2008 IEEE/RSJ International Conference on Intelligent Robots and Systems
Acropolis Convention Center
Nice, France, Sept, 22-26, 2008

Accurate and Robust Ego-Motion Estimation using Expectation Maximization

Gijs Dubbelman[1,2], Wannes van der Mark[1] and Frans C.A. Groen[2]

***Abstract*— A novel robust visual-odometry technique, called EM-SE(3) is presented and compared against using the random sample consensus (RANSAC) for ego-motion estimation. In this contribution, stereo-vision is used to generate a number of minimal-set motion hypothesis. By using EM-SE(3), which involves expectation maximization on a local linearization of the rigid-body motion group SE(3), a distinction can be made between inlier and outlier motion hypothesis. At the same time a robust mean motion as well as its associated uncertainty can be computed on the selected inlier motion hypothesis. The data-sets used for evaluation consist of synthetic and large real-world urban scenes, including several independently moving objects. Using these data-sets, it will be shown that EM-SE(3) is both more accurate and more efficient than RANSAC.**

***Index Terms*— Robust estimation, visual-odometry, Stereo-vision.**

I. Introduction

In this article, the focus is on robust ego-motion estimation of a moving vehicle using an onboard stereo-rig, this is also known as stereo-based visual-odometry. Stereo processing allows estimation of the three dimensional (3D) location and associated uncertainty of landmarks observed by the stereo camera. Subsequently, 3D point clouds can be obtained for each stereo frame. By establishing the frame-to-frame correspondences of landmarks, the point clouds of successive stereo-frames can be related to each other. From these corresponding point clouds the frame-to-frame motion of the stereo-rig can be estimated, Matei and Meer [11], Umeyama [24]. By concatenating all the frame-to-frame motion estimates, the pose of the stereo-rig and vehicle can be tracked. In general, vision based approaches for ego-motion estimation are susceptible to outlier landmarks. Sources of outlier landmarks range from sensor noise, correspondences errors, to independent moving objects such as cars or people that are visible in the camera views.

In the past decade the random sample consensus (RANSAC) method, developed by Fischler and Bolles [3], emerged as the golden standard for robust parameter estimation in computer vision. Hence, it is used in many visual odometry systems, Maimone et al. [9], Nistér et al. [14], [15] and Olson et al. [16]. The idea behind RANSAC is to estimate a large number of minimal-set (containing for example six landmarks) motion hypothesis. For each motion hypothesis also a robust score is calculated, this score is based on the alignment of the motion hypothesis with all landmarks in the set. The best scoring minimal-set motion hypothesis is taken as the final robust estimate. Often, refinement of the initial RANSAC estimate is performed using a more advanced estimator. Since [3], many different RANSAC approaches have emerged, for example LmedS, Rosin [20], for an overview see Torr and Murray [23]. Most notably for the case of visual-odometry is preemptive RANSAC, Nistér [13], which allows evaluating a larger number of minimal-set motion estimates in the same amount of computation time. Using some sort of multi-frame optimization or multi-frame landmark tracking is also frequently used. In the extreme case, this leads to simultaneous localization and mapping (SLAM) approaches such as that of Elinas et al. [2]. In this paper however, the focus is on frame-to-frame approaches only.

Closely related to ego-motion estimation is the problem of pose estimation, which involves estimating the motion of an object from images taken by a (fixed) camera. Recently, alternative approaches to RANSAC have emerged for robust pose estimation, Pennec et al. [19] and Subbarao et al. [22]. These methods use geometrical computation within Riemannian geometry or Lie algebra, Selig [21], together with robust statistics to obtain reliable results.

In line with this work, a novel algorithm called EM-SE(3) is presented. It uses expectation maximization (EM) on a local linearization of the rigid-body motion group i.e. SE(3). In this article it will be compared to RANSAC using both synthetic and real-world data.

II. Expectation maximization in SE(3)

The goal of the EM-SE(3) algorithm is computing a robust mean motion $\bar{M}$ and its covariance Σ from a set of rigid-body transformations $\mathbf{M} = \{M^1 ..... M^n\}$. For this, expectation maximization on the group of 3D Euclidean motion is used. EM is an iterative algorithm often used for finding the parameters of a mixture model. Here, a similar approach is taken for finding the parameters of a two class model i.e. inlier motion hypothesis versus outlier motion hypothesis. The inlier motion hypothesis are minimal-set motion estimates from the true ego-motion perturbed with Gaussian noise. The outlier motion hypothesis are erroneous minimal-set motion estimates, they are caused by inclusion of outlier landmarks into the minimal-set. The class of outlier motions will be modeled with an uniform distribution.

Electro-Optical Systems[1], TNO Defence, Security and Safety, Oude Waalsdorperweg 63, 2509 JG The Hague, The Netherlands, {gijs.dubbelman, wannes.vandermark}@tno.nl

Intelligent Systems Laboratory Amsterdam[2] (ISLA), University of Amsterdam, Kruislaan 403, 1098 SJ Amsterdam, The Netherlands groen@science.uva.nl

A. Representing SE(3)

The EM-SE(3) algorithm requires the ability to compute a weighted mean translation and rotation from a set of rigid-body transformations. Whereas, translations belong to a vector space, more precisely the Euclidean space, rotations do not. Therefore, computing the usual weighted arithmetic mean on a set of rotations is far from appropriate. In fact, for most mathematical representations of rotation, for instance rotation matrices and quaternions, it will not result in a proper rotation, unless the result is orthogonalized in the case of rotation matrices, or normalized in the case of quaternions, Gramkow [4].

