

Fixation Simplifies 3D Motion Estimation

Konstantinos Daniilidis*

Computer Science Institute, University of Kiel, Preusserstrasse 1-9, 24105 Kiel, Germany

Received November 15, 1995; accepted July 15, 1996

Fixation is defined as holding the gaze direction towards the same environmental point through time. It was proven in the past that fixation reduces the number of unknowns in passive visual navigation from five to four. In this paper, we show that fixation further simplifies 3D motion estimation because it decouples the motion parameter space. We project the spherical motion field in two latitudinal directions with respect to two different poles of the image sphere. The first projection enables the computation of the longitude of the translation direction and the torsion. The second projection gives the angle between the direction of fixation and translation, as well as the time to collision to the fixated scene point. Both computational steps are based on one-dimensional searches along meridians of the image sphere. The observer may move with all six degrees of freedom. We do not use the efference copy of the fixational rotation of the camera. Performance of the algorithm is tested on real world sequences with fixation accomplished either off-line or during the recording using an active camera. A comprehensive review of most theories on advantages of fixation clarifies the differences to our approach. © 1997 Academic Press

1. INTRODUCTION

The ability to perceive the three-dimensional motion relative to the environment is crucial for every robot acting in a dynamically changing world. The estimation of 3D motion parameters has been addressed in the past as a reconstruction problem: Given a monocular image sequence the goal was to obtain the relative 3D motion to every scene component as well as a relative depth map of the environment. Solutions given suffer under instability problems and require an immense computational effort which excludes a real time reactive behavior.

In the light of behavior-based active vision [1–3] new approaches were proposed that do not try to recover a complete motion and structure description. Instead, they try to give individual solutions to tasks where motion is involved. Such tasks are the independent motion detection, ego motion computation, time to collision estimation, ob-

stacle detection, convoy following, etc. Researchers realized that the key for a computationally simple solution is in the selection of the appropriate representation for the dynamic imagery. Furthermore, active vision involves the control of the degrees of freedom of image acquisition. In motion related tasks this could mean the pursuing of a moving object, the fixation on a stationary point, or even keeping the gaze aligned with the heading direction.

In this study, we are interested in the computational advantages of fixation on an environmental point. There is a large amount of work in biological and computer vision research on how fixation is achieved [4–6]. Regarding also the other kinds of eye movements (vergence and optokinetic) several approaches studied the image cues guiding the eye movements as well as the underlying feedback loops. The evident advantages of overcoming the field of view, foveal sensing, and reducing the motion blur have been sufficiently, justifying the fixational movements so that only sporadic approaches delved into the computational advantages of fixation.

We show in this paper that the ability to fixate on a stationary point, combined with the appropriate representation of the motion field, enables the decoupling of the 3D motion parameters. We use a spherical image surface which can be mapped 1:1 to the image plane. We do not use any information from the motor encoders or from the input in the fixation feedback loop (called the efference copy in biology). Fixation is formulated only as a constraint on the motion field. This constraint reduces the number of unknowns from five to four. The translation direction remains unknown (two parameters), but instead of the angular velocity (three unknowns) we obtain only the torsion—rotation about the target direction—and the time to collision to the fixated scene point. The new representation for the fixated motion field is based on two projections. Assuming that the fixated target point is the pole of the sphere, we show that the latitudinal projection of the motion field has the property of being constant along a meridian. The constant value is equal to the torsion and the meridian contains the heading direction. Taking as a new pole the normal to this meridian we again project the

* E-mail: kd@informatik.uni-kiel.de.

flow field in the latitudinal direction and obtain a similar pattern: a meridian with respect to the new pole where the new latitudinal projection is constant and equal to the time to impact to the target. This new meridian fully constrains the heading direction. We are, thus, able to compute the heading direction by applying only two one-dimensional searches.

We elaborate the geometry configurations that lead to ambiguity. In case of a heading direction outside the field of view we replace the second projection with the solution of an equation in the two remaining unknowns.

In addition to the new algorithm this paper provides a comprehensive review of previously published results on fixation for egomotion estimation. Before we turn to the review we will first precisely state the problem at hand, and then establish connections to relevant approaches. We then present the novel solution for estimating egomotion in the most general case. We finish the paper with experiments on simulated fixated motion fields as well as optical flow fields obtained from fixated real world sequences.

2. PROBLEM STATEMENT

We assume that the imaging surface is a sphere with unit radius. We denote by $\hat{\mathbf{p}}$ the points on this sphere resulting from the projection $\hat{\mathbf{p}} = \mathbf{P}/\|\mathbf{P}\|$. The mapping of the planar imaging surface to a spherical surface is one to one. Let $\mathbf{x} = \mathbf{P}/\hat{z}^T \mathbf{P}$ be a point on the image plane $Z = 1$ with the optical axis parallel to the Z -axis with unit vector $\hat{\mathbf{z}}$. If $\dot{\mathbf{x}}$ is the motion field on that plane then it can be easily proved that the spherical motion field reads

$$\dot{\hat{\mathbf{p}}} = \frac{1}{\|\dot{\mathbf{x}}\|} (\hat{\mathbf{p}} \times (\dot{\mathbf{x}} \times \hat{\mathbf{p}})). \quad (1)$$

Most of the authors assume that for a small field of view the two fields are approximately equal. However, for a large field of view the above equation should be used. Special care should be taken in the mapping of the planar discretization noise onto the sphere, a problem fully described in [7].

We assume that the observer is moving with instantaneous linear velocity \mathbf{v} and angular velocity $\boldsymbol{\omega}$ relative to the environment so that the velocity of a scene point \mathbf{P} can be written as $\dot{\mathbf{P}} = \mathbf{v} + \boldsymbol{\omega} \times \mathbf{P}$. In case of pure ego motion all equations are valid with the opposite sign for the velocities \mathbf{v} and $\boldsymbol{\omega}$.

The spherical motion field reads

$$\dot{\hat{\mathbf{p}}} = \frac{1}{\|\mathbf{P}\|} (\hat{\mathbf{p}} \times (\mathbf{v} \times \hat{\mathbf{p}})) + \boldsymbol{\omega} \times \hat{\mathbf{p}}, \quad (2)$$

where we can observe the classical decomposition into a translational component depending on the environment ($\|\mathbf{P}\|$) and the rotational term depending only on the image position. The spherical motion field vector lies on the tangential plane at point $\hat{\mathbf{p}}$ so that $\hat{\mathbf{p}}^T \dot{\hat{\mathbf{p}}} = 0$. As we mentioned at the beginning we suppose that a control algorithm exists that makes a target point $\hat{\mathbf{t}}$ on the sphere be fixated which means

$$\dot{\hat{\mathbf{t}}} = 0.$$

From (2) follows that

$$-\frac{\mathbf{v} \times \hat{\mathbf{t}}}{\|\mathbf{T}\|} + \boldsymbol{\omega} \text{ is parallel to } \hat{\mathbf{t}},$$

where $\|\mathbf{T}\|$ is the distance to the target scene point. Hence, the angular velocity in case of fixation reads

$$\boldsymbol{\omega} = \gamma \hat{\mathbf{t}} + \frac{\mathbf{v} \times \hat{\mathbf{t}}}{\|\mathbf{T}\|}. \quad (3)$$

It is constrained to be a function of the linear velocity and possesses only one degree of freedom γ : the torsion around the target point $\hat{\mathbf{t}}$. Thus, after fixation the flow field contains three components (Fig. 1): a translational one due to \mathbf{v} , a fixational equal to the second term $\mathbf{v} \times \hat{\mathbf{t}}/\|\mathbf{T}\|$ of (3), and a torsional component $\gamma \hat{\mathbf{t}}$.

After inserting the fixation angular velocity (3) into (2) the spherical motion field of a point $\hat{\mathbf{p}}$ different from the target reads

$$\dot{\hat{\mathbf{p}}} = \hat{\mathbf{p}} \times \left(\mathbf{v} \times \left(\frac{\hat{\mathbf{p}}}{\|\mathbf{P}\|} - \frac{\hat{\mathbf{t}}}{\|\mathbf{T}\|} \right) \right) + \gamma (\hat{\mathbf{t}} \times \hat{\mathbf{p}}). \quad (4)$$

After eliminating the structure information $\|\mathbf{P}\|$ by taking the scalar product with $\mathbf{v} \times \hat{\mathbf{p}}$ we obtain the ‘‘epipolar’’ equation for the fixated motion field

$$(\mathbf{v} \times \hat{\mathbf{p}})^T \left(\dot{\hat{\mathbf{p}}} - \frac{\hat{\mathbf{t}}}{\|\mathbf{T}\|} \hat{\mathbf{p}}^T \mathbf{v} - \gamma (\hat{\mathbf{t}} \times \hat{\mathbf{p}}) \right) = 0 \quad (5)$$

which corresponds to the instantaneous version of epipolar equation for general motion,

$$(\mathbf{v} \times \hat{\mathbf{p}})^T (\dot{\hat{\mathbf{p}}} - \boldsymbol{\omega} \times \hat{\mathbf{p}}) = 0. \quad (6)$$

We see that the depth-free equation (5) contains three unknowns for the scaled linear velocity $\mathbf{v}/\|\mathbf{T}\|$ plus one unknown for the torsion γ around the target. Furthermore, Eq. (5) is quadratic in the components of \mathbf{v} and bilinear in (\mathbf{v}, γ) .

