On the Consistency of Instantaneous Rigid Motion Estimation

Tong Zhang
Mathematical Sciences Department
IBM T.J. Watson Research Center
Yorktown Heights, NY 10598
tzhang@watson.ibm.com

Carlo Tomasi
Department of Computer Science
Stanford University
Stanford, CA 94305
tomasi@cs.stanford.edu

Abstract

Instantaneous camera motion estimation is an important research topic in computer vision. Although in theory more than five points uniquely determine the solution in an ideal situation, in practice one can usually obtain better estimates by using more image velocity measurements because of the noise present in the velocity measurements. However, the usefulness of using a large number of observations has never been analyzed in detail. In this paper, we formulate this problem in the statistical estimation framework. We show that under certain noise models, consistency of motion estimation can be established: that is, arbitrarily accurate estimates of motion parameters are possible with more and more observations. This claim does not simply follow from the the general consistency result for maximum likelihood estimates. Some experiments will be provided to verify our theory. Our analysis and experiments also indicate that many previously proposed algorithms are inconsistent under even very simple noise models.

1 Introduction

In principle, the field of instantaneous velocities measured in the images produced by a moving camera contains enough information to determine the rotation and direction of translation of the camera, as well as depth, *i.e.* , the distances of scene points from the camera's center of projection.

In the absence of prior information about the unknown depth and motion parameters, this *motion field analysis* problem is most naturally cast as a statistical estimation problem. It turns out that the image motion field carries no information about the absolute depth of points in the world, nor about the magnitude of the translational velocity of the camera. A constraint is consequently imposed on the magnitude of the translational velocity, under the assumption that translation is nonzero. We thus face a nonlinear, constrained, statistical estimation problem.

The mathematics of motion field analysis has been studied thoroughly in the absence of noise, as outlined in section 3. Conditions for the existence and uniqueness of a solution are now well understood. However, this is an inverse problem: the calculation of image velocities from depth and motion is well-behaved, but the converse is ill-conditioned, in that small errors (henceforth referred to as *noise*) in the measurements of velocities can produce large errors in the estimates of depth and motion. One immediate consequence is that outliers in the image motion field measurements cannot be ignored, as a single outlier can play havoc with the solution. In addition, bias and variance can be amplified by inappropriate transformations of the problem formulation. When previous researchers have seeked insights and designed algorithms through linearization or various algebraic transformations of the original problem, they have exposed themselves to this danger.

Although in theory, five points or more can be sufficient to uniquely determine the motion parameters in the ideal situation, in practice, one should use a large number of velocity measurements to "average out" the effect of noise. However, no analysis has been carried out to study whether noise can be truly averaged out as hoped, and to understand the true behavior of motion estimation algorithms with a large number of velocity measurements. This paper addresses this issue. We show that under appropriate noise models, there exists a family of motion parameter estimation methods from sparse measurements of an image motion field that are consistent as the number of measurements increases. The noise model we consider in this paper is non-parametric in that it cannot be specified with a finite number of parameters. In addition, the family of estimation methods we propose do not correspond to maximum likelihood estimates, which require certain parametric assumptions on the noise distribution. It is also important to note that even under the assumption of a parametric noise model, the standard consistency argument for maximum likelihood estimate still cannot be applied directly to the motion problem, as explained in Section 4.

The paper is organized as follows. Section 2 defines the problem of motion field analysis, and section 3 sketches the history of previous work in the area. Section 4 contains the main result of this paper, and provides a rigorous statistical analysis to show the consistency of a class of motion estimation methods under moderate noise assumptions. Experiments will

be given in section 5 to illustrate the theoretical results in this paper. Section 6 summarizes the results and offers some final remarks.

2 Problem Statement

The image velocity caused by the motion of a camera with respect to a rigid scene under perspective projection is given by the following equation (see for instance [24], section 8.2.1):

$$\mathbf{u}(\mathbf{x}) = A(\mathbf{x}) \left(\frac{\mathbf{t}}{Z(\mathbf{x})} + \omega \times \mathbf{x} \right) . \tag{1}$$

In this equation, $\mathbf{u}(\mathbf{x})$ is the image velocity at image position $\mathbf{x} = (x_1, x_2, 1)^T$, $\mathbf{t}$ is the camera's translational velocity, ω is its rotational velocity, $Z(\mathbf{x})$ is the scene depth of the point imaged at $\mathbf{x}$, and the camera's focal length is taken without loss of generality to be 1. The matrix

$$A(\mathbf{x}) = \begin{bmatrix} 1 & 0 & -x_1 \\ 0 & 1 & -x_2 \end{bmatrix} \tag{2}$$

projects three-dimensional velocities onto a plane orthogonal to the camera's optical axis. The magnitude of the rows of $A(\mathbf{x})$ accounts for the variable distance of image points from the center of projection. The image motion analysis problem is to estimate the 3D motion parameters $\mathbf{t}$ and ω and the depths $Z(\mathbf{x})$ from a collection of velocity vectors sampled at some image positions. Because $\mathbf{t}$ and Z appear as a ratio in equation (1), their absolute magnitudes cannot be determined. We therefore add the constraint

$$\|\mathbf{t}\|_2 = 1 \tag{3}$$

under the assumption that $\mathbf{t} \neq 0$.[1]

Assume now that the velocity measurement $\mathbf{u}(\mathbf{x})$ is corrupted with noise $\mathbf{n}(\mathbf{x})$. Then, for convenience, we can rewrite the motion equation in the matrix form:

$$\mathbf{u}(\mathbf{x}) = p(\mathbf{x})A(\mathbf{x})\mathbf{t} + B(\mathbf{x})\omega + \mathbf{n}(\mathbf{x}), \tag{4}$$

[1] The implication of $\mathbf{t} = 0$ will be discussed later in the paper. For example, see discussions at the end of Section 4.3 and the end of Section 4.4, and the second example in Section 5.2.

under the constraint that $\|\mathbf{t}\|_2 = 1$, where

$$B(\mathbf{x}) = \begin{bmatrix} -x_1 x_2 & 1 + x_1^2 & -x_2 \\ -1 - x_2^2 & x_1 x_2 & x_1 \end{bmatrix},$$

and $p(\mathbf{x}) = 1/Z(\mathbf{x})$ is the inverse of depth. The quantities $\mathbf{t}, \omega, p(\mathbf{x})$ are the true parameters, and $\mathbf{n}(\mathbf{x})$ denotes noise. We call

$$\mathbf{r}(\mathbf{x}) = \mathbf{u}(\mathbf{x}) - \mathbf{u}_*(\hat{\mathbf{t}}, \hat{\omega}, \hat{p}(\mathbf{x})) \tag{5}$$

the *residual* between measured velocity $\mathbf{u}(\mathbf{x})$ and predicted velocity $\mathbf{u}_*$ from an estimate $(\hat{\mathbf{t}}, \hat{\omega}, \hat{p}(\mathbf{x}))$ of the true parameters.

In this paper, we regard motion estimation as the following statistical estimation problem: Consider m measured image velocities $\mathbf{u}_i$ at $\mathbf{x}_i$ $(i = 1, \ldots, m)$, where each coordinate $\mathbf{x}_i$ is taken from a fixed but unknown distribution D_{x_i}. At each measurement point $\mathbf{x}_i$, there holds the equality

$$\mathbf{u}_i = p_i A_i \mathbf{t} + B_i \omega + \mathbf{n}_i \qquad (i = 1, \ldots, m) \tag{6}$$

that corresponds to (4), where $\mathbf{n}_i$ is the noise at $\mathbf{x}_i$. We are mainly interested in the behavior of an estimation rule so that when $m \to \infty$, the estimated parameters $\hat{\mathbf{t}}, \hat{\omega}$ approach the true motion parameters $\mathbf{t}, \omega$ in probability under the normalization condition that $\|\hat{\mathbf{t}}\|_2 = \|\mathbf{t}\|_2 = 1$.

To further simplify the notation, we write (4) as

$$\mathbf{u}(\mathbf{x}) = C(\alpha_t, \mathbf{x}) + \mathbf{n}(\mathbf{x}),$$

where α_t denotes the true motion parameter $(\mathbf{t}_t, \omega_t)$, and

$$C(\alpha, \mathbf{x}) = p(\mathbf{x}) A(\mathbf{x}) \mathbf{t} + B(\mathbf{x}) \omega.$$

Figure (1 shows the terminology used in these equations. Note that we treat each $p(\mathbf{x})$ as a hidden variable, which later becomes irrelevant in the computation. Therefore we do not explicitly include $p(\mathbf{x})$ in the vector of unknowns $\alpha = (\mathbf{t}, \omega)$.

Throughout this paper, we use $\|\cdot\|$ to denote the 2-norm $\|\cdot\|_2$. For any two dimensional vector $\mathbf{a} = [a_1, a_2]$, we define its *normalized orthogonal direction* as $Q(\mathbf{a}) = [a_2, -a_1]/\|\mathbf{a}\|$. If $\mathbf{a} = 0$, then we shall just set $Q(\mathbf{a}) = [1, 0]$.

Statistically, it is not easy to analyze the motion equation (6) directly, since for each

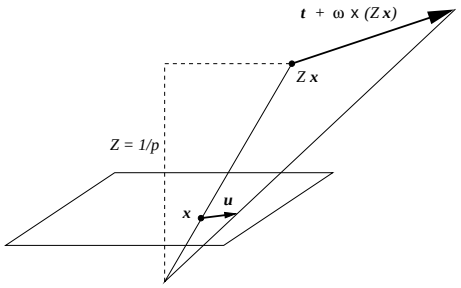


Figure 1: Projection $\mathbf{u}$ of the world motion onto the image. In the absence of noise, $\mathbf{u} = C(\alpha, \mathbf{x})$.

measurement point there is an associated unknown parameter p_i that is sensitive to the corresponding noise $\mathbf{n}_i$. For the same reason, it is not possible to obtain a good estimate for every p_i. If not treated appropriately, the uncertainty of p_i could also lead to additional uncertainty in the estimated motion parameters. This additional uncertainty can be difficult to analyze theoretically. In order to eliminate this undesirable effect, we shall eliminate p_i from (6), so that our analysis only focuses on the motion parameter α. Intuitively, we need to project (6) onto the direction that is orthogonal to $A_i\mathbf{t}$, which eliminates the p_i-dependent term in (6). Formally, we shall define this projection as $h(\alpha, \mathbf{x}, \mathbf{u}) = Q(A(\mathbf{x})\mathbf{t})^T(\mathbf{u} - B(\mathbf{x})\omega)$. It is easy to verify that the following ($p(\mathbf{x})$-independent) relation holds:

$$h(\alpha, \mathbf{x}, C(\alpha, \mathbf{x})) \equiv 0. \tag{7}$$

Furthermore, the function value $h(\alpha, \mathbf{x}, C(\alpha_t, \mathbf{x}))$ can be regarded as a closeness measure of the flow caused by a motion parameter α to the true flow $C(\alpha_t, \mathbf{x})$ of the true motion parameter α_t. To see this, note that $p(\mathbf{x})$ is not observable, therefore the closeness of the flow of α and the true flow $C(\alpha_t, \mathbf{x})$ at a spatial coordinate $\mathbf{x}$ can only be measured by $\inf_p \|pA(\mathbf{x})\mathbf{t} + B(\mathbf{x})\omega - C(\alpha_t, \mathbf{x})\|$. That is, we want to choose p such that the approximate flow $C(\alpha, \mathbf{x})$ and the true flow $C(\alpha_t, \mathbf{x})$ is closest in 2-norm.

This definition of closeness leads to $h(\alpha, \mathbf{x}, C(\alpha_t, \mathbf{x}))$:

$$h(\alpha, \mathbf{x}, C(\alpha_t, \mathbf{x})) = \inf_p \|pA(\mathbf{x})\mathbf{t} + B(\mathbf{x})\omega - C(\alpha_t, \mathbf{x})\|. \tag{8}$$

In summary, $h(\alpha, \mathbf{x}, C(\alpha_t, \mathbf{x}))$ measures the closeness of the flow implied by the computed motion parameter with the optimal depth parameter to the true flow.

Utilizing equation (7), we obtain the following estimation method for motion parameters:

$$\hat{\alpha} = \arg\min_{\alpha} \sum_{i=1}^{m} f(h(\alpha, \mathbf{x}_i, \mathbf{u}_i)), \tag{9}$$

where f is a known function which penalizes large values of $h(\cdot)$. In this paper, we are especially interested in the case that f has a form of $f(x) = |x|^q$ for $q \geq 1$. The idea is to penalize large values of $h(\alpha, \mathbf{x}_i, \mathbf{u}_i)$ so that the equality $h(\alpha, \mathbf{x}, C(\alpha, \mathbf{x})) = 0$ is approximately satisfied on average. The normalized projection used in the definition of h is important. It follows from the basic invariant principle and the philosophy of non-informative assumptions (or priors) in statistics. This philosophy implies that for the motion estimation problem, with minimum prior knowledge, we should not bias toward any specific projection direction in our formulation. Such a bias may lead to an inconsistent estimator, as we show later.

In this paper, we call $f(h(\alpha, \mathbf{x}, \mathbf{u}))$ a *loss function*. Equation (9) is defined through an average of the losses of the empirical (observed) data, therefore we call it an *empirical estimation*. We also call the corresponding empirical average of the losses *empirical risk*:

$$R_{emp}(\alpha) = \frac{1}{m} \sum_{i=1}^{m} f(h(\alpha, \mathbf{x}_i, \mathbf{u}_i)), \tag{10}$$

which can be regarded as an empirical goodness measurement for each potential motion parameter α.

Empirical risk is an approximation to the *true underlying risk*, where the average over observed samples in (10) is replaced by the following expectation over the unknown underlying distribution D of $(\mathbf{x}, \mathbf{u})$:

$$R(\alpha) = E_{(\mathbf{x},\mathbf{u})} f(h(\alpha, \mathbf{x}, \mathbf{u})), \tag{11}$$

where we use $E_{(\mathbf{x},\mathbf{u})}$ to denote the expectation with respect to the joint random variable $(\mathbf{x}, \mathbf{u})$.

The true risk $R(\alpha)$ can be regarded as the risk of a motion parameter associated with infinitely many samples. There is no randomness involved in the definition of $R(\alpha)$. On the other hand, empirical risk $R_{emp}(\alpha)$ can be regarded an approximation of (11) with a finite number of observations. It is a random estimator of $R(\alpha)$ which depends on the choice of empirical samples. In this regard, estimator (9) attempts to minimize approximately the true risk $R(\alpha)$ through a limited number of measurements. Therefore the consistency of this estimator (the formal definition of consistency is given at the beginning of Section 4), which

is the topic we address in this paper, includes the following two aspects:

- If α approximately minimizes (11), then α is close to the true motion parameter α_t. An estimator with this property is said to be *infinite-sample unbiased*.
- With large probability over random image velocity measurements (the probability approaches 1 as the sample size goes to ∞), if α minimizes (10), then α also approximately minimizes (11). An estimator with this property is said to exhibit *finite-sample convergence*.

3 Previous Work

The literature on the computation of camera motion from image velocities is rich in both the psychological and computational literature of vision. In the following, we make a few remarks about the measurement of the image motion field. We then sketch what is known about the motion analysis problem in the absence of noise, and finally survey previous work that explicitly addresses problems related to noise.

3.1 Motion Field Measurements

In a seminal book [3], Gibson coined the term "optical flow" to denote the apparent velocity of image points, which was intended to be the input to the computation of depth and camera motion. More recently, it was recognized that optical flow combines photometric and geometric aspects in complicated ways, and that the former can interfere with the latter. A thorough discussion of these issues can be found in [25]. Assuming that the motion field can be measured, say, by one of the methods surveyed in [1], the problem becomes a purely geometric one. It has been argued (see [19] for one of the first proposals in this sense) that this separation requires an assumption of flow smoothness, usually unwarranted, and that an approach that computes depth and motion directly from image intensities, without intervening flow computation or feature tracking, is preferable ([7] discusses this point in detail). However, direct methods replace flow smoothness with the equally controversial assumption that surface brightness is independent on viewing direction. Because of this, we prefer to separate photometric and geometric aspects, because the problem becomes simpler to understand. Lack of flow smoothness is not an issue if *sparse* motion field measurements are made. The locations for the measurements can be chosen carefully, say by the methods in [16] or [23], so as to lead to good estimates of the motion field.

