Causal Camera Motion Estimation by Condensation and Robust Statistics Distance Measures

Tal Nir and Alfred M. Bruckstein

Computer Science Department, Technion, Haifa 32000, ISRAEL
{taln,freddy}@cs.technion.ac.il

Abstract. The problem of Simultaneous Localization And Mapping (SLAM) originally arose from the robotics community and is closely related to the problems of camera motion estimation and structure recovery in computer vision. Recent work in the vision community addressed the SLAM problem using either active stereo or a single passive camera. The precision of camera based SLAM was tested in indoor static environments. However the extended Kalman filters (EKF) as used in these tests are highly sensitive to outliers. For example, even a single mismatch of some feature point could lead to catastrophic collapse in both motion and structure estimates. In this paper we employ a robust-statistics-based condensation approach to the camera motion estimation problem. The condensation framework maintains multiple motion hypotheses when ambiguities exist. Employing robust distance functions in the condensation measurement stage enables the algorithm to discard a considerable fraction of outliers in the data. The experimental results demonstrate the accuracy and robustness of the proposed method.

1 Introduction

While the vision community struggled with the difficult problem of estimating motion and structure from a single camera generally moving in 3D space (see [5]), the robotics community independently addressed a similar estimation problem known as Simultaneous Localization and Mapping (SLAM) using odometery, laser range finders, sonars and other types of sensors together with further assumptions such as planar robot motion. Recently, the vision community has adopted the SLAM name and some of the methodologies and strategies from the robotics community. Vision based SLAM has been proposed in conjunction with an active stereo head and odometry sensing in [7], where the stereo head actively searched for old and new features with the aim of improving the SLAM accuracy. In [6] the more difficult issue of localization and mapping based on data from a single passive camera is treated. The camera is assumed to be calibrated and some features with known 3D locations are assumed present and these features impose a metric scale on the scene, enable the proper use of a motion model, increase the estimation accuracy and avoid drift. These works on vision based SLAM employ an Extended Kalman Filter (EKF) approach where camera motion parameters are packed together with 3D feature locations to form a large and tightly coupled estimation problem. The main disadvantage of this approach is that even a single outlier in measurement data can lead to a collapse of the whole estimation problem. Although there are means for excluding problematic

T. Pajdla and J. Matas (Eds.): ECCV 2004, LNCS 3023, pp. 119–131, 2004.
© Springer-Verlag Berlin Heidelberg 2004

feature points in tracking algorithms, it is impossible to completely avoid outliers in uncontrolled environments. These outliers may result from mismatches of some feature points which are highly likely to occur in cluttered environments, at depth discontinuities or when repetitive textures are present in the scene. Outliers may exist even if the matching algorithm performs perfectly when some objects in the scene are moving. In this case multiple-hypothesis estimation as naturally provided by particle filters is appropriate. The estimation of the stationary scene structure together with the camera ego-motion is the desired output under the assumption that most of the camera's field of view looks at a static scene. The use of particle filters in SLAM is not new. Algorithms for FastSLAM [19] employed a particle filter for the motion estimation, but their motivation was mainly computational speed and robust estimation methodology was neither incorporated nor tested. In [18] a version of FastSLAM addressing the problem of data association between landmarks and measurements is presented. However, the solution to the data association problem provided there does not offer a solution to the problem of outliers since all landmarks are assumed stationary and every measurement is assumed to correctly belong to one of the real physical landmarks. Other works like e.g. [6] employed condensation only in initialization of distances of new feature points before their insertion into the EKF. However the robustness issue is not solved in this approach since the motion and mapping are still provided by the EKF. In [23] the pose of the robot was estimated by a condensation approach. However, here too the algorithm lacked robust statistics measures to effectively reject outliers in the data. Furthermore the measurements in this work were assumed to be provided by laser range finders and odometric sensors. In this work we propose a new and robust solution to the basic problem of camera motion estimation from known 3D feature locations, which has practical importance of its own. The full SLAM problem is then addressed in the context of supplementing this basic robust camera motion estimation approach for simultaneously providing additional 3D scene information. The paper is organized as follows: Section 2 formulates the basic motion estimation problem. Section 3 presents the proposed framework for robust motion from structure. Section 4 discusses methods for incorporating the proposed framework for the solution of SLAM. Section 5 presents results on both synthetic data and real sequences and compares the performance to that of EKF based methods.