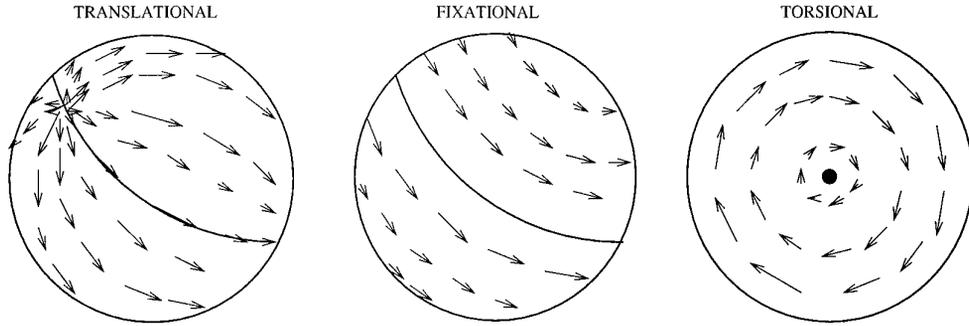


FIG. 1. The three components of a fixated motion field.

3. LITERATURE REVIEW

We will now turn to relevant work and will relate, where possible, the underlying equations used in the problem statement above. The first and most important result obtained by Bandopadhyay and Ballard [8] and by Aloimonos *et al.* [9] was that fixation reduces the number of unknowns from five to four. Their fixation constraint was that $\boldsymbol{\omega} = (v_y, -v_x, \gamma)$ which is a direct implication of (3) if we set the target parallel to the optical axis: $\hat{\mathbf{t}} = \hat{\mathbf{z}}$. The flow on the image plane can be written as

$$\dot{\mathbf{x}} = \frac{1}{Z} (\hat{\mathbf{z}} \times (\mathbf{v} \times \mathbf{x}) + (\hat{\mathbf{z}} \times (\mathbf{x} \times (\mathbf{x} \times \boldsymbol{\omega}))), \quad (7)$$

with $\mathbf{x} = (x, y, 1)$ a point on the image plane $Z = 1$. After elimination of the depth Z we obtain the epipolar equation for the image plane,

$$(\mathbf{v} \times \mathbf{x})^T (\dot{\mathbf{x}} - \boldsymbol{\omega} \times \mathbf{x}) = 0, \quad (8)$$

which is identical to the spherical case (6) if we replace $(\hat{\mathbf{p}}, \hat{\mathbf{p}})$ with $(\mathbf{x}, \dot{\mathbf{x}})$. Let the components of $\boldsymbol{\omega}$ be denoted by (A, B, γ) . Introducing the fixation constraint $A = v_y$, $B = -v_x$ in (8) we obtain

$$\frac{u + Axy - B(x^2 + 1) + \gamma y}{v + A(1 + y^2) - Bxy - \gamma x} = -\frac{B + xW'}{A - yW'}, \quad (9)$$

which is found in [8, Eq. (12); 9, Eq. (5.11)]. The above equation is identical to our epipolar constraint in the fixation case (5) if we replace (v_x, v_y) by $(-B, A)$.

In the work of Fermüller and Aloimonos [10, 11] fixation is exploited to compute the line on the image which passes through the FOE. Using only normal flow the location of the FOE on this line is found by matching patterns to the repeatedly detranslated flow. The line containing the FOE passes through the fixated origin and has slope (U/V) if $\mathbf{v} = (U, V, W)$. In order to pursue their analysis we need

to introduce the difference between the rotation that the observer undergoes independent of fixation ($\boldsymbol{\omega}_{\text{obs}}$) and the control rotation of the camera necessary to obtain fixation ($\boldsymbol{\omega}_{\text{f}}$). The sum of the two rotations is the rotation that gives the rotational component of the motion field under fixation. That is,

$$\boldsymbol{\omega} = \boldsymbol{\omega}_{\text{obs}} + \boldsymbol{\omega}_{\text{f}}. \quad (10)$$

If the flow in the center is

$$\dot{x} = U/Z + \omega_y, \quad \dot{y} = V/Z - \omega_x \quad (11)$$

and the control rotation $\boldsymbol{\omega}_{\text{f}} = (c_x, c_y, 0)$ is such that it introduces the opposite flow at the center then

$$-c_y = U/Z + \omega_y, \quad c_x = V/Z - \omega_x. \quad (12)$$

Since the pair of equations have five unknowns, they assume that these variables remain constant at two time instants at which the control rotation is known, thereby obtaining U/V as

$$U/V = \frac{c_{y,2} - c_{y,1}}{c_{x,1} - c_{x,2}}. \quad (13)$$

This approach makes use of the control signals of the camera movements $(c_x, c_y, 0)$ and makes the assumption that translation direction is almost constant despite fixation. It should be noticed that the stepwise compensation of the translation in our algorithm can also be found as the process of detranslation in [11, 12].

The equation of fixational motion field (4) is used by Taalebinezhad [13]. The flow field in the brightness change constraint equation (BCCE) is substituted by the fixational motion field. As the BCCE at every pixel introduces a new unknown (depth) an additional assumption of minimal variation of depth near the fixation point is

added. To convert the resulting minimization into an eigenvalue problem it is further assumed that the torsion γ is already computed in a preceding step. This step is solved, assuming local frontoparallel patches. However, this assumption enables a local and linear computation of rotation and translation without fixation [14].

Raviv and Herman [15] study the surfaces in the world that produce constant flow in the image. In case of rotation axis perpendicular to translation they regard the rotation axis as the sphere pole. Then they show that the level sets of equal latitudinal flow are cylinders and that the longitudinal flow is zero along two planes. The intersection of these planes with the cylinder corresponds to the points in the world that produce zero flow. The equal flow circles can be used to analyze the space around the fixation point and to predict the optical flow in case of fixation.

The first part in [16] is identical to the work by Raviv and Herman [15]. They derive the equal flow cylinders and planes. However, Thomas *et al.* [16] apply their findings of zero longitudinal flow to determine the angle between the target and the velocity \mathbf{v} . This plane always appears in the image as a line, provided that the FOV is 180° . These results are tested using a novel 180° field of view camera. The original idea of using the entire spherical field of view to recover 3D motion is attributed to Nelson and Aloimonos [17].

Raviv and Ozery [18] assume a restricted motion model: the rotation axis is orthogonal to the optical axis, the relative motion of the camera with respect to the object is purely rotational about the fixation point, and the rotation rate is constant. They assume further scaled orthographic projection. Under these limiting assumptions, they determine the magnitude of angular velocity from the image positions of two distinct points at two time instances.

In [19] fixation is combined with the log-polar transformation. Using the second-order spatial derivatives of the fixated log-polar field it is shown that the time to collision can be computed using only the radial component of the velocity. Advantages of the polar transformation in case of fixation are also shown in [20], where the heading direction is computed using two specific lines through the center of the image.

The work of Barth and Tsuji [21] addresses the issue of how to fixate in the direction of the translation. Their technique is based on the following heuristic. They group the flow vectors near the point of fixation into two groups: positive and negative flows. The difference in the average of the flow values at these groups indicates the direction of translation with respect to the current fixation direction. Based on this value the robot is controlled to turn toward the direction of translation. The same issue is addressed in [22] using an affine model for the optical flow field. Servoing toward the heading direction is achieved by min-

imizing the lateral translational components by means of a task function.

4. PROJECTIONS OF THE FIXATED MOTION FIELD

We proceed by projecting the fixated spherical motion field (4) into two different orthogonal basis systems of the tangential plane at an arbitrary point on the sphere. The first projection assumes that the target direction $\hat{\mathbf{t}}$ is the pole of the sphere defining thus a latitudinal and a longitudinal unit vector

$$\hat{\phi}_1 = \frac{\hat{\mathbf{t}} \times \hat{\mathbf{p}}}{\|\hat{\mathbf{t}} \times \hat{\mathbf{p}}\|}, \quad \hat{\theta}_1 = \hat{\phi}_1 \times \hat{\mathbf{p}},$$

respectively, lying in the tangential plane of point $\hat{\mathbf{p}}$.