Once motion field measurements are available, a further distinction is useful. Namely, image motion field analysis can be divided into the recovery of camera motion alone, followed by depth computation. Few image points (see subsection below) are sufficient for the recovery of camera motion, and less expensive and more reliable methods are then available for the subsequent computation of a denser depth map given the now-known camera motion, possibly over several frames (see for instance [14]). Nalwa [17] calls this two-stage approach the *bootstrap approach*. To appreciate the importance of this point, notice that one frame provides only two scalar measurements (the two components of image velocity) for each unknown depth value, so depth is only weakly constrained by a single frame. Depth values are more strongly constrained when several frames are used, assuming that the scene is stationary. A large number of velocity vectors from an individual motion field, on the other hand, can be used to determine camera motion.

In our work, we track point features from one frame to the next to measure approximate image velocity. In particular, we use the algorithm in [12] for tracking, and the method in [23] for selecting good measurement locations. Still, some features yield possibly grossly mistaken motion field estimates. This occurs particularly at depth discontinuities, where false features occur frequently. Unless better feature selection methods are devised, a motion field analysis method has to be able to cope with outliers like these.

3.2 Existence, Uniqueness, and Ambiguity of Solutions

In the absence of noise, image velocities at five points yield a finite number of depth and motion solutions. This follows by a simple limit argument from an early result by Kruppa [11], who gave a proof for finite image displacements rather than image velocities. With more points, the solution is essentially unique, except when points in the world are on a certain (possibly degenerate) hyperboloid of one sheet that depends on camera translation, called a *critical surface* [15], [6], [18]. In this case, two solutions are possible from the given flow.

Of course, all points in a scene are hardly ever on a critical surface, but the probability that they are in some sense close to one may not be negligible, especially because hyperboloids can degenerate into cylinders or pairs of planes, which are frequent surfaces in the world. In such a case, an algorithm may fail in two different ways. If two possible approximate solutions are similar to each other, their basins of attraction in the estimation may merge, and lead to a shallow extremum. If the two solutions are different, the optimization algorithm may find the wrong solution. This is a lesser problem, since the relation between the two solutions is known, and one can check both solutions to find the one with lower residual.

3.3 Previous Algorithms for Noisy Data

Image velocity measurements at more than five points are necessary if the data are noisy. Several algorithms have been proposed to solve this over-constrained, nonlinear minimization problem, and most transform the problem algebraically in order to obtain solutions that are either in semi-closed form or more efficient than a brute-force search over the set of all possible solutions. In this section, we briefly review a small but representative sample of these solutions. Since we consider algorithms based on image velocities, no mention is made of algorithms that assume point correspondences between possibly widely separated viewing positions.

Bruss and Horn [2] applied a simple algebraic manipulation to remove depth from the estimation problem, and obtained a residual $\mathbf{r}(\mathbf{x})$ that is bilinear in camera rotation and translation. However, they then simplified the expression of the residual for computational purposes. Their simplification is equivalent to replacing the residual term $\mathbf{r}(\mathbf{x})$ in equation (5) with

$$\mathbf{r}'(\mathbf{x}) = \mathbf{r}(\mathbf{x}) \| A(\mathbf{x}) \hat{\mathbf{t}} \| \tag{12}$$

where $\hat{\mathbf{t}}$ is the unit-norm camera velocity vector and $A(\mathbf{x})$ is the scaled projection matrix in equation (2). We will see in section 4 and 5 that this simplification can introduce severe bias into the solution. Later MacLean and Jepson [13] derived exactly the same bilinear residual as Bruss and Horn, but by applying a different algebraic manipulation. In either case, a least-squares estimate of both depth and rotation can be obtained as a function of translation. These estimates are substituted back into the bilinear residual to obtain a nonlinear function of translation alone. Translation is estimated by minimizing this nonlinear residual over all image velocities, subject to the constraint that the translational velocity has unit norm.

Rieger and Lawton [22] proposed a method based on motion parallax. If two 3D points have the same image location but are at different depths, then the vector difference between the two flow vectors is oriented toward the focus of expansion (FOE). The Rieger-Lawton algorithm locates the FOE from the local flow-vector differences. Hildreth [5] later modified the Rieger-Lawton algorithm to improve its performance. An obvious problem with both versions of this algorithm is that it is particularly difficult to measure flow vectors near occlusion boundaries.

Motion parallax is more general than the constraint used by Rieger and Lawton. Prazdny [20], for example, noted that the difference between any two (not necessarily adjacent) flow vectors gives a constraint on translation, independent of rotation. Jepson and Heeger built

upon these previous efforts and proposed a series of subspace methods for estimating egomotion [4, 8, 9]. The simplest of these is the so-called linear subspace method [8, 9]. Given optical flow sampled at N discrete points in the image, one can construct a set of constraint vectors that are shown to be orthogonal to the camera translation velocity. For N image velocity samples, there are $N-6$ constraint vectors. The translational velocity turns out to be the eigenvector corresponding to the smallest eigenvalue of a matrix suitably constructed from all the constraint vectors.

The advantage of the linear subspace method is that translational velocity is computed directly without requiring iterative numerical optimization. The disadvantage is that this method does not make use of all of the available information ($N-6$ linear constraints versus N bilinear constraints).

From a problem formulation equivalent to equation (1), Zhuang, Thomas, Ahuja, and Haralick [29] derived the so-called *epipolar constraint* on the velocity field:

$$\mathbf{t}^T(\mathbf{x} \times \mathbf{u}) + \mathbf{x}^T K \mathbf{x} = 0 \tag{13}$$

where K is a symmetric matrix with eigenvalues 1,1,0, related to camera motion by the following equation:

$$K = \omega^T \mathbf{t}\, I - \frac{1}{2}(\omega \mathbf{t}^T + \mathbf{t}\omega^T) .$$

This constraint can be shown [10] to mean that camera motion and the two projection rays of any given scene point before and after motion must be coplanar. Based on this instantaneous-time epipolar constraint, Zhuang, Thomas, Ahuja, and Haralick [29] proposed a linear algorithm for egomotion estimation.

Since camera motion and scene depth are nonlinear functions of image measurements, their estimates are systematically biased. In fact, a simple argument based on the Taylor series expansion shows that zero mean noise is almost invariably transformed into nonzero-mean noise by a nonlinear transformation. To see this, write the nonlinear transformation $T(n)$ from input noise to output noise as

$$T(n) = \sum_{k=0}^{\infty} c_k n^k \approx c_0 + c_1 n + c_2 n^2,$$

where the c_k are Taylor coefficients. We assume that the noise is relatively large so that second order term $O(n^2)$ is not negligible. In this case, even when $c_0 = 0$ and $E[n] = 0$, we

have

$$E[T(n)] \approx E[c_1 n + c_2 n^2] = c_2 E[n^2] \ ,$$

which is nonzero if $c_2 \neq 0$. Thus, nonlinear transformations introduce bias into the results. Kanatani [10] analyzed the statistical bias of image motion analysis with an argument essentially equal to the one above, and proposed a method (called *renormalization*) that subtracts an estimate of the output bias from the solution. The quality of Kanatani's results is hard to evaluate from that paper, since only one simulation is presented. More results are shown in section 5 below.

The somewhat disappointing results obtained in the literature motivated researchers to analyze the efficiency[2] of motion estimation [26] based on the Cramer-Rao lower bound [21]. This analysis led some researchers to believe [26] that computation of camera motion from instantaneous image velocities is unlikely to succeed. In this paper we show that this is not necessarily so. While the bias and variance problems discussed above do render motion estimation essentially impossible in some cases, this is shown to occur only at relatively narrow field-of-view angles, and with a limited number of motion field measurements, as in the simulations in [26]. As we will show next, under appropriate noise models, estimators based on (9) converge to the true motion parameters in probability, when the sample size m increases.

4 Consistency

In the previous section, we have argued that in general, with a limited number of image velocity measurements, the computed motion parameters $\mathbf{t}$ and ω will be systematically biased. However, this bias can decrease as the sample size m tends to infinity. To characterize the performance of a statistical estimation algorithm, it is thus important to study its large-sample behavior. In this regard, a fundamental problem for a statistical estimation algorithm is its consistency. For motion estimation problems, we can formally define consistency as follows.

Definition 1 *Let $\hat{\alpha}(\{(\mathbf{x}_i, \mathbf{u}_i)\}_{i=1,\ldots,m})$ be any estimator of the true motion parameter α_t based on m random velocity measurements $(\mathbf{x}_i, \mathbf{u}_i)$ $(i = 1, \ldots, m)$. We say that $\hat{\alpha}$ is consistent if for any $\epsilon > 0$, and $\eta > 0$, there exists a number M such that as long as the sample size $m > M$, then with probability at least $1 - \eta$ over m random velocity measurements $(\mathbf{x}_i, \mathbf{u}_i)$*

[2]Efficiency can be interpreted as the average closeness of an estimator to the true motion parameter.

($i = 1, \ldots, m$), the estimated motion parameter is within distance ϵ of the true motion parameter α_t:

$$\mathcal{P}\left[\|\hat{\alpha}(\{(\mathbf{x}_i, \mathbf{u}_i)\}_{i=1,\ldots,m}) - \alpha_t\| \leq \epsilon\right] \geq 1 - \eta.$$

Intuitively speaking, the consistency of a motion estimation algorithm implies that when we use more and more image velocity measurements, we have greater and greater chance to obtain an accurate estimate of motion parameter. As mentioned at the end of Section 2, consistency includes two components: infinite-sample unbiasedness and finite-sample convergence.

4.1 Noise model

From the statistical point of view, if our goal is to estimate the parameter of a distribution in a parametric family (where the number of parameters should be fixed) based on random samples drawn from the distribution, it is well-known that under quite moderate assumptions, the maximum likelihood estimate (MLE) for the corresponding parametric distribution is consistent. Furthermore, it is also asymptotically a most efficient unbiased estimator as far as the Cramer-Rao lower bound is concerned. The motion estimation problem can be casted as a maximum-likelihood estimate problem if the distribution of noise $\mathbf{n}(\mathbf{x})$ has a known form. For example, if we assume an iid [3] noise model with density proportional to $\exp(-\lambda f(\|\mathbf{n}\|))$ for a fixed parameter $\lambda > 0$, then it can be verified that equation (9) is equivalent to the MLE for this noise model, which is of the following form:

$$\inf_{p, \mathbf{t}, \omega} f(\|\mathbf{u}_i - (p_i A_i \mathbf{t} + B_i \omega)\|). \tag{14}$$

Unfortunately, even under this noise model assumption, the standard conditions that ensure the consistency of maximum likelihood estimate are not satisfied for the motion estimation problem. The reason is that in the MLE formulation (14), each p_i appears as a distribution-related unknown. This implies that in the original formulation, the number of unknowns is proportional to the number of samples. On the other hand, formulation (9) itself, which eliminates p_i, cannot be regarded directly as a maximum likelihood estimate. Otherwise, the corresponding noise distribution would have a density of the form $\propto \exp(\lambda f(Q(A(\mathbf{x})\mathbf{t}_t)^T \mathbf{n}))$, which has a strange dependency on the true motion parameter $\mathbf{t}_t$ [4]. Such a noise model is not even well-defined mathematically since it would imply the

[3] Independent, identically distributed.

[4] This density follows from the fact that the inverse depth p is chosen so as to make the residual orthogonal

unsatisfiable condition that $\mathbf{n}$ and $\mathbf{n} + sA(\mathbf{x})\mathbf{t}_t$ have the same density for all real numbers s.

From a more technical point of view, even if we accept the above ill-defined noise model and regard equation (9) as the corresponding MLE in a non-natural way, the standard consistency proof of maximum-likelihood estimate still cannot be applied to (9), due to the discontinuity of $Q(A(\mathbf{x})\mathbf{t})$ — this technical issue has been carefully treated in the proof of Theorem 4.

Another important difference between our analysis and that of MLE is that we consider a relatively general non-parametric noise model, which cannot be directly handled by the maximum likelihood method. In our model, we do not assume any fixed parametric form for the noise distribution. We only impose certain restrictions on the noise distribution. This is useful for two reasons: the noise distribution is more general, which means that results we obtain will be more general; we are able to derive a family of different consistent motion estimation algorithms under the same noise model.

A reasonable noise model is necessary for the consistency analysis. For example, assume that α_t is the true motion, but noise is added so that the observed "noisy" flow exactly matches that of a different motion parameter α_w. In this case, without any prior knowledge of the noise, it is perfectly valid to consider the estimate α_w as the true motion parameter. This suggests that if the noise is added so as to bias toward a specific motion configuration, then without any prior knowledge, it is impossible to obtain consistent estimators. We will thus need to make reasonable and moderate assumptions on the noise model so that it does not introduce a bias toward any specific motion configuration.

Specifically, the following assumptions are essential in our analysis: each noise term $\mathbf{n}_i$ in (6) is taken from a rotationally symmetric (isotropic) distribution, and independent from one another. However, we do not assume that noise distributions at different image coordinates are identical. Under this noise model (together with a number of less important technical assumptions which will be introduced later), we show that (9) is consistent. We also demonstrate that the formulations used by Bruss and Horn as well as some others are inconsistent under this simple noise model. This analysis demonstrates that although numerous methods can be proposed through algebraic manipulations to the rigid motion equation under the noiseless assumption, these manipulations can be potentially very dangerous when we take the effects of noise into consideration.

Before we go into the technical details of the proofs, we would like to briefly justify the assumptions we have imposed on the noise distribution. Although, strictly speaking, the independence assumption is not exactly realistic, it is a quite standard assumption to

to $A(\mathbf{x})\mathbf{t}_t$, that is, parallel to the normalized orthogonal direction $Q(A(\mathbf{x})\mathbf{t}_t)$.

simplify statistical analysis since statistical principles such as the law of large numbers can be applied. In practice, even if there exists moderate dependency among observations, similar statistical principles may still be valid (approximately), and hence our analysis can still be applied to provide useful insights.

The more subtle problem is the assumption of rotational symmetry. Because of the aperture problem, positional uncertainty in images is not truly isotropic. However, a feature detector like the one in [23] leads to uncertainties that are very close to isotropic. Also from a statistical point of view, even when the noise is biased toward a certain direction locally, the derivations in this section will be valid as long as the bias directions are globally isotropic on average. Practically, this can be non-rigorously interpreted in the following way: if the overall effect of local noise biases does not favor any specific motion parameter, then the noise has an effect similar to that of the isotropic model.

We emphasize that the rotational symmetry of noise is very important in our analysis: it is used to demonstrate that the proposed formulation is infinite-sample unbiased. One explanation of the inconsistency of the formulation used by Bruss and Horn is that such a formulation breaks this symmetry. Even though it is possible to accommodate non-isotropic noise models into our formulation, provided that such information is known, it is reasonable to assume that noise is isotropic without any prior knowledge. As we have pointed out earlier, this conforms with the philosophy of non-informative assumptions in Bayesian statistics.

Under the isotropic noise model, in the following, we prove the consistency of motion estimation based on (9) by analyzing its two components that are mentioned at the end of Section 2: infinite-sample unbiasedness and finite-sample convergence.

4.2 Infinite-sample unbiasedness

In this section, we investigate the first component of consistency of the proposed motion estimation formulation (9) — infinite-sample unbiasedness. That is, we show that as long as we can find α that approximately minimize the true risk (11), α is close to α_t.

We start our analysis with the following theorem showing that with isotropic noise and a convex function f, the true motion parameter achieves the minimum true risk defined in (11). This is the first step to demonstrate that the proposed schemes are infinite-sample unbiased.

Theorem 1 *Assume that the noise distribution of* $\mathbf{n}(\mathbf{x})$ *is rotational symmetric for any given* $\mathbf{x}$*, and assume that* f *is symmetric and convex. Then, the true motion parameter* α_t *achieves the minimum of* $R(\alpha) = E_{(\mathbf{x},\mathbf{u})} f(h(\alpha, \mathbf{x}, \mathbf{u}))$.

Proof. See Appendix A. □

The above theorem implies that the true motion parameter also minimizes the true risk with respect to the true distribution defined in (11). This is a very important step in our analysis since if the true motion parameter α_t does not even minimize the true risk, then a scheme based on the minimization of (11) will not be consistent.

In order to show infinite-sample unbiasedness, we shall introduce a *noise-free risk* with respect to the true flow $C(\alpha_t, \mathbf{x})$ as:

$$R_{free}(\alpha) = E_{(\mathbf{x})} f(h(\alpha, \mathbf{x}, C(\alpha_t, \mathbf{x}))). \tag{15}$$

Note that by (7), the minimum of $R_{free}(\alpha)$ is $f(0)$ which can be achieved at α_t. This definition is very useful since we can now separate the problem into two parts as a noise dependent analysis plus a noise free analysis:

1. We would like to show that if α approximately minimizes the true risk $R(\alpha)$ in (11) which is noise-model dependent, then α also approximately minimizes the noise-free $R_{free}(\alpha)$ which is not noise-model dependent.