2 Problem Formulation

Throughout this work it is assumed that the camera is calibrated. This assumption is commonly made in previous works on vision based SLAM. A 3D point indexed by i in the camera axes coordinates, $(X_i(t)\ Y_i(t)\ Z_i(t))^T$ projects to the image point $(x_i(t)\ y_i(t))^T$ at frame time t via some general projection function Π as follows:

$$\begin{pmatrix} x_i(t) \\ y_i(t) \end{pmatrix} = \Pi \begin{pmatrix} X_i(t) \\ Y_i(t) \\ Z_i(t) \end{pmatrix} \tag{1}$$

The camera motion between two consecutive frames is represented by a rotation matrix R(t) and a translation vector V(t). Hence for a static point in the scene:

$$\begin{pmatrix} X_i(t) \\ Y_i(t) \\ Z_i(t) \end{pmatrix} = R(t) \begin{pmatrix} X_i(t-1) \\ Y_i(t-1) \\ Z_i(t-1) \end{pmatrix} + V(t) \tag{2}$$

The rotation is represented using the exponential canonical form $R(t) = e^{\hat{\omega}(t)}$ where $\omega(t)$ represents the angular velocity between frames t-1 and t, and the exponent denotes the matrix exponential. The hat notation for some 3D vector q is defined by:

$$q = \begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix} \; ; \; \hat{q} = \begin{pmatrix} 0 & -q_3 & q_2 \\ q_3 & 0 & -q_1 \\ -q_2 & q_1 & 0 \end{pmatrix}$$

The matrix exponential of such skew-symmetric matrices may be computed using the Rodrigues' formula:

$$e^{\hat{\omega}} = I + \frac{\hat{\omega}}{\|\omega\|}\sin(\|\omega\|) + \frac{\hat{\omega}^2}{\|\omega\|^2}\left(1 - \cos(\|\omega\|)\right)$$

Let us denote by $\Omega(t)$ and T(t) the overall rotation and translation from some fixed world coordinate system to the camera axes:

$$\begin{pmatrix} X_i(t) \\ Y_i(t) \\ Z_i(t) \end{pmatrix} = e^{\hat{\Omega}(t)} \begin{pmatrix} X_i^{World} \\ Y_i^{World} \\ Z_i^{World} \end{pmatrix} + T(t) \tag{3}$$

Equation (3) describes the pose of the world relative to the camera. The camera pose relative to the world is given by: $\Omega_{(t)}^{Camera} = -\Omega(t)$; $T_{(t)}^{Camera} = -e^{-\hat{\Omega}(t)}T(t)$

Using equations (2), (3) and (3) written one sample backward:

$$\begin{aligned} \Omega(t) &= \log_{S0(3)}\left(e^{\hat{\omega}(t)}e^{\hat{\Omega}(t-1)}\right) \\ T(t) &= e^{\hat{\omega}(t)}T(t-1) + V(t) \end{aligned} \tag{4}$$

Where, $q = \log_{S0(3)}(A)$ denotes the inverse of the matrix exponential of the skew symmetric matrix A such that $A = e^{\hat{q}}$ (i.e. inverting Rodrigues' formula). Let us define the robust motion from structure estimation problem: given matches of 2D image feature points to known 3D locations, estimate the camera motion in a robust framework accounting for the possible presence of outliers in measurement data.