The second projection assumes as a pole the unit vector in the direction of $\mathbf{v} \times \hat{\mathbf{t}}$ yielding a latitudinal and a longitudinal unit vector

$$\hat{\phi}_2 = \frac{(\mathbf{v} \times \hat{\mathbf{t}}) \times \hat{\mathbf{p}}}{\|(\mathbf{v} \times \hat{\mathbf{t}}) \times \hat{\mathbf{p}}\|}, \quad \hat{\theta}_2 = \hat{\phi}_2 \times \hat{\mathbf{p}},$$

respectively. Through the course of exposition the reader may consult Fig. 2, where the projections are illustrated.

The latitudinal projection using the target direction $\hat{\mathbf{t}}$ as a pole reads

$$\hat{\mathbf{p}}^T \hat{\phi}_1 = \frac{1}{\|\hat{\mathbf{t}} \times \hat{\mathbf{p}}\|} \mathbf{v}^T (\hat{\mathbf{t}} \times \hat{\mathbf{p}}) \left(\frac{1}{\|\mathbf{P}\|} - \frac{\hat{\mathbf{p}}^T \hat{\mathbf{t}}}{\|\mathbf{T}\|} \right) + \gamma \|\hat{\mathbf{t}} \times \hat{\mathbf{p}}\|.$$

Because the angle between the target $\hat{\mathbf{t}}$ and the considered point is known we divide by its sine which is equal to $\|\hat{\mathbf{t}} \times \hat{\mathbf{p}}\|$:

$$\frac{\hat{\mathbf{p}}^T \hat{\phi}_1}{\|\hat{\mathbf{t}} \times \hat{\mathbf{p}}\|} - \gamma = \frac{1}{\|\hat{\mathbf{t}} \times \hat{\mathbf{p}}\|^2} \mathbf{v}^T (\hat{\mathbf{t}} \times \hat{\mathbf{p}}) \left(\frac{1}{\|\mathbf{P}\|} - \frac{\hat{\mathbf{p}}^T \hat{\mathbf{t}}}{\|\mathbf{T}\|} \right). \quad (14)$$

We see that the latitudinal component minus the torsion vanishes if the considered point lies on the plane spanned by the target and the translation direction. Thus, we are able to constrain the translation direction if we find the meridian with longitude η , where the term $\hat{\mathbf{p}}^T \hat{\phi}_1 / \|\hat{\mathbf{t}} \times \hat{\mathbf{p}}\|$ is constant, independent of the latitude $\|\hat{\mathbf{t}} \times \hat{\mathbf{p}}\|$. Unfortunately this is not the only case where this term becomes constant. Suppose that a part of the environment is planar. Let the equation of the plane be $\hat{\mathbf{N}}^T \mathbf{X} = d$ and assume that the target is on the optical axis. If the plane normal reads $\hat{\mathbf{N}} = (\cos \alpha \sin \beta, \sin \alpha \sin \beta, \cos \beta)$ then it can be easily proved that

$$\frac{1}{\|\hat{\mathbf{t}} \times \hat{\mathbf{p}}\|^2} \mathbf{v}^T (\hat{\mathbf{t}} \times \hat{\mathbf{p}}) \left(\frac{1}{\|\mathbf{P}\|} - \frac{\hat{\mathbf{p}}^T \hat{\mathbf{t}}}{\|\mathbf{T}\|} \right) = \frac{1}{d} \mathbf{v}^T \hat{\phi}_1 \sin \beta \cos(\alpha - \eta), \quad (15)$$

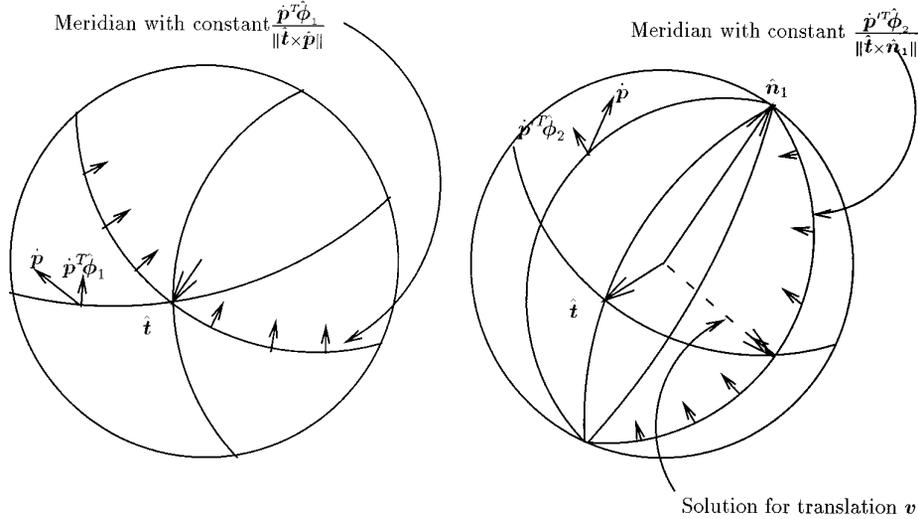


FIG. 2. The meridians with respect to the target pole \hat{t} are drawn on the left sphere. The spherical flow \hat{p} is projected on the latitudinal direction. The first step of the algorithm is a 1D search for the meridian with constant $\hat{p}^T \hat{\phi}_1 / \|\hat{t} \times \hat{p}\|$. In the second step (see the right sphere) the pole is \hat{n}_1 perpendicular to the meridian found in the first step. The flow without torsion $\hat{p}' = \hat{p} - \gamma(\hat{t} \times \hat{p})$ is projected on the new latitudinal directions. A 1D search among the meridians with respect to pole \hat{n}_1 for the meridian with constant $\hat{p}'^T \hat{\phi}_2 / \|\hat{t} \times \hat{n}_1\|$ yields a second big circle. The intersection of the big circles found in the two steps gives the solution for the desired translation direction \mathbf{v} .

which is independent of the latitude $\|\hat{t} \times \hat{p}\|$. Hence, all meridians that are projections of lines on planes in the scene will have a constant latitudinal projection independent of the colatitude angle. Furthermore, the right-hand side of (14) will vanish on the meridians that are projections of infinite depths ($1/\|\mathbf{P}\| = 0$) and on the entire field of view if the translation is parallel to the target direction: $\mathbf{v} \times \hat{t} = 0$.

To summarize the defeating configurations:

1. There may exist meridians with constant latitudinal projection if these meridians are projections of planar parts of the environment or of scene points at infinity.
2. The latitudinal projection is everywhere constant if

we fixate on the translation direction or if translation does not exist at all.

Suppose now that the unit vector \hat{n}_1 in the direction $\mathbf{v} \times \hat{t}$ is given and let it be the new pole. The new pole introduces new meridians and latitudes. Since torsion can be computed in the first projection above we consider the latitudinal projection of the torsion-free flow,

$$(\hat{p} - \gamma(\hat{t} \times \hat{p}))^T \hat{\phi}_2 = \frac{1}{\|\mathbf{P}\| \|(\mathbf{v} \times \hat{t}) \times \hat{p}\|}$$

$$(\hat{p} \times (\mathbf{v} \times \hat{p}))^T ((\mathbf{v} \times \hat{t}) \times \hat{p}) + \frac{\|\mathbf{v} \times \hat{t}\|}{\|\mathbf{T}\|} \|\hat{p} \times \hat{n}_1\|,$$

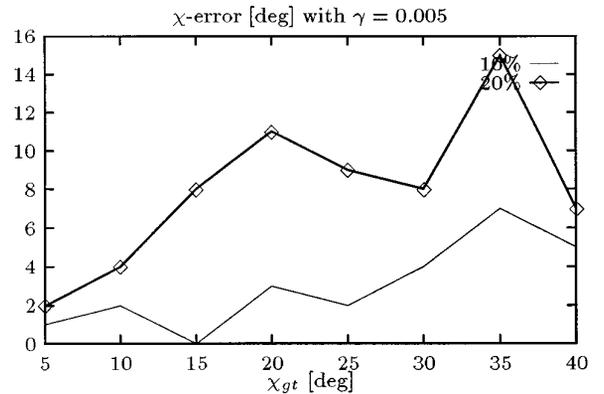
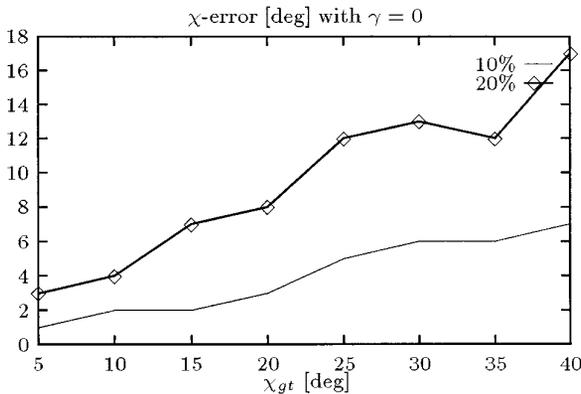


FIG. 3. The error in the χ -angle as a function of the translation direction for a field of view of 90° and two values of torsion: 0 (left) and 0.005 (right). The motion field is corrupted by gaussian relative error with standard deviation of 10% and 20%.

