Available online at www.sciencedirect.com
SCIENCE DIRECT®
Computer Vision and Image Understanding xxx (2006) xxx–xxx
Computer Vision and Image Understanding
www.elsevier.com/locate/cviu
ELSEVIER

Optimal instantaneous rigid motion estimation insensitive to local minima

Karl Pauwels *, Marc M. Van Hulle

K.U. Leuven, Laboratorium voor Neuro- en Psychofysiologie, Herestraat 49-bus 1021, B-3000 Leuven, Belgium

Received 19 October 2005; accepted 4 July 2006

Abstract

A novel method is introduced for optimal estimation of rigid camera motion from instantaneous velocity measurements. The error surface associated with this problem is highly complex and existing algorithms suffer heavily from local minima. Repeated minimization with different random initializations and selection of the minimum-cost solution are a common (albeit *ad hoc*) procedure to increase the likelihood of finding the global minimum. We instead show that the optimal estimation problem can be transformed into one of arbitrary complexity, which allows for a gradual regularization of the error function. A simple reweighting scheme is presented that smoothly increases the problem complexity at each iteration. We show that the resulting method retains all the desirable properties of optimal algorithms, such as unbiasedness and minimal variance of the parameter estimates, but is substantially more robust to local minima. This robustness comes at the expense of a slightly increased computational complexity.

Keywords: Egomotion; Optic flow; Calibrated camera; Local minima; Reweighting

1. Introduction

The instantaneous velocity or optic flow field encountered by a moving observer contains an enormous amount of information related to the three dimensional (3D) structure of the environment and to the presence and motion of independently moving objects. Knowledge of the egomotion or self-motion of the observer is a necessary prerequisite to obtain this valuable information. Since small observer motions can have large effects on the optic flow field, it is advisable to extract the egomotion parameters from the optic flow field itself. This, however, is non-trivial and an active topic of research.

The field has matured a lot over the years and a number of 'optimal' algorithms (unbiased and minimal variance of the estimates) have appeared [1,2]. The error function of the optimal problem formulation is however highly nonlinear and contains a large number of local minima [3,4], which renders these algorithms unreliable and hard to use in practical applications. The earlier approaches [5–8], which operate on a linearization of the problem, are no valid alternative. Compared to optimal algorithms, they are extremely sensitive to noise [1,2,9] and the estimates they provide are unsuitable, even as initializations for the optimal methods.

As an alternative to the time-consuming process of repeatedly minimizing with different, random initializations and selection of the minimum-cost solution, we propose to regularize the error function. We reformulate the problem in such a way that the complexity of the error function (the likelihood that algorithms end up in local minima) is controlled by a single parameter. We propose a reweighting scheme that gradually increases the problem complexity during the minimization, until the optimal problem formulation is obtained. We demonstrate, both in simulation and on real data, that the proposed method retains the accuracy of optimal algorithms, but is much less sensitive to local minima. On the extensive set of data investigated, these

* Corresponding author. Fax: +32 16 345960.
E-mail addresses: karl.pauwels@med.kuleuven.be (K. Pauwels), marc.vanhulle@med.kuleuven.be (M.M. Van Hulle).

1077-3142/$ - see front matter © 2006 Published by Elsevier Inc.
doi:10.1016/j.cviu.2006.07.001

improvements come at the cost of less than a doubling in computation time compared to previous optimal algorithms.

2. Problem statement

Under a static environment assumption, the motion of all points in space, relative to a coordinate system centered in the nodal point of the observer's eye, is determined by the translational velocity, $\mathbf{t} = (t_x, t_y, t_z)^{\mathrm{T}}$, and rotational velocity, $\boldsymbol{\omega} = (\omega_x, \omega_y, \omega_z)^{\mathrm{T}}$, of the moving observer. The 3D velocity, $\mathbf{v} = (v_x, v_y, v_z)^{\mathrm{T}}$, of a point in space, $\mathbf{x} = (x, y, z)^{\mathrm{T}}$, is then [10]

$$\mathbf{v} = -\mathbf{t} - \boldsymbol{\omega} \times \mathbf{x}. \tag{1}$$

Under perspective projection and assuming, without loss of generality, a focal length equal to unity, these 3D motion vectors are transformed into a two dimensional velocity or optic flow field. At feature location $\mathbf{x} = (x, y, 1)^{\mathrm{T}}$, the observed flow $\mathbf{u}(\mathbf{x}) = (u_x, u_y)^{\mathrm{T}}$ equals

$$\mathbf{u}(\mathbf{x}) = d(\mathbf{x})A(\mathbf{x})\mathbf{t} + B(\mathbf{x})\boldsymbol{\omega} + \mathbf{n}(\mathbf{x}), \tag{2}$$

where

$$A(\mathbf{x}) = \begin{bmatrix} -1 & 0 & x \\ 0 & -1 & y \end{bmatrix}, \tag{3}$$

$$B(\mathbf{x}) = \begin{bmatrix} xy & -1 - x^2 & y \\ 1 + y^2 & -xy & -x \end{bmatrix}. \tag{4}$$

The observed flow consists of three parts: a component due to the observer's translation (which also depends on the inverse depth $d(\mathbf{x}) = 1/z$), a component due to the observer's rotation, and $\mathbf{n}(\mathbf{x}) = (n_x, n_y)^{\mathrm{T}}$, which is assumed to be independently and identically distributed zero mean Gaussian noise. These different components are illustrated in Fig. 1. Also indicated is $\boldsymbol{\tau}(\mathbf{x}, \mathbf{t}, 1)$, a unit length vector orthogonal to the translational component of the flow:

$$\boldsymbol{\tau}(\mathbf{x}, \mathbf{t}, 1) = \frac{1}{\|A(\mathbf{x})\mathbf{t}\|}\left([A(\mathbf{x})\mathbf{t}]_y, -[A(\mathbf{x})\mathbf{t}]_x\right)^{\mathrm{T}}, \tag{5}$$

where $[\mathbf{p}]_x$ and $[\mathbf{p}]_y$ refer to the x- and y-components of the vector $\mathbf{p}$ respectively. The meaning of the third parameter (equal to unity in Eq. (5)) is explained in Section 4. When depth is eliminated from Eq. (2), the well-known bilinear constraint [11] on translation and rotation is obtained at each location $\mathbf{x}$