2. We would like to demonstrate that if α approximately minimizes the noise-free risk $R_{free}(\alpha)$ in (15), then α is also close to α_t. This case can in fact be regarded as the noise-free scene ambiguity analysis, which has been well studied in the literature.

The first step of the remaining analysis requires us to define the residual risk of a parameter α as $\Delta R(\alpha) = R(\alpha) - R(\alpha_t)$. By Theorem 1, we have an inequality $\Delta R(\alpha) \geq 0$ for all α. The following theorem bounds the noise-free risk of a motion parameter α in terms of the residual risk of α:

Theorem 2 *Under the assumptions of Theorem 1, then* $\forall \epsilon_0 > f(0)$,

$$R_{free}(\alpha) \leq \epsilon_0 + \frac{\Delta R(\alpha)}{\inf_{\mathbf{X}} P(f(\mathbf{n}_1) \leq \epsilon_0 | \mathbf{x})},$$

where $\mathbf{n}_1$ *is the first component of the noise.*

Proof See Appendix B. □

Intuitively speaking, the quantity $\inf_{\mathbf{X}} P(f(\mathbf{n}_1) \leq \epsilon_0 | \mathbf{x})$ in Theorem 2 is related to the probability of noise $\mathbf{n}$ that is small: $P(\|\mathbf{n}\| \leq f^{-1}(\epsilon))$ where $f^{-1}(\epsilon)$ is the largest z such that $f(z) \leq \epsilon$.

Definition 2 *We say that small noise is uniformly nonvanishing if* $\forall \epsilon > 0, \inf_{\mathbf{X}} P(\|\mathbf{n}\| \leq \epsilon | \mathbf{x}) > 0$.

Thus, small noise is uniformly nonvanishing if the probability of finding arbitrarily small noise values anywhere in the image is nonzero. In this case, the right hand side of the bound in Theorem 2 can approach the minimum $f(0)$ as $\Delta R(\alpha) \to 0$, since $\forall \epsilon_0 \geq f(0)$ $\inf_{\mathbf{x}} P(f(\mathbf{n}_1) \leq \epsilon_0 | \mathbf{x}) > 0$.

Since we deal with a general convex function f in Theorem 2, the property of nonvanishing small noise is required. The reason can be illustrated through the following intuitive example: consider a simple regression model: $x_i = \theta_t + n_i$ where $\theta_t = 0$ and n_i is symmetric noise. If we estimate θ_t by minimizing $\sum_i |x_i - \theta|$, then if small noise n_i vanishes, say $n_i = \pm 1$, then clearly the minimum is achieved for all $\theta \in [-1, 1]$.

However, this example also implies that this condition is required only for convex functions f that contain flat segments, such as the 1-norm $f(x) = |x|$. For a convex function that does not contain a flat segment, (such as $f(x) = |x|^q$ for $q > 1$ which is what we are interested in), the condition of nonvanishing small noise can be removed. However, for simplicity, we shall not go into such technical details which are not essential in our discussion. One shall just keep in mind that although we state our results with the assumption of non-vanishing small noise, this assumption is not essential. As a simple example, we consider the square loss function $f(x) = x^2$. Observe that $[f(a+b)+f(a-b)]/2 = f(a)+f(b)$. We now consider Lemma 1 in Appendix B, and can set $k(a) = 1$ and $\rho(b) = f(b)$. Since equality holds in this case, the following result follows directly from the proof of Lemma 1: *Under the assumptions of Theorem 1, and for the least squares formulation $f(x) = x^2$, then $R_{free}(\alpha) = \Delta R(\alpha)$.*

Theorem 2 essentially relates the approximate minimization of (11) with the noisy flow to the approximate minimization of R_{free} with the noise-free image velocity measurement $\mathbf{u}_t = C(\alpha_t, \mathbf{x})$. As we have pointed out earlier, in addition to Theorem 2, in order to provide conditions so that a formulation based on minimizing (11) is infinite-sample unbiased, we need to demonstrate that motion estimation based on the minimization of the noise-free risk (15) is numerically stable. This is equivalent to say that if $h(\alpha, \mathbf{x}, C(\alpha_t, \mathbf{x})) \to 0$ throughout coordinates, then $\alpha \to \alpha_t$.

Since the corresponding analysis is independent of any noise model assumption, in this work, we regard it as an aspect of the noise-free scene ambiguity analysis which has been widely studied in the earlier literature, referenced in Section 3.2. In order to emphasize the main contributions of our work, which is noise dependent, we shall skip further investigation on this noise-free ambiguity issue, and simply list the relevant assumptions on lack of ambiguity below.

Definition 3 *We call a scene* non-ambiguous *if $E_{\mathbf{x}} f(h(\alpha, \mathbf{x}, C(\alpha_t, \mathbf{x})) = f(0)$ (with constraint $\|\mathbf{t}\| = 1$) has a unique solution.*

We call a scene stably non-ambiguous *if* $\forall \epsilon > 0$, $\exists \delta > f(0)$ *such that* $E_{\mathbf{X}} f(h(\alpha, \mathbf{x}, C(\alpha_t, \mathbf{x}))) < \delta$ *(with* $\|\mathbf{t}\| = 1$*) implies that* $\|\alpha - \alpha_t\| < \epsilon$.

We call a scene absolutely non-ambiguous *if for all camera motion* $[\mathbf{t}_0, \omega_0]$ *(with* $\|\mathbf{t}_0\| = 1$*),* $E_{\mathbf{X}} f(h(\alpha, \mathbf{x}, C([\mathbf{t}_0, \omega_0], \mathbf{x})) = f(0)$ *has a unique solution* $\alpha = [\mathbf{t}_0, \omega_0]$.

Note that if we make a reasonable choice of f such that $f(x) > f(0)$ for all $x \neq 0$, then the condition

$$E_{\mathbf{X}} f(h(\alpha, \mathbf{x}, C(\alpha_t, \mathbf{x})) = f(0)$$

is equivalent to the condition that $h(\alpha, \mathbf{x}, C(\alpha_t, \mathbf{x})) = 0$ almost everywhere. Such a choice of f is also required for the validity of stable non-ambiguity, which is an essential requirement in our analysis. Since $h(\alpha, \mathbf{x}, C(\alpha_t, \mathbf{x}))$ measures the closeness of the flow with motion α and the true flow $C(\alpha_t, \mathbf{x})$, as outlined in (8), the stable non-ambiguity requirement of the scene is equivalent to saying that if a flow generated from some camera motion is close to the true motion flow throughout the scene, then the motion itself is also close to the true motion.

The following theorem summarizes the basic results of this section. As we have pointed out earlier, the nonvanishing small noise assumption is not essential.

Theorem 3 *Under the assumptions of Theorem 1, and with the further assumption that*

1. *The small noise is uniformly nonvanishing.*

2. *The scene is stably non-ambiguous.*

Then motion estimation based on minimization of (11) is infinite-sample unbiased: $\forall \epsilon > 0$*, there exists* $\eta > 0$ *such that if* α *approximately minimizes (11) as* $\Delta R(\alpha) \leq \eta$*, then* $\|\alpha - \alpha_t\| \leq \epsilon$.

4.3 Finite-sample convergence

In this section, we investigate the second component of consistency of the proposed motion estimation formulation (9) — finite-sample convergence. That is, with large probability over random image velocity measurements (the probability approaches 1 as the sample size goes to ∞), if α minimizes (10), then α also approximately minimizes the true risk (11). We shall first introduce the following definition which is used in Theorem 4.

Definition 4 *We call a scene* non-degenerate *if there exists no solution* $\alpha = [\mathbf{t}, \omega]$ *of* $E_{\mathbf{X}} f(h(\alpha, \mathbf{x}, 0)) = f(0)$ *such that* $\mathbf{t} \neq 0$ *and* $\omega \neq 0$.

In other words, the only acceptable explanation of zero flow (see the third argument of h) for a non-degenerate scene is zero camera motion. Note that this definition is also independent of any noise model assumption, thus can be regarded as a specific aspect of the noise-free scene ambiguity analysis which we do not consider in this paper. It is also easy to see from the definition that if a scene is absolutely non-ambiguous, then it is also non-degenerate. We are now ready to prove the finite-sample convergence result: under certain technical assumptions, the true risk $R(\hat{\alpha})$ of $\hat{\alpha}$ obtained from (9) with a finite number of samples converges to the minimum true risk $R(\alpha_t)$ in probability, as the sample size $m \to \infty$. For simplicity, we shall assume that the loss function is of the form $f(x) = |x|^q$ $(q \geq 1)$.

Theorem 4 *Assume that the scene is non-degenerate, and*

1. $\lim_{\epsilon \to 0} \sup_{\mathbf{y}} P(\mathbf{x} : \|\mathbf{x} - \mathbf{y}\| < \epsilon) = 0$.
2. $E_{\mathbf{x}} f(\|\mathbf{x}\|^2) < \infty$.
3. $\sup_{\mathbf{x}} E_{\mathbf{n}|\mathbf{x}} f(\|\mathbf{n}\|) < \infty$.
4. $\sup_{\mathbf{x}} p(\mathbf{x}) < \infty$.

Then $\forall \epsilon > 0$, $\lim_{m \to \infty} P(\Delta R(\hat{\alpha}) < \epsilon) = 1$, *where* $\hat{\alpha}$ *is obtained from (9), with* $f(x) = |x|^q$ $(q \geq 1)$.

Proof. See Appendix C. □

Next, we would like to discuss the assumptions in Theorem 4. Assumption 1 is non-essential. It can in fact be removed by a more careful analysis. It is provided only to simplify the proof. This assumption requires that the density of $\mathbf{x}$ cannot be concentrated (like a delta function) at a single point. This in turn guarantees that any "bad" feature point won't contribute too much to the solution. Assumption 2 is important in our current proof since it prevents the dominance of large $\mathbf{x}$ which causes large variance in the solution. An intuitive interpretation of this condition is to avoid the large field-of-view situation which is known to cause problems in motion estimation. Assumption 3 is quite natural. It prevents the noise at a particular point from overly influencing the solution. This assumption is related to the robustness of the motion estimator in (9): by Jensen's inequality, the condition $\sup_{\mathbf{x}} E_{\mathbf{n}|\mathbf{x}} \|\mathbf{n}\|^{q_1} < \infty$ is weaker than $\sup_{\mathbf{x}} E_{\mathbf{n}|\mathbf{x}} \|\mathbf{n}\|^{q_2} < \infty$ when $q_1 < q_2$. This implies that it is more robust to use a smaller q in $f(x) = |x|^q$ in the sense that the resulting method tolerates larger noise. However, similarly to Assumption 1, condition 3 can be weakened with a more careful analysis. Assumption 4 prevents a 3D-world point to be too close to the center of projection, and cause an exceedingly large motion field – note that this situation

may happen with a small field of view. In summary, assumptions in Theorem 4 are intuitively very sensible: they reflect conditions such as the effective field of view used in the estimation and the robustness of the estimator.

It is also important to mention that in Theorem 4 the restriction of setting $f(x)$ in the form $|x|^q$ ($q \geq 1$) is non-essential. This assumption only simplifies certain stages of the proof. Furthermore, the convexity condition on $f(x)$ can also be removed if we make additional assumptions on the noise distribution.

At this stage, we can combine the results of infinite-sample unbiasedness and finite-sample convergence to prove the consistency of motion estimation based on (9).

Theorem 5 *Under the assumptions of Theorem 1 and Theorem 4, assume further that*

1. *small noise is uniformly nonvanishing.*

2. *the scene is stably non-ambiguous.*

Then $\forall \epsilon > 0$, $\lim_{m\to\infty} P(\|\hat{\alpha} - \alpha_t\| < \epsilon) = 1$, *where* $\hat{\alpha}$ *is obtained from (9), with* $f(x) = |x|^q$ ($q \geq 1$).

Proof. $\forall \epsilon > 0$, by Assumption 1, Assumption 2, and Theorem 2, $\exists \delta > 0$ such that $\|\hat{\alpha} - \alpha_t\| > \epsilon$ implies that $\Delta R(\hat{\alpha}) > \delta$. Therefore $P(\|\hat{\alpha} - \alpha_t\| > \epsilon) \leq P(\Delta R(\hat{\alpha}) > \delta)$. Now by Theorem 4, we know that $\lim_{m\to\infty} P(\Delta R(\hat{\alpha}) > \delta) = 0$, therefore $\lim_{m\to\infty} P(\|\hat{\alpha} - \alpha_t\| > \epsilon) = 0$. □

Although we have considered the consistency problem, we have not studied efficiency and robustness issues in depth. A detailed analysis requires additional information on the noise distribution besides our simple isotropic assumption. Such a study is beyond the scope of this paper. We shall simply mention that if we let $f(x) = |x|^q$, then a smaller q leads to more a more robust estimator since large noise (outliers) has less malicious impact. As we have mentioned earlier, this point is already reflected in Assumption 3 (and Assumption 2 to a lesser degree) of Theorem 4. We include a simulation example in Section 5 to illustrate this assertion. Furthermore, real image sequence examples in Section 5 also indicate that in the presence of outliers, using a small q in $f(x) = |x|^q$ can make a significant difference.

Finally, we discuss the role of $\mathbf{t}_t \neq 0$ in our analysis. Obviously, if $\mathbf{t}_t = 0$, then it does not have a well-defined direction. Therefore it is impossible to obtain an estimate $\hat{\mathbf{t}}$ that is consistent. However, the estimate of the rotational parameter ω will still be consistent. To see this, we shall note that the only part in our analysis that truly requires the assumption of $\mathbf{t}_t \neq 0$ is in the definition of stable non-ambiguity of the scene, which is the noise-free part of the analysis. In the case of $\mathbf{t}_t = 0$, as long as the scene is stably non-ambiguous with respect to ω, then the estimation of ω in (9) will still be consistent.

4.4 Inconsistency of some previous motion estimation methods

Compared with (9), the Bruss-Horn approach is equivalent to the minimization of

$$\sum_{i=1}^{m} f(h(\alpha, \mathbf{x}_i, \mathbf{u}_i)\|A(\mathbf{x})\mathbf{t}\|_2)$$

with $f(x) = x^2$. Instead of using the normalized definition of $Q(\mathbf{a})$ as in our formulation, they essentially employed an unnormalized definition of orthogonal projection so that $\|Q(\mathbf{a})\|_2 = \|\mathbf{a}\|_2$, which breaks the rotational symmetry. The fact that our formulation is infinite-sample unbiased (and in addition, consistent) implies that the Bruss-Horn formulation will bias toward a translation direction so that $\|A(\mathbf{x})\mathbf{t}\|_2$ is small with the constraint $\|\mathbf{t}\|_2 = 1$, even in the infinite-sample case, (and more so in the finite sample case). This implies that their formulation will be inconsistent.

By the definition of $A(\mathbf{x})$, if the selected coordinates (x_1, x_2) are small compared to 1, then increasing the third component of $\mathbf{t}$ tends to give a smaller value of $\|A(\mathbf{x})\mathbf{t}\|_2$; on the other hand, if many of the selected coordinates (x_1, x_2) are large compared to 1, then decreasing the third component of $\mathbf{t}$ tends to give a smaller value of $\|A(\mathbf{x})\mathbf{t}\|_2$. Since the former situation occurs when the camera's field-of-view (fov) is small, and the second situation occurs when the fov is large, the translation direction computed by the Bruss-Horn formulation will bias toward forward motion with a small fov and will bias toward side motion with a large fov.

The above argument relies on the assumption that the modified formulation (9) is infinite-sample unbiased and consistent, which we have just proved. The conclusion of this analysis will be verified in Section 5 by experiments.

Since the linear subspace method introduced by Jepson and Heeger[8, 9] starts with the same formulation as that of Bruss and Horn, it suffers from the same bias behavior as the latter. It is also not hard to see that the formulation (13) by Zhuang, et. al. in [29] also suffers from the same problem, due to the term $\mathbf{t}^T(\mathbf{x} \times \mathbf{u})$, which again breaks the rotational symmetry.

Although the method obtained by Kanatani in [10] is statistically more sensible, it starts from an inconsistent formula. It is hard to analyze how good the bias corrected formulation is, but we shall demonstrate by experiments in Section 5 that while the bias is not totally removed, the formulation introduces a larger variance than methods based on (9).

In our experiments, one can notice that the effect of any bias in the translation parameter will be compensated by a bias in the rotation parameter, so that the overall motion field remains similar. It is also interesting to mention that if the camera translation is small compared with the camera rotation, then the translation bias of the above algorithms will

have a very small impact on the optical flow field, hence the corresponding rotation bias will be small. In the extreme case of zero-translation, the bias caused by the translation could have a negligible effect on the rotation estimate. Therefore rotation estimate in this case could be unaffected by the potential translation bias. This claim will be verified by a synthetic-image experiment.