2.1 Dynamical Motion Model

One can address the camera motion estimation problem with no assumptions on the dynamical behavior of the camera (motion model), thus using only the available geometric information in order to constrain the camera motion. This is equivalent to assuming independent and arbitrary viewpoints at every frame. In most practical applications though, physical constraints result in high correlation of pose between adjacent frames. For example, a camera mounted on a robot traveling in a room produces smooth motion trajectories unless the robot hits some obstacle or collapses. The use of a proper motion model accounts for uncertainties, improves the estimation accuracy, attenuates the influence of measurement noise and helps overcome ambiguities (which may occur if at some time instances, the measurements are not sufficient to uniquely constrain camera pose, see [5] and [6]). Throughout this work, the motion model assumes constant velocity with acceleration disturbances, as follows:

$$\begin{aligned} \omega(t) &= \omega(t-1) + \dot{\omega}(t) \\ V(t) &= V(t-1) + \dot{V}(t) \end{aligned} \tag{5}$$

If no forces act on the camera the angular and translation velocities are constant. Accelerations result from forces and moments which are applied on the camera, and these being unknown are treated as disturbances (recall that the vectors $\omega(t), V(t)$ are velocity terms and the time is the image frame index).
Acceleration disturbances are modeled here probabilistically by independent white Gaussian noises:

$$\begin{aligned} \dot{\omega}(t) &\sim N(0, \sigma_{\dot{\omega}}) \\ \dot{V}(t) &\sim N(0, \sigma_{\dot{V}}) \end{aligned} \tag{6}$$

where $\sigma_{\dot{\omega}}$ and $\sigma_{\dot{V}}$ denote expected standard deviations of the angular and linear acceleration disturbances.

3 Robust Motion from Structure by Condensation

In this section we present the proposed condensation based algorithm designed for robust camera 3D motion estimation. A detailed description of condensation in general and its application to contour tracking can be found in [12] and [13]. The state vector of the estimator at time t, denoted by s_t, includes all the motion parameters:

$$s_t = \left(\Omega(t) \quad T(t) \quad \omega(t) \quad V(t)\right)^T$$

The state vector is of length 12. The state dynamics are generally specified in the condensation framework by the probability distribution function $p(s_t \mid s_{t-1})$. Our motion model is described by equations (4),(5),(6). All measurements at time t are denoted compactly as z(t). The camera pose is defined for each state s_t separately, with the corresponding expected projections being tested on all the visible points in the current frame:

$$\begin{pmatrix} x_i(t) \\ y_i(t) \end{pmatrix} = \Pi \left(e^{\hat{\Omega}(t)} \begin{pmatrix} X_i^{World} \\ Y_i^{World} \\ Z_i^{World} \end{pmatrix} + T(t) \right)$$

The influence of the measurements is quantified by $p(z(t) \mid s_t)$. This is the conditional Probability Distribution Function (PDF) of measuring the identified features z(t) when the true parameters of motion correspond to the state s_t. The conditional PDF is calculated as a function of the geometric error, which is the distance denoted by d_i between the projected 3D feature point location on the image plane and the measured image point. If the image measurement errors are statistically independent random variables with zero mean Gaussian PDF, then up to a normalizing constant:

$$p(z \mid s) = \exp\left(- \frac{\sum_{i=1}^{N_{\text{points}}} d_i^2}{2\sigma^2 N_{\text{points}}} \right)$$

Where N_{points} is the number of visible feature points and σ is the standard deviation of the measurement error (about 1 pixel). Since outliers have large d_i values even for the correct motion, the quadratic distance function may be replaced by a robust distance function $\rho\left(d_i^2\right)$ see e.g. [20]:

$$p(z \mid s) = \exp\left(- \frac{\sum_{i=1}^{N_{\text{points}}} \rho\left(d_i^2\right)}{2\sigma^2 N_{\text{points}}} \right) \tag{7}$$

$$\rho\left(d^2\right) = \frac{d^2}{1 + d^2 / L^2} \tag{8}$$

If some feature point is behind the camera (this occurs when its 3D coordinates expressed in camera axes have a negative Z value), clearly this feature should not have been visible and hence its contribution to the sum is set to the value:

$$\lim_{d_i \to \infty} \rho\left(d_i^2\right) = L^2$$

The influence of every feature point on the PDF is now limited by the parameter L. The choice of L reflects a threshold value between inliers and outliers. In order to understand why robustness is achieved using such distance functions, let us consider the simpler robust distance function, the truncated quadratic:

$$\rho\left(d^2\right) = \begin{cases} d^2 & d^2 < A^2 \\ A^2 & Otherwise \end{cases}$$

where, A is the threshold value between inliers and outliers. Using this ρ function in equation (7) yields:

$$p(z \mid s) = \exp\left(-\frac{\sum_{i \in \text{Inlier points}} d_i^2 + \sum_{i \in \text{Outlier points}} A^2}{2\sigma^2 N_{\text{points}}} \right) = \exp\left(-\frac{\sum_{i \in \text{Inlier points}} d_i^2 + A^2 \cdot (\#Outliers)}{2\sigma^2 N_{\text{points}}} \right)$$

Maximizing this PDF (a maximum likelihood estimate) is equivalent to minimizing the sum of the two terms, the first is the sum of the quadratic distances at the inlier points and the second term is proportional to the number of outliers. The robust distance function of equation (8) is similar to the truncated quadratic, with a smoother transition between the inliers and outliers (see [2] and [3] for an analysis of ρ functions used in robust statistics and their use for image reconstruction and for the calculation of piecewise-smooth optical flow fields). Let us summarize the proposed algorithm for robust 3D motion estimation from known structure:

Initialization- Sample N states $s_0^{(n)}, n = 1...N$ from the prior PDF of $\omega(0)$, V(0) and $\Omega(0); T(0)$. Initialize $\pi_0^{(n)}$ with the PDF corresponding to each state.

At every time step t=1,2, ... :

- Sample N states $\tilde{s}_{t-1}^{(n)}$ copied from the states $s_{t-1}^{(n)}$ with probabilities $\pi_{t-1}^{(n)}$.
- Propagate the sampled states using equations (6),(5),(4) to obtain $s_t^{(n)}$.
- Incorporate the measurements to obtain $\pi_t^{(n)} = p\left(z_t \mid s_t^{(n)}\right)$ using equations (7),(8). Then normalize by the appropriate factor so that: $\sum_{n=1}^{N} \pi_t^{(n)} = 1$
- Extract the dominant camera motion from the state $s_t^{(n)}$ corresponding to the maximum of $\pi_t^{(n)}$: $\Omega_{(t)}^{Camera} = -\Omega_{(t)}^{(n)}$; $T_{(t)}^{Camera} = -e^{-\hat{\Omega}_{(t)}^{(n)}} T_{(t)}^{(n)}$

Code written in C++ implementing the algorithm of this section can be found in [25]. It can run in real time on a Pentium 4, 2.5GHz processor, with 30Hz sampling rate, 1000 particles and up to 200 instantaneously visible feature points.

4 Application to SLAM

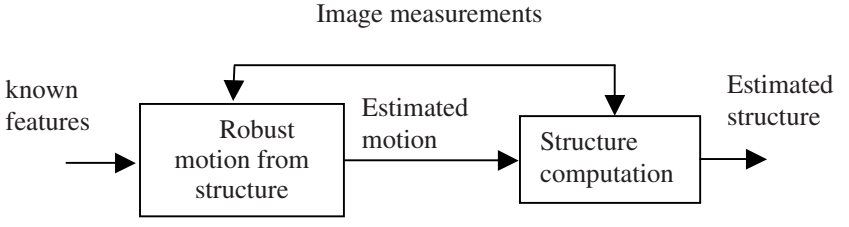
This section describes various possible solutions to the robust SLAM problem.

4.1 SLAM in a Full Condensation Framework

The most comprehensive solution to robust SLAM is the packing of all the estimated parameters into one large state and solve using a robust condensation framework. The state is composed of the motion parameters and each feature contributes three additional parameters for its 3D location. As stated in [7], this solution is very expensive computationally due to the large number of particles required to properly sample from the resulting high dimensional space.