Because a rigid-body transformation often includes rotation, it also does not belong to a vector space. Therefore, the problem to be solved is that of mapping the space of rigid-body transformations to a space that is as close to a vector space as possible. Solutions for this problem can be found in the field of differentiable manifolds, Loring [7], particularly in Lie algebra, Riemannian geometry and Geometric algebra Dorst et al. [1]. These fields of mathematics are still actively studied and an introduction is beyond the scope of this text. Because the complete group structure of SE(3) is irrelevant when computing the mean motion, the problem becomes more straightforward. Note that the set of motion hypothesis are estimates for the same real-world motion at one particular point in time. Therefore, the rotations do not transform any of the other rotations or translations, as would be the case when the motion hypothesis formed a chain of rigid-body transformations. This allows processing the translations separately from the rotations, hence only the rotational part has to be mapped. Keeping the rotations and translations separated when applying a metric on SE(3) is in line with the work by Park [17].

In this work rotations will represented using unit quaternions. Quaternions will be denoted as $\mathbf{q}$ and consists of a one dimensional real part q and a three dimensional spatial part $\vec{\mathbf{q}} = (q_i\, i + q_j\, j + q_k\, k)$ given on the imaginary basis i, j, k ($i^2 = j^2 = k^2 = ijk = -1$). Thus $\mathbf{q} = (q + \vec{\mathbf{q}})$. Quaternion addition is simply the pairwise addition of their elements. The inverse of a unit quaternion is given by its conjugate $\mathbf{q}^* = (q - \vec{\mathbf{q}})$. The identity quaternion is defined as $\mathbf{e} = (1 + 0i + 0j + 0k)$. A rotation around a normalized axis $\vec{\mathbf{r}}$ with angle θ can be put into quaternion form with $\mathbf{q} = (\cos(\theta/2) + \sin(\theta/2)\vec{\mathbf{r}})$. To rotate an arbitrary vector $\vec{\mathbf{v}}$ it can be represented as a quaternion i.e. $\mathbf{v} = (0 + \vec{\mathbf{v}})$. Given the quaternion $\mathbf{q}$ we apply its rotation on $\mathbf{v}$ by using $\mathbf{v} = \mathbf{q}\mathbf{v}\mathbf{q}^*$. Where the quaternion product is defined as $\mathbf{q_1 q_2} = (q_1 q_2 - \vec{\mathbf{q}}_\mathbf{1} \cdot \vec{\mathbf{q}}_\mathbf{2} + q_1 \vec{\mathbf{q}}_\mathbf{2} + q_2 \vec{\mathbf{q}}_\mathbf{1} + \vec{\mathbf{q}}_\mathbf{1} \times \vec{\mathbf{q}}_\mathbf{2})$ and the dot and cross product are defined as usual. Note that the quaternion product is anti-commutative, associative and left/right distributive over addition.

The space of unit quaternions can be thought of as the three dimensional surface of an unit sphere in four dimensional space. Because, rotations only have three degrees of freedom the manifold only has three dimensions. Clearly, the rotation manifold is not a vector space and therefore the Euclidean distance between quaternions is far from appropriate. Using Riemannian geometry the points, i.e. rotations on the manifold, can be mapped to a vector space, Pennec [18]. This can be thought of as locally "linearizing" the unit sphere in 4D space. Therefore, this vector space is often called the tangent space. To go from the group of rotations $SO(3)$ to its tangent space $\mathfrak{so}(3)$ at the identity the so called logarithmic mapping $\log : \mathrm{SO}(3) \rightarrow \mathfrak{so}(3), \quad \log(\mathbf{q}) = \mathfrak{q}$ can be used. Mapping back from the tangent space $\mathfrak{so}(3)$ to $SO(3)$ can be done using the exponential mapping $\exp : \mathfrak{so}(3) \rightarrow \mathrm{SO}(3), \quad \exp(\mathfrak{q}) = \mathbf{q}$. For unit quaternions these mappings are

$$\mathfrak{q} = \log(\mathbf{q}) = \begin{cases} (\arccos(q)\frac{\vec{\mathbf{q}}}{\|\vec{\mathbf{q}}\|}), & q \neq 0 \\ (0,\ 0,\ 0)\,, & q = 0 \end{cases} \tag{1}$$

and

$$\mathbf{q} = exp(\mathfrak{q}) = \begin{cases} (\cos(\|\mathfrak{q}\|) + \sin(\|\mathfrak{q}\|)\ \frac{\mathfrak{q}}{\|\mathfrak{q}\|}), & \|\mathfrak{q}\| \neq 0 \\ \mathbf{e}, & \|\mathfrak{q}\| = 0 \end{cases}. \tag{2}$$