5 Experiments

We would like to verify our theoretical results by some experiments, and to demonstrate the importance of consistency by comparing different camera motion estimation methods with each other. In this section, the following algorithms will be compared:

1. Bruss-Horn: the Algorithm in [2] with least square minimization.
2. Jepson-Heeger: the linear subspace method in [8, 9].
3. Kanatani: the renormalization method in [10].
4. RM-L2: the least square formulation with $f(z) = z^2$ in (9).
5. RM-L1.2: the robust formulation with $f(z) = |z|^{1.2}$ in (9).

The algorithm we used to solve RM-L2 and RM-L1.2 has been described in [28].

In the figures that follow, the directions of camera translation and the axes of camera rotation are plotted on a hemisphere and projected to a circle, which is scaled to be uniform from 0° – 90° in the radial direction. The true camera translation and rotation directions are denoted by symbol $\circ$ and $\times$ respectively on each plot. The estimated camera translation directions are plotted as black dots (one dot per run). The cluster size of the black dots (individual translation direction estimates) thus gives a good indication of the variance of the underlying motion estimation method. Individual estimated rotation directions are not plotted. This makes the figures appear less cluttered. In addition, the cluster center of the estimated translation directions is denoted as $\triangleright$; the cluster center of the estimated rotation directions is denoted as $+$. They can be compared with the true translation and rotation directions to illustrate the corresponding biases. Magnitudes of the rotations in degrees are also reported ("$|R|$" as the computed average over all runs in each plot, "true $|R|$" as the true magnitude of rotation). We have also included an instance of optical flow field to illustrate the observed camera motion field.

The tables report estimation errors using the format of "error mean $\pm$ error standard deviation". Error for translation direction is defined as the angle in degrees between the

estimated direction and the true direction. Error for rotation is defined as the 2-norm of the difference of the estimated rotation and the true rotation, measured in degrees.

5.1 Simulated flows

In each experiment, 100 random features are used, unless otherwise stated. The scene contains randomly generated 3D positions with depth uniformly distributed between 1 and 4 units of focal length. The projected image coordinate $\mathbf{x}$ is uniformly distributed in the image of size 512×512 pixels. A hundred runs (each run with 100 different random features) are made in each experiment, and the reported results are highly repeatable. For all experiments, we fix the true translation direction as $[4, -3, 5]$ and the true rotation as $[-1, 2, 0.5]$ which corresponds to an angular velocity of 2.39°/frame. Independent Gaussian noise is added unless otherwise stated. The flows are very similar in these experiments. This implies that it is useful to measure the size of noise by signal to noise ratio (SNR): $(E\|\mathbf{u}_t\|_2^2)^{1/2} : (E\|\mathbf{n}\|_2^2)^{1/2}$.

Our first experiment assumes fov=50°. Noise $\sigma = 0.5$ pixels, which leads to a 6:1 SNR. The simulation results are reported in Table 1 and Figure 2. In this experiment, we verify our assertion that with relatively small fov, the Bruss-Horn formulation computes a translation direction biased towards the center. Note that this bias, which is in accord with our analysis, supports our explanation of the inconsistency of the Bruss-Horn (and related Jepson-Heeger) method. The other three algorithms give comparable results. This shows that with a fov of 50°, the Kanatani's method successfully corrected the inherent bias.

The second experiment changes the fov to 150°. Noise standard deviation is still $\sigma = 0.5$ pixel, with SNR around 10:1. The results are reported in Table 2 and Figure 3. In this case, Except RM-L2 and RM-L1.2, all algorithms give translation estimates that are biased towards a more lateral motion. Although 150° is a rather extreme case, this experiment effectively demonstrates fundamental flaws in the previous approaches, which do not appear in our formulation.

The third experiment illustrates the robustness of the methods. We still consider the case with 50° fov. The noise is generated as a mixture of Gaussians, so that 90% has a 6:1 SNR, and the other 10% has 1:1 SNR. The results are reported in Table 3 and Figure 4. In this case, the Bruss-Horn and Jepson-Heeger methods are still biased, while the other three methods are relatively unbiased. However, it is clear from the results that Kanatani's method gives a larger variance than RM-L2, which in turn gives a larger variance than RM-L1.2.

The fourth experiment verifies our theoretical results of consistency. The data are generated in exactly the same way as in the third experiment, except that the sample size increases from 100 to 2000. The results are reported in Table 4 and Figure 5. It can be seen that the

bias of Bruss-Horn and Jepson-Heeger methods can not be corrected by adding more data, which is in accord with our theoretical analysis. On the other hand, variances for the other methods reduced significantly and all of them show close to zero biases. Graph containing expected error for each algorithm against sample size is reported in Figure 6.

5.2 Synthetic sequences

To check whether our analysis is consistent with image velocity measurements obtained by tracking image features, we test the performance of the algorithms on some synthetic image sequences. In the literature, synthetic images have been used extensively for evaluating various optical flow algorithms since they resemble real images and the true flow fields are known. For motion estimation experiments, synthetic image sequences can be even more attractive since the exact ground truth information is available. On the other hand, accurate ground truth information is in general difficult to obtain for images taken by a camera (see section 5.3).

The first experiment is done with a sequence of twenty ray-traced images of a chessboard (courtesy of Andy Kniveton). This dataset (chessboard) is generated with a 45° field of view. The image sequence resolution is 512×512 pixels. The true translation direction is $[-2.28, 4.03, -4.94]$ which is randomly generated; and the rotation is $[0.0357, -0.2642, 0.2586]$, which corresponds to an angle velocity of 0.371°/frame. Figure 10 shows an image in the sequence. Figure 7 and Table 5 report results of the algorithms on this dataset. The computation is based on 100 automatically tracked features throughout the sequence, using the algorithm in [12].

In accord with our theoretical analysis and results obtained from the simulated flow experiments, the translation directions of Bruss-Horn and Jepson-Heeger are biased toward the forward direction, because of the narrow field of view. As we have pointed out in Section 4.4, this bias also leads to a bias in the estimated rotation parameter: The estimated translations are smaller than truth, which causes a smaller translational flow field. As a consequence, the estimated rotation parameters are larger than the true ones, so as to compensate the effect on the flow field. This phenomenon can also be observed in our simulation experiments (the effect is reversed with a large fov). Interestingly, for this particular image sequence, the Kanatani method shows a significant bias. Although we do not full understand its true cause, we conjecture that the compensation used in the Kanatani formulation is not suitable for the noise model in this image sequence. This also shows why it is useful to relax the underlying noise model assumption as much as possible, as we try to achieve in our analysis.

The second experiment is to illustrate that all algorithms perform well with zero trans-

	translation error	rotation error
Bruss-Horn	30.2 ± 7.2	0.16 ± 0.03
Jepson-Heeger	29.0 ± 4.7	0.16 ± 0.02
Kanatani	6.7 ± 5.5	0.09 ± 0.04
RM-L2	6.3 ± 3.6	0.05 ± 0.03
RM-L1.2	6.9 ± 4.1	0.05 ± 0.04

Table 1: Bias: fov = 50°

	translation error	rotation error
Bruss-Horn	14.5 ± 5.5	0.07 ± 0.03
Jepson-Heeger	45.4 ± 16.4	0.08 ± 0.04
Kanatani	38.1 ± 19.9	0.19 ± 0.09
RM-L2	7.8 ± 7.9	0.08 ± 0.10
RM-L1.2	8.6 ± 7.8	0.08 ± 0.10

Table 2: Bias: fov = 150°

	translation error	rotation error
Bruss-Horn	39.3 ± 7.8	0.20 ± 0.04
Jepson-Heeger	39.0 ± 7.6	0.19 ± 0.04
Kanatani	32.1 ± 24.9	0.17 ± 0.10
RM-L2	22.0 ± 19.9	0.15 ± 0.12
RM-L1.2	9.5 ± 5.2	0.08 ± 0.05

Table 3: Robustness: fov= 50°

	translation error	rotation error
Bruss-Horn	43.4 ± 3.2	0.20 ± 0.01
Jepson-Heeger	39.6 ± 1.7	0.19 ± 0.01
Kanatani	7.6 ± 5.5	0.04 ± 0.02
RM-L2	5.3 ± 3.0	0.03 ± 0.02
RM-L1.2	3.0 ± 1.8	0.02 ± 0.01

