T (\dot{T}_{ej}^k - \dot{T}_{pj} - \widehat{w}_{ej}^k (T_{ej}^k - T_{pj})) \quad (13)$$

$$R_{ej}^k = R_{pj}^T R_{ej}^k R_{p0}^T \quad (14)$$

$$\lambda^k T_j^k = R_{pj}^T (T_{ej}^k - T_{pj} - R_{ej}^k R_{e0}^k (T_{e0}^k - T_{p0})). \quad (15)$$

Again, we observe that the motion of the evaders can be recovered up to a  $7n_e$ -parameter family given by the initial pose of each evader and the unknown scales lost under perspective projection.

In order to resolve the translation ambiguity  $T_{e0}^k$ , we assume that image points corresponding to each evader are concentrated in a specific region of the image. Therefore, the average of the 3D points associated to those image points well approximates the position of the evader relative to the pursuer (up to scale). We then approximate the initial position of each evader as:

$$T_{e0}^k \approx \lambda^k R_{p0} \sum_{i \in \mathcal{I}^k} \frac{\mathbf{x}_0^i Z_0^i}{n^k} + T_{p0} \quad (16)$$

Combining (14), (15) and (16), the position of the evaders in the remaining frames is given by:

$$T_{ej}^k \approx \lambda^k R_{pj} \left( T_j^k + R_{ej}^k \sum_{i \in \mathcal{I}^k} \frac{\mathbf{x}_0^i Z_0^i}{n^k} \right) + T_{pj} \quad (17)$$

In relation to the rotation ambiguity, we observe that it is not possible to estimate  $R_{e0}^k$ . One can only estimate  $R_{ej}^k R_{e0}^k$  which is the orientation of the evader relative to its initial configuration.

Finally, assuming that the initial orientation of the evaders is known,  $(\omega_{ej}^k, \dot{T}_{ej}^k)$  can be trivially obtained from (12) and (13). We conclude that, given the assumptions, the motion of each evader can be completely solved with  $(R_{ej}^k, \omega_{ej}^k)$  obtained uniquely, and  $(T_{ej}^k, \dot{T}_{ej}^k)$  obtained up to a scale  $\lambda^k$ .

In order to determine the unknown scales, some additional information is needed. In (Vidal *et al.*, 2001) we assumed a flat terrain with known ground level. Here, we show how this can be done when two pursuers observe the same evader, which can be detected when the estimates of the two pursuers for  $(R_{ej}, \omega_{ej})$  match. Since  $T_{e0}^k$  in (16) must be the same for both pursuers, one can solve for the 2 unknown scales from those 3 equations.

## 6. SIMULATIONS AND EXPERIMENTS

We evaluate the performance of our motion algorithm in a simulated game with one aerial pursuer and one ground evader. The pursuer observes and tracks 12 static points plus 12 points on the moving evader for 105 seconds. Random noise with 1 pixel standard deviation is added to the image measurements. Figure 2 shows the motion esti-

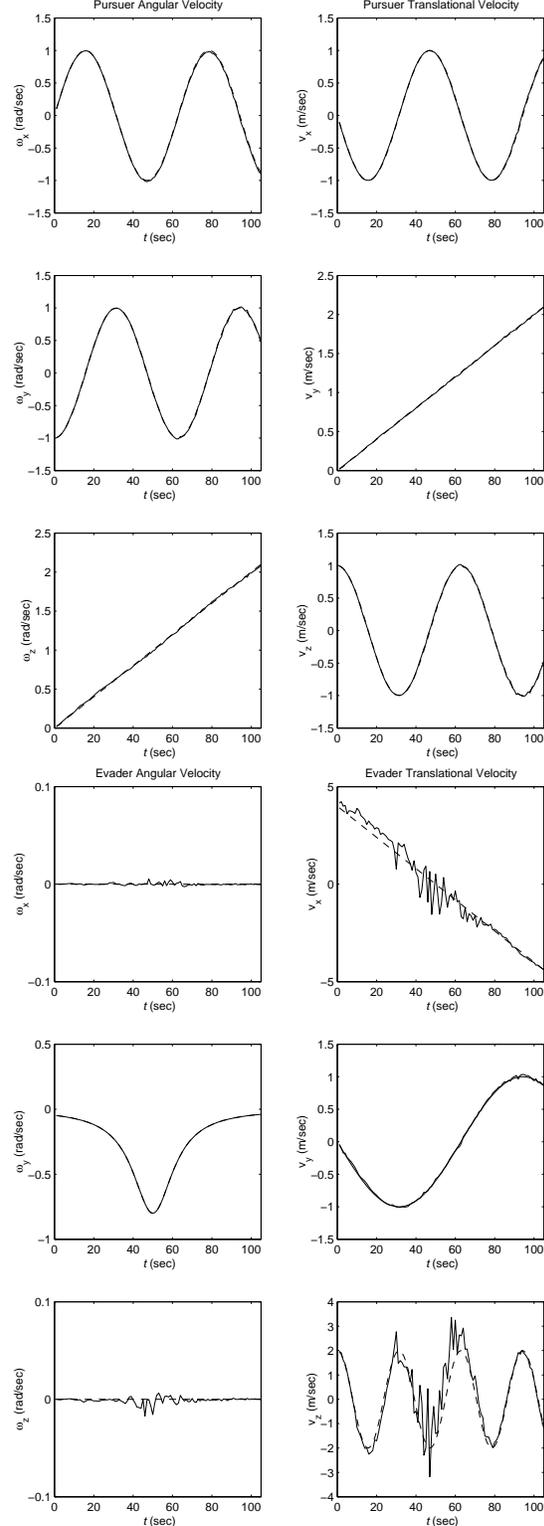


Fig. 2. Pursuer and evader motion estimates.

mates for a particular game. The dashed trajectories correspond to the ground truth while the solid trajectories correspond to the estimates given by our vision-based algorithm. We observe that there is no noticeable difference between the motion estimates for the pursuer and the ground truth. In relation to the motion of the evader, there is a maximum error of 0.017 rad/sec for angular velocity and 2.384 m/sec for translational velocity. This latter error corresponds to high frequency components in the estimates and can be easily eliminated with a low pass filter.