Treating SO(3) as a vector space is only appropriate close to the mapping point (as is also the case with regular linearization). Therefore, it is useful to be able to define the tangent space at other points than the identity. For this a placement function, which is composed of quaternion operations, can be used. Placing $\mathbf{q}_2$ in the tangent space at $\mathbf{q}_1$ can be done using

$$\mathfrak{q} = \log_{\mathbf{q_1}}(\mathbf{q_2}) = \log(\mathbf{q}_\mathbf{1}^* \mathbf{q_2}) \tag{3}$$

and retrieving $\mathbf{q}_2$ from the tangent space defined at $\mathbf{q}_1$ can be accomplished with

$$\mathbf{q}_2 = \exp_{\mathbf{q}_1}(\mathfrak{q}_2) = \mathbf{q}_1 \exp(\mathfrak{q}_2). \tag{4}$$

These operations are important because they allow linearization of SO(3) at a certain point of interest, for instance at the mean rotation $\bar{\mathbf{q}}$. This causes less linearization errors than always mapping at the identity.

The idea behind locally mapping SO(3) to a vector space is allowing to treat $\mathfrak{so}(3)$ as if it were $\mathbb{R}^3$ and subsequently use statistical methods designed for $\mathbb{R}^3$ on $\mathfrak{so}(3)$. The problem that is solved here is that $SO(3)$ is not an Euclidean space and therefore the Euclidean distance in not appropriate. Because the Euclidean distance is the basis for most statistical methods for geometric computation, it is assumed that by transforming $SO(3)$ to $\mathfrak{so}(3)$ and treat it as $\mathbb{R}^3$ statistical inferences can be made about $SO(3)$.

For notational convenience an element of SE(3) consisting of a translation $\vec{\mathbf{v}}$ and a rotation embedded in an unit quaternion $\mathbf{q}$ is given by $M = (\vec{\mathbf{v}}, \mathbf{q})$. Furthermore, the localized logarithmic and exponential mappings on the used representation of SE(3) are defined as

$$log_{M_1}(M_2) \equiv (\vec{\mathbf{v}}_2, \log_{\mathbf{q}_1}(\mathbf{q}_2)) = (\vec{\mathbf{v}}_2, \mathfrak{q}_2) \equiv \mathfrak{m}_2, \tag{5}$$

$$exp_{M_1}(\mathfrak{m}_2) \equiv (\vec{\mathbf{v}}_2, \exp_{q_1}(\mathfrak{q}_2)) = (\vec{\mathbf{v}}_2, \mathbf{q}_2) \equiv M_2. \tag{6}$$

Putting the translation and rotation together in a seven dimensional vector allows for compact formulas and efficient computation. However, one has to note that they are distinct elements, and for example, computing the correlation between a dimension in the translational part and a dimension in the rotational part is prohibited.

B. Expectation maximization in $\mathfrak{se}(3)$

In order to perform EM on SE(3), the weighted mean of a set of motion hypothesis $\{M_1.....M_n\}$ must be calculated. Let us start by considering the unweighed case. Given the logarithmic and exponential mappings defined in section II-A the mean motion $\bar{M}$ is obtained by minimizing the following function:

$$\underset{\bar{M} \in SE(3), \hat{M} \in SE(3)}{\arg\min} \quad \sum_{i=1}^{n} \left\| log_{\hat{M}}(\bar{M}) - log_{\hat{M}}(M_i) \right\|. \tag{7}$$

The challenge here is that two goals have to be met simultaneously. That is, minimizing the summed squared difference between $\bar{M}$ and the motion hypothesis in the set $\{M_1.....M_n\}$. Secondly, choosing a mapping point $\hat{M}$ such that treating $\mathfrak{so}(3)$ as $\mathbb{R}^3$ is as appropriate as possible. If a mapping point with minimal summed squared distance to the motion hypothesis in the set is used, the optimal linearization point and the mean motion are the same. This problem can be solved by using an iterative approach, see eq. 8. Firstly, linearization of SE(3) is performed at the identity (or the motion estimate from the previous time-step) by using the logarithmic map. Then, the mean rotation is computed in $\mathfrak{se}(3)$ and mapped back to SE(3) using the exponential map. In the next iteration, linearization of SE(3) is performed using the new mean. Again, the mean is computed in $\mathfrak{se}(3)$ and mapped back to SE(3). This continues until convergence or until j has reached a maximum.

$$\bar{M}^j = exp_{\bar{M}^{j-1}} \left(\frac{\sum_{i=1}^{n} w_n \, log_{\bar{M}^{j-1}}(M_i)}{\sum_{i=1}^{n} w_n} \right). \tag{8}$$