Table 4: Consistency: sample size= 2000

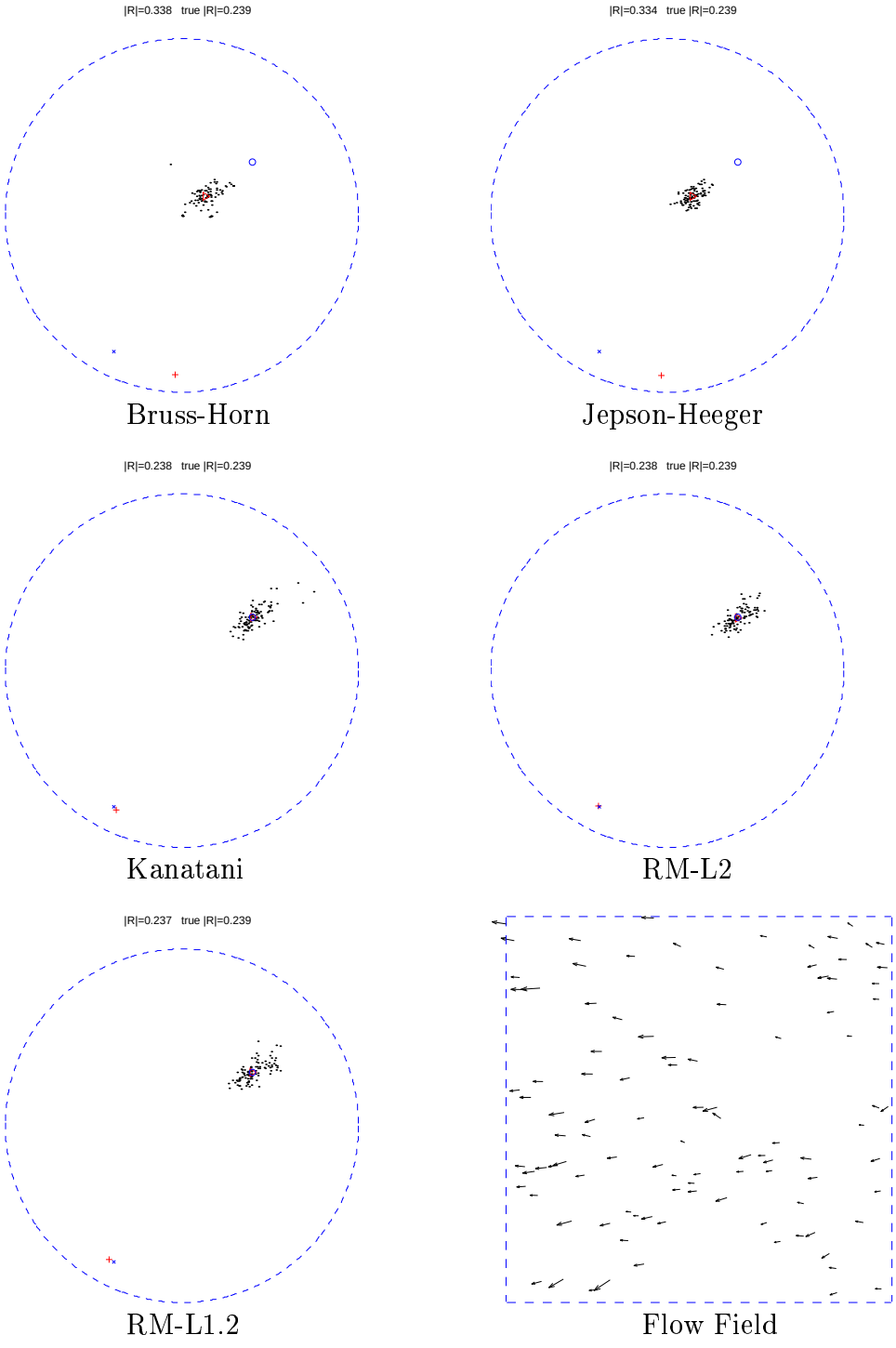


Figure 2: Bias: fov = 50°

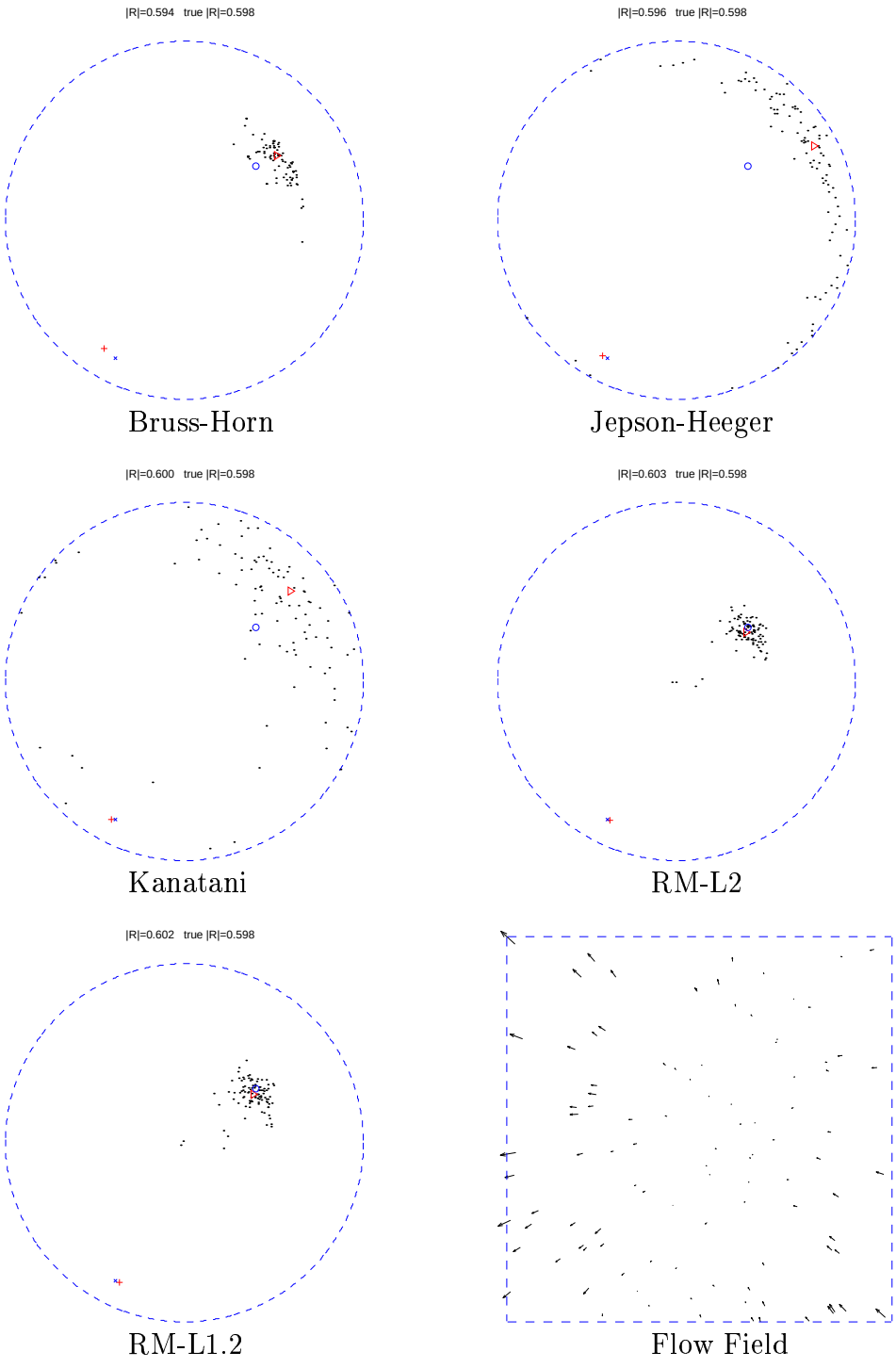


Figure 3: Bias: fov = 150°

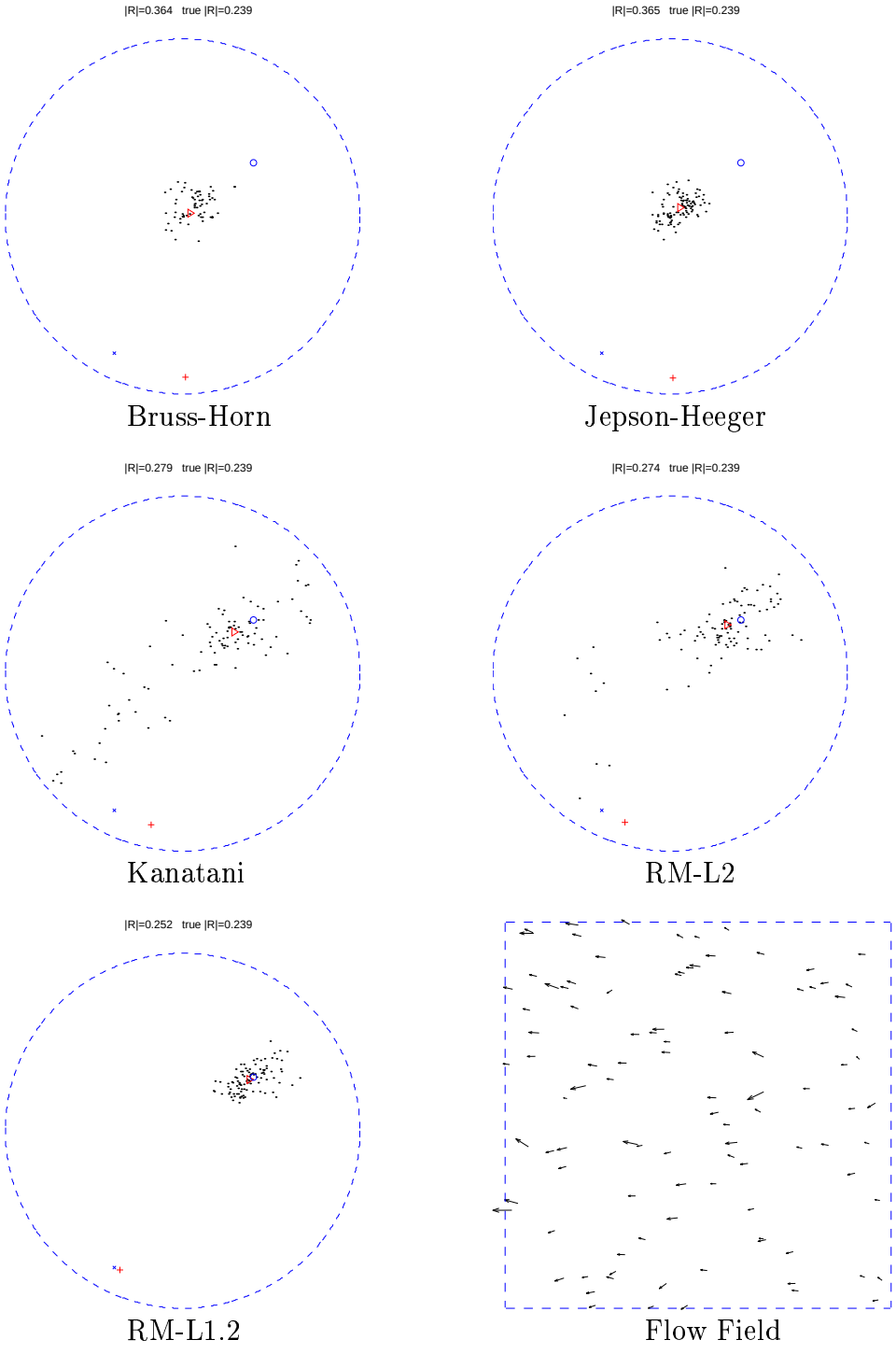


Figure 4: Robustness: fov= 50°

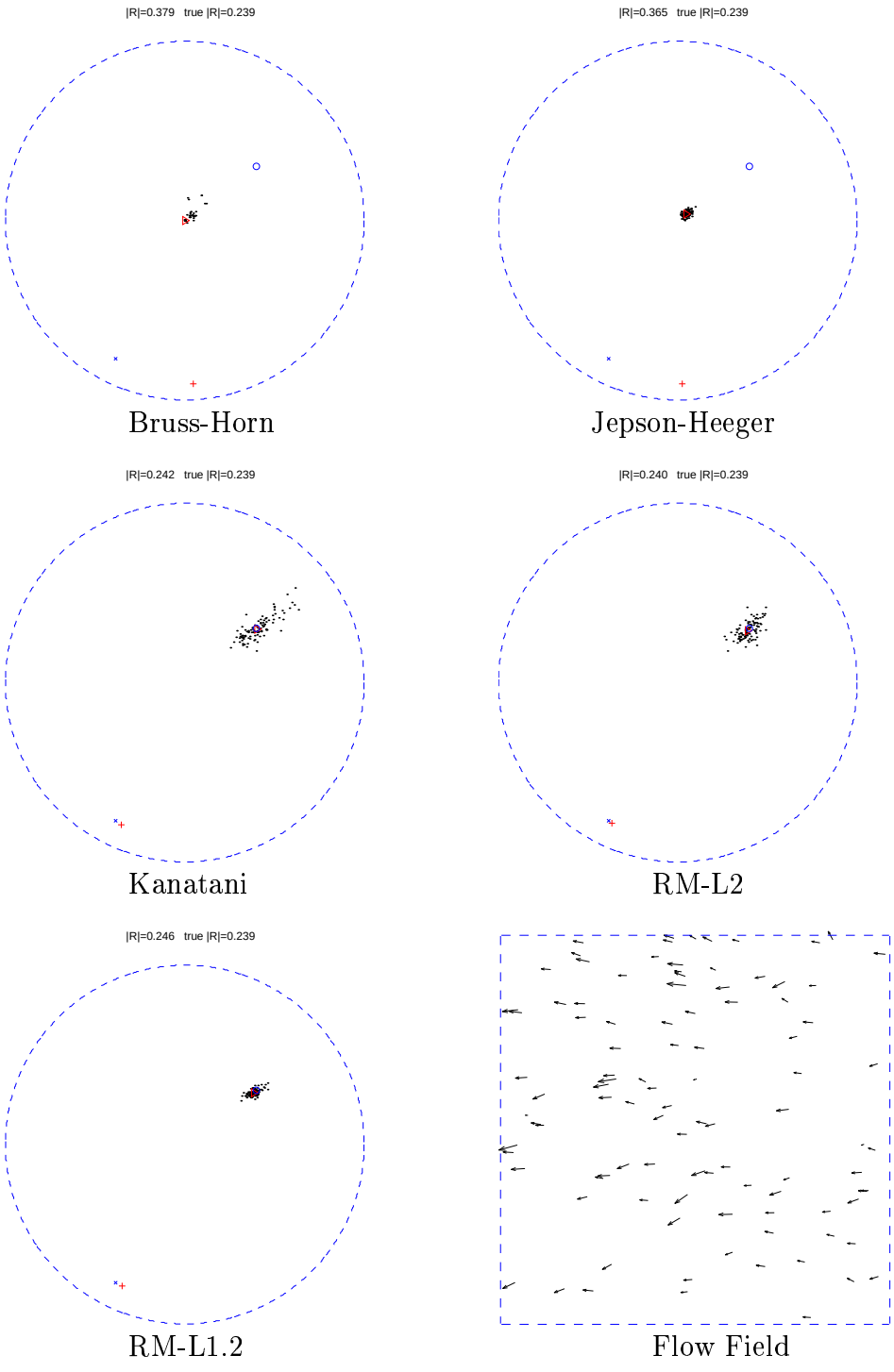
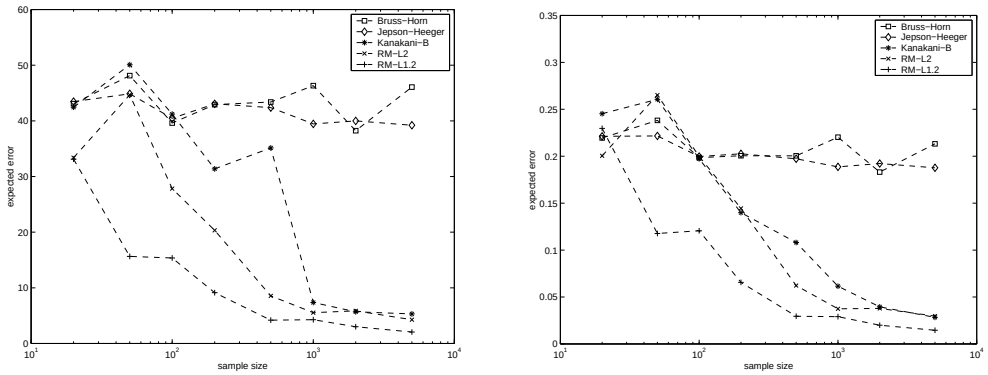


Figure 5: Consistency: sample size= 2000



expected translation error expected rotation error

Figure 6: Consistency: average error v.s. sample size

	translation error	rotation error
Bruss-Horn	17.6 ± 5.5	0.21 ± 0.08
Jepson-Heeger	31.2 ± 1.6	0.35 ± 0.04
Kanatani B	21.4 ± 3.6	0.13 ± 0.04
RM-L2	3.5 ± 2.0	0.07 ± 0.03
RM-L1.2	1.1 ± 0.6	0.02 ± 0.01

Table 5: Chessboard: sample size= 100

lation. We use another image sequence containing twenty images of an Amiga computer (courtesy of Marvin Landis). The resolution is still 512×512 pixels, with a 40° field of view. Figure 11 shows an image in the sequence. The translation is zero, and the rotation is $[0.3910, 0.0081, -0.0163]$, which corresponds to an angle velocity of 0.391°/frame. We have intentionally chosen a rotational direction that is almost parallel to the x-axis, since the flow field is very similar to that from a translation in the y-axis. However, as we can observe from Figure 8 and Table 6, with 100 tracked features, all algorithms perform well on this data, although it is clear that estimated translation directions are meaningless with all alpgorithms.

5.3 A lab sequence

We use an image sequence obtained from a camera in a controlled lab (the "lab" dataset) to further illustrate that algorithms based on our theoretical analysis perform well in practice.

This sequence is provided to us by John Zhang, who used the data in [27]. The images are taken with a 5.5mm lens on a Sony XC-77 CCD Video Camera. The sensor area is

	translation error	rotation error
Bruss-Horn	82.7 ± 0.9	0.00 ± 0.00
Jepson-Heeger	82.5 ± 0.8	0.00 ± 0.00
Kanatani B	80.8 ± 5.5	0.07 ± 0.09
RM-L2	75.2 ± 16.9	0.01 ± 0.03
RM-L1.2	74.1 ± 17.0	0.01 ± 0.01

Table 6: Amiga: sample size= 100

6.6mm $\times$ 8.8mm, which gives images of size 486 $\times$ 640 pixels. The camera is mounted on a precision positioning device, which allows measuring the true camera motion. The detailed lab set-up and camera calibration procedure can be found in [27].

As in [27], we use the first 51 frames in the sequence, which contain different motion directions from one frame to the next frame without rotation. The computation is based on 150 automatically tracked features, which are identical to features used in [27]. However, in this experiment, we do not perform any outlier removal, as opposed to [27]. Figure 12 shows an image in the sequence. Figure 9 and Table 7 contain results of the algorithms on this dataset. These results include 50 runs, where each run corresponds to the motion from one frame to the next frame. Since the motion directions vary throughout the sequence, in Figure 9, we plot all 50 true motion directions, each denoted by a symbol $\circ$. The estimated motion directions are still plotted as dots. The rotations are not illustrated in Figure 9, to make the plots appear less congested. Clearly, outliers cause problems for this sequence. As we can observe from the sample flow plot (from frame 30 to frame 31) in Figure 9, although most tracked features look reasonable, there are some obviously incorrect ones. This example shows that outliers can really become a problem in practice. It is not surprising to see that the robust formulation RM-L1.2, which in this example gives results comparable to those of the more complex scheme proposed in [27], is much less sensitive to outliers than the other methods.

	translation error	rotation error
Bruss-Horn	33.9 ± 27.5	0.49 ± 0.39
Jepson-Heeger	37.0 ± 22.5	0.60 ± 0.40
Kanatani B	32.1 ± 23.3	0.38 ± 0.31
RM-L2	16.3 ± 18.4	0.25 ± 0.28
RM-L1.2	2.2 ± 1.7	0.03 ± 0.02

Table 7: Lab dataset

6 Conclusion

In this paper, we have investigated the consistency of the instantaneous camera motion estimation problem. The theory we have developed shows that under certain moderate noise assumptions, we can construct consistent motion parameter estimators. One interesting consequence of this theory is that although the depth information cannot be recovered, it does not cause problems in motion parameter estimation. The theory also implies that any careless algebraic manipulation of the standard ridge motion formulation can lead to both bias and excessive variance in the results. These assertions have also been verified by experiments.

Although the theory implies that under our noise model a family of estimators are consistent, some of the estimators may be more efficient (in terms of convergence rate) or more robust (to outliers) than others. Due to limitations of space, we have skipped a detailed theoretical analysis. However, since outliers play an important role in motion parameter estimation, it is important to understand that robustness can be enhanced by using a q-norm with $q < 2$ rather than the least squares formulation with $q = 2$. This is because with a smaller q, an outlier has a smaller impact on the estimation. We have verified this point through experiments.

The consistency analysis has added a compelling argument for the use of more velocity measurements for camera motion estimation in the presence of noise, even if five points are sufficient in the noiseless situation. This argument shows that our intuition that noise can be "averaged out" with more measurements is valid under moderate assumptions. Furthermore, we have also shown that an arbitrary algebraic manipulation under the noiseless assumption should be avoided. Since our noise model is relatively simple, it is reasonable to expect a good motion estimation algorithm to be (at least approximately) consistent under this model. All of our experimental results, including simulation, synthetic image sequences, and a real image sequence, confirm our theoretical analysis. This suggests that our analysis, which is based on the isotropic flow noise model assumption, can provide useful insights in practical situations. It could also be interesting to apply the principle of consistency to more general structure-from-motion algorithms.

Acknowledgement

The authors are grateful to John Zhang for providing the "lab" dataset from his thesis [27]. We would also like to thank suggestions from the anonymous referees that helped significantly to improve the presentation of this paper.

A Proof of Theorem 1

Let α be any parameter, and denote $b(\alpha, x) = h(\alpha, \mathbf{x}, C(\alpha_t, \mathbf{x}))$, then

$$h(\alpha, \mathbf{x}, \mathbf{u}) = b(\alpha, \mathbf{x}) + g(\alpha, \mathbf{x}, \mathbf{u}),$$

where $g(\alpha, \mathbf{x}, \mathbf{u}) = Q(A(\mathbf{x})\mathbf{t})^T\mathbf{n}$ has symmetric and α independent distribution. Intuitively, $b(\alpha, \mathbf{x})$ is the component of $h(\alpha, \mathbf{x}, \mathbf{u})$ with the corrupted flow $\mathbf{u}$ replaced by the true flow, and $g(\alpha, \mathbf{x}, \mathbf{u})$ is the component of $h(\alpha, \mathbf{x}, \mathbf{u})$ caused by noise $\mathbf{n}$. Therefore this decomposition separates the effect of noise from the effect of true flow. Utilizing this decomposition, we obtain

$$\begin{aligned}
R(\alpha) &= \int dP(\mathbf{x}) \int f(h(\alpha, \mathbf{x}, \mathbf{u}))\, dP(\mathbf{u}|\mathbf{x}) \\
&= \int dP(\mathbf{x}) \int f(b(\alpha, \mathbf{x}) + g(\alpha, \mathbf{x}, \mathbf{u}))\, dP(\mathbf{u}|\mathbf{x}) \\
&= \int dP(\mathbf{x}) \int \frac{1}{2}[f(b(\alpha, \mathbf{x}) + g(\alpha, \mathbf{x}, \mathbf{u})) + f(b(\alpha, \mathbf{x}) - g(\alpha, \mathbf{x}, \mathbf{u}))]\, dP(\mathbf{u}|\mathbf{x}) \\
&\geq \int dP(\mathbf{x}) \int f(g(\alpha, \mathbf{x}, \mathbf{u}))\, dP(\mathbf{u}|\mathbf{x}) \\
&= \int dP(\mathbf{x}) \int f(g(\alpha_t, \mathbf{x}, \mathbf{u}))\, dP(\mathbf{u}|\mathbf{x}) = R(\alpha_t).
\end{aligned}$$

Note that in the above derivation, the third equality uses the fact that noise is symmetric. The inequality uses the assumption that f is symmetric and convex. It is a direct consequence of the Jensen's inequality of convex functions: the inequality can be geometrically interpreted as that in the graph of a convex function, the set above the function is convex, thus the middle point of any line-segment connecting two points on the boundary (line 3 in the above derivation) is in the set, so that it is above the boundary below it (line 4 in the above derivation). The last equality uses the fact that the noise distribution is isotropic, so that for any normalized direction Q, $Q^T\mathbf{n}$ has the same distribution. □

B Proof of Theorem 2

We need a Lemma in order to prove Theorem 2.

