Egomotion Estimation Using Quadruples of Collinear Image Points*

Manolis I.A. Lourakis

INRIA Sophia-Antipolis, 2004 route des Lucioles
BP 93 06902 Sophia-Antipolis Cedex, FRANCE
mlouraki@sophia.inria.fr
http://www-sop.inria.fr/robotvis/personnel/Manolis.Lourakis/

Abstract. This paper considers a fundamental problem in visual motion perception, namely the problem of egomotion estimation based on visual input. Many of the existing techniques for solving this problem rely on restrictive assumptions regarding the observer's motion or even the scene structure. Moreover, they often resort to searching the high dimensional space of possible solutions, a strategy which might be inefficient in terms of computational complexity and exhibit convergence problems if the search is initiated far away from the correct solution. In this work, a novel linear constraint that involves quantities that depend on the egomotion parameters is developed. The constraint is defined in terms of the optical flow vectors pertaining to four collinear image points and is applicable regardless of the egomotion or the scene structure. In addition, it is exact in the sense that no approximations are made for deriving it. Combined with robust linear regression techniques, the constraint enables the recovery of the FOE, thereby decoupling the 3D motion parameters. Extensive simulations as well as experiments with real optical flow fields provide evidence regarding the performance of the proposed method under varying noise levels and camera motions.

1 Introduction

Knowledge of the velocity of a mobile system with respect to its environment is essential for various servoing tasks that are based on visual feedback, e.g. collision avoidance, docking, image stabilization, etc. Given a sequence of images acquired by a monocular observer pursuing unrestricted rigid motion, the problem of egomotion estimation can be defined as the problem of recovering the linear and angular velocities comprising the motion of the observer. Although simply stated, the problem of estimating egomotion using visual input is particularly difficult. This difficulty primarily stems from the fact that the only information available from images is related to the observed 2D motion of image points, which depends both on the sought egomotion and the unknown 3D structure of the viewed scene.

* This work has been carried out while the author was with the Computer Science Dept, Univ. of Crete and the Inst. of Computer Science, FORTH, Heraklion, Crete, Greece. Funding was partially supplied by the VIRGO research network of the TMR Programme (EC Contract No ERBFMRX-CT96-0049).

D. Vernon (Ed.): ECCV 2000, LNCS 1843, pp. 834–848, 2000.

Since the dependence of the 2D image motion on the scene structure is nonlinear, small errors in the estimates of 2D motion can have a significant impact on the accuracy of the recovered 3D motion [4]. In addition, the confounding of translation and rotation makes the problem of estimating unrestricted egomotion much harder compared to the problem of estimating pure translation or rotation [4]. Due to its importance, many algorithms dealing with the problem of estimating egomotion have appeared in the literature. The following paragraphs provide a short review of a few representative methods; more detailed discussions can be found in [7,8,10]. Most of the methods reviewed here rely on the availability of a dense optical flow field to describe 2D motion. Prazdny [19], for example, assumes that surfaces in the viewed scene are smooth and recovers rotation through numerical optimization techniques using a set of nonlinear equations that are independent of translation. Prazdny [20] and later Burger and Bhanu [2] also suggested solving for rotation first and employed a search in the space of rotational parameters. For each hypothesized rotation, the corresponding rotational field was subtracted from the optical flow and the remaining field was tested for conformance to a purely translational flow field. Bruss and Horn [1] combine information from the whole visual field to determine the 3D motion that is the best least squares fit to the observed velocity field. They developed three different algorithms, the first two of which give closed form solutions for translation and rotation when the motion is purely translational or rotational respectively. The third algorithm applies to the case of general motion and estimates translation by minimizing an appropriate residual function using iterative numerical procedures. Reiger and Lawton [21] solve for translation by exploiting the phenomenon of *motion parallax*. By subtracting the optical flow vectors at two image locations whose corresponding 3D points have sufficiently different depths, a flow vector that is approximately pointing towards the FOE[1] is obtained. The main drawback of this approach stems from the fact that most optical flow algorithms cannot give accurate estimates of optical flow in areas with large depth variations. Recently, Irani et al [9] alleviated some of the difficulties related to the estimation of motion parallax by decomposing image motion into the sum of the motion of a planar surface and a residual planar parallax field that is purely translational.

Heeger and Jepson [7] also make use of the residual function introduced in [1] and propose an efficient search technique for locating its minimum. Hummel and Sundareswaran [8] present an algorithm for finding the rotational motion and one for locating the FOE. The first algorithm is based on the observation that the curl of the optical flow field is approximately a linear function whose coefficients are proportional to the desired rotational parameters of motion. The algorithm for locating the FOE extends the work of Heeger and Jepson [7] by considering for each candidate FOE the projection of the optical flow along vectors emanating from the former. Da Vitoria Lobo and Tsotsos [10] develop a constraint (the *Collinear Point Constraint - CPC*) involving flow projections at three collinear image points, which provides a means for canceling rotation and at the same time

[1] The FOE gives the direction of translation and is defined more rigorously in the following.

constraining the FOE to lie on the line defined by the collinear points. The CPC is discussed in more detail in Section 3. Optical flow projections are also used in [13] and the FOE is recovered through their pairwise differences. Daniilidis [3] employs fixation on a scene point to reduce the number of motion parameters to be estimated from five to four. The associated spherical motion field is projected on two latitudinal directions and the motion parameters are then found by two one-dimensional searches along meridians of the image sphere.

In this paper, it is assumed that either the viewed scene is static or the independently moving objects have been identified and masked out [14]. The motivation behind our egomotion estimation method is twofold. First, we are interested in estimating egomotion by means of linear constraints. Second, we want to avoid making any restrictive assumptions regarding the egomotion or the scene structure. Hence, we have developed a novel linear constraint regarding the motion parameters, defined in terms of four collinear image points. The constraint is applicable regardless of the egomotion or the scene structure and combined with robust linear regression techniques, permits the recovery of the direction of translation, thereby decoupling the 3D motion parameters. The rest of this paper is organized as follows. Section 2 presents an overview of some preliminary results that are essential for the development of the proposed method. Section 3 develops the proposed constraint and shows how it can be employed to recover egomotion. Experimental results from an implementation of the method are presented in Section 4. The paper is concluded with a brief discussion in Section 5. A more detailed version can be found in [12].