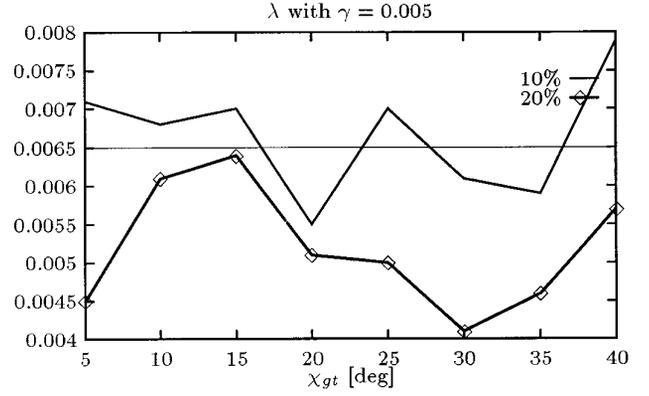
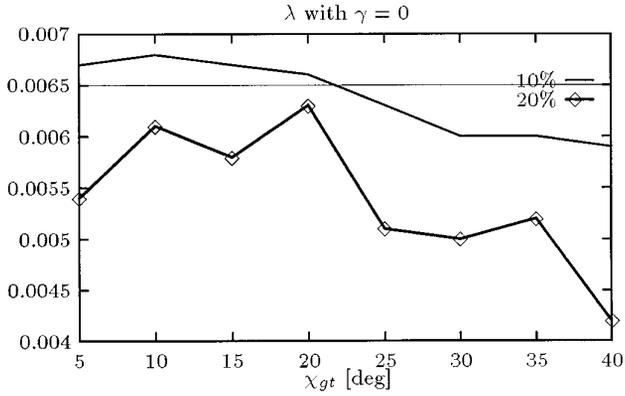


FIG. 4. The inverse of the time to collision $\lambda = \|\mathbf{v}\|/\|\mathbf{T}\|$ as a function of the translation direction for a field of view of 90° and two values of torsion: 0 (left) and 0.005 (right). The motion field is corrupted by gaussian relative error with standard deviation of 10% and 20%. The ground truth value of λ is 0.0065.

where $\hat{\mathbf{n}}_1$ is the unit vector $\mathbf{v} \times \hat{\mathbf{i}}/\|\mathbf{v} \times \hat{\mathbf{i}}\|$ known from the first projection. Hence, we can divide the left-hand side and rewrite the right-hand side as

$$\frac{(\hat{\mathbf{p}} - \gamma(\hat{\mathbf{i}} \times \hat{\mathbf{p}}))^T \hat{\Phi}_2}{\|\hat{\mathbf{p}} \times \hat{\mathbf{n}}_1\|} = \frac{1}{\|\mathbf{P}\| \|(\mathbf{v} \times \hat{\mathbf{i}}) \times \hat{\mathbf{p}}\|} \hat{\mathbf{p}}^T (\mathbf{v} \times (\mathbf{v} \times \hat{\mathbf{i}})) + \frac{\|\mathbf{v}\| \|\hat{\mathbf{v}} \times \hat{\mathbf{i}}\|}{\|\mathbf{T}\|}.$$

Considering now meridians through the pole $\mathbf{v} \times \hat{\mathbf{i}}$ we obtain the following cases, where the torsion-free latitudinal component will be constant:

1. On the meridian with normal $\mathbf{v} \times (\mathbf{v} \times \hat{\mathbf{i}})$.
2. On the meridians containing points with infinite depth.

The detection of the meridian with normal

$$\hat{\mathbf{n}}_2 = \frac{\mathbf{v} \times (\hat{\mathbf{v}} \times \hat{\mathbf{i}})}{\|\hat{\mathbf{v}} \times \hat{\mathbf{i}}\|}$$

allows the full computation of the translation direction

$$\hat{\mathbf{v}} = \hat{\mathbf{n}}_1 \times \hat{\mathbf{n}}_2.$$

Having obtained the heading direction we know the sine of the angle between the heading direction and the target $\|\hat{\mathbf{v}} \times \hat{\mathbf{i}}\|$. The remaining constant after the vanishing of the first term in (4) yields $\lambda = \|\mathbf{v}\|/\|\mathbf{T}\|$ which is the fourth and last unknown of the motion problem in case of fixation. The inverse of it can be interpreted as the time to collision to an object at the same distance as the target in the motion direction.

To find meridians of constant value in the first and the second latitudinal projections we compute for every meridian the mean and the variance over the latitude. Then, we

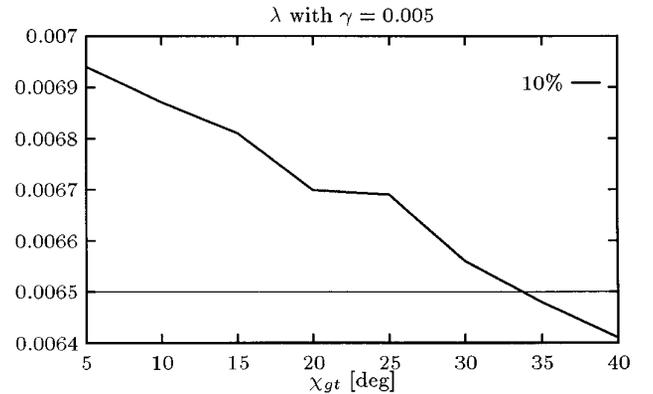
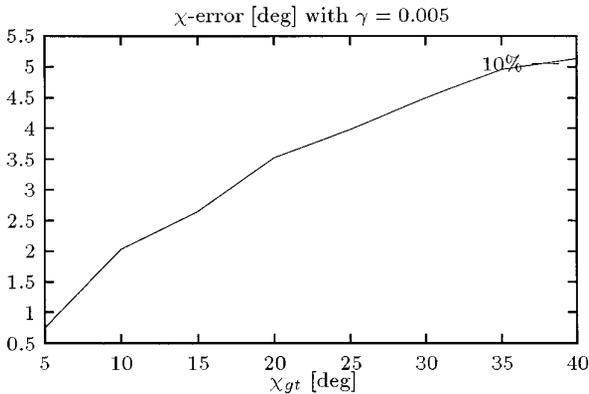


FIG. 5. The error in the χ -angle (left) and the inverse of the time to collision $\lambda = \|\mathbf{v}\|/\|\mathbf{T}\|$ as computed by the alternative to the second step. The field of view is 45° and the relative error in the optical flow is 10%.