Lemma 1 *Under the assumptions of Theorem 1, and if we further assume that* $[f(a+b)+$

$f(a-b))]/2 \geq f(a)+k(a)\rho(b)$ *where* $k(a),\rho(b)\geq 0$, *then* $\forall\alpha$,

$$E_{\mathbf{x}}\ \rho(h(\alpha,\mathbf{x},C(\alpha_t,\mathbf{x}))) \leq \frac{\Delta R(\alpha)}{\inf_{\mathbf{x}} E_{\mathbf{n}_1|\mathbf{x}}\, k(\mathbf{n}_1)}.$$

Proof. We still use the notations from the proof of Theorem 1. Let $b(\alpha,\mathbf{x})=h(\alpha,\mathbf{x},C(\alpha_t,\mathbf{x}))$, then $h(\alpha,\mathbf{x},\mathbf{u})=b(\alpha,\mathbf{x})+g(\alpha,\mathbf{x},\mathbf{u})$:

$$\begin{aligned}&\frac{1}{2}[f(b(\alpha,\mathbf{x})+g(\alpha,\mathbf{x},\mathbf{u}))+f(b(\alpha,\mathbf{x})-g(\alpha,\mathbf{x},\mathbf{u}))]-f(g(\alpha,\mathbf{x},\mathbf{u}))\\ \geq& k(g(\alpha,\mathbf{x},\mathbf{u}))\rho(b(\alpha,\mathbf{x})).\end{aligned}$$

It follows from the proof of Theorem 1 that

$$\begin{aligned}\Delta R(\alpha) &= \int [\frac{f(b(\alpha,\mathbf{x})+g(\alpha,\mathbf{x},\mathbf{u}))+f(b(\alpha,\mathbf{x})-g(\alpha,\mathbf{x},\mathbf{u}))}{2}-f(g(\alpha,\mathbf{x},\mathbf{u}))]\ dP(\mathbf{x},\mathbf{u})\\ &\geq \int k(g(\alpha,\mathbf{x},\mathbf{u}))\rho(b(\alpha,\mathbf{x}))\ dP(\mathbf{x},\mathbf{u})\\ &\geq \int \rho(b(\alpha,\mathbf{x}))\ dP(\mathbf{x})\ \inf_{\mathbf{x}}\int k(g(\alpha,\mathbf{x},\mathbf{u}))\ dP(\mathbf{u}|\mathbf{x})\\ &= E_{\mathbf{x}}\rho(h(\alpha,\mathbf{x},C(\alpha_t,\mathbf{x})))\ \inf_{\mathbf{x}} E_{\mathbf{u}|\mathbf{x}}k(g(\alpha_t,\mathbf{x},\mathbf{u})).\end{aligned}$$

We thus obtain the lemma by noting that $g(\alpha_t,\mathbf{x},\mathbf{u})$ and $\mathbf{n}_1|\mathbf{x}$ have identical distributions (due to the isotropic noise assumption). □

Proof of Theorem 2. Let $k(a)=I(\{a:f(a)\leq\epsilon_0\})$ where I is the set indicator function and $\rho(b)=\max(0,f(b)-\epsilon_0)$, then there are two possibilities:

1. $f(a)>\min(f(b),\epsilon_0)$: in this case, either $k(a)=0$ or $\rho(b)=0$, thus

$$\frac{1}{2}[f(a+b)+f(a-b))]\geq f(a)=f(a)+k(a)\rho(b).$$

2. $f(a)\leq\min(f(b),\epsilon_0)$: in this case, $f(b)-f(a)\geq 0$ and $f(b)-f(a)\geq f(b)-\epsilon_0$, thus

$$\frac{1}{2}[f(a+b)+f(a-b))]\geq f(b)\geq f(a)+\max(0,f(b)-\epsilon_0)=f(a)+k(a)\rho(b).$$

By Lemma 1, we obtain

$$E_{\mathbf{x}}\, f(h(\alpha,\mathbf{x},C(\alpha_t,\mathbf{x})))-\epsilon_0\leq\frac{\Delta R(\alpha)}{\inf_{\mathbf{x}} E_{\mathbf{u}|\mathbf{x}}\, k(g(\alpha_t,\mathbf{x},\mathbf{u}))}=\frac{\Delta R(\alpha)}{\inf_{\mathbf{x}} P(f(\mathbf{n}_1)\leq\epsilon_0|\mathbf{x})}.$$

□

C Proof of Theorem 4

The proof of finite-sample convergence in Theorem 4 is very technical. Therefore it is useful to explain the ideas hidden in the proof. Intuitively, for each fixed α, by the law of large numbers, with sufficiently many samples m, the empirical risk $R_{emp}(\alpha)$ is close to the true risk $R(\alpha)$ with high probability. If in addition we can show that $R_{emp}(\alpha)$ is uniformly smooth in α (in a domain), then with sufficiently many samples, $R_{emp}(\alpha)$ is close to $R(\alpha)$ for all α (in the domain) with high probability. If so, we know that a parameter α that approximately minimizes the empirical risk $R_{emp}(\alpha)$ also approximately minimizes the true risk $R(\alpha)$ with high probability (in the domain). Essentially, inequality (34) characterizes what we mean by uniform smoothness in α. Therefore it plays a very important role in the proof. The remaining proof after (34) is more or less standard manipulations in statistics. Inequality (34) follows from (30) and (31) which decomposes the uniform smoothness in α defined in (34) into the uniform smoothness in $\mathbf{t}$ defined in (30) and the uniform smoothness in ω defined in (31). The derivations of these two inequalities are specially tailored to the motion estimation problem. In particular, the first half proof of Theorem 4 establishes relevant inequalities to show that under appropriate technical assumptions, the discontinuity caused by the normalized projection operator Q does not introduce smoothness problems. As we shall point out, the term smoothness used above does not refer to the smoothness of the optical flow itself (which will be discontinuous when the depth is discontinuous), but rather to the projection of the optical flow in the direction of $Q(A(\mathbf{x})\mathbf{t})$.

The following proof of Theorem 4 is divided into 10 steps to improve its readability. In addition, we summarize the goal of each step with an English sentence at the beginning of the step. These sentences allow a read who is not interested in the mathematical details to have a high level understanding of the proof.

Proof. Before the main steps of the proof, we would like to introduce a number of inequalities and constants that characterize a number of regularity conditions used later in the proof.

Firstly, it is easy to verify that with fixed $\|\mathbf{x}_0\| = 1$, the function $\overline{\lim}_{\mathbf{x}\to\mathbf{x}_0}\|Q(\mathbf{x}) - Q(\mathbf{x}_0)\|/\|\mathbf{x} - \mathbf{x}_0\|$ is bounded. Thus by symmetry and boundedness of Q, there exists a constant c_0 such that $\forall\mathbf{x}_1$ and $\forall\mathbf{x}_2 \neq 0$,

$$\|Q(\mathbf{x}_1) - Q(\mathbf{x}_2)\| \leq c_0\|\mathbf{x}_1 - \mathbf{x}_2\|/\|\mathbf{x}_2\|. \tag{16}$$

We define $\bar{A}(\mathbf{x}) = \frac{1}{\|\mathbf{x}\|+1} A(\mathbf{x})$. It is easy to see that $Q(\bar{A}(\mathbf{x})\mathbf{t}) \equiv Q(A(\mathbf{x})\mathbf{t})$ and $\exists$ constant $c_1 > 0$ s.t.

$$\|\bar{A}(\mathbf{x})\mathbf{t}\| \leq c_1 \|\mathbf{t}\|. \tag{17}$$

There exist constants $c_2, c_3, c_4 > 0$ such that

$$\|B(\mathbf{x})\| \leq c_2(\|\mathbf{x}\|^2 + 1), \tag{18}$$

$$f(a+b) \leq c_3(f(a) + f(b)), \tag{19}$$

$$|a+b|^{q-1} \leq c_4(1 + f(a) + |b|^{q-1}). \tag{20}$$

Also it is easy to verify the following inequality:

$$|f(a) - f(b)| \leq q|a-b|(|a|^{q-1} + |b|^{q-1}). \tag{21}$$

For clarity, we divide the rest of the proof into 10 steps:

Step 1: In this step, we write down a few more regularity conditions formalizing some assumptions of the theorem.

Since $|Q(A(\mathbf{x})\mathbf{t})^T B(\mathbf{x})\omega| \leq c_2(1+\|\mathbf{x}\|^2)\|\omega\|$, it follows from Assumption 2 that $E_{\mathbf{X}} f(Q(A(\mathbf{x})\mathbf{t})^T B(\mathbf{x})\omega)$ exists for all $\mathbf{t}$ and ω. Also $\forall M > 0$, we have constant c_5 that only depends on M:

$$c_5(M) = E_{\mathbf{X}} \sup_{\|\mathbf{t}\|=1, \|\omega\| \leq M} f(Q(A(\mathbf{x})\mathbf{t})^T B(\mathbf{x})\omega) < \infty. \tag{22}$$

It is clear from Assumptions 3 and 4 that

$$E_{\mathbf{X},\mathbf{U}} f(\|\mathbf{u}\|) < \infty, \tag{23}$$

and $\forall \gamma > 0$,

$$c_6(\gamma) = \sup_{\|\mathbf{x}\| < \gamma} E_{\mathbf{U}|\mathbf{X}} f(\|\mathbf{u}\|) < \infty. \tag{24}$$

We thus also have

$$E_{\mathbf{X},\mathbf{U}} \sup_{\|\mathbf{t}\|=1, \|\omega\| \leq M} f(h([\mathbf{t}, \omega], \mathbf{x}, \mathbf{u})) < \infty. \tag{25}$$

Step 2: The goal of this step is to derive equation (28), which means that the discontinuity

caused by $Q(A(\mathbf{x})\mathbf{t})$ (with $\bar{A}(\mathbf{x})\mathbf{t}$ small) is negligible.

For all $\delta \in (0, 0.1)$ and $\epsilon \in (0, 0.1\delta)$. Consider an arbitrary choice of $\|\mathbf{t}\| = 1$ and $\mathbf{x}$ such that $\|\bar{A}(\mathbf{x})\mathbf{t}\| < \epsilon$. We obtain $|t_1 - x_1 t_3| < (1 + \|\mathbf{x}\|)\epsilon$ and $|t_2 - x_2 t_3| < (1 + \|\mathbf{x}\|)\epsilon$. Consider the following two situations:

a. $|t_3| > \delta$: in this case, let $\mathbf{y} = [t_1, t_2]/t_3$, then $\|\mathbf{y}\| \leq 1/\delta$. $\|\bar{A}(\mathbf{x})\mathbf{t}\| < \epsilon$ implies that $\|\mathbf{x} - \mathbf{y}\| \leq 2(1 + \|\mathbf{x}\|)\epsilon/\delta$, therefore $\|\mathbf{x}\| \leq 2(1 + \|\mathbf{y}\|) \leq 2(1 + 1/\delta)$. We obtain

$$\begin{aligned} &E_{(\mathbf{x},\mathbf{u}):\|\bar{A}(\mathbf{x})\mathbf{t}\|<\epsilon} f(\|\mathbf{x}\|^2 + \|\mathbf{u}\| + 1) \\ \leq &P(\mathbf{x} : \|\mathbf{x} - \mathbf{y}\| \leq \frac{8}{\delta}(1 + 1/\delta)\epsilon) \sup_{\|\mathbf{x}\| \leq 2(1+1/\delta)} E_{\mathbf{u}|\mathbf{x}} f(\|\mathbf{x}\|^2 + \|\mathbf{u}\| + 1), \end{aligned}$$

where $\|\mathbf{y}\| \leq 1/\delta$. Since Assumption 1 implies

$$\lim_{\epsilon \to 0} \sup_{\|\mathbf{y}\| \leq 1/\delta} P(\mathbf{x} : \|\mathbf{x} - \mathbf{y}\| \leq \frac{8}{\delta}(1 + 1/\delta)\epsilon) = 0,$$

and since (19) and (24) imply

$$\sup_{\|\mathbf{x}\| \leq 2(1+1/\delta)} E_{\mathbf{u}|\mathbf{x}} f(\|\mathbf{x}\|^2 + \|\mathbf{u}\| + 1) < \infty,$$

therefore we obtain

$$\lim_{\epsilon \to 0} \sup_{|t_3| \geq \delta} E_{(\mathbf{x},\mathbf{u}):\|\bar{A}(\mathbf{x})\mathbf{t}\|<\epsilon} f(\|\mathbf{x}\|^2 + \|\mathbf{u}\| + 1) = 0, \tag{26}$$

b. $|t_3| \leq \delta$: in this case, either $|t_1| > 0.5$ or $|t_2| > 0.5$, therefore $0.5 - |t_3| \cdot \|\mathbf{x}\| < (1 + \|\mathbf{x}\|)\epsilon$. We have $\|\mathbf{x}\| > (0.5 - \epsilon)/(|t_3| + \epsilon) > 0.1/\delta$. Now, from Assumption 2 and equation (23), we obtain

$$\lim_{\delta \to 0} \overline{\lim}_{\epsilon \to 0} \sup_{|t_3| \leq \delta} E_{(\mathbf{x},\mathbf{u}):\|\bar{A}(\mathbf{x})\mathbf{t}\|<\epsilon} f(\|\mathbf{x}\|^2 + \|\mathbf{u}\| + 1) = 0. \tag{27}$$

$\forall \Delta > 0$, by (26) and (27), we can find $\epsilon > 0$ such that

$$\sup_{\|\mathbf{t}\|=1} E_{(\mathbf{x},\mathbf{u}):\|\bar{A}(\mathbf{x})\mathbf{t}\|<\epsilon} f(\|\mathbf{x}\|^2 + \|\mathbf{u}\| + 1) < \Delta.$$

Thus

$$\lim_{\epsilon \to 0} \sup_{\|\mathbf{t}\|=1} E_{(\mathbf{x},\mathbf{u}):\|\bar{A}(\mathbf{x})\mathbf{t}\|<\epsilon} f(\|\mathbf{x}\|^2 + \|\mathbf{u}\| + 1) = 0. \tag{28}$$

Step 3: In this step, we would like to show that $E_{\mathbf{X}} f(Q(A(\mathbf{x})\mathbf{t})^T B(\mathbf{x})\omega)$ is continuous in $\mathbf{t}$ and ω. This result will be used in Step 4.

To prove this, $\forall \|\mathbf{t}_0\| = 1, \omega_0$ and $\epsilon > 0$, by (28), we can decompose R^2 as $S \cup T$ such that $E_{\mathbf{X}\in S} f(\|B(\mathbf{x})\|(\|\omega_0\| + 1)) < \epsilon$ and $Q(A(\mathbf{x})\mathbf{t})$ is continuous when $\mathbf{x} \in T$. Since $f(Q(A(\mathbf{x})\mathbf{t})^T B(\mathbf{x})\omega)$ is bounded by the integrable function $f(c_2(\|\mathbf{x}\|^2 + 1)\|\omega\|)$, we obtain

$$\lim_{(\mathbf{t},\omega)\to(\mathbf{t}_0,\omega_0)} E_{\mathbf{X}\in T} f(Q(A(\mathbf{x})\mathbf{t})^T B(\mathbf{x})\omega) = E_{\mathbf{X}\in T} f(Q(A(\mathbf{x})\mathbf{t}_0)^T B(\mathbf{x})\omega_0).$$

Also note that $E_{\mathbf{X}\in S} f(Q(A(\mathbf{x})\mathbf{t})^T B(\mathbf{x})\omega) < \epsilon$ when $\|\omega - \omega_0\| < 1$, therefore

$$|\overline{\lim}_{(\mathbf{t},\omega)\to(\mathbf{t}_0,\omega_0)} E_{\mathbf{X}} f(Q(A(\mathbf{x})\mathbf{t})^T B(\mathbf{x})\omega) - E_{\mathbf{X}} f(Q(A(\mathbf{x})\mathbf{t}_0)^T B(\mathbf{x})\omega_0)| \le 2\epsilon.$$

Since ϵ is arbitrary, we conclude that $E_{\mathbf{X}} f(Q(A(\mathbf{x})\mathbf{t})^T B(\mathbf{x})\omega)$ is continuous in $\mathbf{t}$ and ω. Similarly, $E_{\mathbf{X},\mathbf{U}} f(Q(A(\mathbf{x})\mathbf{t})^T(\mathbf{u} - B(\mathbf{x})\omega))$ is continuous in $\mathbf{t}$ and ω.