4.2 SLAM in a Partially Decoupled Scheme

The research on vision based SLAM tends to incorporate features with known 3D locations in the scene. The simplest method for incorporating the proposed robust motion from structure algorithm into SLAM is in a partially decoupled block scheme in which the features with known 3D locations are the input to the robust motion estimation block of section 3. Structure of other features in the scene can be recovered using the estimated motion and the image measurements. Assuming known 3D motion, the structure of each feature can be estimated using an EKF independently for each feature (similar to FastSlam in [19]). If enough features with known structure are available in the camera field of view at all times (few can be enough as shown in the experiments section), then this method can work properly. It may be practical for robots moving in rooms and buildings to locate known and uniquely identifiable features (fiducials) at known locations. When the motion estimation is robust, the independence of the estimators for the structure of the different features guarantees the robustness of the structure recovery as well.



4.3 SLAM with Robust Motion Estimation and Triangulation

In this section we propose a solution to the robust SLAM problem in a condensation framework with a state containing motion parameters only. In the measurement phase, features with known locations have their 3D structure projected on the image plane, features with unknown structure have their 3D structure reconstructed using triangulation (see [9] chapter 11) and the geometric error is measured by projecting this structure back on the image plane. The information regarding the camera pose in the current and previous frames is embedded in each state hypothesis of the condensation algorithm which together with the corresponding image measurements form the required information for the triangulation process. Triangulation can be performed from three views, where the third view is the first appearance of the feature.

5 Experimental Results

5.1 Synthetic Tests

It has been experimentally found using synthetic tests that robustness with the proposed method is maintained with up to about 33% of outliers. The proposed

algorithm is compared with the results of the EKF approach in [5] which is run with the code supplied in [15]. The robust motion estimation used triangulation in three frames as described in section 4.3. The 3D structure was unknown to both algorithms. The outlier points are randomly chosen and remain fixed throughout the sequence, these points are given random image coordinates uniformly distributed in the image range (see examples in [25]). The rotation errors are compactly characterized by:

$$\left\| I - \left(e^{\hat{\Omega}_{True}} \right)^T e^{\hat{\Omega}_{Estimated}} \right\|^2_{Frobenius}$$

The estimation results are shown in Fig. 1. With 33% of outliers, the EKF errors are unacceptable while the proposed method maintains reasonable accuracy.

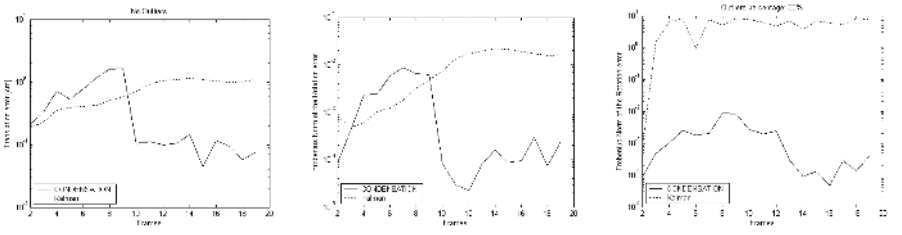


Fig. 1. The translation (*left*) and rotation (*middle*) errors with no outliers in the data. Rotation errors with 33% of outliers (*right*)

5.2 Real Sequence Example

In this section a test consisting of 340 frames is described in detail. More sequences can be found in [25]. Small features (fiducials) were placed in known 3D locations (see Table 1) on the floor and on the walls of a room (see Fig. 2). Distances were measured with a tape having a resolution of 1 millimeter (0.1 cm). The fixed world coordinate system was chosen with its origin coinciding with a known junction on the floor tiles, the X and Z axes on the floor plane and parallel to the floor tiles and the Y axis pointing downwards (with -Y measuring the height above the floor). The balls are 1.4 and the circles are 1cm in diameter, the tiles are squares of 30x30cm.