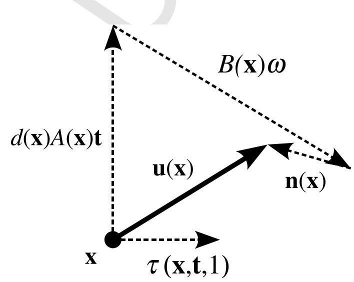


Fig. 1. Optic flow components.

$$\|A(\mathbf{x})\mathbf{t}\|\,\boldsymbol{\tau}(\mathbf{x}, \mathbf{t}, 1)^{\mathrm{T}}(\mathbf{u}(\mathbf{x}) - B(\mathbf{x})\boldsymbol{\omega}) = 0. \tag{6}$$

This particular notation is chosen since it highlights that the constraint is weighted by $\|A(\mathbf{x})\mathbf{t}\|$. This weight term renders the constraints much simpler algebraically but, in the absence of prior knowledge, it is incorrect to weight the different constraints unequally. Instead, the parameters $(\hat{\mathbf{t}}, \hat{\boldsymbol{\omega}})$ should be estimated using the unweighted constraints [2]

$$(\hat{\mathbf{t}}, \hat{\boldsymbol{\omega}}) = \underset{\mathbf{t},\boldsymbol{\omega}}{\operatorname{argmin}} \sum_{\mathbf{x}} [\boldsymbol{\tau}(\mathbf{x}, \mathbf{t}, 1)^{\mathrm{T}}(\mathbf{u}(\mathbf{x}) - B(\mathbf{x})\boldsymbol{\omega})]^2. \tag{7}$$

These constraints represent the normalized, orthogonal deviations from the epipolar lines, and the estimates obtained from Eq. (7) minimize the least-squares image-reprojection error [4]. Since algorithms that operate on this error function obtain the most accurate parameter estimates, they are commonly referred to as 'optimal' [1,2].

3. Previous algorithms

A wide variety of egomotion-estimation methods have been proposed in the past. An important distinction can be made between the earlier approaches, which suffer from biased and/or widely varying estimates, and the more recent optimal algorithms.

3.1. Non-optimal algorithms

One of the first egomotion algorithms has been introduced by Bruss and Horn [11] and consists of a straightforward minimization of the bilinear constraints (Eq. (6)) using nonlinear optimization techniques. Heeger and Jepson (H&J) [5] have proposed a method to compute the heading (normalized translation) without iterative numerical optimization. Their linear subspace method is based on the construction of a set of constraint vectors that are independent of camera rotation. Another linear algorithm has been recently proposed by Ma et al. [6] and is conceptually similar to methods that operate on the discrete epipolar constraint. The heading estimates computed with this algorithm have been shown to be identical to those obtained with H&J but the rotation estimates are better.

The heading estimates obtained with the aforementioned algorithms are all systematically biased. Different bias correction procedures can be found in the literature. Kanatani [7] has introduced a method that subtracts an estimate of the bias from the solution. A second correction procedure has been introduced more recently by Maclean (MAC) [8] as an adaptation to H&J. Contrary to Kanatani's method, this procedure does not require an estimate of the noise variance.

3.2. Optimal algorithms

An optimal, nonlinear algorithm has been introduced by Chiuso et al. (CHI) [1]. This algorithm involves a sequence of fixed-point iterations where each part of the sequence requires solving a linear least-squares problem. Chiuso et al. have proposed iterating between estimates of $\mathbf{t}$ and $\{d(\mathbf{x}), \boldsymbol{\omega}\}$. Since a spherical projection model has been used in their formulation and the other algorithms assume a traditional pin-hole model, we have modified the formulation and implemented the algorithm as follows. Starting from an initial heading estimate $\mathbf{t}^{(1)}$, a rotation estimate $\boldsymbol{\omega}^{(1)}$ is obtained as the linear least-squares solution to Eq. (7). Using both estimates, the least-squares relative inverse depth estimates are obtained at each location $\mathbf{x}$ as

$$d^{(1)}(\mathbf{x}) = \frac{(\mathbf{u}(\mathbf{x}) - B(\mathbf{x})\boldsymbol{\omega}^{(1)})^{\mathrm{T}} A(\mathbf{x})\mathbf{t}^{(1)}}{\|A(\mathbf{x})\mathbf{t}^{(1)}\|^2}, \tag{8}$$

Next, the estimates $\{d^{(1)}(\mathbf{x}), \boldsymbol{\omega}^{(1)}\}$ are used to compute a new translation estimate $\mathbf{t}^{(2)}$ as the linear least-squares solution to the system of Eq. (2). After normalization, the sequence is repeated until the estimates converge. The iterations are stopped when the magnitude of the translation update, $\|\Delta\mathbf{t}\|$, drops below a certain tolerance level ϵ, which is equal to 10^{-13} in all our simulations.

Zhang and Tomasi (Z&T) [2] have introduced a second optimal algorithm. By exploiting the separability of the parameters, a very fast algorithm is obtained that performs Gauss–Newton updates in $\mathbf{t}$. The relative inverse depth estimates $d^{(i)}(\mathbf{x})$ are computed in the same way as CHI Eq. (8) but the heading and rotation estimates are updated as

$$(\Delta\mathbf{t}^{(i+1)}, \boldsymbol{\omega}^{(i+1)}) = \operatorname*{argmin}_{\Delta\mathbf{t},\boldsymbol{\omega}} \sum_{\mathbf{x}} [\boldsymbol{\tau}(\mathbf{x}, \mathbf{t}^{(i)}, 1)^{\mathrm{T}}(\mathbf{u}(\mathbf{x}) - d^{(i)}(\mathbf{x})A(\mathbf{x})\Delta\mathbf{t} - B(\mathbf{x})\boldsymbol{\omega})]^2. \tag{9}$$

Since $\mathbf{t}$ and $d(\mathbf{x})$ appear as a product in Eq. (2), their absolute magnitudes cannot be determined. To remove this ambiguity, the translation update is constrained to be orthogonal to the current estimate: $(\mathbf{t}^{(i)})^{\mathrm{T}}\Delta\mathbf{t}^{(i+1)} = 0$. From Eq. (9), only the translation update is used:

$$\mathbf{t}^{(i+1)} = \mathbf{t}^{(i)} + \Delta\mathbf{t}^{(i+1)}, \tag{10}$$

the rotation estimate is recomputed as the least-squares solution to Eq. (7) (with fixed $\mathbf{t}^{(i+1)}$). This way, more accurate estimates are obtained. The translation estimate is normalized to unit length only after the algorithm has converged.