Because the minimal-set motion hypothesis can contain outliers, it is inappropriate to use eq. 8 directly. Therefore, eq. 8 is embedded in an EM algorithm. The benefit of EM over other robust statistical methods, such as M-estimators or mean shift, is that besides a robust motion estimate also a Monte Carlo estimate of the covariance is returned. Here, the primary goal of the EM algorithm is to make a distinction between inlier and outlier minimal-set motion hypothesis and subsequently computing the mean on the inliers only. Inlier motion hypothesis are modeled with a single Gaussian distribution. The outlier motion hypothesis are modeled with a fixed uniform distribution. The mixing weights of the inlier and outlier classes at iteration k of the EM algorithm are given by $p(I)^k$ and $p(O)^k = 1 - p(I)^k$ respectively. Furthermore, $p(\mathfrak{m}_i|\theta_I^k)$ is the probability that motion hypothesis M_i is an inlier given the inlier parameters $\theta_I^k = \{\bar{\mathfrak{m}}, \Sigma_{\mathfrak{m}}\}$ i.e. the mean motion with its covariance at iteration k. Similar notation is used for the constant pdf. of the outliers i.e. $p(\mathfrak{m}_i|\theta_O)$. Note, that these probabilities are calculated on the logarithmic mappings of the motion samples. The parameters that will optimized are the inlier mixing weight and the inlier density parameters i.e. $\Psi^k = \left\{p(I)^k, \theta_I^k\right\}$. Given these quantities the log expectation of the motion samples can be written as

$$Q(\Psi^k, \Psi^{k-1}) = \sum_{i=1}^{n} I_i \log(p(I)^k p(\mathfrak{m}_i|\theta_I^k)) + O_i \log(p(O)^k p(\mathfrak{m}_i|\theta_O)) \tag{9}$$

, where the inlier weights I_i and outlier weights O_i are expressed as

$$\begin{aligned} I_i &= \frac{p(I)^{k-1} p(\mathfrak{m}_i|\theta_I^{k-1})}{p(I)^{k-1} p(\mathfrak{m}_i|\theta_I^{k-1}) + p(O)^{k-1} p(\mathfrak{m}_i|\theta_O)} \\ O_i &= \frac{p(O)^{k-1} p(\mathfrak{m}_i|\theta_O)}{p(I)^{k-1} p(\mathfrak{m}_i|\theta_I^{k-1}) + p(O)^{k-1} p(\mathfrak{m}_i|\theta_O)} \end{aligned}. \tag{10}$$

For clarity, log and exp have their usual meaning in formulas concerning probabilities. Since, $p(\mathfrak{m}_i|\theta_I)$ is Gaussian it is given by

$$p(\mathfrak{m}_i|\theta_I) = \frac{1}{(2\pi)^{1/6}\sqrt{|\Sigma_{\mathfrak{m}}|}} \, e^{-\frac{1}{2}(\mathfrak{m}_i - \bar{\mathfrak{m}})^{\mathrm{T}} \, \Sigma_{\mathfrak{m}} \, (\mathfrak{m}_i - \bar{\mathfrak{m}})}. \tag{11}$$

Note that $\Sigma_{\mathfrak{m}}$ must have the form

$$\Sigma_{\mathfrak{m}} = \begin{bmatrix} \Sigma_{\vec{\mathfrak{v}}} & 0 \\ 0 & \Sigma_{\mathfrak{q}} \end{bmatrix}.$$

Thus, the translation and rotation are independent and $p(\mathfrak{m}) = p(\vec{\mathfrak{v}})p(\mathfrak{q})$. The goal is to maximize Q with respect to θ_I^k i.e. the new mean motion $\bar{\mathfrak{m}}^k$ and its covariance $\Sigma_{\mathfrak{m}}^k$. Taking the derivative of Q to $\bar{\mathfrak{m}}^k$, setting it to zero and solving for $\bar{\mathfrak{m}}^k$ gives

$$\bar{\mathfrak{m}}^k = \frac{\sum_{i=1}^{n} I_i \mathfrak{m}_i}{\sum_{i=1}^{n} I_i}. \tag{12}$$

This equation does not take into account the errors introduced by the logarithmic mapping. Therefore, we choose to use eq. 8 for iteratively computing the weighted mean motion during every EM iteration. Next optimizing Q with respect to $\Sigma_{\mathfrak{m}}^k$ based on the new mean results in

$$\Sigma_{\mathfrak{m}}^k = \frac{\sum_{i=1}^{n} I_i (\mathfrak{m}_i - \bar{\mathfrak{m}}^k)(\mathfrak{m}_i - \bar{\mathfrak{m}}^k)^{\mathrm{T}}}{\sum_{i=1}^{n} I_i}. \tag{13}$$

Clearly, computing the elements of the upper right block and the lower left block of $\Sigma_{\mathfrak{m}}$ is not necessary and can be set to zeros. Finally, Q will be optimized with respect to the inlier mixing weight $p(I)^k$, this gives

$$p(I)^k = \frac{1}{n} \sum_{i=1}^{n} I_i. \tag{14}$$

The EM algorithm iterates between computing the weights with eq. 10 given the current parameters Ψ^{k-1}, i.e. the

expectation step, and computing the new parameters Ψ^k with eq. 8,13,14 given the new weights, i.e. the maximization step. This goes on until convergence or k has reached a maximum.