2 Visual Motion Representation

Before proceeding with the description of the proposed method, issues related to motion representation are discussed. Consider a coordinate system $OXYZ$ positioned at the optical center (nodal point) of a pinhole camera, such that the OZ axis coincides with the optical axis. Suppose that the camera is moving rigidly with respect to its 3D static environment with translational motion (U, V, W) and rotational motion (α, β, γ). Under perspective projection, the 3D point $P(X, Y, Z)$ projects to image point $p(x, y)$ which moves on the image plane with velocity (u, v), given by [11]:

$$\begin{aligned} u &= \frac{(-Uf + xW)}{Z} + \alpha\frac{xy}{f} - \beta\left(\frac{x^2}{f} + f\right) + \gamma y \\ v &= \frac{(-Vf + yW)}{Z} + \alpha\left(\frac{y^2}{f} + f\right) - \beta\frac{xy}{f} - \gamma x \end{aligned} \tag{1}$$

Equations (1) describe the optical flow field, which relates the 3D motion of points to their projected 2D motion on the image plane. The problem of estimating the optical flow from an image sequence is fundamental to motion analysis. However, due to space limitations, it will not be discussed further here. An excellent introduction to the problem as well as a review of the state of the art can be found

in [18]. Several observations regarding Eqs. (1) can be made. First, the effect of translation on the observed 2D motion is independent from that of rotation, i.e. the translational and rotational components of motion are separable. Second, the rotational component of motion is independent of scene structure, since the depth Z influences the translational component only. Third, the vectors defined by the translational components of the motion field, lie on lines going through the point $(x_0, y_0) \equiv (Uf/W, Vf/W)$, which is known as the *Focus Of Expansion* (FOE). The FOE defines the direction of the translational motion, and is of central importance for several motion analysis problems. Finally, if the quantities W and Z are multiplied by the same scale factor, the flow defined by Eqs. (1) remains the same. In other words, there exists a scale ambiguity that prevents us from differentiating between a close object moving slowly and a distant one that is moving fast. Thus, the information related to the translational component of egomotion that can be recovered from Eqs. (1) is at most its direction, i.e. the FOE. The ratio $\frac{Z}{W}$ is often referred to as the *time-to-contact* [16].

3 Using Quadruples of Collinear Points to Constrain the FOE

In the following, it is assumed that the camera has been intrinsically calibrated, so that the retinal transformations among pixel and image coordinate systems are known [15]. Before proceeding to the description of the proposed method, we state two theorems which are essential for its derivation. The proofs, which are omitted due to space limitations, can be found in [12].

3.1 Two Precursory Theorems

Theorem 1 *Suppose that two image points* $\mathbf{p_1} = (x_1, y_1)$ *and* $\mathbf{p_2} = (x_2, y_2)$ *lie on a line that goes through the origin of the image coordinate system (i.e. the principal point). The difference of the projections of their corresponding optical flow vectors along the direction* $\mathbf{n} = (n_x, n_y)$ *that is normal to the line is equal to*

$$un_1 - un_2 = DW(\frac{1}{Z_1} - \frac{1}{Z_2}) + \frac{\gamma}{n_y}(x_2 - x_1), \tag{2}$$

where $un_i = u_i n_x + v_i n_y$, $i = 1,\ 2$ *and* $D = (x_1 - x_0)n_x + (y_1 - y_0)n_y$.

Theorem 2 *Let* $\mathbf{p_1} = (x_1, y_1)$, $\mathbf{p_2} = (x_2, y_2)$ *and* $\mathbf{p_3} = (x_3, y_3)$ *be three collinear image points lying on a line whose equation is* $y = \kappa x + \nu$. *Let also* (x_0, y_0) *be the FOE and assume that* $\mathbf{p_2}$ *divides the line segment* $\overrightarrow{\mathbf{p_1}\,\mathbf{p_3}}$ *in ratio* λ. *For the projections* $un_i, i = 1 \ldots 3$ *of the optical flow vectors at points* $\mathbf{p_1}, \mathbf{p_2}$ *and* $\mathbf{p_3}$ *along an arbitrary direction* (n_x, n_y), *the following holds*

$$un_2 - \frac{1}{1+\lambda}un_1 - \frac{\lambda}{1+\lambda}un_3 = D_2 W(\frac{1}{Z_2} - \frac{1}{1+\lambda}\frac{1}{Z_1} - \frac{\lambda}{1+\lambda}\frac{1}{Z_3}) + \frac{d_{21}}{1+\lambda}W(\frac{1}{Z_1} - \frac{1}{Z_3}) + \frac{\kappa d_{21}(x_2 - x_3)}{f}\alpha - \frac{d_{21}(x_2 - x_3)}{f}\beta \tag{3}$$

In the above equation, $D_2 = (x_2 - x_0)n_x + (y_2 - y_0)n_y$ *and* $d_{21} = (x_2 - x_1)n_x + (y_2 - y_1)n_y$.

By inspecting Eq. (3), it can easily be seen that in the case that the direction of projection (n_x, n_y) is perpendicular to the line defined by the points $\mathbf{p_i}$, the term d_{21} is zero, thus the sum of the rotational components vanishes. The remaining terms are identical to the expression for the Collinear Point Constraint (CPC) that was derived by Da Vitoria Lobo and Tsotsos in [10]. The CPC states that when an appropriate linear combination of the projections of optical flow vectors in the direction perpendicular to the line joining them is zero, there exist two possible situations. Either the three 3D points whose projections form the collinear triplet are also collinear in the scene (i.e. $\frac{1}{Z_2} - \frac{1}{1+\lambda}\frac{1}{Z_1} - \frac{\lambda}{1+\lambda}\frac{1}{Z_3} = 0$), or the line defined by the collinear triplet passes through the FOE (i.e. $D_2 = 0$). By employing a voting scheme to differentiate between these two cases, the CPC has been combined in [10] with exhaustive image based search for locating the FOE.