4. Proposed method

As mentioned in the introduction, the optimal algorithms suffer heavily from local minima. These minima are due to singularities in the unweighted error function that arise from the normalization of the bilinear constraints Eq. (6) by $\|A(\mathbf{x})\mathbf{t}\|$. As a consequence, a singularity exists for each feature where $\mathbf{t} \propto (x, y, 1)^{\mathrm{T}}$. Under certain conditions, which are not uncommon in real-world optic flow fields, these singularities interact and influence larger regions of heading space [3,4]. Optimal algorithms initialized with a heading estimate in these regions are then likely to get trapped in a non-optimal local minimum. The weighted (bilinear) constraints on the other hand do not suffer from these singularities and consequently fewer local minima exist. Only minima due to the so-called bas-relief ambiguity persist (for details, see [1,3]) and these are fewer in number (typically two). However, since the different features are incorrectly weighted, algorithms operating on this error function are not optimal.

We propose a novel method that arrives at optimal estimates by gradually 'unweighting' the bilinear constraints until the unweighted error function is obtained. The method is illustrated for Z&T but can be applied to other optimal algorithms as well. The relative inverse depth estimates are again computed using Eq. (8) but the heading and rotation updates now equal

$$(\Delta\mathbf{t}^{(i+1)}, \boldsymbol{\omega}^{(i+1)}) = \operatorname*{argmin}_{\Delta\mathbf{t},\boldsymbol{\omega}} \sum_{\mathbf{x}} [\boldsymbol{\tau}(\mathbf{x}, \mathbf{t}^{(i)}, \rho^{(i)})^{\mathrm{T}}(\mathbf{u}(\mathbf{x}) - d^{(i)}(\mathbf{x})A(\mathbf{x})\Delta\mathbf{t} - B(\mathbf{x})\boldsymbol{\omega})]^2, \tag{11}$$

where

$$\boldsymbol{\tau}(\mathbf{x}, \mathbf{t}, \rho) = \frac{1}{\|A(\mathbf{x})\mathbf{t}\|^{\rho}}([A(\mathbf{x})\mathbf{t}]_y, -[A(\mathbf{x})\mathbf{t}]_x)^{\mathrm{T}}, \tag{12}$$

Note that the constraint weighting now depends on the value of ρ, which we define as the regularization parameter. When ρ equals zero, Eq. (11) minimizes the weighted (bilinear) constraints and few local minima will be encountered. However, when ρ equals unity, the unweighted error function is minimized (Z&T) and local minima are plentiful. The novelty of our method consists of a gradual increase of ρ (and hence of the complexity of the associated error function) from zero to unity during the Gauss–Newton iterations. Different update schemes are possible, but we use the following in all our experiments. At iteration i, the regularization parameter is updated as follows:

$$\rho^{(i)} = \min\left(1, \rho^{(i-1)} + \lambda\left[\frac{\log_{10}\|\Delta\mathbf{t}^{(i)}\|}{\log_{10}\epsilon}\right]^{+}\right), \tag{13}$$

where $[x]^{+} = \max(x, 0)$ and ϵ equals 10^{-13} (note that $\|\Delta\mathbf{t}\| \approx \epsilon$ at convergence). The parameter λ, the adaptation parameter, determines the adaptation speed and its value is set to 1/4. The choice of this parameter is discussed further in Section 5.4. Since ρ is non-decreasing and upper-bounded, the scheme is guaranteed to converge. In the remainder, we refer to the proposed regularized algorithm (the adaptation scheme from Eq. (13) applied to the heading and rotation updates from Z&T) as REG. Some typical convergence traces for both Z&T (dotted line) and REG (dashed line) are shown in Figs. 2(A) and (B), with the evolution of ρ overlaid (solid line). The traces of Fig. 2(A) have been obtained on a typical problem from Section 5.1 whereas those of Fig. 2(B) have resulted from solving a difficult problem, involving very noisy optic flow. The simple update scheme from Eq. (13) smoothly increases the regularization parameter. If the update magnitude exceeds unity, ρ is left unchanged. Otherwise, ρ is updated proportionally to the size of the update; the smaller the update (indicating that a solution is close by), the stronger ρ is

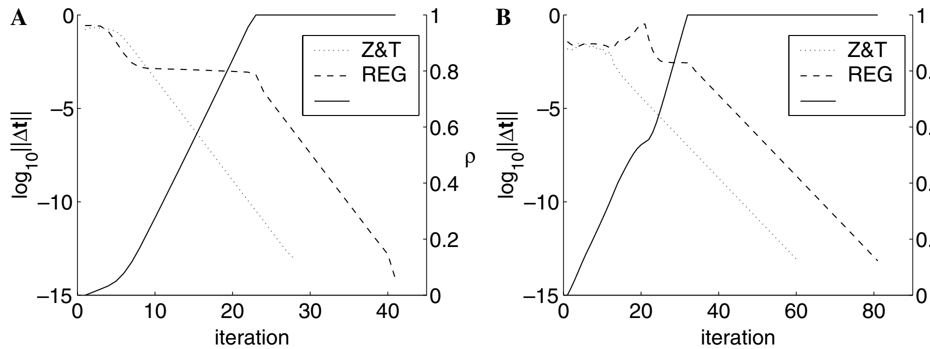


Fig. 2. Convergence traces (left y-axes) for Z&T (dotted lines) and REG (dashed lines) together with the evolution of the regularization parameter ρ (solid line, right y-axes) for two different problems; (A) a typical problem and (B) a problem with very noisy optic flow.

increased. This has a stabilizing effect on the algorithm, as exemplified by the traces of ρ and REG in Fig. 2(B) around iteration 20. As a result of the increased update magnitude at that point, ρ is increased more slowly. This in turn stabilizes the algorithm, as can be seen from the subsequent drop in the update magnitude. This increased stability warrants the slightly increased complexity of the adaptation scheme as compared to one that simply increases ρ with a fixed value at each iteration. The regularization parameter is increased until its maximum value of unity is reached. From that point on, until convergence, ρ is kept fixed and the updates are identical to those of Z&T. The convergence traces from Fig. 2 show that Z&T converges quadratically and that the regularized algorithm converges somewhat slower but still very smoothly. In the experiments performed here, the proposed method requires less than twice the number of iterations needed by Z&T (see below). Since updating ρ creates little overhead, one iteration takes an equal amount of time in both algorithms.