III. Stereo vision based motion estimation

The motion hypothesis $\{M_1.....M_n\}$ needed for the EM-SE(3) algorithm are estimated using stereo vision. It is assumed that stereo images are rectified according to the epipolar geometry of the used stereo-rig, Hartley and Zisserman [5]. To obtain the landmarks needed for motion estimation, image feature correspondences must be established between successive stereo-frames and between the images in the stereo-frames themselves. To this purpose the Scale Invariant Feature Transform (SIFT), Lowe [8], is used. A threshold is applied on the distance between SIFT descriptors to ensure reliable matches between image features. Furthermore, the epipolar constraint, back-and-forth and left-to-right consistency are enforced. From an image point in the left image $\vec{v}_l = [x_l, y_l]'$ and its corresponding point in the right image $\vec{v}_r = [x_r, y_r]'$ their disparity can be obtained with sub-pixel accuracy $d = x_l - x_r$. Using the disparity d, the focal length of the left camera f and the stereo base line b, the 3D position of the landmark imaged by $\vec{v}_l$ and $\vec{v}_r$ relative to the left camera can be recovered with

$$\vec{\mathbf{v}} = \left[v_x = \frac{x_l b}{d}, \quad v_y = \frac{y_l b}{d}, \quad v_z = \frac{fb}{d} \right]^{\mathrm{T}}. \tag{15}$$

For each reconstructed landmark we also estimate its three dimensional uncertainty as covariance matrix $\Sigma_{\vec{\mathbf{v}}}$. This uncertainty is based on error-propagation of the image feature position uncertainty using the Jacobian J of the reconstruction function, Matthies and Shafer [12],

$$\Sigma_{\vec{\mathbf{v}}} = J \begin{bmatrix} \Sigma_{\vec{v}_l} & 0 \\ 0 & \Sigma_{\vec{v}_r} \end{bmatrix} J^{\top}, \tag{16}$$

$$J = \begin{bmatrix} \frac{-x_l b}{d^2} + \frac{b}{d} & 0 & \frac{x_l b}{d^2} & 0 \\ \frac{-y_l b}{d^2} & \frac{b}{d} & \frac{y_l b}{d^2} & 0 \\ \frac{-fb}{d^2} & 0 & \frac{fb}{d^2} & 0 \end{bmatrix}. \tag{17}$$

Here, $\Sigma_{\vec{v}_l}$ and $\Sigma_{\vec{v}_r}$ are the image feature covariance matrices in the left and right images respectively. For our purposes, only the shape and relative magnitude of image feature position uncertainty is important. Therefore, it suffices to estimate $\Sigma_{\vec{v}}$ with

$$\Sigma_{\vec{v}} = s \frac{G^{-1}}{\sqrt{|G^{-1}|}}. \tag{18}$$

Where s is the scale of the image feature in scale-space and G is the gradient Grammian at $\vec{v}$, Zhou et al. [25]. The stereo reconstruction procedure results in normally distributed uncertainty in landmark positions. The uncertainty is significantly larger in the direction from the optical center to the landmark position. Furthermore, the magnitude of the uncertainty increases with the distance from the landmark to the camera. Clearly, the noise in landmark position is anisotropic and inhomogeneous. For estimating motion parameters for these kind of error distributions the HEIV motion estimator was developed by Matei and Meer [10], [11]. For successive stereo frames, the process described above can be used to obtain a set of corresponding landmarks with their 3D position and related uncertainty. This results in a set $\mathbf{L}$ of corresponding 3D points i.e. $\mathbf{L} = \{L_1.....L_m\}$, where $L_i = (\vec{\mathbf{v}}_{i,t-1}, \Sigma_{i,t-1}, \vec{\mathbf{v}}_{i,t}, \Sigma_{i,t})$. Then the HEIV motion estimator is used to generate n minimal-set motion hypothesis $\{M_1.....M_n\}$. Each motion hypothesis M is estimated on six landmarks from the set $\mathbf{L}$. The number of motion hypothesis n required to attain a probability of p that there is at least one motion hypothesis based on only inlier landmarks, can be computed with

$$n = \log(1-p)/\log(1-(1-\varepsilon)^6). \tag{19}$$

Where ε is the probability of selecting an outlier landmark.

Often, it is useful to measure how well two time-corresponding 3D points align with a motion estimate. For this we use the Bhattacharyya distance. Given two points $\vec{\mathbf{v}}_{t-1}$ and $\vec{\mathbf{v}}_t$ with their reconstruction uncertainties Σ_t and Σ_{t-1}, we apply the estimate motion M on $\vec{\mathbf{v}}_t$ and $\vec{\Sigma}_t$. The Bhattacharyya distance is then computed as

$$\begin{array}{c} D = \frac{1}{4}(\vec{\mathbf{v}}_t - \vec{\mathbf{v}}_{t-1})(\Sigma_t + \Sigma_{t-1})(\vec{\mathbf{v}}_t - \vec{\mathbf{v}}_{t-1}) \\ + \frac{1}{2}\log(\frac{|\Sigma_t + \Sigma_{t-1}|}{2\sqrt{|\Sigma_t||\Sigma_{t-1}|}}) \end{array}. \tag{20}$$