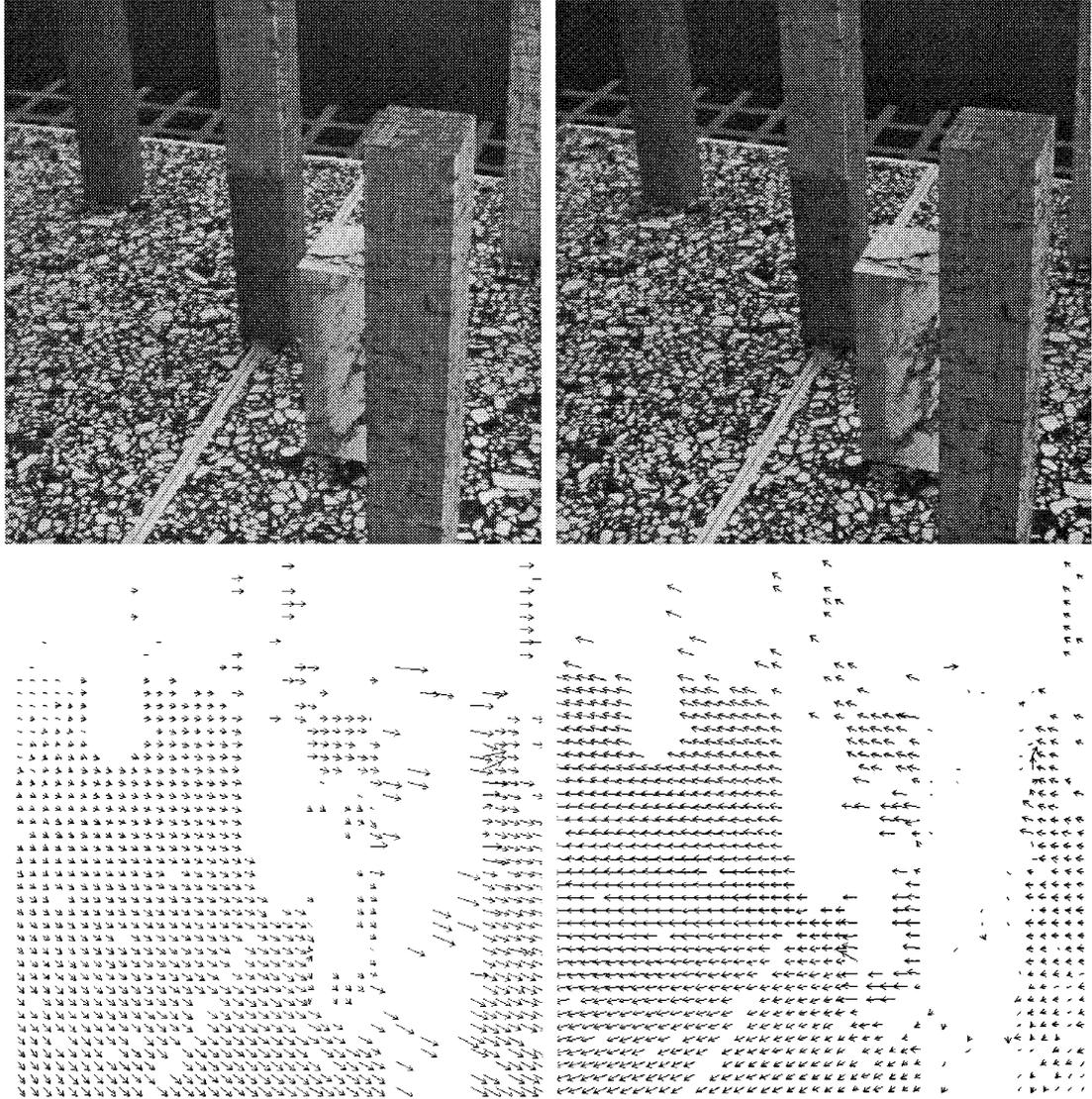


FIG. 6. The 1st and the 12th images of the “marbled block” sequence (above), the original flow field (bottom left), and the fixated flow field (bottom right).

search for the meridians on which this variance is minimized. The means on these meridians yield the torsion and the inverse of the time to collision, in the first and second projections, respectively.

Although in the first projection all meridians—or sectors of them—were contained in the field of view this is not the case in the second projection, where the meridians are with respect to the new pole \hat{n}_1 . It is very easy to imagine this case if, for example, $\hat{n}_1 = (0, 1, 0)$. We will see in the experiments that in such a case the variance of the second latitudinal projection gets its minimum at the border of the field of view. A corrective saccade can then shift the focus of expansion inside the field of view and the process can be continued with a refixation on a new point. If we want to avoid a corrective saccade we must replace the

second search with a procedure as follows. The first step constrains the translational velocity to the plane with normal \hat{n}_1 . Thus, we can write

$$\hat{v} = \cos \chi \hat{i} + \sin \chi (\hat{n}_1 \times \hat{i}), \quad (16)$$

where χ is the remaining degree of freedom of the translation direction or, in the terms of the formulation above, the longitude of the searched meridian in the second step. We rewrite (5) as

$$(\hat{v} \times \hat{p})^T (\hat{p}' - \lambda \hat{i} \hat{p}^T \hat{v}) = 0, \quad (17)$$

where $\hat{p}' = \hat{p} - \gamma (\hat{i} \times \hat{p})$ is known from the second step and $\lambda = \|\mathbf{v}\|/\|\mathbf{T}\|$ is the inverse of the time to collision.

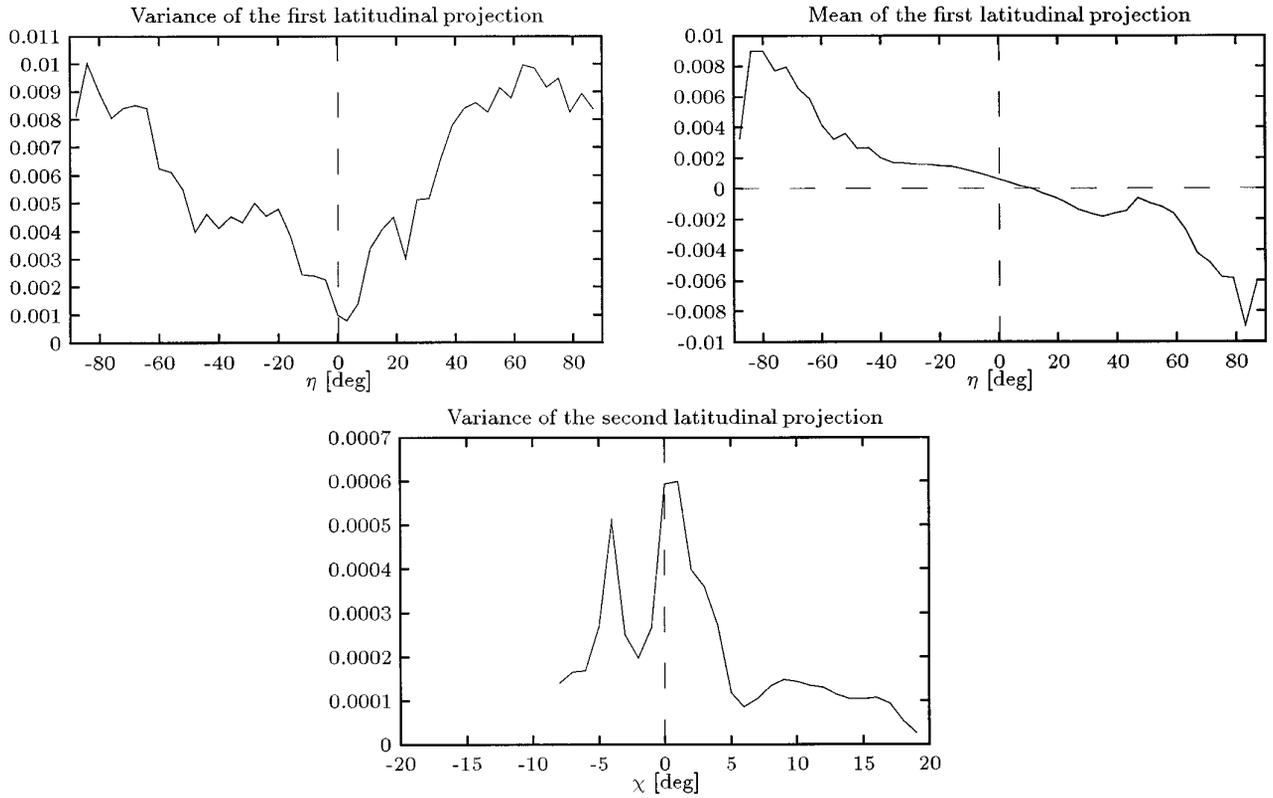


FIG. 7. The variance (top left) and the mean (top right) of the first latitudinal projection for the “marbled block” sequence. The minimum of the variance gives the angle η and the mean for this η gives the torsion. The variance of the second latitudinal projection has its minimum at the left bound of χ indicating a focus of expansion outside the field of view.

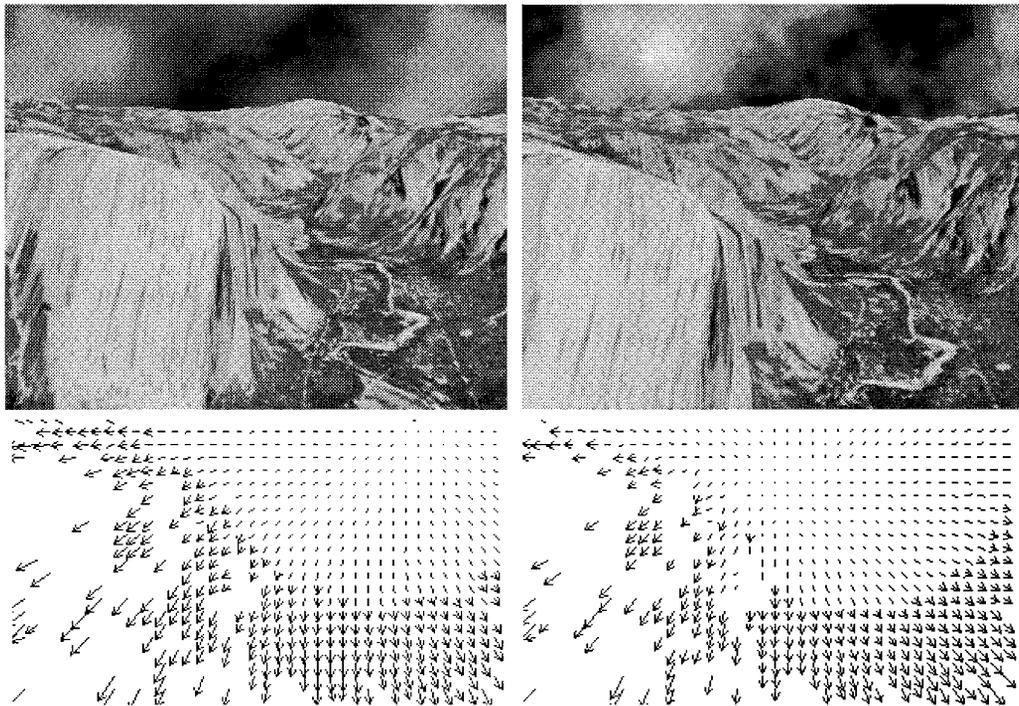


FIG. 8. The 1st and the 14th images of the Yosemite sequence (above), the original flow field (bottom left), and the fixated flow field (bottom right).