5. Experiments

In this section, the proposed method is extensively compared to some of the algorithms discussed in Section 3. First, in Section 5.1, the algorithms are compared in terms of accuracy of the parameter estimates. This evaluation involves synthetic data only and is applied to both optimal and non-optimal algorithms. Next, in Section 5.2, the proposed method's superior robustness to local minima as compared to other optimal algorithms is demonstrated. For this purpose, a synthetic problem is specifically designed so that the unweighted error function is highly complex. In Section 5.3 the proposed method's robustness is also demonstrated on the well-known real-world NASA-sequence [12]. Finally, Section 5.4 discusses the choice of the adaptation parameter λ.

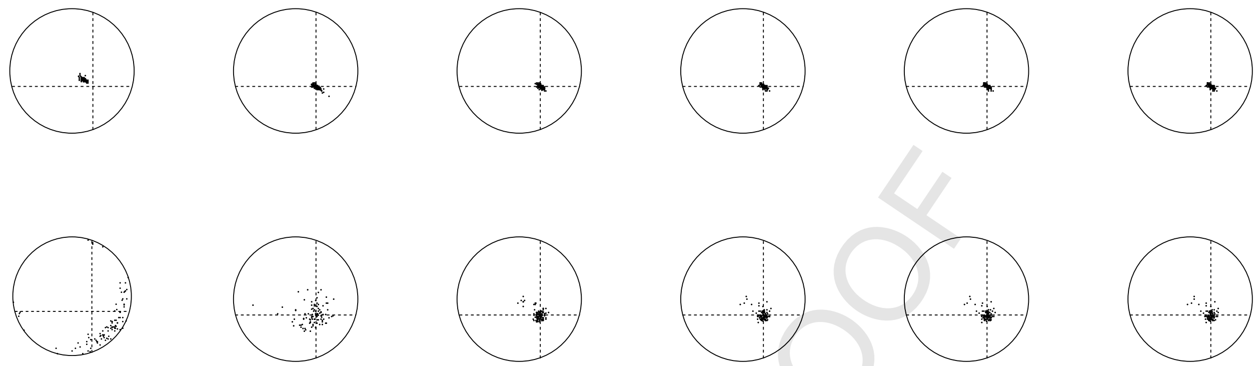
5.1. Bias/variance

We compare H&J, MAC, CHI and Z&T to the proposed method REG in terms of the bias and variance of the heading estimates. Also included is an algorithm identical to REG but with the regularization parameter ρ fixed to zero. This algorithm (BIL) effectively minimizes the weighted (bilinear) constraints. We use implementations provided by Tian et al. [9] for H&J, our own implementations for MAC, BIL, CHI and REG and an implementation provided by Dr. Tong Zhang for Z&T. We have not included the algorithms by Ma et al. [6] (the heading estimates of which are identical to H&J's) and by Kanatani [7] (which fails to provide unbiased estimates consistently throughout this dataset [2]). The rotation estimates are not analyzed since the bias is entirely due to heading estimation and the heading estimates can be visualized and interpreted more easily. We examine the same configuration of translation and rotation as Zhang and Tomasi [2], namely a translation and rotation direction equal to $(4,-3,5)^{\mathrm{T}}$ and $(-1,2,0.50)^{\mathrm{T}}$ respectively. The rotation rate is fixed to 0.23°/frame and the translational magnitude is chosen so that the speeds of the translational and rotational flow components are identical in the center of the random depth cloud. In each experiment, 100 feature locations are randomly chosen and uniformly distributed over the image. The focal length is set to unity. The depth of the features is uniformly distributed between 1 and 4 units of focal length. Independently and identically distributed zero mean Gaussian noise is added to the flow vectors. The signal-to-noise ratio (SNR), defined as: $(\mathrm{E}\{\|\mathbf{u}\|^2\}/\mathrm{E}\{\|\mathbf{n}\|^2\})^{1/2}$, is varied between 10 and 30. For each algorithm, 100 trials are performed, in which the feature locations, depth and noise are randomized. For the nonlinear algorithms (BIL, CHI, Z&T and REG), 15 heading initializations, evenly spread on the unit sphere, are used and the solution with the smallest residual error is retained.

Table 1 contains the heading estimates obtained with all algorithms, for a SNR equal to 10. The field of view (FOV) is equal to 50° and 150° in the top and bottom rows respectively. The estimates are mapped to the upper hemisphere and projected onto a circle. The dashed cross marks the true heading. Example flow fields for the two conditions are shown in Fig. 3. For each algorithm and noise level, the bias, defined as the angular difference between the mean heading estimate and the actual heading, and a 95% confidence cone (measured in degrees), closely related to the

Table 1
Heading estimates obtained with six different algorithms on 100 random trials

FOV	H&J	MAC	BIL	CHI	Z&T	REG
50°						
150°						



The FOV is equal to 50° and 150° in the top and bottom rows respectively (the SNR is equal to 10 for both). Example flow fields for these two conditions are shown in Fig. 3.