We evaluate the performance of the segmentation part of the algorithm in a real video sequence with one pursuer and two evaders. Figure 3(a) shows one frame of the sequence with the corresponding optical flow superimposed. Optical flow is computed using Black's algorithm (Black, 1996). Each frame of the sequence has  $200 \times 150$  pixels. Their optical flows are used to build the segmentation matrix and find the independent motions. Figures 3(b)-(c) show the results of the segmentation. Groups 1 and 2 correspond to the each one of the moving evaders, while group 3 corresponds to the background, which is the correct segmentation.

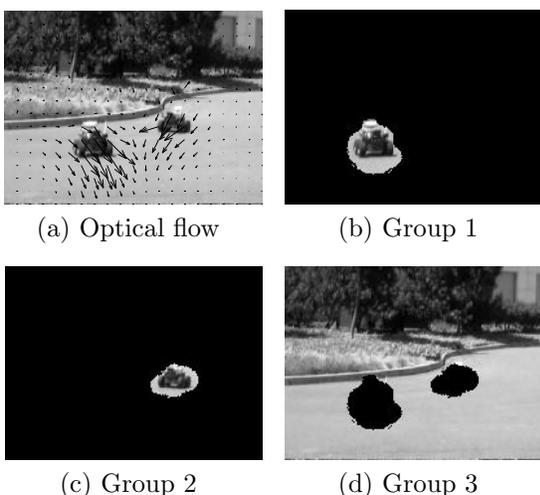


Fig. 3. Segmentation results for a pursuit-evasion game between 1 pursuer and 2 evaders.

## 7. CONCLUSIONS

We proposed a vision-based algorithm for detecting multiple moving evaders as observed by multiple moving pursuers in multiple frames. The proposed algorithm does not assume prior segmentation of the image points or previous knowledge of the number of evaders. Through simulations and experiments we showed the applicability of our algorithm to the control of multiple autonomous vehicles for a pursuit-evasion game scenario. Future work will include real-time implementation of the algorithm and experimental results on the Berkeley fleet of autonomous vehicles.

## 8. REFERENCES

- Black, M. (1996). <http://www.cs.brown.edu/people/black/ignc.html>.
- Costeira, J. and T. Kanade (1995). Multi-body factorization methods for motion analysis. In: *IEEE International Conference on Computer Vision*. pp. 1071–1076.
- Faugeras, O. and Q.-T. Luong (2001). *Geometry of Multiple Images*. The MIT Press.
- Han, M. and T. Kanade (2000). Reconstruction of a scene with multiple linearly moving objects. In: *International Conference on Computer Vision and Pattern Recognition*. Vol. 2. pp. 542–549.
- Hartley, R. and A. Zisserman (2000). *Multiple View Geometry in Computer Vision*. Cambridge.
- Hespanha, J., H.J. Kim and S. Sastry (1999). Multiple-agent probabilistic pursuit-evasion games. In: *IEEE Conference on Decision and Control*. pp. 2432–2437.
- Irani, M. (1999). Multi-frame optical flow estimation using subspace constraints. In: *IEEE International Conference on Computer Vision*. pp. 626–633.
- Kim, H.J., R. Vidal, D. Shim, O. Shakernia and S. Sastry (2001). A hierarchical approach to probabilistic pursuit-evasion games with unmanned ground and aerial vehicles. In: *IEEE Conference on Decision and Control*. pp. 1243–1248.
- Ma, Y., R. Vidal, K. Huang and S. Sastry (2001). New rank deficiency condition for multiple view geometry of point features. *UIUC, CSL Tech. Report, UILU-ENG 01-2208 (DC-200)*.
- Murray, R. M., Z. Li and S. S. Sastry (1994). *A Mathematical Introduction to Robotic Manipulation*. CRC press Inc.
- Shashua, A. and A. Levin (2001). Multi-frame infinitesimal motion model for the reconstruction of (dynamic) scenes with multiple linearly moving objects. In: *IEEE International Conference on Computer Vision*. Vol. 2. pp. 592–599.
- Soatto, S., R. Frezza and P. Perona (1996). Motion estimation via dynamic vision. *IEEE Transactions on Automatic Control* **41**(3), 393–413.
- Tomasi, C. and T. Kanade (1992). Shape and motion from image streams under orthography. *International Journal of Computer Vision* **9**(2), 137–154.
- Torr, P. H. S. (1998). Geometric motion segmentation and model selection. *Phil. Trans. Royal Society of London A* **356**(1740), 1321–1340.
- Vidal, R., S. Rashid, C. Sharp, O. Shakernia, H.J. Kim and S. Sastry (2001). Pursuit-evasion games with unmanned ground and aerial vehicles. In: *IEEE International Conference on Robotics and Automation*. pp. 2948–2955.