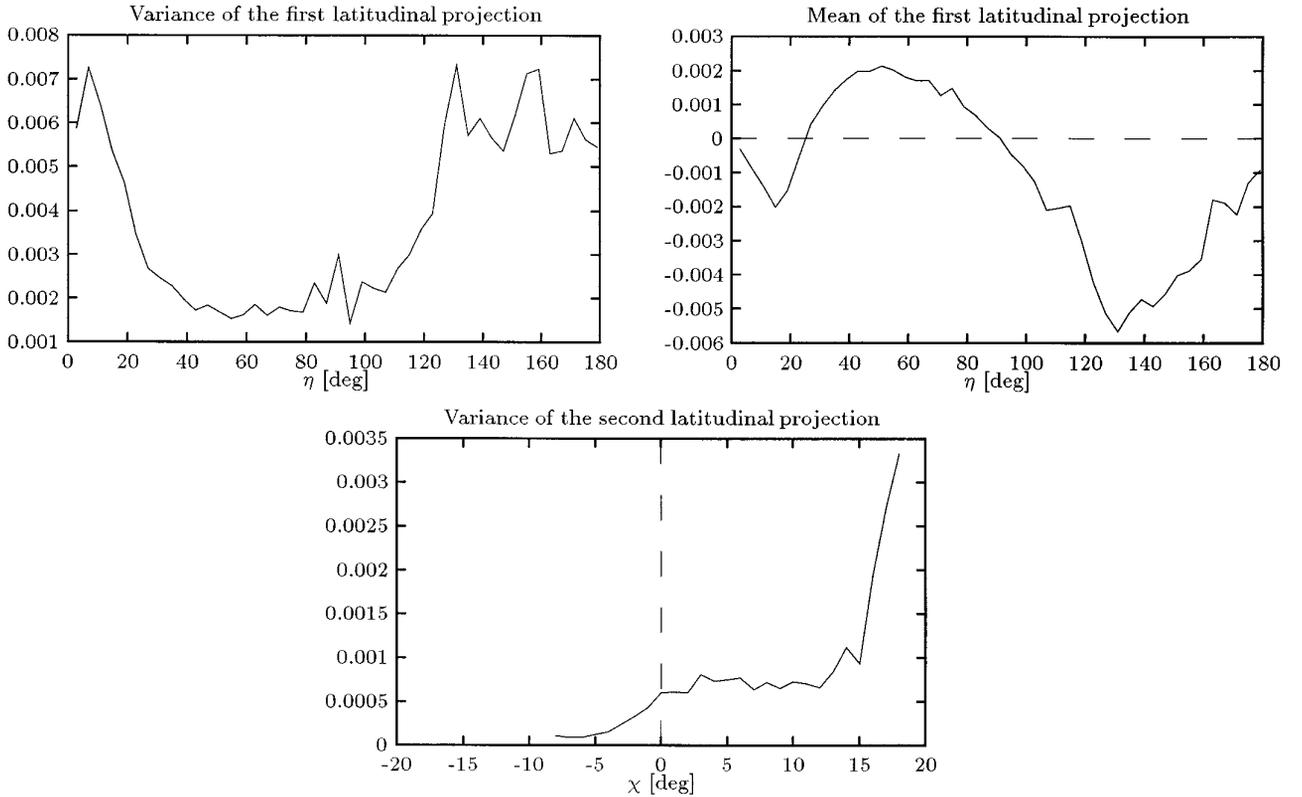


FIG. 9. The variance (top left) and the mean (top right) of the first latitudinal projection for the Yosemite sequence. The minimum of the variance gives the angle η and the mean for this η gives the torsion. The variance of the second latitudinal projection has its minimum at the right bound of χ indicating a focus of expansion outside the field of view.

If we insert $\hat{\mathbf{v}}$ from (16) in (17) in the above equation we obtain

$$\begin{aligned} \cos \chi (\hat{\mathbf{i}} \times \hat{\mathbf{p}})^T \hat{\mathbf{p}}' + \sin \chi ((\hat{\mathbf{n}}_1 \times \hat{\mathbf{i}}) \times \hat{\mathbf{p}})^T \hat{\mathbf{p}}' \\ = \lambda \sin \chi \hat{\mathbf{p}}^T \hat{\mathbf{n}}_1 (\cos \chi \hat{\mathbf{p}}^T \hat{\mathbf{i}} + \sin \chi (\hat{\mathbf{n}}_1 \times \hat{\mathbf{i}})^T \hat{\mathbf{p}}). \end{aligned} \quad (18)$$

This is a nonlinear equation in the two unknowns χ and λ which can be solved numerically with nonlinear minimization.

To summarize, we present the algorithmic steps of our method:

1. Choose a sampling step for the longitude angle η with respect to pole $\hat{\mathbf{i}}$ —in reality being always the optical axis if we fixate on the center. Divide the optical flow field in groups with the same longitude η corresponding to meridians. Compute for every group the mean and the variance of

$$\frac{\hat{\mathbf{p}}^T \hat{\phi}_1}{\|\hat{\mathbf{i}} \times \hat{\mathbf{p}}\|}.$$

Carry out an 1D-search for the minimum η_{\min} of the variance. The new pole $\hat{\mathbf{n}}_1$ reads $(\sin \eta_{\min}, -\cos \eta_{\min}, 0)$ if $\hat{\mathbf{i}}$ is the optical axis.

2. Compute for all points the longitude angle χ with

respect to the new pole $\hat{\mathbf{n}}_1$ and group the vectors with the same χ . Compute for every group the mean and the variance of

$$\frac{(\hat{\mathbf{p}} - \gamma(\hat{\mathbf{i}} \times \hat{\mathbf{p}}))^T \hat{\phi}_2}{\|\hat{\mathbf{p}} \times \hat{\mathbf{n}}_1\|}$$

and search for the minimum χ_{\min} of the variance. Divide the mean by $\|\hat{\mathbf{v}} \times \hat{\mathbf{i}}\|$ in order to obtain the inverse of the time to collision $\|\mathbf{v}\|/\|\mathbf{T}\|$. If χ_{\min} is near the border of the field of view then either carry out a saccade towards χ_{\min} or apply the nonlinear minimization described above.

5. EXPERIMENTAL RESULTS

We tested the proposed algorithms with synthetic as well as real data. Real data experiments were carried out using sequences recorded by passive as well as active cameras. In the nonfixated sequences we emulated the fixation by appropriately rotating the optical flow field. The fixated sequences were recorded using the TRC binocular camera mount. In all the experiments, the 1D search of the first step runs over 45 samples of the 180° η -range. The sampling interval for χ in the 1D search of the second step is 1°. If