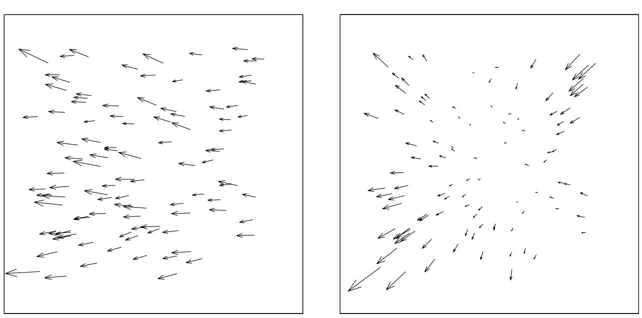


Fig. 3. Example noisy flow fields (magnified 10 times) corresponding to a FOV of 50° (left) and 150° (right). The SNR is equal to 10 in both cases.

variance of the estimates, are computed using techniques from the domain of spherical statistics [13]. Contrary to the bias/variance measures used in previous studies [1,2,9], this more sophisticated analysis clearly brings out the bias in the estimates obtained with H&J. Table 2 contains the variance measure for all algorithms, SNRs and FOVs. The value is underlined in the table if the mean heading estimate is contained within the confidence cone (unbiased). With FOV equal to 50°, this is the case for all algorithms and noise levels except, as expected, for H&J. We also see that the variance in the estimates is much smaller for the nonlinear algorithms than for the linear ones, as observed in other studies [1,2,9]. Note that the variance for CHI, Z&T and REG is nearly identical for all configurations. However, when the constraints are weighted (BIL) the variance is about 10% larger on all occasions, which clearly demonstrates the non-optimality of this approach. Table 3 contains the median number of iterations required by the nonlinear algorithms to reach convergence for the different configurations of Table 2. The median is used since CHI and Z&T are less stable than REG and sometimes fail to converge within the maximum number of iterations (1000) allowed in our experiments. Consequently, the mean would give misleading results in favor of the proposed method. REG needs less than twice the number of iterations required by Z&T to reach convergence. The alternation steps are probably responsible for the slow convergence of CHI. Since alternation methods perform coordinate-descent, flatlining often occurs in valleys of the error surface [14]. The Gauss–Newton algorithm

Table 2
Radii of the 95% confidence cones (in degrees) of the heading estimates obtained with all six algorithms tested for different FOVs and SNRs

FOV	SNR	Non-optimal			Optimal		
		H&J	MAC	BIL	CHI	Z&T	REG
50°	30	0.29	0.28	0.25	0.23	0.23	0.23
	20	0.45	0.43	0.38	0.35	0.35	0.35
	10	0.86	0.97	0.77	0.74	0.74	0.74
150°	30	1.25	1.05	0.45	0.41	0.41	0.41
	20	2.57	1.65	0.69	0.62	0.62	0.62
	10	6.10	4.13	2.25	2.02	2.03	2.02

The value is underlined if the mean heading estimate falls within the confidence cone.

Table 3
Median number of iterations required by the nonlinear algorithms to reach convergence in the simulations of Table 2

FOV	SNR	BIL	CHI	Z&T	REG
50°	30	13	365	16	29
	20	15	368	19	32
	10	20	391	30	41
150°	30	11	118	16	33
	20	13	132	19	36
	10	16	168	26	45

on the contrary, is much faster since translation and rotation are updated simultaneously.

In summary, REG performs equally well as the optimal algorithms CHI and Z&T in terms of unbiasedness and variance of the estimates and requires less than twice the number of iterations to reach convergence as compared to Z&T.

5.2. Local minima

In the previous section we have shown that the accuracy of the proposed method is similar to that of optimal algorithms. Here, we demonstrate the greatly increased robustness to local minima that is achieved by gradually increasing the regularization parameter ρ. The error surface associated with the optimal egomotion problem is known to become flatter in a situation of lateral translation and the number of local minima increases when the feature locations are clustered together, even in the noiseless case [3]. Using this information we have constructed a particularly difficult scenario that enables us to investigate the robustness to local minima of the optimal algorithms: CHI, Z&T and REG. The egomotion consists of a translation and rotation direction equal to $(1,0,0.1)^{\mathrm{T}}$ and $(0,1,0)^{\mathrm{T}}$ respectively. The depth, translation and rotation magnitudes are chosen as in Section 5.1 and the FOV is set to 100°. A total of 500 features are used but, contrary to Section 5.1, they are not uniformly distributed in the image. Instead, their locations are drawn from 20 spatially distinct clusters, the centers of which are uniformly distributed over the image. The cluster centers are indicated with circles in the rightmost figure of Fig. 4. Also shown in this figure is the (subsampled and scaled) flow field used. No noise is added to the computed flow vectors. Each algorithm is run with the same 50,000 heading initializations, randomly sampled from the unit sphere, and is allowed a maximum of 1000 iterations to reach convergence. This large number of initializations allows for a detailed account of the behaviors of the algorithms over the entire heading space.

The first three figures of Fig. 4 contain the estimated headings (black circles) together with the normalized feature locations $\mathbf{x}/\|\mathbf{x}\|$ (black dots). As before, the dashed cross marks the actual heading. It is apparent from these figures that both CHI and Z&T suffer from a large number of local minima, located near clusters of image pixels, whereas REG does not suffer from this problem at all and only finds one additional local minimum besides the global minimum (labeled A in Fig. 4). This second minimum is located near the image center and labeled B in Fig. 4. This minimum is also found by the other algorithms and is a consequence of the bas-relief ambiguity. Techniques have been proposed to discriminate between these two strong minima and to quickly find the other once one is known [1]. In the remainder, we refer to local minima different from these dominant minima as undesired local minima, and to the corresponding heading initializations as undesired initializations. The fact that all undesired local minima are related to clusters of feature locations clearly indicates that they are caused by the singularities in the unweighted error function.

We repeat the experiment for different noise levels and summarize the results in Table 4: the undesired initializations (gray dots) are shown in relation to feature locations (black dots) with the number of undesired initializations underneath each instance. Besides the optimal algorithms CHI, Z&T ($\rho = 1$), and the proposed method REG, we also include a number of algorithms with different, fixed, values of ρ, namely 0.9, 0.8 and 0 (BIL). Each row in Table 4 corresponds to a different noise level. In general, we observe that the number of undesired initializations increases with increasing noise. The fact that noise further increases the error surface complexity and the likelihood of convergence into a local minimum has also been observed by Oliensis [4]. As expected, the locations of these undesired initializations are related to the feature locations. It is notable that the feature clusters have a rather large spatial extent over which they exert their influence and interactions between clusters are clearly visible. The larger number of local minima of CHI is due to flatlining [14]. For all three noise levels, we see that the number of local minima gradually decreases as ρ goes to zero. When ρ equals zero, no undesired local minima are found on any occasion. This nicely illustrates how the problem complexity decreases with decreasing ρ. From the rightmost column of Table 4 it is clear that the proposed method does not suffer from undesired local minima at all, no matter the noise level.