IV. Experimental setup

A synthetic data-set, consisting of uniformly distritbuted 3D points, is used to evaluate algorithm performances under specific outlier conditions. A percentage of the 3D landmarks is transformed according to the true ego-motion $\hat{M} = (\hat{\vec{v}}, \hat{q})$, whereas the other landmarks are transformed with a random motion each to create outliers. The motion $\hat{M}$ consist of translation and rotation along/around all axis. The 3D landmarks from before and after the motion are projected onto the imaging planes of a modeled stereo camera. Subsequently, Gaussian noise is added to the locations of the landmark projections. These image points are used as input to the EM-SE(3) and RANSAC algorithms. The process described above is repeated for $k = 500$ times for several outlier percentages. For each of the successive 500 trials the translation and rotation along/around each axis is increased, this simulates an accelerating platform. For the synthetic data-set, the robustly estimated ego-motion M can be compared against the true ego-motion $\hat{M}$. The used performance metric is the error expressed as a percentage of the groundtruth. For translation this gives

$$\frac{100\%}{k} \sum_{n=1}^{k} \frac{\left\| \vec{\mathbf{v}}_n - \hat{\vec{\mathbf{v}}} \right\|}{\left\| \hat{\vec{\mathbf{v}}} \right\|}. \tag{21}$$

An appropriate single valued distance metric between rotations $\mathbf{q}$ and $\hat{\mathbf{q}}$ is the angle of the rotation $\mathbf{q}^*\hat{\mathbf{q}}$, this rotation brings the estimated rotation $\mathbf{q}$ onto the groundtruth rotation $\hat{\mathbf{q}}$, Gramkow [4]. For rotations this results in the performance metric

$$\frac{100\%}{k} \sum_{n=1}^{k} \frac{\arccos(\mathbf{q}_n \cdot \hat{\mathbf{q}})}{\arccos(\hat{q})}. \tag{22}$$

Also, a real-world automotive data-set containing approximately 2800 images and spanning about 600 meters was used. It was recorded using a stereo camera with a baseline of approximately 40 centimeters. The stereo camera was mounted on our test vehicle "RoboJeep". Furthermore, an image resolution of 640 by 480 pixels is used at 30 frames per second. An impression of the stereo-images in the data-set is given in fig. 1. By concatenating the frame-to-frame

Fig. 1. Two left images extracted from their stereo-pairs both containing an independently moving object. In both instances this is an approaching car.

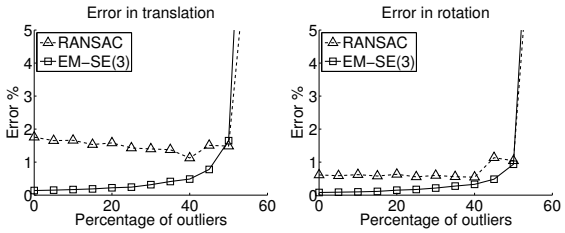
ego-motion estimates, the vehicle pose can be tracked. This enables that the complete driven trajectory can be reconstructed. In parallel to the stereo-frames also differential GPS (DGPS) positions were recorded. Both the DGPS readouts and the stereo-frames have synchronized time-stamps. This could be exploited for comparing the estimated motion at any time with the DGPS based ground truth. Because of the inaccuracy and sparsity of the DGPS readouts however, a different performance metric is used. The data-set encompasses an almost exact loop. Ideally, the final estimated pose should be near the starting pose. Therefore, the performance metric is the distance between the final estimated position and the starting position.

In realistic conditions the number of motion hypothesis that can be generated is limited. Therefore, both algorithms only generate 293 motion hypothesis. When assuming an outlier percentage of 50, this ensures with a probability of 0.99 that at least one motion hypothesis was estimated on inlier landmarks only, eq. 11. The pseudo code of the EM-SE(3) and RANSAC algorithms are given in appendix A.

V. Results

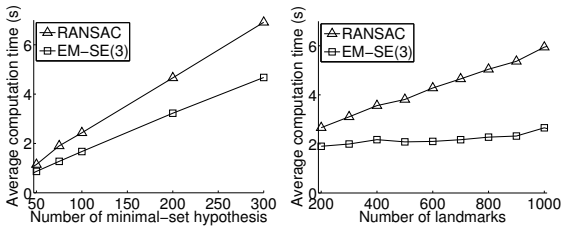
Fig. 2 shows the performance of both algorithms on the synthetic data-set. It can be observed that the EM-SE(3) is more accurate for outlier percentages below 50 percent. While these differences may seem marginal, keeping the error as low as possible, especially for rotations, is crucial for estimating large real-world trajectories. Also, note the breakdown of both methods at an outlier percentage of 50. This is caused by the fact that only 293 motion hypothesis were used.

To measure the computational complexity, both methods were evaluated using various parameter settings, the results are plotted in fig. 3. It can be seen that the novel EM-SE(3) algorithm requires less computational time and scales more favorably in the number of landmarks. An algorithm that



(a) (b)

Fig. 2. Error on synthetic data using 293 motion hypothesis, translation (a), rotation (b).



(a) (b)

Fig. 3. Time complexity in the number of minimal-set motion hypothesis (a), in the number of landmarks (b).

is faster can generate more minimal-set motion hypothesis in a given time-span making it potentially more robust and accurate.