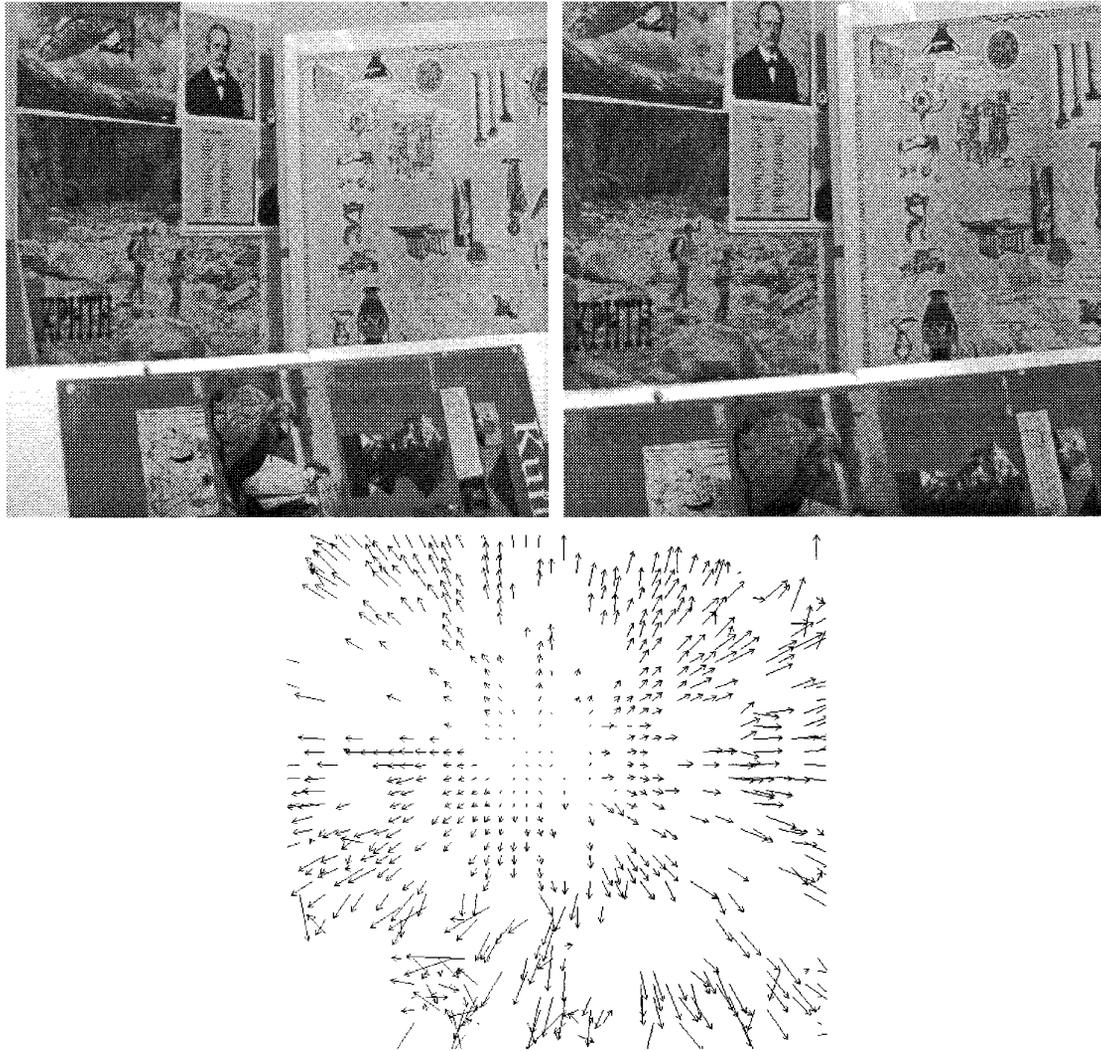


FIG. 10. The 1st and the 10th of the real fixated sequences (above) and the computed optical flow field (below).

the focus of expansion lies outside the field of view we replace the second step with the alternative nonlinear minimization method. This is done with the Levenberg–Marquardt method as implemented in the routine LMDER of the Netlib library.

We produce synthetic motion fields assuming a scene looking like a corridor. In the first experiment we assume a wide field of view of 90° and we apply translations $\mathbf{v} = (\sin \chi_{\text{gt}}, 0, \cos \chi_{\text{gt}})$, where χ_{gt} is the ground truth angle between translation and target direction. The latter is assumed to coincide with the optical axis. In this as well as all subsequent simulations it turns out that the error in the azimuthal angle η of the translation direction was under 2° and the relative error in the torsion γ is under 3%. Therefore, we will plot only the error in the χ angle and the inverse of the time to collision λ . In Fig. 3 we show the error in the angle χ for translation directions deviating from 5 to 40° from the target direction. The motion field

is corrupted by Gaussian noise with relative standard deviation of 10% and 20%. We tested for two torsion values 0 and 0.005, shown in the left and right sides of Fig. 3, respectively. We observe that the error increases with the deviation of the translation from the target direction and its behavior is not smooth in the presence of torsion. The same qualitative behavior is observed for the inverse of the time to collision $\lambda = \|\mathbf{v}\|/\|\mathbf{T}\|$ in Fig. 4.

The second synthetic experiment concerns a smaller field of view (45°) in the presence of torsion and relative optical flow error of 10%. Since the second step can be applied only for $\chi < 20^\circ$ in all steps we applied the nonlinear minimization with respect to χ and λ . The results (Fig. 5) are significantly better than the 1D search for even a larger field of view (see above) but with the additional cost of an iterative method. The same initial values were used in the nonlinear minimization for all translation directions.

In the following image sequences we computed the opti-

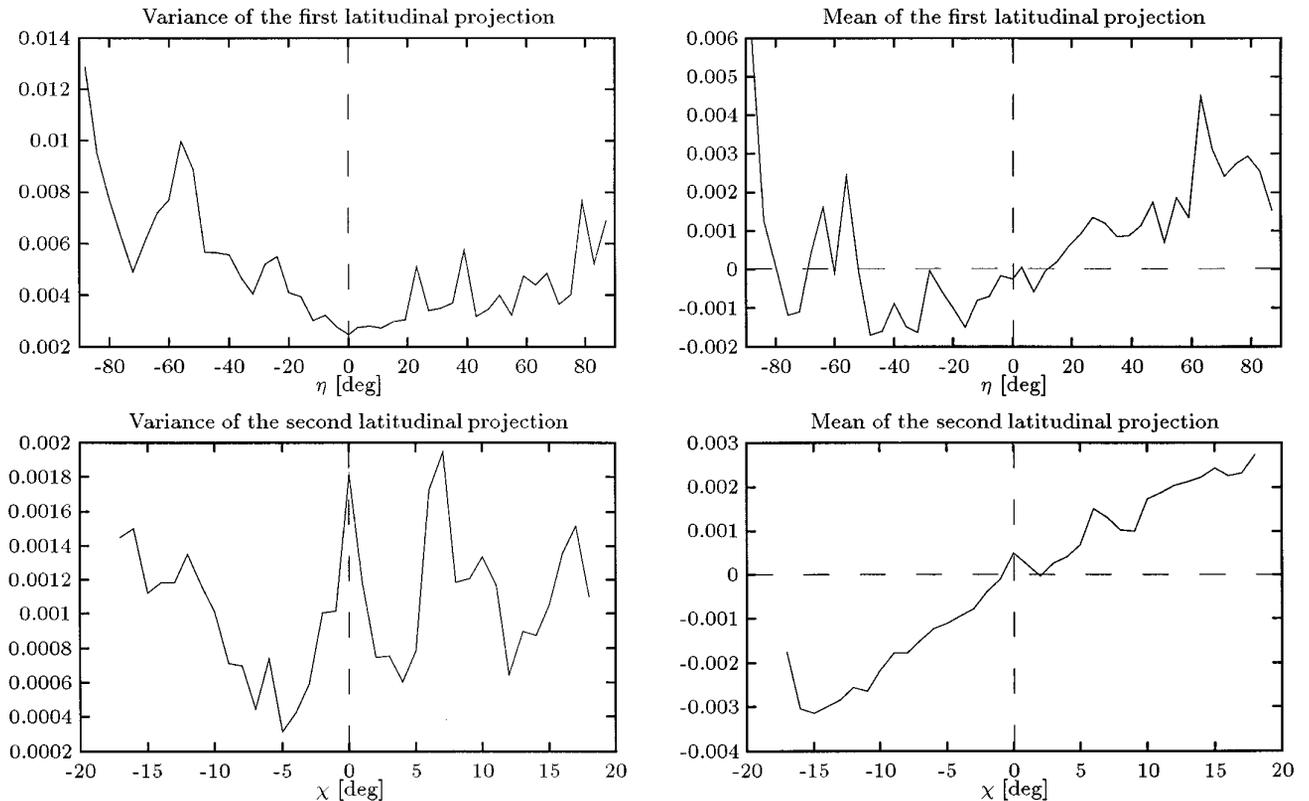


FIG. 11. The variance (top left) and the mean (top right) of the first latitudinal projection for the real fixated sequence. The minimum of the variance gives the angle η and the mean for this η gives the torsion. The variance (bottom left) and the mean (top right) of the second latitudinal projection for the real fixated sequence giving the angle χ at the minimum of the variance and the inverse of the time to collision, respectively.

cal flow with a standard differential method [23] which assumes a constant flow field in the local neighborhood of every pixel. The spatiotemporal derivatives are computed with binomial kernels which are approximations of the first derivative of a Gaussian. The computed flow field is first mapped to the plane $Z = 1$ using the intrinsic parameters and then transformed to a spherical flow field using (1).