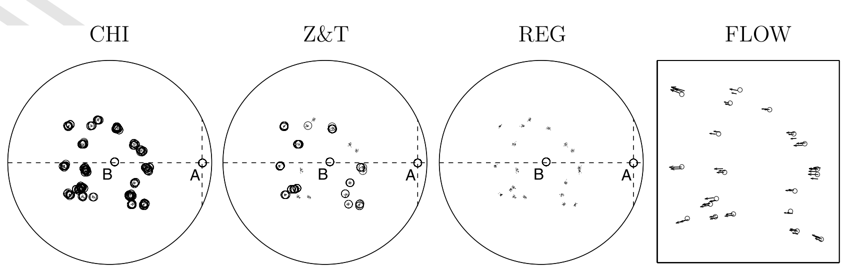
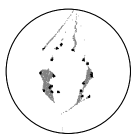
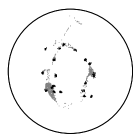
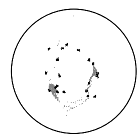
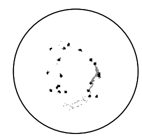
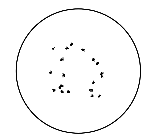
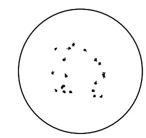
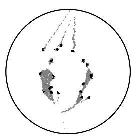
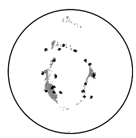
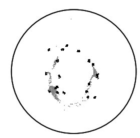
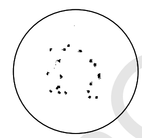
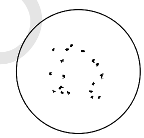
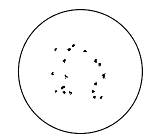
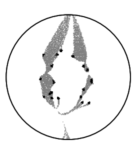
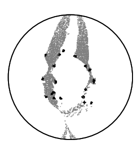
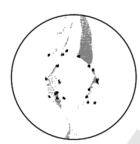
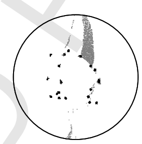
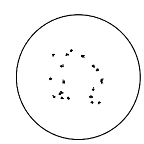
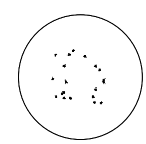


Fig. 4. Small circles in the leftmost figures correspond to heading estimates obtained with the optimal algorithms when initializing with 50,000 distinct random headings. The global minimum is labeled A and the local minimum due to the bas-relief ambiguity is labeled B. Feature locations are indicated with small black dots. The rightmost figure contains the noiseless flow field used (subsampled and magnified 10 times). In this figure, the small circles indicate the feature cluster centers.

Table 4
Undesired initializations (gray dots) in relation to feature locations (black dots) for a number of different algorithms

SNR	CHI	Z&T	$\rho = 0.9$	$\rho = 0.8$	BIL	REG
∞	 2734	 1136	 827	 334	 0	 0
10	 2993	 1633	 847	 20	 0	 0
5	 7599	 7329	 2935	 2215	 0	 0

The results are shown for three noise levels. The number of undesired initializations is shown underneath each instance.

The median number of iterations for these simulations are shown in Table 5. We again see less than a doubling in computation time for REG as compared to Z&T.

Fig. 5 contains error functions of the noiseless local minima problem discussed in this section for different values of the regularization parameter ρ. The error is evaluated over an area of the image similar to Fig. 4 (rightmost). At each location (x, y) the error has been obtained by computing the least-squares rotation estimate (using Eq. (7) with the current value of ρ) assuming a candidate heading $\mathbf{t} \propto (x, y, 1)^{\mathrm{T}}$. It is clear from this figure that the complexity of the error function smoothly increases with increasing ρ.

5.3. Real-world data

We repeat the analysis from the previous section on a real-world image sequence and show that the problem characteristics are not specific to our engineered data set. We use the well-known NASA-sequence [12], the center frame of which is shown in Fig. 6 (left), and compute optic flow using a phase-based algorithm [15]. Since the obtained flow field is very dense (around 50,000 vectors), we randomly select 500 flow vectors to keep the computation times reasonable. This subsampled flow field is shown in Fig. 6 (right). Next, as in Section 5.2, we run the optimal algorithms with 50,000 heading initializations, randomly sampled from the unit sphere, and allow each algorithm a maximum of 1000 iterations to converge. As before, two dominant minima are obtained for all algorithms, one of which is the global optimum (roughly forward translation). These minima are then used to identify the undesired local minima and corresponding initializations. The results are shown in Fig. 7 for CHI, Z&T and REG. Black dots again mark the feature locations (note the small FOV) and gray dots the undesired initializations. The results are in accordance with those obtained on the synthetic datasets: REG clearly shows a superior robustness to local minima. The number of undesired initializations is 10,856 for CHI, 5018 for Z&T and only 4 for REG. The median number of iterations is 1000 for CHI, 48 for Z&T and 58 for REG. Although CHI failed to converge in more than half the trials on this very hard problem, the two dominant minima were clearly dis-

Table 5
Median number of iterations to reach convergence in the simulations of Table 4

SNR	CHI	Z&T	$\rho = 0.9$	$\rho = 0.8$	BIL	REG
∞	138	7	7	7	7	7
10	144	17	17	17	17	30
5	157	32	32	32	30	45