3.2 The Proposed Constraint on Egomotion

Assume now a mobile observer undergoing rigid motion in a static environment. Let $\mathbf{p_1} = (x_1, y_1)$, $\mathbf{p_2} = (x_2, y_2)$ and $\mathbf{p_3} = (x_3, y_3)$ be three collinear image points lying on a line $\mathcal{L}$ through the image principal point. Let also (n_x, n_y) be the direction normal to $\mathcal{L}$ and $(n_x^{'}, n_y^{'})$ and $(n_x^{''}, n_y^{''})$ two other directions that are not perpendicular to $\mathcal{L}$. According to Theorem 2, for the projections of the optical flow vectors along the direction $(n_x^{'}, n_y^{'})$ the following holds

$$un_2^{'} - \frac{1}{1+\lambda}un_1^{'} - \frac{\lambda}{1+\lambda}un_3^{'} = D_2^{'}W(\frac{1}{Z_2} - \frac{1}{1+\lambda}\frac{1}{Z_1} - \frac{\lambda}{1+\lambda}\frac{1}{Z_3}) + \frac{d_{21}^{'}}{1+\lambda}W(\frac{1}{Z_1} - \frac{1}{Z_3}) + (\kappa\alpha - \beta)\,\frac{d_{21}^{'}(x_2 - x_3)}{f}, \tag{4}$$

where the primed terms are defined analogously to the unprimed ones in Eq. (3). Similarly, for the projections along the normal direction (n_x, n_y), Eq. (3) gives

$$un_2 - \frac{1}{1+\lambda}un_1 - \frac{\lambda}{1+\lambda}un_3 = D_2W(\frac{1}{Z_2} - \frac{1}{1+\lambda}\frac{1}{Z_1} - \frac{\lambda}{1+\lambda}\frac{1}{Z_3}) \tag{5}$$

Dividing Eq. (4) with Eq. (5) yields

$$\frac{un_2^{'} - \frac{1}{1+\lambda}un_1^{'} - \frac{\lambda}{1+\lambda}un_3^{'}}{un_2 - \frac{1}{1+\lambda}un_1 - \frac{\lambda}{1+\lambda}un_3} = \frac{D_2^{'}}{D_2} + \frac{d_{21}^{'}}{1+\lambda}\;\frac{\frac{1}{Z_1} - \frac{1}{Z_3}}{D_2(\frac{1}{Z_2} - \frac{1}{1+\lambda}\frac{1}{Z_1} - \frac{\lambda}{1+\lambda}\frac{1}{Z_3})} + (\kappa\alpha - \beta)\,\frac{d_{21}^{'}(x_2 - x_3)}{f}\;\frac{1}{un_2 - \frac{1}{1+\lambda}un_1 - \frac{\lambda}{1+\lambda}un_3} \tag{6}$$

Applying Eq. (2) for points $\mathbf{p_1}$ and $\mathbf{p_3}$ results in $un_1 - un_3 = D_2W(\frac{1}{Z_1} - \frac{1}{Z_3}) + \frac{\gamma}{n_y}(x_3 - x_1)$. Solving this equation for $\frac{1}{Z_1} - \frac{1}{Z_3}$, dividing in terms by Eq. (5) and

substituting the result into Eq. (6) yields

$$\frac{un_2^{'}-\frac{1}{1+\lambda}un_1^{'}-\frac{\lambda}{1+\lambda}un_3^{'}}{un_2-\frac{1}{1+\lambda}un_1-\frac{\lambda}{1+\lambda}un_3}\frac{1}{d_{21}^{'}}=\frac{D_2^{'}/d_{21}^{'}}{D_2}+\frac{1}{1+\lambda}\frac{un_1-un_3-\frac{x_3-x_1}{n_y}\gamma}{D_2(un_2-\frac{1}{1+\lambda}un_1-\frac{\lambda}{1+\lambda}un_3)}+(\kappa\alpha-\beta)\frac{(x_2-x_3)}{f}\frac{1}{un_2-\frac{1}{1+\lambda}un_1-\frac{\lambda}{1+\lambda}un_3} \quad (7)$$

Let now $\mathbf{p_4}=(x_4,y_4)$ be a fourth point collinear with the triplet $\mathbf{p_1},\mathbf{p_2}$ and $\mathbf{p_3}$ and such that point $\mathbf{p_2}$ divides the segment $\overrightarrow{\mathbf{p_1}\,\mathbf{p_4}}$ in ratio μ. Eq. (7) gives for the projections along the direction $(n_x^{''},n_y^{''})$

$$\frac{un_2^{''}-\frac{1}{1+\mu}un_1^{''}-\frac{\mu}{1+\mu}un_4^{''}}{un_2-\frac{1}{1+\mu}un_1-\frac{\mu}{1+\mu}un_4}\frac{1}{d_{21}^{''}}=\frac{D_2^{''}/d_{21}^{''}}{D_2}+\frac{1}{1+\mu}\frac{un_1-un_4-\frac{x_4-x_1}{n_y}\gamma}{D_2(un_2-\frac{1}{1+\mu}un_1-\frac{\mu}{1+\mu}un_4)}+(\kappa\alpha-\beta)\frac{(x_2-x_4)}{f}\frac{1}{un_2-\frac{1}{1+\mu}un_1-\frac{\mu}{1+\mu}un_4} \quad (8)$$