The first sequence is the pure translational “marbled block” sequence [24] (Fig. 6). Fixation is achieved by adding a rotational flow field so that the flow in the image center vanishes (Fig. 6, bottom). We should note that the image center is given by the intrinsic calibration and does not coincide with the apparent center in the figure. The first step of our algorithm (Fig. 7, top) gives an η estimate of 6° and a torsion of 0.00032. The second step (Fig. 7, bottom) gives a minimum of the variance of the second latitudinal projection at the right limit of the interval, indicating, thus, a focus of expansion outside the field of view. Applying the nonlinear minimization we obtain the estimates $\chi = 42.14^\circ$ and $\lambda = 0.00387$. The ground truth values are $\eta = 7^\circ$ and $\chi = 35^\circ$.

The second sequence is the well-known synthetic Yosemite sequence (courtesy of Lynn Quam at SRI) which

contains both translation with ground truth ($\eta = 90^\circ$, $\chi = -9.84^\circ$) and rotation with ground truth $\omega = (0.00023, 0.00162, 0.00028)$. The original and the fixated flow field (Fig. 8) are computed only for the part of the image that contains ego motion (the clouds area is excluded). The minimum of the variance of the first latitudinal projection (Fig. 9, top) gives an η estimate of 97.37° and a torsion estimate of -0.00063 (the opposite sign is due to our formulation of scene motion). Since the minimum of the variance of the second latitudinal projection (Fig. 9, bottom) is at the limit of the field of view we again apply the nonlinear minimization for the second step and obtain $\chi = -5.96$ and $\lambda = 0.00145$.

The last sequence is already fixated during its recording with an active camera (Fig. 10). Up to the fixational movement the motion of the observer is pure translational with ground truth measured manually ($\eta_{\text{gt}} = 0^\circ$ and $\chi_{\text{gt}} = 9.2^\circ$). Because the focus of expansion is inside the field of view, solutions are obtained by applying both steps of our algorithm, yielding the estimates $\eta = -2^\circ$ and $\chi = 5^\circ$ (Fig. 11).

We emphasize here that no special effort was applied on optimizing the estimation process. In particular, no smoothing or weighting with measurement variances was

employed in the search along the meridians. Furthermore, the computation of the torsion and the time to collision can lead to instabilities due to the small amount of data contributing to the mean along each meridian. Because the main objective of this paper was the computation of the heading direction, the employment of a stable estimator over a larger area is a part of the planned extensions to this work. As already observed in the simulations the main error is in the deviation of the translation from the target direction in the second step. The observed robustness of the first step is consistent with the theoretical results by Maybank and Jepson [25, 26] in the case of general motion. Both proved that if the observed surface is irregular the line through the center and the focus of expansion can be robustly estimated.

6. CONCLUSION

It was proven in the past that fixation reduces the number of unknowns in the structure from motion problem from five to four. We showed in this paper that fixation can further simplify the computation of 3D motion parameters from a monocular sequence. Appropriate projections of the spherical flow field enable the decoupling of the motion parameters in two groups: The first contains the azimuthal angle of the translation and the torsion. The second parameter group contains the polar angle of the translation direction and the time to collision to the fixated target. Two 1D searches in each of the two projections yield all four unknowns. In contrast to other algorithms, we do not make any use of the measurements of camera movements necessary for fixation. We assume for the second search that the focus of expansion is inside the field of view. If this is not the case we can apply a two-unknowns nonlinear minimization or, even better, carry out a correcting saccade that will bring the focus of expansion inside the field of view. The algorithm was tested in three real world sequences fixated off-line or actively. Without applying any special method for accurate computation of the flow we obtained very promising results.

ACKNOWLEDGMENTS

The contribution of Inigo Thomas from the GRASP Laboratory, University of Pennsylvania, during his visit in Kiel is highly appreciated. I gratefully acknowledge the constructive discussions with Ruzena Bajcsy and Gerald Sommer.

REFERENCES

1. R. Bajcsy, Active perception, *Proc. IEEE* **76**, 1988, 996–1005.
2. Y. Aloimonos, Z. Duric, C. Fermüller, L. Huang, E. Rivlin, and R. Sharma, Behavioral motion analysis, in *DARPA Image Understanding Workshop, San Diego, CA, Jan. 26–29, 1992*, pp. 521–541.
3. D. Ballard and C. Brown, Principles of animate vision, *CVGIP: Image Understanding* **56**, 1992, 3–21.
4. R. H. S. Carpenter, *Movements of the Eyes*, Pion Press, London, 1988.
5. K. J. Bradshaw, P. F. McLauchlan, I. D. Reid, and D. W. Murray, Saccade and pursuit on an active head-eye platform, *Image Vision Comput.* **12**, 1994, 155–163.
6. D. Coombs and C. Brown, Real-time binocular smooth pursuit, *Int. J. Comput. Vision* **11**, 1993, 147–164.
7. K. Kanatani, *Geometric Computation for Machine Vision*, Oxford University Press, Oxford, UK, 1993.
8. A. Bandopadhyay and D. H. Ballard, Egomotion perception using visual tracking, *Comput. Intell.* **7**, 1990, 39–47.
9. Y. Aloimonos, I. Weiss, and A. Bandyopadhyay, Active vision, in *Proc. Int. Conf. on Computer Vision, London, UK, June 8–11, 1987*, pp. 35–54.
10. C. Fermüller and Y. Aloimonos, Tracking facilitates 3-D motion estimation, *Biol. Cybern.* **67**, 1992, 259–268.
11. C. Fermüller and Y. Aloimonos, The role of fixation in visual motion analysis, *Int. J. Comput. Vision* **11**, 1993, 165–186.
12. C. Fermüller and Y. Aloimonos, Qualitative egomotion, *Int. J. Comput. Vision* **15** 1995, 7–29.
13. M. A. Taalebinezhad, Direct recovery of motion and shape in the general case by fixation, *IEEE Trans. Pattern Anal. Mach. Intell.* **14**, 1992, 847–853.
14. K. J. Hanna, Direct multi-resolution estimation of ego-motion and structure from motion, in *Proc. IEEE Workshop on Visual Motion, Princeton, NJ, Oct. 7–9, 1991*, pp. 156–163.
15. D. Raviv and M. Herman, A unified approach to camera fixation and vision-based road following, *IEEE Trans. Systems, Man, and Cybernetics* **24**, 1994, 1125–1141.
16. I. Thomas, E. Simoncelli, and R. Bajcsy, Spherical retinal flow for a fixating observer, in *Proc. IEEE Workshop on Visual Behaviors, 1994*, pp. 37–41.
17. R. C. Nelson and J. Aloimonos, Finding motion parameters from spherical motion fields (Or the advantages of having eyes in the back of your head), *Biol. Cybernet.* **58**, 1988, 261–273.
18. D. Raviv and N. Ozery, A visual-motion fixation invariant, in *IEEE Conf. Computer Vision and Pattern Recognition, Seattle, WA, June 21–23, 1994*, pp. 188–193.
19. M. Tistarelli and G. Sandini, Dynamic aspects in active vision, *CVGIP: Image Understanding* **56**, 1992, 108–129.
20. K. Daniilidis, Computation of 3D-motion parameters using the log-polar transform, in V. Hlavac et al. (Ed.), *Proc. Int. Conf. Computer Analysis of Images and Patterns CAIP, Prag, 1995*, pp. 82–89.
21. M. J. Barth and S. Tsuji, Egomotion determination through an intelligent gaze control strategy, *IEEE Trans. Systems, Man, and Cybernetics* **23**, 1993, 1424–1432.
22. V. Sundareswaran, P. Bouthemy, and F. Chaumette, Active camera self-orientation using dynamic image parameters, in *Proc. Third European Conference on Computer Vision, Stockholm, Sweden, May 2–6 (J. O. Eklundh, Ed.)*, pp. 111–115. [LNCS, Vol. 800, Springer, 1994]
23. B. Lucas and T. Kanade, An iterative image registration technique with an application to stereo vision, in *DARPA Image Understanding Workshop, 1981*, pp. 121–130.
24. M. Otte and H.-H. Nagel, Optical flow estimation: advances and comparisons, in *Proc. Third European Conference on Computer Vision, Stockholm, Sweden, May 2–6 (J. O. Eklundh, Ed.)*, pp. 51–60. [LNCS, Vol. 800, Springer, 1994].
25. S. Maybank, *Theory of Reconstruction from Image Motion*, Springer-Verlag, Berlin, 1993.
26. A. D. Jepson and D. J. Heeger, *Subspace Methods for Recovering Rigid Motion II: Theory*, Technical Report RBCV-TR-90-36, University of Toronto, 1990.