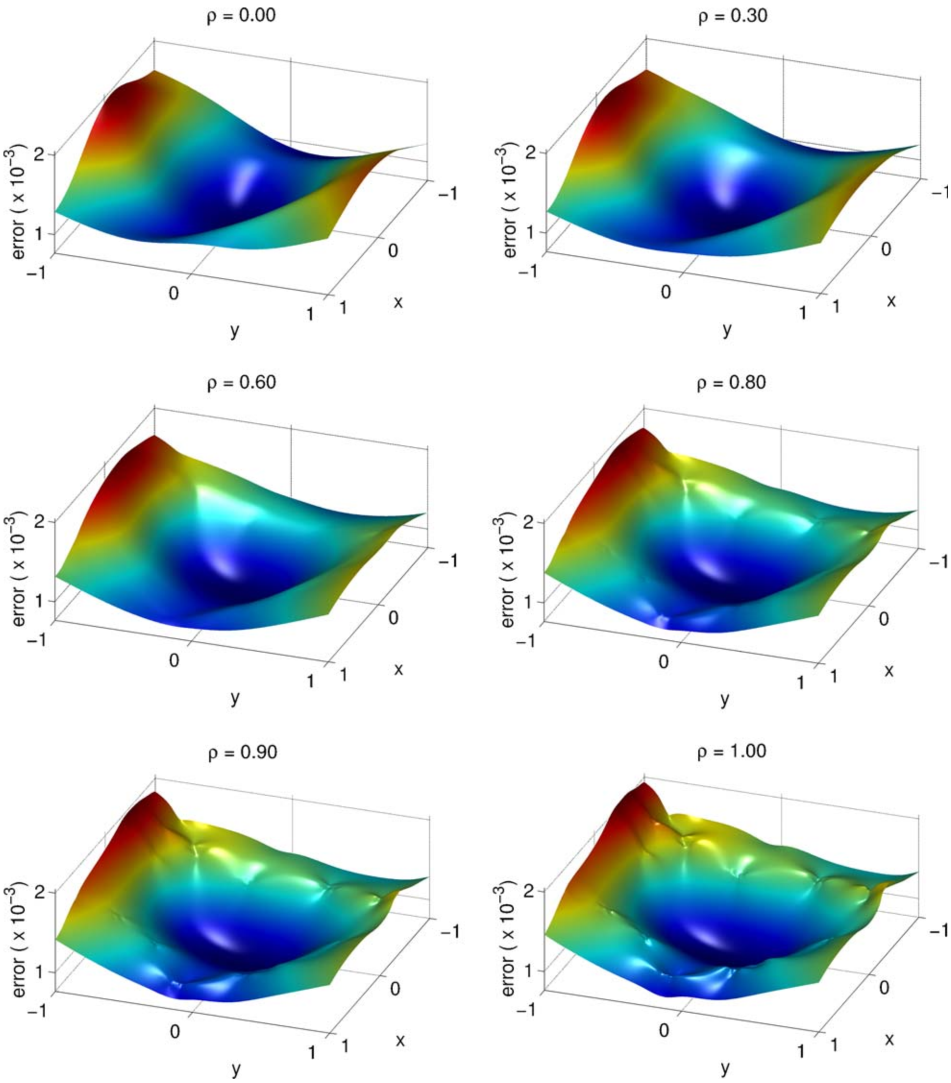


Fig. 5. Error functions of the noiseless local minima problem of Fig. 4 for different values of the regularization parameter ρ. The complexity of the error surface smoothly increases with increasing ρ.

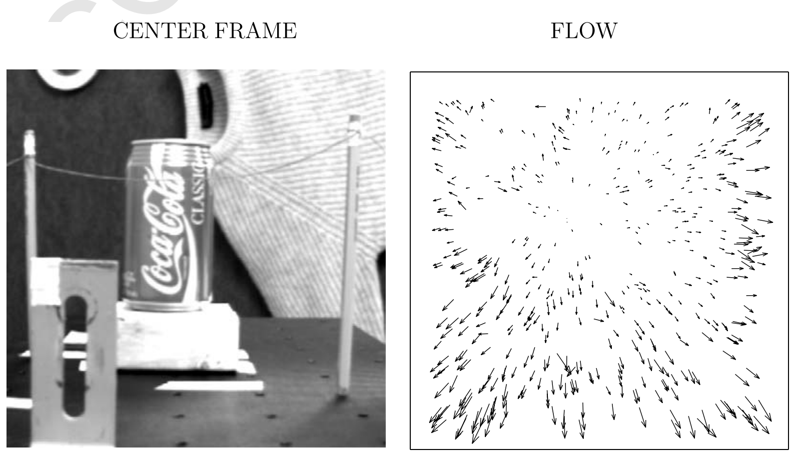


Fig. 6. The center frame of the well-known NASA-sequence (left) and 500 flow vectors (scaled) randomly selected from the complete flow field extracted from this sequence (right).

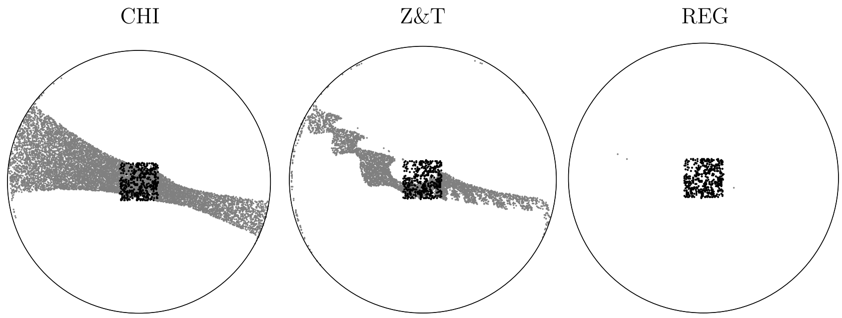


Fig. 7. Undesired initializations (gray dots) in relation to feature locations (black dots) for CHI, Z&T and REG.

cernible. The results are consistent with those of the previous section: the reweighting scheme offers a largely increased robustness to local minima at a relatively small computational cost.

5.4. Choice of adaptation parameter

The parameter λ in the reweighting scheme Eq. (13) controls the speed at which the regularization parameter ρ increases during the Gauss–Newton iterations. The larger its value, the sooner ρ reaches unity and, consequently, the sooner the algorithm starts minimizing the unweighted error function. To examine the influence of the adaptation parameter on the proposed method, we ran the algorithm on the local minima problem of Section 5.2 for different values of λ. The SNR is fixed and equal to five on all occasions. The results are shown in Fig. 8.