Subtracting Eq. (8) from Eq. (7) and noting that $\frac{x_1-x_3}{1+\lambda}=x_2-x_3$ and $\frac{x_1-x_4}{1+\mu}=x_2-x_4$, results in

$$\frac{un_2^{'}-\frac{1}{1+\lambda}un_1^{'}-\frac{\lambda}{1+\lambda}un_3^{'}}{un_2-\frac{1}{1+\lambda}un_1-\frac{\lambda}{1+\lambda}un_3}\frac{1}{d_{21}^{'}}-\frac{un_2^{''}-\frac{1}{1+\mu}un_1^{''}-\frac{\mu}{1+\mu}un_4^{''}}{un_2-\frac{1}{1+\mu}un_1-\frac{\mu}{1+\mu}un_4}\frac{1}{d_{21}^{''}}=\frac{D_2^{'}/d_{21}^{'}-D_2^{''}/d_{21}^{''}}{D_2}+\frac{1}{D_2}\Big(\frac{1}{1+\lambda}\frac{un_1-un_3}{un_2-\frac{1}{1+\lambda}un_1-\frac{\lambda}{1+\lambda}un_3}-\frac{1}{1+\mu}\frac{un_1-un_4}{un_2-\frac{1}{1+\mu}un_1-\frac{\mu}{1+\mu}un_4}\Big)+\Big(\frac{\gamma f}{D_2n_y}+\kappa\alpha-\beta\Big)\Big(\frac{x_2-x_3}{f(un_2-\frac{1}{1+\lambda}un_1-\frac{\lambda}{1+\lambda}un_3)}-\frac{x_2-x_4}{f(un_2-\frac{1}{1+\mu}un_1-\frac{\mu}{1+\mu}un_4)}\Big) \quad (9)$$

The term $\frac{D_2^{'}/d_{21}^{'}-D_2^{''}/d_{21}^{''}}{D_2}$ in Eq. (9) is independent of the FOE and can be computed using the point retinal coordinates only. Indeed, it can be shown that

$$\frac{D_2^{'}/d_{21}^{'}-D_2^{''}/d_{21}^{''}}{D_2}=\frac{(n_x^{''}n_y{}'-n_x^{'}n_y^{''})n_y}{(n_xn_y^{'}-n_x^{'}n_y)(n_xn_y^{''}-n_x^{''}n_y)(x_2-x_1)} \quad (10)$$

Equation (9) is independent of the scene depths and linear in the two unknowns $\frac{1}{D_2}$ and $\frac{\gamma f}{D_2n_y}+\kappa\alpha-\beta$, therefore forms the basis for the development of the proposed egomotion estimation method: Given a line $\mathcal{L}$ through the image principal point, Eq. (9) is employed for estimating the term $\frac{1}{D_2}$ corresponding to $\mathcal{L}$. In theory, two quadruples of image points lying on $\mathcal{L}$ suffice to provide estimates of the unknown parameters $\frac{1}{D_2}$ and $\frac{\gamma f}{D_2n_y}+\kappa\alpha-\beta$. However, to enhance noise immunity, multiple quadruples of points on $\mathcal{L}$ are selected at random and robust estimates of the two unknowns are computed using the LMedS robust estimator [22]. Knowledge of the term $D_2^{\mathcal{L}}$ for a line $\mathcal{L}$ provides one constraint on the location of the FOE, namely

$$x_0n_x^{\mathcal{L}}+y_0n_y^{\mathcal{L}}=x^{\mathcal{L}}n_x^{\mathcal{L}}+y^{\mathcal{L}}n_y^{\mathcal{L}}-D_2^{\mathcal{L}}, \quad (11)$$

where (x_0, y_0) is the sought FOE, $(n_x^{\mathcal{L}}, n_y^{\mathcal{L}})$ is the unit normal for line $\mathcal{L}$ and $(x^{\mathcal{L}}, y^{\mathcal{L}})$ is a point on $\mathcal{L}$. Noting that each line $\mathcal{L}$ through the image principal point supplies one constraint of the form of Eq. (11) regarding the FOE, the constraints arising from multiple such lines can be combined to yield the FOE. More specifically, using many lines through the image principal point, robust estimates of the corresponding distances $\frac{1}{D_2^{\mathcal{L}}}$ are obtained as previously outlined. For each of the obtained distance estimates, Eq. (11) gives rise to a linear constraint regarding the FOE. The LMedS estimator is then applied once again on these constraints to give a robust estimate of the FOE. If required, estimates of the rotational velocity can be obtained in a similar manner by employing robust regression for (α, β, γ) on the constraints derived from the terms $\frac{\gamma f}{D_2^{\mathcal{L}} n_y} + \kappa\alpha - \beta$ computed for each line through the image principal point. Alternatively, rotation can be estimated using optical flow projections along directions that are normal to lines through the estimated FOE and therefore are independent of translation.

4 Experimental Results

The proposed method has been extensively tested with the aid of simulated and real flow fields. Representative results from these experiments are given in this section. In all the experiments reported here, at most 180 lines through the image principal point and 400 quadruples of points along each line have been employed.

4.1 Synthetic Flow Fields

The use of simulated data is justified by the fact that knowledge of the ground truth facilitates a quantitative assessment of the accuracy of the results. Besides, simulation enables us to vary in a controlled manner subsets of the parameters involved in the problem of egomotion estimation and then study their effect on the recovered motion. Therefore, a simulator has been constructed, which given appropriate values for the intrinsic parameters of the simulated camera (focal length and principal point), the translational and rotational motion parameters, the dimensions of the retina and the depth corresponding to each image point, employs Eqs. (1) to synthesize an optical flow field. The depths of image points are generated by random variables following various distributions. For the experiments reported here, a uniform distribution in the range $[Z_{min}, Z_{max}]$ and a Gaussian distribution with nonzero mean have been employed. All distances and sizes used by the simulator are specified in units of pixels. To account for the fact that optical flow fields might be sparse, their *density*, i.e. a percentage specifying the fraction of image points for which optical flow vectors have been computed, can be supplied. To make the simulated optical flow fields more realistic, noise is added to the synthetic optical flows. The noise we employ is generated according to the model suggested in [10]:

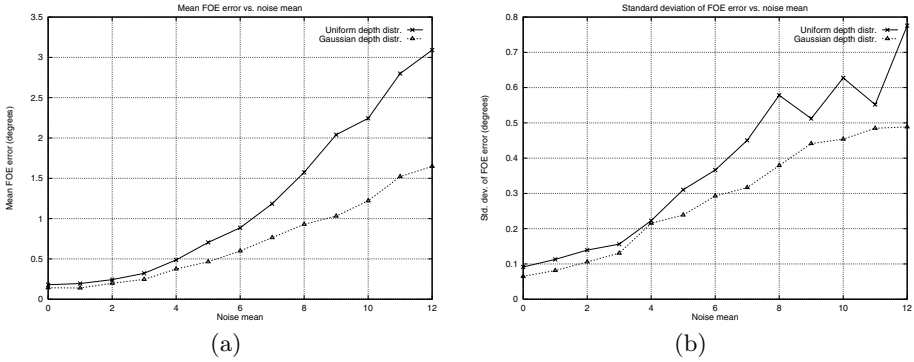
$$u_{noisy} = u + sign_1 * N(a, b) * 0.01 * u \quad , \quad v_{noisy} = v + sign_2 * N(a, b) * 0.01 * v$$

where $sign_1$ and $sign_2$ are binary values (i.e. 1 or -1) that are randomly chosen with equal probability and $N(a, b)$ is a Gaussian random variable with mean a and standard deviation b. This noise model is referred to as "Gaussian noise with mean a% and $\sigma = b$%". As noted in [10], 8% and 2% are realistic values for the noise mean and the standard deviation respectively, accounting for most of the errors observed in actual flow fields.

Throughout all experiments, image size was 512×512 pixels and the principal point was assumed to be in the center of the image. Also, in all but the third set of experiments, the focal length was 256 pixels, amounting to a field of view of 90 degrees. The density of the optical flow fields was 70%. Two different scenarios for the scene depth were simulated. The first uses a random variable that is uniformly distributed in the range $[10000, 50000]$ pixels to model the depth of a scene with large depth variations. The second scenario employs a Gaussian distribution with mean 15000 pixels and standard deviation 3000, to emulate a scene with less depth variation, in which the majority of the points lie at a dominant depth rather close to the camera. To ensure that the results are independent of the exact depth values used to synthesize the optical flow field, each experiment was run 100 times, each time using a different depth population drawn from the distributions described above.

In the first set of experiments, the effect of noise on the accuracy of the estimated FOE is examined. Employing increasing noise levels, Figures 1 (a) and (b) illustrate the mean and the standard deviation respectively of the FOE error for both depth distributions. Each point in the plots summarizes error statistics computed from 100 runs. If f is the focal length and the true FOE is at (x_0, y_0) while the estimated is at $(\hat{x_0}, \hat{y_0})$, the error in the FOE estimate is defined as the angle between the vectors (x_0, y_0, f) and $(\hat{x_0}, \hat{y_0}, f)$, given by $cos^{-1}(\frac{(x_0,y_0,f)\cdot(\hat{x_0},\hat{y_0},f)}{||(x_0,y_0,f)||\ ||(\hat{x_0},\hat{y_0},f)||})$. The 3D motion parameters used to synthesize flow were $(U, V, W) = (-120, 100, 150)$ (measured in pixels per frame) and $(\alpha, \beta, \gamma) = (0.005, 0.004, 0.002)$ (measured in radians per frame). The egomotion parameters and the depth values are such that the magnitude of the average translational component of the flow fields is comparable to that of the average rotational component. The angle between the direction of translation and the optical axis is about 46 degrees. The noise mean was increased to 12% in steps of 1% and the standard deviation was kept equal to 2%. As expected, the error increases with noise but remains acceptable even with very large amounts of noise. The error in the case of Gaussian depths is smaller since in this case the translational component of motion is larger than that in the case of uniformly distributed depths; this is further explained in the discussion of the experiments related to the magnitude of translation below.

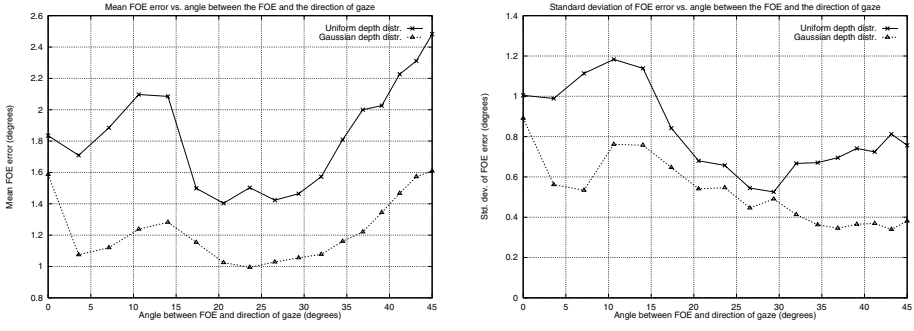
It has been observed in previous work on egomotion estimation that the error of the estimated FOE increases with the angle between the direction of translation and the direction of gaze (i.e. the direction defined by the optical axis) [4]. The second set of experiments studies the dependence of the FOE error on this angle for the proposed method. Figures 2 (a) and (b) show the mean and standard deviation of the FOE error with respect to the angle between the direction of



(a) (b)

Fig. 1. (a) Mean FOE error versus noise and (b) Standard deviation of FOE error versus noise.

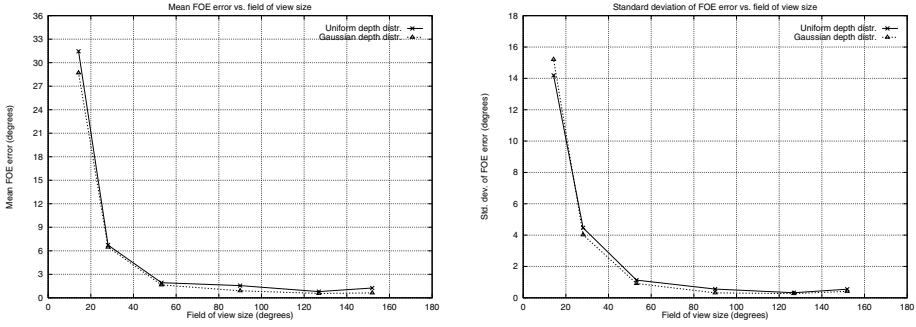
translation and the direction of gaze. The direction of translation was varied from $(0, 0, f)$ to $(f, 0, f)$, where f is the focal length. In other words, the translations considered range from a straight ahead motion to a sideways motion forming an angle of 45 degrees with the direction of gaze. The rotation parameters were again equal to $(\alpha, \beta, \gamma) = (0.005, 0.004, 0.002)$ and the magnitude of translation has been kept constant, equal to 216.565 pixels per frame, which is the magnitude of translation used in the first set of experiments. Each point in the graphs has been computed from 100 trials, performed with Gaussian noise of mean 8% and standard deviation of 2%. As can be seen from Fig. 2 (a), the FOE error does not vary considerably when the angle between the direction of translation and the direction of gaze is increased. This is a desirable characteristic of the proposed method, since it implies that the observer does not need to fixate on the estimated FOE to ensure small errors in the FOE estimates.