To evaluate the real-world applicability of the proposed methods an automotive data-set is used. A top view of the estimated trajectories of both methods as well as DGPS is given in fig 4. The estimated trajectories show that EM-SE(3) is more accurate than RANSAC. The 3D (thus including height) difference in starting and ending location is for EM-SE(3) approximately 3 m and for RANSAC 10 m. As percentage of the traveled distance this gives 0.5% and 1.6% respectively.

For one particular estimate, the rotation covariance matrix Σ_q is visualized in fig. 5. It can be seen that the largest uncertainty is in rotation around an axis almost parallel to the Z axis i.e. roll. By converting the diagonal of Σ_q to a rotation vector and extracting its Euler angles, the standard deviation in pitch, heading and roll can be approximated. Table I shows these standard deviations in rotation and translation for several segments of the data-set. It can clearly be seen

TABLE I

Standard deviation in rotation and translation estimated on-line by the EM-SE(3) algorithm

	σ rotation (degrees)			σ translation (millimeters)		
	Pitch	Heading	Roll	X	Y	Z
1 Accelerating	0.0151°	0.0088°	0.0491°	1.6	4.2	10.0
2 Cornering	0.0221°	0.0157°	0.0333°	1.9	3.1	4.3
3 Accelerating	0.0089°	0.0073°	0.0355°	3.1	3.6	13.6
4 Cornering	0.0179°	0.0179°	0.0304°	2.8	2.8	5.2

Fig. 4. Satellite view of the driven trajectory with the test vehicle in an urban environment. The DGPS coordinates are indicated in red. Results for both methods using 293 minimal-set motion estimates are indicated for EM-SE(3) in green and for RANSAC in blue.

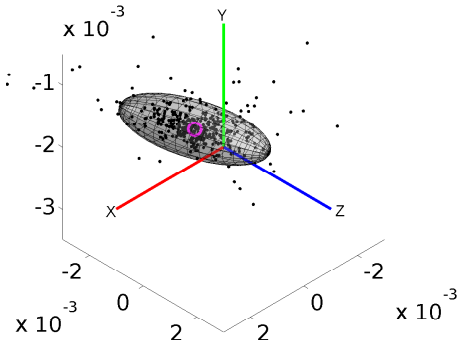


Fig. 5. Rotation error ellipsoid in $\mathfrak{so}(3)$, rotation samples are given with black dots, mean rotation is given by the magenta circle.

that most uncertainty occurs in the estimated roll angle. Also note the larger deviations in translation over the Z axis when the vehicle is accelerating. Knowing the magnitude of these errors is crucial for optimally fusing the motion estimate with other motion or pose sensors. It should be noted that for this application the covariance Σ_q can be used directly instead of the approximated standard deviations in Euler angles.

VI. Conclusion

A novel robust visual-odometry technique, called EM-SE(3), has been presented and compared against RANSAC. EM-SE(3) utilizes an expectation maximization algorithm over a local linearization of the rigid-body motion group SE(3). The results on the used synthetic and real-world data-sets show that EM-SE(3) is both more efficient and more accurate.

This contribution shows that robust and accurate statistical inferences can be made on elements of SE(3). This is achieved through the interplay of robustly discarding outlier minimal-set hypothesis, by means of expectation maximization, and the ability of computing an accurate weighted mean motion, by using Riemannian geometry. Because this technique works directly with the motion hypothesis it is more efficient and scales better if the number of landmarks is increased.

For this research a basic RANSAC approach is used. It has to be noted that both efficiency and accuracy of RANSAC can be improved by, for example, using advanced preemptive methods or post optimization. In this work however, the choice has been made to compare both methods in their purest forms. Nevertheless, the results show that EM-SE(3) is a powerful alternative.

In future work, the EM-SE(3) algorithm will be evaluated on even more challenging urban and outdoor data-sets. Furthermore, guided sampling based on landmark classification and tracking is a promising extension.

REFERENCES

[1] L. Dorst, D. Fontijne, and S. Mann, *Geometric Algebra for Computer Science (An Object-Oriented Approach to Geometry)*. Morgan Kaufmann, 2007.

[2] P. Elinas, R. Sim, and J. J. Little, "σslam: Stereo vision slam using the rao-blackwellised particle filter and a novel mixture proposal distribution," in *IEEE International Conference on Robotics and Automation*, 2006, pp. 1564–1570.

[3] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381 – 395, June 1981.

[4] C. Gramkow, "On averaging rotations," *International Journal of Computer Vision*, 2001.

[5] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, 2nd ed. Cambridge University Press, ISBN: 0521540518, 2004.

[6] K. Kanatani, *Statistical Optimization for Geometric Computation: Theory and Practice*. Dover Publications, ISBN: 0486443086, 1996.

[7] W. T. Loring, *An Introduction to Manifolds*, S. Axler and K. Ribet, Eds. Springer, 2007.

[8] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[9] M. Maimone, Y. Cheng, and L. Matthies, "Two years of visual odometry on the mars exploration rovers: Field reports," *Journal of Field Robotics*, vol. 24, no. 3, pp. 169 – 186, March 2007.