Fig. 8(A) shows the number of undesired initializations as a function of λ. As expected, this number increases with increasing λ. In the limit ($\lambda = \infty$, which implies switching to Z&T after one iteration) 5008 undesired initializations are obtained. This is still smaller than the 7329 obtained by Z&T (see Table 4) since in the proposed reweighting scheme, the first iteration is always performed using the weighted (bilinear) constraints ($\rho = 0$). Fig. 8(B) contains the median number of iterations required, as a function of λ. Since the reweighting process slows down when λ is decreased, the number of iterations increases with decreasing λ. However, even at the smallest value of λ shown here (1/8), the number of iterations is still less than twice the number required by Z&T.

We can summarize that, as long as the adaptation parameter λ is between zero and one, the method is relatively insensitive to its value. In this range, a reasonable tradeoff between robustness to local minima and computational requirements is obtained.

6. Discussion

We have presented a novel method that reduces the sensitivity to local minima of optimal egomotion-estimation algorithms by gradually increasing the problem complexity during the optimization. We have demonstrated that the local minima encountered by these algorithms are related to the feature (or feature cluster) locations and, as such, their values can be arbitrary and unrelated to the true solution. This makes these algorithms hard to use in practical applications.

As a remedy, it has been previously suggested to initialize the optimal algorithms with estimates obtained by simplified (linear) algorithms. As shown in Section 5.1 however, noise has a detrimental effect on the accuracy of linear algorithms. We have nevertheless examined this alternative and verified that REG still outperforms Z&T in terms of robustness to local minima, even when the latter is initialized with solutions obtained by BIL (results not shown). Since the variance of all linear algorithms tested is larger than BIL, it is unlikely that their estimates will

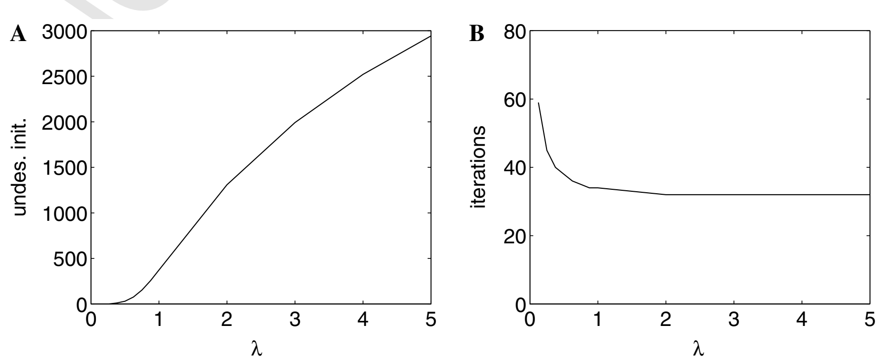


Fig. 8. Number of undesired initializations (A) and required number of iterations (B) to reach convergence on the local minima problem (SNR = 5) as a function of the adaptation parameter λ.

prove better initializations. An alternative way to deal with local minima is to perform multiple runs with different random initializations and retain the solution with the smallest residual. To achieve in this way the same robustness as the proposed method, a large number of runs are necessary and since our method uses fewer than twice the number of iterations required by the fastest optimal algorithm (Z&T), it is computationally more efficient.

Finally, we have shown that the proposed method behaves very similar to BIL in terms of the number of local minima found (typically two). By exploiting the relationship between these minima, the global minimum can thus be found with high certainty in only one or two runs of our method.

Acknowledgments

Thanks to Dr. Tong Zhang and Dr. Tina Y. Tian and coworkers for providing the source code of some of the egomotion algorithms used in this paper. Thanks also to Dr. Temujin Gautama for helpful suggestions on the manuscript. Karl Pauwels and Marc M. Van Hulle are supported by the Belgian Fund for Scientific Research—Flanders (G.0248.03, G.0234.04), the Flemish Regional Ministry of Education (Belgium) (GOA 2000/11), the Belgian Science Policy (IUAP P5/04), and the European Commission (NEST-2003-012963, IST-2002-016276, IST-2004-027017).

References

[1] A. Chiuso, R. Brockett, S. Soatto, Optimal structure from motion: local ambiguities and global estimates, International Journal of Computer Vision 39 (3) (2000) 195–228.

[2] T. Zhang, C. Tomasi, On the consistency of instantaneous rigid motion estimation, International Journal of Computer Vision 46 (2002) 51–79.

[3] T. Xiang, L. Cheong, Understanding the behavior of SFM algorithms: a geometric approach, International Journal of Computer Vision 51 (2) (2003) 111–137.

[4] J. Oliensis, The least-squares error for structure from infinitesimal motion, International Journal of Computer Vision 61 (3) (2005) 1–41.

[5] D. Heeger, A. Jepson, Subspace methods for recovering rigid motion I: Algorithm and implementation, International Journal of Computer Vision 7 (2) (1992) 95–117.

[6] Y. Ma, J. Kosecka, S. Sastry, Linear differential algorithm for motion recovery: a geometric approach, International Journal of Computer Vision 36 (1) (2000) 71–89.

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

[8] W. MacLean, Removal of translation bias when using subspace methods, in: Proceedings of the Eight International Conference on Computer Vision, IEE Computer Society Press, Corfu, Greece, 1999, pp. 753–758.

[9] Y. Tian, C. Tomasi, D. Heeger, Comparison of approaches to egomotion computation, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, USA, 1996, pp. 315–320.

[10] H. Longuet-Higgins, K. Prazdny, The interpretation of a moving retinal image, Proceedings of the Royal Society of London Biology 208 (1980) 385–397.

[11] A. Bruss, B. Horn, Passive navigation, Computer Graphics and Image Processing 21 (1983) 3–20.

[12] J. Barron, D. Fleet, S. Beauchemin, Performance of optical flow techniques, International Journal of Computer Vision 12 (1) (1994) 43–77.

[13] N.I. Fisher, T. Lewis, B.J.J. Embleton, Statistical Analysis of Spherical Data, Cambridge University Press, Cambridge, 1987.

[14] A. Buchanan, A. Fitzgibbon, Damped Newton algorithms for matrix factorization with missing data, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, San Diego, 2005, vol. 2, pp. 316–322.

[15] T. Gautama, M. Van Hulle, A phase-based approach to the estimation of the optical flow field using spatial filtering, IEEE Transactions on Neural Networks 13 (5) (2002) 1127–1136.