(a) (b)

Fig. 2. (a) Mean FOE error versus the angle between the direction of translation and the direction of gaze and (b) Standard deviation of FOE error versus the angle between the direction of translation and the direction of gaze.

The third set of experiments investigates the dependence of the FOE error on the field of view size. Figures 3 (a) and (b) show the mean and standard deviation of the FOE error with respect to the size of the field of view. The field of view size was varied by adjusting the focal length while keeping the image size constant. More specifically, the former was decreased by a multiplicative factor of 0.5 from 2048 to 64 pixels while the image size remained equal to 512×512 pixels. This change of the focal length amounts to the field of view being increased from 14.250 to 151.927 degrees. Recall that a focal length of 256 pixels used in the previous experiments corresponds to a field of view equal to 90 degrees. The simulated 3D velocity was identical to that of the first set of experiments, i.e. translation was equal to $(-120, 100, 150)$ and rotation to $(0.005, 0.004, 0.002)$. Gaussian noise of mean 8% and standard deviation of 2% was added to the simulated flows and each point in the graphs was again computed from 100 trials. As can be seen from Figs. 3, the error in the recovered FOE is almost identical for both depth distributions. More specifically, the FOE error is very large for small fields of view but becomes acceptable when the latter are larger than 25 degrees. This observation agrees with the theoretical findings of [6,5], which conclude that the inhomogeneous flow characteristics of a large field of view make it more helpful for determining the singularities of the flow field (i.e. the FOE and axis of rotation) compared to a narrow field of view. This conclusion holds independently of the particular algorithm that is employed to recover 3D motion.

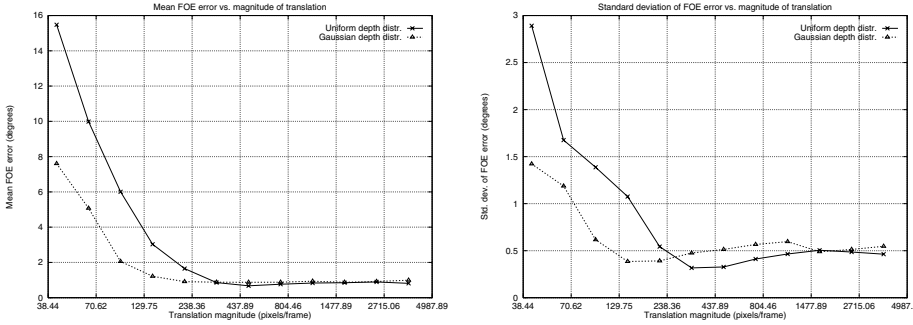


(a) (b)

Fig. 3. (a) Mean FOE error versus the size of the field of view and (b) Standard deviation of FOE error versus the size of the field of view.

The last set of experiments evaluates the performance of the method when the ratio between the magnitude of translation and that of rotation is varied. More specifically, assuming that the rotation is constant, Figures 4 (a) and (b) depict the effect of variable translation magnitude on the mean and the standard deviation of the FOE error. In this series of experiments, the direction of translation is identical to that defined by $(U, V, W) = (-120, 100, 150)$, but its magnitude is increased by a multiplicative factor of 1.5 between successive experiments. The

rotation has been kept constant at $(\alpha, \beta, \gamma) = (0.005, 0.004, 0.002)$ and 100 runs were made for each set of motion parameters. The noise was Gaussian with mean 8% and standard deviation 2%. As can be clearly seen from the plots, the FOE error is significant when the translation magnitude is small (less than 130 pixels per frame in Fig. 4 (a)). This is due to the fact that in this case, the translational components of the optical flow vectors are negligible compared to the rotational ones. Therefore, noise has a more pronounced effect on the translational components from which the FOE is recovered. However, as the magnitude of translation increases beyond 130 pixels per frame, the translational parts become comparable or even larger than the rotational ones. Thus, the translational parts are more immune to noise, giving rise to small FOE errors which are almost constant with respect to the magnitude of translation. Assuming constant translation, Figu-

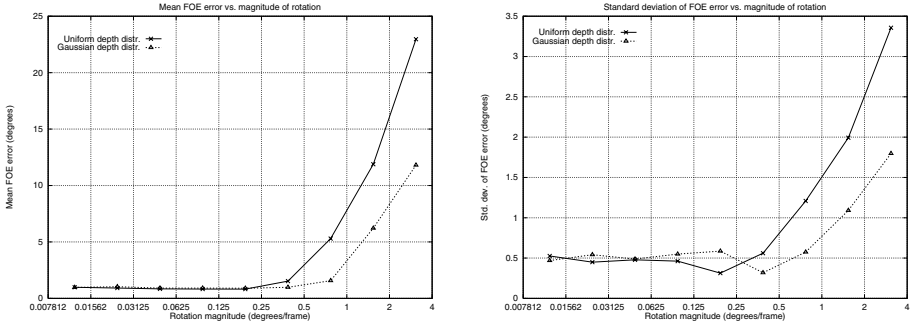


(a) (b)

Fig. 4. (a) Mean FOE error versus magnitude of translation (b) Standard deviation of FOE error versus magnitude of translation. Note that the scale on the horizontal axes is logarithmic with base 1.5.

res 5 (a) and (b) show the effects on the mean and the standard deviation of the FOE error induced by altering the rotation magnitude. Here, the behavior of the method is the converse of that observed in the case of constant rotation investigated in the previous paragraph. As can be seen from Fig. 5 (a), the error in the FOE estimates is almost constant for realistic amounts of rotation (less than 0.5 degrees per frame). When the rotation increases too much, the flow field becomes mainly rotational, with the rotational components accounting for a large fraction of the full flow field. Thus, noise has an increased impact on the translational parts, resulting in large errors for the FOE estimates. During the experiments outlined in Fig. 5, translation was kept fixed at $(U, V, W) = (-120, 100, 150)$, the rotation magnitude was increased by a multiplicative factor of 2.0 between successive experiments and 100 runs were made for each experiment. As before, the noise was Gaussian with mean 8% and standard deviation 2%. Note that a rotation of $(\alpha, \beta, \gamma) = (0.005, 0.004, 0.002)$ has a magnitude of 0.3845 degrees. When assuming continuous image motion (i.e. fine time sampling), rotations having

magnitudes larger than one degree per frame are very large and thus unrealistic.