Step 4: In this step, we derive another regularity condition that formalizes the assumption of non-degeneracy of the scene. This condition will be used in Step 10.

The non-degeneracy of the scene implies that $\forall \|\mathbf{t}\|$ and $\|\omega\| = 1$,

$$E_{\mathbf{X}} f(Q(A(\mathbf{x})\mathbf{t})^T B(\mathbf{x})\omega) > 0.$$

By the continuity of $E_{\mathbf{X}} f(Q(A(\mathbf{x})\mathbf{t})^T B(\mathbf{x})\omega)$ in Step 3, we obtain

$$c_7 = \inf_{\|\mathbf{t}\|=1, \|\omega\|=1} E_{\mathbf{X}} f(Q(A(\mathbf{x})\mathbf{t})^T B(\mathbf{x})\omega) > 0. \tag{29}$$

Step 5: We would like to show that the motion estimation formulation is uniformly smooth in $\mathbf{t}$. That is, for all $\epsilon, \gamma, M > 0$, we want to show that there exists $\delta > 0$ such that $\forall \|\mathbf{t}_0\| = 1$ and $T_0 = \{\mathbf{x} : \|\mathbf{x}\| \le \gamma, \|\bar{A}(\mathbf{x})\mathbf{t}_0\| > \epsilon\}$:

$$E_{\mathbf{X}\in T_0,\mathbf{U}} \sup_{\|\mathbf{t}-\mathbf{t}_0\|\le\delta, \|\omega\|\le M} |f(h([\mathbf{t},\omega],\mathbf{x},\mathbf{u})) - f(h([\mathbf{t}_0,\omega],\mathbf{x},\mathbf{u}))| < \epsilon. \tag{30}$$

Note that by (16) and (17), we obtain $\|Q(A(\mathbf{x})\mathbf{t}) - Q(A(\mathbf{x})\mathbf{t}_0)\| \le c_0 c_1 \delta/\epsilon$ when $\mathbf{x} \in T_0$ and $\|\mathbf{t} - \mathbf{t}_0\| \le \delta$. Therefore

$$|h([\mathbf{t},\omega],\mathbf{x},\mathbf{u})) - h([\mathbf{t}_0,\omega],\mathbf{x},\mathbf{u})| \le c_0 c_1 \frac{\delta}{\epsilon}(\|\mathbf{u}\| + c_2(\gamma^2 + 1)M).$$

Together with (18) (20) and (21), we know that $\forall \|\mathbf{t}-\mathbf{t}_0\| \leq \delta$,

$$\begin{aligned}
&|f(Q(A(\mathbf{x})\mathbf{t})^T(\mathbf{u}-B(\mathbf{x})\omega)) - f(Q(A(\mathbf{x})\mathbf{t}_0)^T(\mathbf{u}-B(\mathbf{x})\omega))| \\
\leq & qc_0c_1\frac{\delta}{\epsilon}c_4[\|\mathbf{u}\| + c_2(\gamma^2+1)M][2+2\|\mathbf{u}\|^{q-1} + f(h([\mathbf{t},\omega],\mathbf{x},0)) + f(h([\mathbf{t}_0,\omega],\mathbf{x},0))].
\end{aligned}$$

By (22), there exists a constant c which depends on M:

$$c = E_{\mathbf{X}}(2 + f(h([\mathbf{t},\omega],\mathbf{x},0)) + f(h([\mathbf{t}_0,\omega],\mathbf{x},0))) < \infty.$$

Now,

$$\begin{aligned}
&E_{\mathbf{x},\mathbf{u}}[\|\mathbf{u}\| + c_2(\gamma^2+1)M][2+2\|\mathbf{u}\|^{q-1} + f(h([\mathbf{t},\omega],\mathbf{x},0)) + f(h([\mathbf{t}_0,\omega],\mathbf{x},0))] \\
\leq & \sup_{\|\mathbf{x}\|\leq M} E_{\mathbf{u}|\mathbf{x}}[\|\mathbf{u}\| + c_2(\gamma^2+1)M]c + \sup_{\|\mathbf{x}\|\leq M} E_{\mathbf{u}|\mathbf{x}}[2f(\|\mathbf{u}\|) + 2c_2(\gamma^2+1)M\|\mathbf{u}\|^{q-1}] \\
\leq & c(1 + c_6(M) + c_2(\gamma^2+1)M) + [2c_6(M) + 2c_2(\gamma^2+1)M(1+c_6(M))] \leq c',
\end{aligned}$$

where c' is a constant which depends on γ and M. Therefore

$$E_{\mathbf{x}\in T_0,\mathbf{u}}|f(Q(A(\mathbf{x})\mathbf{t})^T(\mathbf{u}-B(\mathbf{x})\omega)) - f(Q(A(\mathbf{x})\mathbf{t}_0)^T(\mathbf{u}-B(\mathbf{x})\omega))| \leq (qc_0c_1c_4c'/\epsilon)\delta.$$

(30) follows from any choice of $\delta < \epsilon^2/(qc_0c_1c_4c')$.

Step 6: We would like to show that the motion estimation formulation is uniformly smooth in ω. That is, for all $\epsilon, \gamma, M > 0$, we want to show that there exists $\delta > 0$ such that $\forall \|\omega_1\|, \|\omega_2\| \leq M$ and $\|\omega_1 - \omega_2\| < \delta$:

$$E_{\|\mathbf{x}\|\leq\gamma,\mathbf{u}} \sup_{\|\mathbf{t}\|=1,\|\omega_1-\omega_2\|\leq\delta} |f(h([\mathbf{t},\omega_1],\mathbf{x},\mathbf{u})) - f(h([\mathbf{t},\omega_2],\mathbf{x},\mathbf{u}))| < \epsilon, \tag{31}$$

where $\|\omega_1\|, \|\omega_2\| \leq M$. Similar to the proof of (30), we note that

$$|h([\mathbf{t},\omega_1],\mathbf{x},\mathbf{u}) - h([\mathbf{t},\omega_2],\mathbf{x},\mathbf{u})| \leq c_2(1+\gamma^2)\delta.$$

The rest of the proof is the same as that of (30).

Step 7: In this step, we combine results from the previous two steps to show that the motion estimation formulation is uniformly smooth in α, as formalized in (34).

$\forall \epsilon, \gamma, M > 0$, by (28), we can find $\epsilon' > 0$ such that we can decompose $\{\mathbf{x} : \|\mathbf{x}\| \leq \gamma\}$ as

$S(\mathbf{t}_0) \cup T(\mathbf{t}_0)$ for each $\|\mathbf{t}_0\| = 1$, such that $\|\bar{A}(\mathbf{x})\mathbf{t}_0\| > \epsilon'$ when $\mathbf{x} \in T(\mathbf{t}_0)$ and

$$E_{\mathbf{x}\in S(\mathbf{t}_0),\mathbf{u}} \sup_{\|\omega\|\le M} c_3[f(c_2(\|\mathbf{x}\|^2+1)\|\omega\|) + f(\|\mathbf{u}\|)] < \epsilon,$$

which implies that

$$E_{\mathbf{x}\in S(\mathbf{t}_0),\mathbf{u}} \sup_{\|\mathbf{t}\|=1,\|\omega\|\le M} f(h([\mathbf{t},\omega],\mathbf{x},\mathbf{u})) < \epsilon.$$

Now by (30), we can find δ and k quantities, $\mathbf{t}_1, \ldots, \mathbf{t}_k$ with corresponding decomposition $S_i \cup T_i$ for each $\mathbf{t}_i$, such that $\forall \|\mathbf{t}\| = 1$, $\|\mathbf{t} - \mathbf{t}_i\| \le \delta$ for some i, and

$$\sup_i E_{\mathbf{x}\in S_i,\mathbf{u}} \sup_{\|\mathbf{t}\|=1,\|\omega\|\le M} f(h([\mathbf{t},\omega],\mathbf{x},\mathbf{u})) < \epsilon, \tag{32}$$

and $\forall i$,

$$E_{\mathbf{x}\in T_i,\mathbf{u}} \sup_{\|\mathbf{t}-\mathbf{t}_i\|\le\delta,\|\omega\|\le M} |f(h([\mathbf{t},\omega],\mathbf{x},\mathbf{u})) - f(h([\mathbf{t}_i,\omega],\mathbf{x},\mathbf{u}))| < \epsilon. \tag{33}$$

It follows from (32) and (33) that $\forall \epsilon, \gamma, M > 0$, $\exists$ a finite partition of $\|\mathbf{t}\| = 1$ into the union of sets $D_1, \ldots D_k$ and $\exists \mathbf{t}_i \in D_i$ $(i = 1, \ldots, k)$ such that

$$\sup_i E_{\|\mathbf{x}\|\le\gamma,\mathbf{u}} \sup_{\mathbf{t}\in D_i,\|\omega\|\le M} |f(h([\mathbf{t},\omega],\mathbf{x},\mathbf{u})) - f(h([\mathbf{t}_i,\omega],\mathbf{x},\mathbf{u}))| < \epsilon.$$

From this inequality and (31), we know that $\forall \epsilon, \gamma, M > 0$, $\exists$ a finite partition of $\|\mathbf{t}\| = 1$ into the union of sets $D_1, \ldots D_k$ with $\mathbf{t}_i \in D_i$ $(i = 1, \ldots, k)$; and a finite partition of $\|\omega\| \le M$ into the union of sets $O_1, \ldots O_\ell$ with $\omega_j \in O_j$ $(j = 1, \ldots, \ell)$ such that

$$\sup_{i,j} E_{\|\mathbf{x}\|\le\gamma,\mathbf{u}} \sup_{\mathbf{t}\in D_i,\omega\in O_j} |f(h([\mathbf{t},\omega],\mathbf{x},\mathbf{u})) - f(h([\mathbf{t}_i,\omega_j],\mathbf{x},\mathbf{u}))| < \epsilon. \tag{34}$$

Step 8: In this step, we apply the law of large numbers to (34) to derive (37), which implies that the empirical risk restricted to $\|\mathbf{x}\| \le \gamma$ converges uniformly in probability to the true underlying risk restricted to $\|\mathbf{x}\| \le \gamma$, in the domain $\|\omega\| \le M$.

By the law of large numbers, we have

$$\lim_{m\to\infty} P(\sup_{i,j} |E_{\mathbf{x},\mathbf{u}} f(h([\mathbf{t}_i,\omega_j],\mathbf{x},\mathbf{u})) - E_{emp,\mathbf{x},\mathbf{u}} f(h([\mathbf{t}_i,\omega_j],\mathbf{x},\mathbf{u}))| > 2\epsilon) = 0, \tag{35}$$

where we use E_{emp} to denote the empirical expectation with m iid samples of $(\mathbf{x}, \mathbf{u})$ (the same convention will be employed through the rest of the proof):

$$E_{emp,\mathbf{x},\mathbf{u}} f(h([\mathbf{t}_i, \omega_j], \mathbf{x}, \mathbf{u})) = \frac{1}{m} \sum_{k=1}^{m} f(h([\mathbf{t}_i, \omega_j], \mathbf{x}_k, \mathbf{u}_k).$$

Applying the law of the large numbers again to (34), we obtain

$$\lim_{m\to\infty} P(\sup_{i,j} E_{emp,\|\mathbf{x}\|\leq\gamma,\mathbf{u}} \sup_{\mathbf{t}\in D_i,\omega\in O_j} |f(h([\mathbf{t}, \omega], \mathbf{x}, \mathbf{u})) - f(h([\mathbf{t}_i, \omega_j], \mathbf{x}, \mathbf{u}))| > 2\epsilon) = 0. \quad (36)$$

By combining (34), (35) and (36), we obtain that $\forall \epsilon, \gamma, M > 0$:

$$\lim_{m\to\infty} P(\sup_{\|\mathbf{t}\|=1,\|\omega\|\leq M} |E_{emp,\|\mathbf{x}\|\leq\gamma,\mathbf{u}} f(h([\mathbf{t}, \omega], \mathbf{x}, \mathbf{u})) - E_{\|\mathbf{x}\|\leq\gamma,\mathbf{u}} f(h([\mathbf{t}, \omega], \mathbf{x}, \mathbf{u}))| > 6\epsilon) = 0. \quad (37)$$

Step 9: In this step, by refining inequality (37) in the previous step, we derive inequality (38), which implies that the empirical risk converges uniformly in probability to the true underlying risk in the domain $\|\omega\| \leq M$.

From (25), we obtain

$$\lim_{\gamma\to\infty,\mathbf{u}} E_{\|\mathbf{x}\|>\gamma} \sup_{\|\mathbf{t}\|=1,\|\omega\|\leq M} f(Q(A(\mathbf{x})\mathbf{t})^T B(\mathbf{x})\omega) = 0.$$

Therefore $\forall \epsilon > 0$, by the law of large numbers, $\exists \gamma > 0$, such that

$$\lim_{m\to\infty} P(\sup_{\|\mathbf{t}\|=1,\|\omega\|\leq M} |E_{emp,\|\mathbf{x}\|>\gamma,\mathbf{u}} f(h([\mathbf{t}, \omega], \mathbf{x}, \mathbf{u})) - E_{\|\mathbf{x}\|>\gamma,\mathbf{u}} f(h([\mathbf{t}, \omega], \mathbf{x}, \mathbf{u}))| > \epsilon) = 0.$$

Combining this inequality with (37), we have: $\forall \epsilon, M > 0$,

$$\lim_{m\to\infty} P(\sup_{\|\mathbf{t}\|=1,\|\omega\|\leq M} |E_{emp,\mathbf{x},\mathbf{u}} f(h([\mathbf{t}, \omega], \mathbf{x}, \mathbf{u})) - E_{\mathbf{x},\mathbf{u}} f(h([\mathbf{t}, \omega], \mathbf{x}, \mathbf{u}))| > \epsilon) = 0. \quad (38)$$

Step 10: In this step, we would like show that there exists M such that $\lim_{m\to\infty} P(\|\hat{\omega}\| \leq M) = 1$. Then together with the uniform convergence of empirical risk in (38), and the fact that $E_{emp,\mathbf{x},\mathbf{u}} f(\hat{\alpha}, \mathbf{x}, \mathbf{u}) \leq E_{emp,\mathbf{x},\mathbf{u}} f(\alpha_t, \mathbf{x}, \mathbf{u})$, we are able to obtain the theorem.

Now, we set $\mathbf{u} = 0$ (obviously in this case, the assumptions in the Theorem are still

satisfied with $p(x) = 0$, $\mathbf{n} = 0$ and $\omega_t = 0$) and $\epsilon = c_7/2$ in (38):

$$\lim_{m\to\infty} P(\inf_{\|\mathbf{t}\|=1,\|\omega\|=1} E_{emp,\mathbf{x}} f(Q(A(\mathbf{x})\mathbf{t}_1)^T B(\mathbf{x})\omega) \geq c_7/2) = 1,$$

which implies that

$$\lim_{m\to\infty} P(\inf_{\|\mathbf{t}\|=1,\|\omega\|\geq M} E_{emp,\mathbf{x}} f(Q(A(\mathbf{x})\mathbf{t}_1)^T B(\mathbf{x})\omega) \geq M^q c_7/2) = 1.$$

Therefore we obtain $\forall M > 0$,

$$\lim_{m\to\infty} P(\inf_{\|\mathbf{t}\|=1,\|\omega\|\geq M} E_{emp,\mathbf{x},\mathbf{u}} f(h([\mathbf{t},\omega],\mathbf{x},\mathbf{u})) \geq M^q c_7/2c_3 - E_{emp} f(\|\mathbf{u}\|)) = 1.$$

Note that $\hat{\omega}$ is obtained by minimization of $E_{emp,\mathbf{x},\mathbf{u}} f(h([\mathbf{t},\omega],\mathbf{x},\mathbf{u}))$, and since if we choose M large enough, then

$$\lim_{m\to\infty} P(E_{emp,\mathbf{x},\mathbf{u}} f(\alpha_t,\mathbf{x},\mathbf{u}) < M^q c_7/2c_3 - E_{emp} f(\|\mathbf{u}\|)) = 1,$$

thus if we denote $\hat{\alpha} = [\hat{\mathbf{t}},\hat{\omega}]$, then $\lim_{m\to\infty} P(\|\hat{\omega}\| \leq M) = 1$. Combining this with (38), and note that $E_{emp,\mathbf{x},\mathbf{u}} f(\hat{\alpha},\mathbf{x},\mathbf{u}) \leq E_{emp,\mathbf{x},\mathbf{u}} f(\alpha_t,\mathbf{x},\mathbf{u})$, we obtain the theorem. □

References

[1] S. Beauchemin and J. Barron. The computation of optical flow. *ACM Computing Surveys*, 27(3):433–467, 1996.