[10] B. C. Matei and P. Meer, "Optimal rigid motion estimation and performance evaluation with bootstrap," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, no. 1, 1999, p. 345.

[11] ——, "Estimation of nonlinear errors-in-variables models for computer vision applications," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 10, pp. 1537 – 1552, October 2006.

[12] L. Matthies and S. A. Shafer, "Error modeling in stereo navigation," *IEEE Journal of Robotics and Automation*, vol. 3, no. 3, pp. 239 – 248, June 1988.

[13] D. Nistér, "Preemptive ransac for live structure and motion estimation," *Machine Vision and Applications*, vol. 16, no. 5, pp. 321–329, November 2005.

[14] D. Nistér, O. Naroditsky, and J. Bergen, "Visual odometry," in *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 1, 2004, pp. 652–659.

[15] ——, "Visual odometry for ground vehicle applications," *Journal of Field Robotics*, vol. 23, no. 1, pp. 3–20, January 2006.

[16] C. F. Olson, L. H. Matthies, M. Schoppers, and M. W. Maimoneb, "Rover navigation using stereo ego-motion," *Robotics and Autonomous Systems*, vol. 43, no. 4, pp. 215–229, June 2003.

[17] F. Park, "Distance metrics on the rigid-body motions with applications to mechanism design," *Transactions of the ASME*, vol. 117, pp. 48–54, March 1995.

[18] X. Pennec, "Computing the mean of geometric features: Application to the mean rotation," INRIA, Tech. Rep. 3371, March 1998.

[19] X. Pennec, C. R. G. Guttmann, and J.-P. Thirion, "Feature-based registration of medical images: Estimation and validation of the pose accuracy," *Lecture Notes in Computer Science*, vol. 1496/1998, pp. 1107–1114, 2006.

[20] P. L. Rosin, "Robust pose estimation," *IEEE Transaction on Systems, Man, and Cybernetics*, vol. 29, no. 2, pp. 297–303, April 1999.

[21] J. M. Selig, *Geometrical Methods in Robotics*, D. Gries and F. B. Schneider, Eds. Springer, 1996.

[22] R. Subbarao, Y. Genc, and P. Meer, "Nonlinear mean shift for robust pose estimation," in *IEEE Workshop on Applications of Computer Vision*, February 2007, pp. 6–6.

[23] P. Torr and D. Murray, "The development and comparison of robust methods for estimating the fundamental matrix," *International Journal of Computer Vision*, vol. 24, no. 3, pp. 271 – 300, 1997.

[24] S. Umeyama, "Least-squares estimation of transformation parameters between twopoint patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 4, pp. 376 – 380, 1991.

[25] X. S. Zhou, D. Comaniciu, and A. Gupta, "An information fusion framework for robust shape tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 27, no. 1, pp. 1–15, January 2005.

APPENDIX A

Algorithm 1 EM-SE(3)

1) Use the SIFT matching procedure as described in III and create the set $\mathbf{L}$ of time corresponding three dimensional points with their uncertainty i.e. $\mathbf{L} = \{L^1 .....L^n\}$ using eq. 15,16,17 and 18.

2) Predict the current motion based on the previous motion with $\bar{M}_t = \bar{M}_{t-1}$ and $\Sigma_t = \Sigma_{t-1} + \Sigma$. This will be the starting point for EM optimization.

3) Create 293 minimal-set motion hypothesis with:
for 293 iterations **do**
- Take at random 6 points from $\mathbf{L}$.
- Estimate the motion on these six points using the HEIV estimator [10].

end for
This results in the set of motion hypothesis $\mathbf{M} = \{M_1 .....M_{293}\}$.

7) Start the EM algorithm:
for 40 iterations **do**
- Compute the weights I_i and O_i for each motion sample given the current mean motion using eq. 10.
- Compute a new mean motion $\bar{M}$ based on the new weights using eq. 8.
- Compute the motion covariance matrix Σ_{m} based on the new mean and the weights using eq. 13.
- Compute the inlier mixing weight $p(I)$ based on the new mean and covariance using eq. 14.

end for

8) Return the mean motion $\bar{M}$ and its covariance matrix Σ_{m}.

Algorithm 2 RANSAC

1) Use the SIFT matching procedure as described in III and create the set $\mathbf{L}$ of time corresponding three dimensional points with their uncertainty i.e. $\mathbf{L} = \{L^1 .....L^n\}$ using eq. 15,16,17 and 18.

2) Create 293 minimal subset motion hypothesis with:
for 293 iterations **do**
- Take at random 6 points from $\mathbf{L}$.
- Estimate the motion on these six points using the HEIV estimator [10].
- Calculate the Bhattacharyya distance given the current motion hypothesis for each point in $\mathbf{T}$. The score of the current motion hypothesis is the number of landmarks with a Bhattacharyya distance less than 1.5.

end for
This results in the set of motion hypothesis $\mathbf{M} = \{M_1 .....M_{293}\}$ with their scores.

3) Select from $\mathbf{M}$ that motion estimate M with the highest score as the final robust estimate.