(a) (b)

Fig. 5. (a) Mean FOE error versus magnitude of rotation and (b) Standard deviation of FOE error versus magnitude of rotation. Note that the scale on the horizontal axes is logarithmic with base 2.0.

4.2 Real Image Sequences

The method has also been tested using flow fields computed from real imagery for which the ground truth was known a priori. Throughout all experiments, optical flow was computed using an implementation of the Lucas & Kanade algorithm [17]. The first experiment employed the "yosemite" image sequence, one frame of which is shown in Fig. 6 (a). This sequence contains both translation and rotation and depicts a flight through Yosemite valley. Since the clouds are moving independently, only the optical flow vectors computed at the lower portion of the images have been employed. This portion of the original images corresponds to a field of view equal to 49.6 degrees horizontally and 29 degrees vertically. The true FOE is rather close to the center of the field of view, namely at (0, 58)[2] while the estimate computed by the proposed method was (-17.3, 72.3), a value that corresponds to an error of 22.4 pixels or 3.7 degrees. This amount of error compares favorably to errors in the "yosemite" FOE estimates appearing in the literature. More specifically, Heeger and Jepson [7] report an error of 3.5 degrees for the "yosemite" sequence and Daniilidis [3] reports an error of 4.0 degrees. The rotation recovered by the proposed method using robust regression on projections of flow vectors that are perpendicular to lines through the recovered FOE, was equal to $(0.000906, 0.002116, 0.000481)$ (in radians/frame). As mentioned in [7], the actual rotational velocity for the "yosemite" sequence is $(0.00023, 0.00162, 0.00028)$.

[2] These are "calibrated" image coordinates, defined with respect to the image principal point.

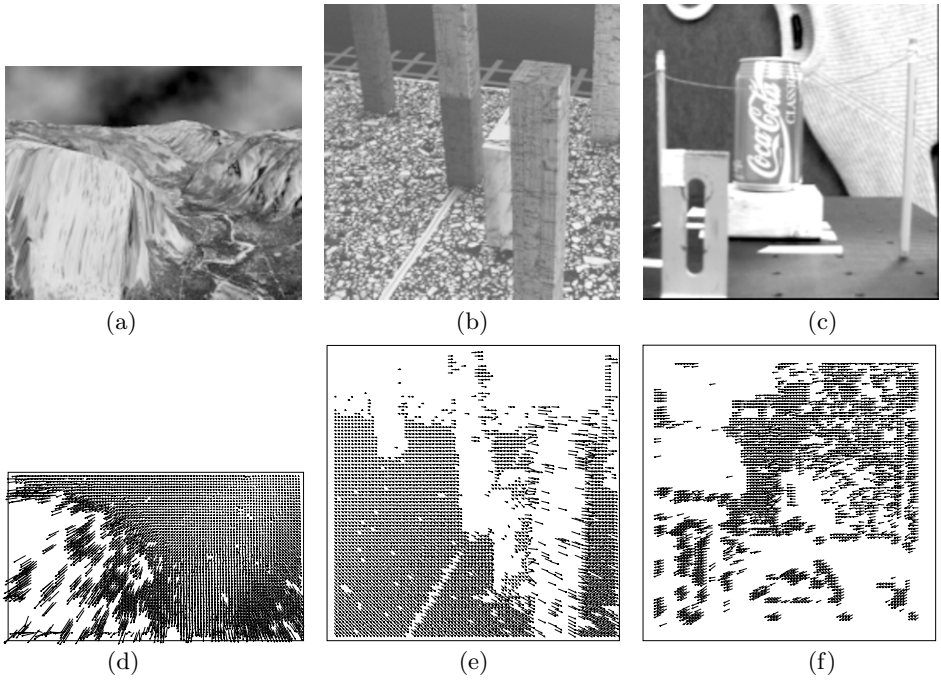


Fig. 6. (a)-(d) the "yosemite" image sequence, (b)-(e) the "marbled block" image sequence and (c)-(f) the "nasa" image sequence. One frame from each sequence is shown in the top row, while the optical flow fields used for egomotion estimation are shown in the bottom row.

The second experiment refers to the "marbled block" sequence, one frame of which is shown in Fig. 6 (b). The sequence was captured by a translating camera mounted on a robot arm that was moving above a textured floor in a right to left direction and contains many sharp discontinuities in depth and motion. The four dark blocks that lie on the floor are stationary, while the white block in the middle of the scene is moving independently with a right to left direction. The images of the "marbled block" sequence subtend 25.6 degrees of visual angle. The primary difficulty when estimating the egomotion for this sequence stems from the fact that the true FOE is outside the field of view, specifically at (777, 95.6). Thus, the angle between the direction of translation and the optical axis is about 35 degrees. The proposed method estimated the FOE at (625.0, 111.4), in error by 152.7 pixels or 5.65 degrees. For comparison, the FOE estimate reported by Daniilidis in [3] amounts to an error of 7.17 degrees. The rotation estimated by the proposed method was equal to $(-0.000748, 0.000291, 0.000031)$, close to being zero as expected.