[2] A. Bruss and B. Horn. Passive navigation. *Computer Graphics and Image Processing*, 21:3–20, 1983.

[3] J. Gibson. *The senses considered as perceptual systems.* Houghton Mifflin, Boston, MA, 1966.

[4] D. J. Heeger and A. D. Jepson. Subspace methods for recovering rigid motion I: Algorithm and implementation. *International Journal of Computer Vision*, 7(2):95–118, 1992.

[5] E. C. Hildreth. Recovering heading for visually-guided navigation. *Vision Research*, 32(6):1177–1192, 1992.

[6] B. K. P. Horn. Motion fields are hardly ever ambiguous. *International Journal of Computer Vision*, 1:259–274, 1987.

[7] B. K. P. Horn and E. J. Weldon Jr. Direct methods for recovering motion. *International Journal of Computer Vision*, 2:51–76, 1988.

[8] A. D. Jepson and D. J. Heeger. A fast subspace algorithm for recovering rigid motion. In *Proceedings of IEEE Workshop on Visual Motion*, 124–131, Princeton, NJ, 1991.

[9] A. D. Jepson and D. J. Heeger. Linear subspace methods for recovering translation direction. In L. Harris and M. Jenkin, editors, *Spatial Vision in Humans and Robots*, pages 39–62, New York, 1993. Cambridge University Press.

[10] K. Kanatani. 3-d interpretation of optical flow by renormalization. *International Journal of Computer Vision*, 11(3):267–282, 1993.

[11] E. Kruppa. Zur ermittlung eines objektes aus zwei perspektiven mit innerer orientierung. *Sitz.-Ber. Akad. Wiss., Wien, Math. Naturw. Kl., Abt. IIa.*, 122:1939–1948, 1913.

[12] B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. *IJCAI*, 1981.

[13] W. J. MacLean, A. D. Jepson, and R. C. Frecker. Recovery of egomotion and segmentation of independent object motion using the em algorithm. In *Proceedings of the 5th British Machine Vision Conference*, pages 13–16, York, UK, 1994.

[14] L. Matthies, T. Kanade, and R. Szeliski. Kalman filter-based algorithms for estimating depth from image sequences. *International Journal of Computer Vision*, 3(3):209–236, 1989.

[15] S. Maybank. The angular velocity associated with the angular flowfield arising from motion through a rigid environment. *Proceedings of the Royal Society of London*, A-401:317–326, 1985.

[16] H. P. Moravec. Towards automatic visual obstacle avoidance. In *Proceedings of the 5th International Joint Conference on Artificial Intelligence*, page 584, Cambridge, MA, 1977.

[17] V. S. Nalwa. *A Guided Tour of Computer Vision*. Addison-Wesley, Reading, MA, 1993.

[18] S. Negahdaripour. Critical surface pairs and triplets. *International Journal of Computer Vision*, 3:293–312, 1989.

[19] S. Negahdaripour and B. K. P. Horn. Direct passive navigation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 9(1):168–176, 1987.

[20] K. Prazdny. On the information in optical flows. *Computer Graphics and Image Processing*, 22:239–259, 1983.

[21] C. R. Rao. *Linear Statistical Inference and Its Applications.* J. Wiley and Sons, New York, NY, 1973.

[22] J. H. Rieger and D. T. Lawton. Processing differential image motion. *Journal of the Optical Society of America*, A((2)):254–360, 1985.

[23] J. Shi and C. Tomasi. Good features to track. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR94)*, pages 593–600, Seattle, WA, 1994.

[24] E. Trucco and A. Verri. *Introductory techniques for 3-D computer vision.* Prentice Hall, Upper Saddle River, NJ, 1998.

[25] A. Verri and T. Poggio. Motion field and optical flow: Qualitative properties. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 11(5):490–498, 1989.

[26] J. Weng, N. Ahuja, and T. Huang. Optimal motion and structure estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 15(9):864–884, 1993.

[27] J. Zhang. *Computing Camera Heading: A Study.* PhD thesis, Stanford University, 1999.

[28] T. Zhang and C. Tomasi. Fast, robust, and consistent camera motion estimation. In *Proc. CVPR 99*, volume 1, pages 164–170, 1999.

[29] X. Zhuang, T. Huang, N. Ahuja, and R. Haralick. A simplified linear optic flow-motion algorithm. *Computer Vision, Graphics, and Image Processing*, 42(3):334–344, 1988.

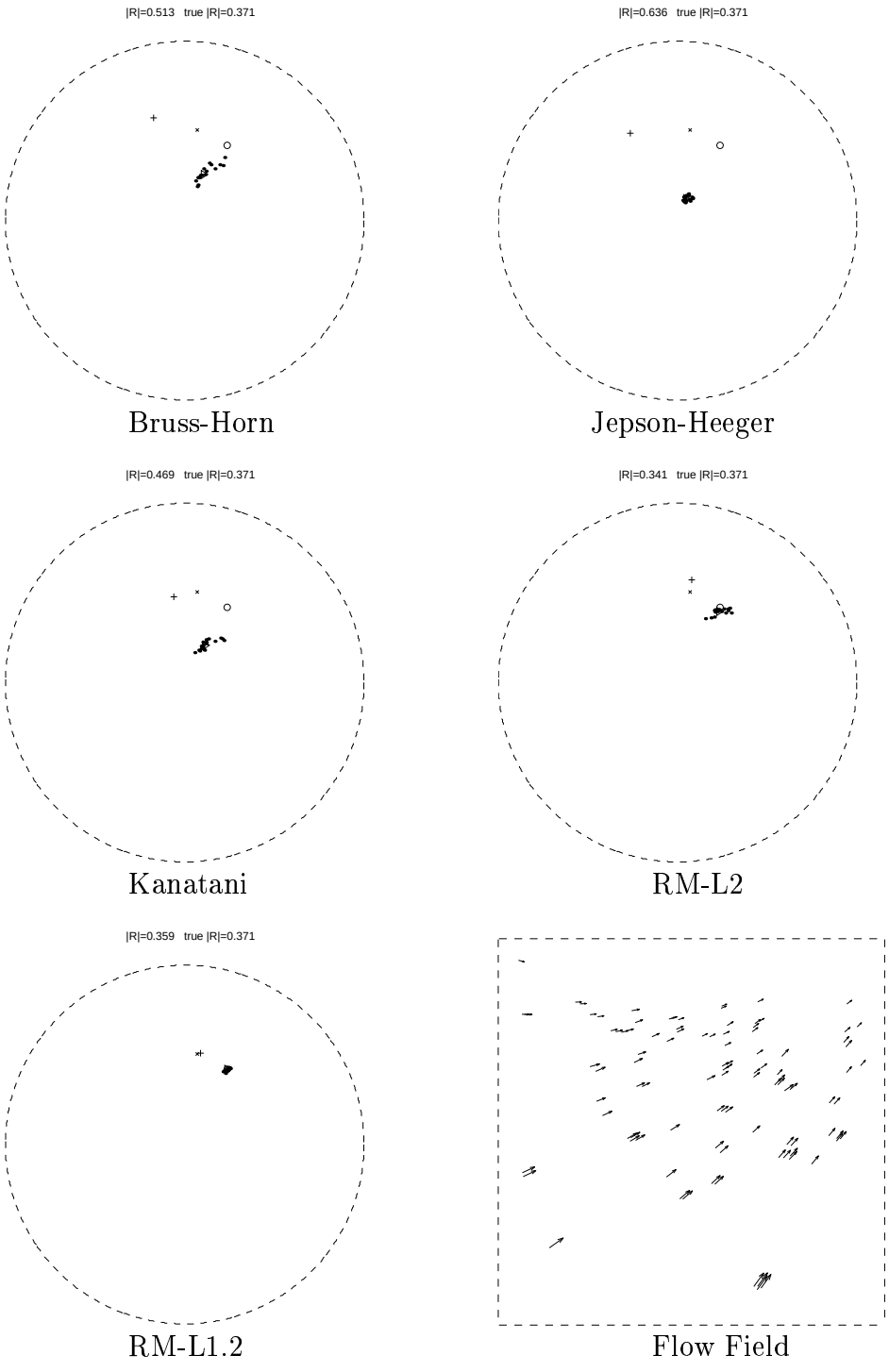


Figure 7: Chessboard: sample size= 100

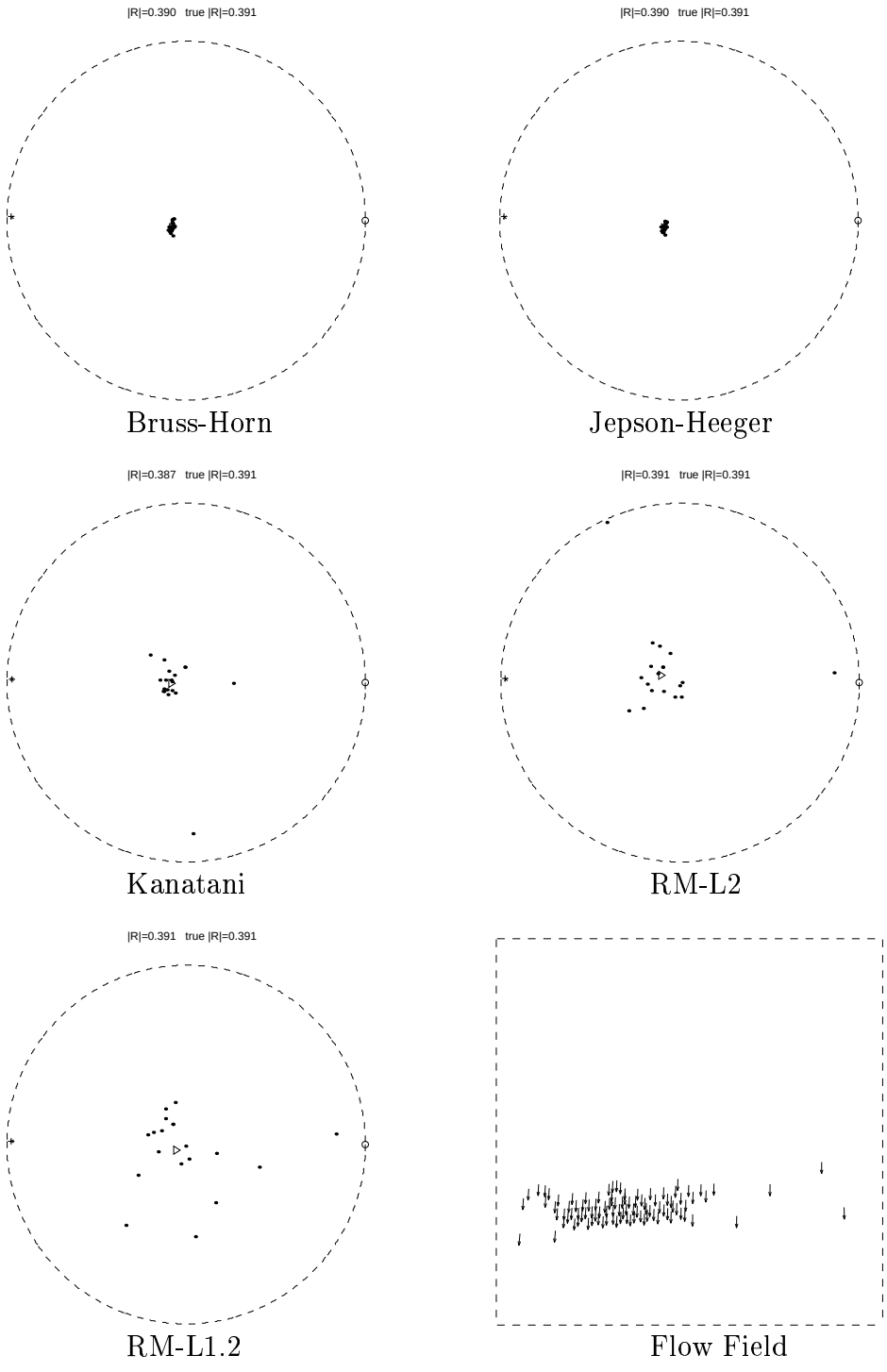
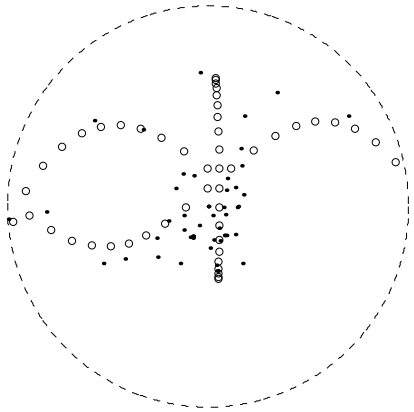
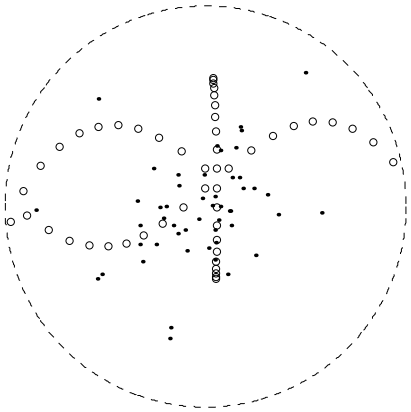


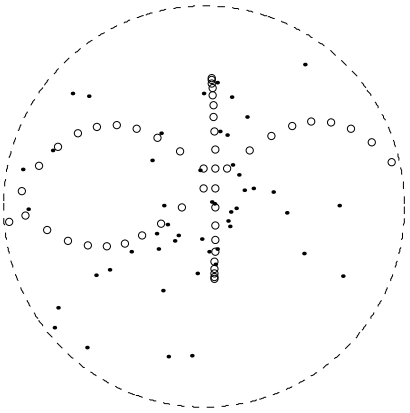
Figure 8: Amiga: sample size= 100



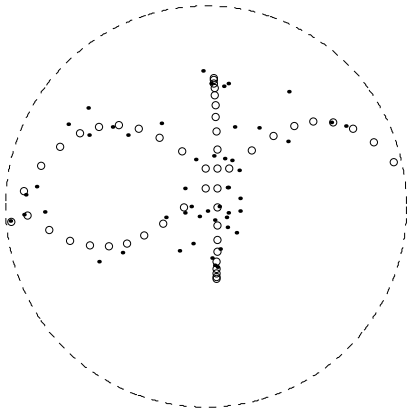
Bruss-Horn



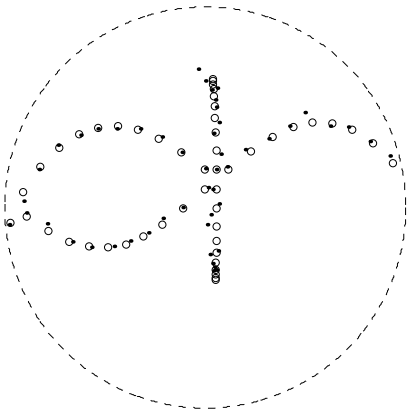
Jepson-Heeger



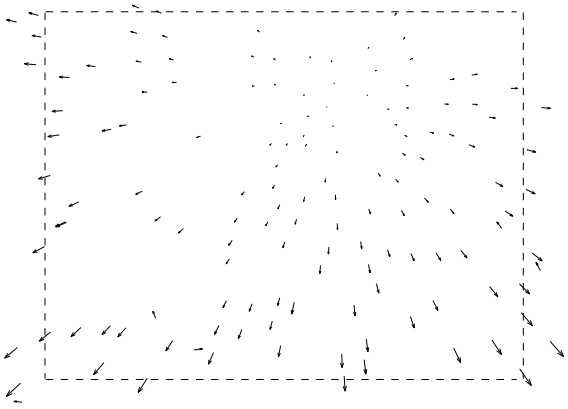
Kanatani



RM-L2



RM-L1.2



Flow Field

Figure 9: Lab dataset

Figure 10: Chessboard Image

Figure 11: Amiga Image

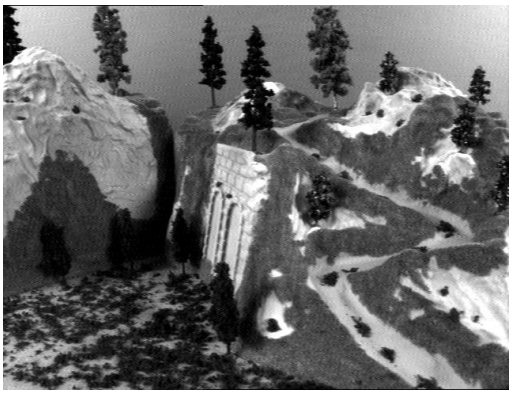


Figure 12: The Block Image