The last experiment is based on the "nasa" image sequence, shown in Fig. 6 (c). Since the camera undergoes a purely translational motion, a rotation of $(\alpha, \beta, \gamma) = (-0.00025, -0.0018, 0.00030)$ was added synthetically in order to

make the experiment more challenging[3]. The ground truth for the FOE is (-5, -8) while the recovered FOE was (2.21, 49.29), in error by 57.74 pixels or 5.5 degrees. For reference, the images of the "nasa" sequence subtend 24 degrees of visual angle. The rotation estimated by the proposed method was equal to $(-0.000176, -0.001918, 0.000138)$. The rather large error in the recovered FOE for the "nasa" sequence is due to the proximity between the true FOE and the image principal point. Therefore, in this case, the distance D_2 (see Eq. (9)) of the FOE from every line through the principal point is very small and thus difficult to estimate accurately.

5 Conclusions

Accurate estimation of camera motion is important for many vision based tasks. In this paper, a novel constraint regarding the parameters of 3D motion has been presented. This constraint was used to develop a method for egomotion estimation that has several advantages. First, the method does not impose any constraints on the egomotion that can be recovered or on the structure of the viewed scene. Second, egomotion is computed through closed form solutions of linear equations, avoiding searching the space of possible solutions. The use of such linear constraints permits the exploitation of overdetermined linear systems through the application of robust linear regression techniques. The egomotion estimate computed by the proposed method can either be used as is, or, optionally, for bootstrapping more elaborate, iterative nonlinear egomotion estimation methods for refining it. Third, instead of employing local information derived from small image regions, redundancy is exploited by combining information across the whole visual field. Fourth, the method does not assume the availability of a dense optical flow field. This is very important for practical applications, since image sequences often have uniform, textureless areas that give rise to sparse optical flow fields. Finally, the use of a robust estimator such as LMedS safeguards against errors in the input, which could otherwise have a significant effect on the accuracy of the computations. Experimental results collected from extensive simulations as well as real image sequences indicate the effectiveness and robustness of the proposed method.

References

1. A. R. Bruss and B.K.P. Horn. Passive Navigation. *CVGIP*, 21:3–20, 1983.
2. W. Burger and B. Bhanu. Estimating 3D Egomotion from Perspective Image Sequences. *IEEE Trans. on PAMI*, 12(11):1040–1058, Nov. 1990.
3. K. Daniilidis. Fixation Simplifies 3D Motion Estimation. *CVIU*, 68(2):158–169, Nov. 1997.

[3] This is possible since the rotational component of motion is independent of the (unknown) scene structure; see Eqs. (1).

4. K. Daniilidis and M.E. Spetsakis. Understanding Noise Sensitivity in Structure From Motion. In Y. Aloimonos, editor, *Visual Navigation: From Biological Systems to Unmanned Ground Vehicles*, chapter 4. Lawrence Erlbaum Associates, Hillsdale, NJ, 1997.
5. S. Fejes and L.S. Davis. Direction-Selective Filters for Egomotion Estimation. Technical Report CS-TR-3814, University of Maryland, Jul. 1997.
6. C. Fermüller and Y. Aloimonos. The Confounding of Translation and Rotation in Reconstruction from Multiple Views. In *Proceedings of CVPR'97*, pages 250–256, 1997.
7. D.J. Heeger and A.D. Jepson. Subspace Methods for Recovering Rigid Motion I: Algorithm and Implementation. *IJCV*, 7(2):95–117, 1992.
8. R. Hummel and V. Sundareswaran. Motion Parameter Estimation from Global Flow Field Data. *IEEE Trans. on PAMI*, 15(5):459–476, May 1993.
9. M. Irani, B. Rousso, and S. Peleg. Recovery of Ego-Motion Using Region Alignment. *IEEE Trans. on PAMI*, 19(3):268–272, Mar. 1997.
10. N. V. Lobo and J. K. Tsotsos. Computing Egomotion and Detecting Independent Motion from Image Motion Using Collinear Points. *CVIU*, 64(1):21–52, July 1996.
11. H.C. Longuet-Higgins and K. Prazdny. The Interpretation of a Moving Retinal Image. In *Proceedings of the Royal Society*, pages 385–397. London B, 1980.
12. M.I.A. Lourakis. Egomotion Estimation Using Quadruples of Collinear Image Points. Technical Report 240, ICS/FORTH, Greece, Dec. 1998. Available at ftp://ftp.ics.forth.gr/tech-reports/1998.
13. M.I.A. Lourakis. Using Constraint Lines for Estimating Egomotion. In *Proc. of ACCV'2000*, volume 2, pages 971–976, Taipei, Taiwan, Jan 2000.
14. M.I.A. Lourakis, A.A. Argyros, and S.C. Orphanoudakis. Independent 3D Motion Detection Using Residual Parallax Normal Flow Fields. In *Proceeedings of ICCV'98*, pages 1012–1017, Bombay, India, Jan. 1998.
15. M.I.A. Lourakis and R. Deriche. Camera Self-Calibration Using the Singular Value Decomposition of the Fundamental Matrix: From Point Correspondences to 3D Measurements. Research Report 3748, INRIA Sophia-Antipolis, Aug. 1999.
16. M.I.A. Lourakis and S.C. Orphanoudakis. Using Planar Parallax to Estimate the Time-to-Contact. In *Proc. of CVPR'99*, volume 2, pages 640–645, Fort Collins, CO, June 1999.
17. B.D. Lucas and T. Kanade. An Iterative Image Registration Technique with an Application to Stereo Vision. In *Proceedings DARPA IU Workshop*, pages 121–130, 1981.
18. A. Mitiche and P. Bouthemy. Computation and Analysis of Image Motion: A Synopsis of Current Problems and Methods. *IJCV*, 19(1):29–55, Jul. 1996.
19. K. Prazdny. Egomotion and Relative Depth from Optical Flow. *Biological Cybernetics*, 36:87–102, 1980.
20. K. Prazdny. Determining the Instantaneous Direction of Motion From Optical Flow Generated by a Curvilinearly Moving Observer. *CVGIP*, 17:238–248, 1981.
21. J.H. Reiger and D.T. Lawton. Processing Differential Image Motion. *Journal of the Optical Society of America A*, 2:354–359, 1985.
22. P.J. Rousseeuw. Least Median of Squares Regression. *Journal of American Statistics Association*, 79:871–880, 1984.