Table 1. Scene fiducial geometry

Serial number	Type	Color	World axes location [cm]		
			X	Y	Z
1	Ball	Blue	30	-0.7	180
2	Ball	Green	30	-0.7	210
3	Ball	Yellow	-60	-0.7	240
4	Ball	Light blue	30	-0.7	240
5	Ball	Black	0	-0.7	270
6	Ball	Red	-30	-0.7	330
7	Ball	Orange	60	-0.7	360
8	Circle	Light blue	-31	-100.3	388
9	Circle	Light blue	29	-120.7	492.5

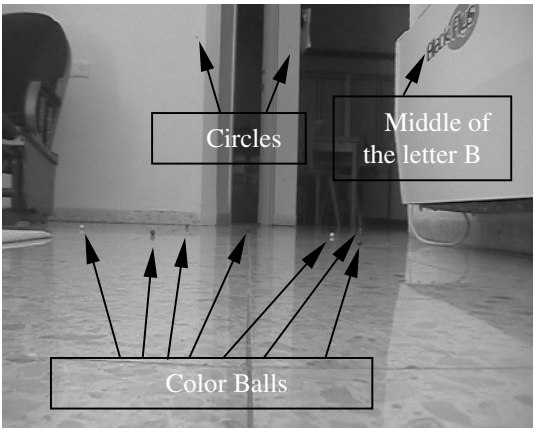


Fig. 2. First frame of the sequence

5.2.1 Camera Setup and Motion

The camera was a Panasonic NV-DS60 PAL color camera with a resolution of 720x576 pixels. The camera zoom was fixed throughout the test at the widest viewing angle. A wide field of view reduces the angular accuracy of a pixel, but enables the detection of more features (overall, [5] has experimentally found that a wide viewing angle is favorable for motion and structure estimation). The camera projection parameters at this zoom were obtained from a calibration process:

$$x = 938\,X/Z + 360.5 \quad ; \quad y = 1004\,Y/Z + 288.5$$

The camera was initially placed on the floor with the optical axis pointing approximately in the Z direction of the world. The camera was moved backwards by hand on the floor plane with the final orientation approximately parallel to the initial (using the tile lines). The comparison between the robust and the EKF approach is made with both having the same motion parameters in the estimated state, the same measurements and the same knowledge of the 3D data of table 1. The acceleration disturbance parameters for both methods are: $\sigma_{\dot{\omega}} = 0.003$, $\sigma_{\dot{V}} = 0.0005$. The number of particles is 2000 and the robust distance function parameter is L=4 pixels.

5.2.2 Feature Tracking

The features were tracked with a Kanade-Lucas-Tomasi (KLT) type feature tracker (see [21]). The tracker was enhanced for color images by minimizing the sum of squared errors in all three RGB color channels (the standard KLT is formulated for grayscale images). The tracking windows of size 9x9 pixels were initialized in the first frame at the center of each ball and circle by manual selection. To avoid the fatal effect of interlacing, the resolution was reduced in the vertical image plane by sampling every two pixels (processing one camera field), the sub-pixel tracking results were then scaled to the full image resolution.

5.2.3 Motion Estimation Results

The results obtained by the proposed robust approach and the EKF approach are shown in Fig. 3. Most of the motion is in the Z direction. The final position was at

approximately Z=-60cm. The robust approach estimates the value of Z=-61cm at the end of the motion (there is some uncertainty regarding the exact location of the camera focal center), the estimated Y coordinate is almost constant and equal to the camera lens center height above the floor (about -7.4 cm). The trajectory estimated by the EKF is totally different with $Z \approx -80\,cm$ and $Y \approx 30\,cm$ at the end of the motion.

The deviation from the expected final camera position is by two orders of magnitude higher than the expected experimental accuracy, the EKF estimation is therefore erroneous. After observing the tracking results of all the feature points, the points 1,2,4,5,6,8 were manually selected as the inlier points (those which reasonably track the appropriate object throughout the sequence). Running again both estimators with only the inlier points, the proposed approach results are almost unchanged, while the EKF estimation changes drastically, now producing a trajectory similar to the robust approach (see Fig. 4). It should be noted that the EKF estimation produces a smoother trajectory. Image plane errors between the measurements and the projected 3D structure are shown in Fig. 5 (corresponding to the motion estimation of Fig. 3). The robust method exhibits low errors for most of the features and allows high errors for the outliers (this implies that algorithm can automatically separate the inliers from the outliers by checking the projection errors). The EKF approach on the other hand exhibits large errors for both inlier and outlier features. It should be noted that the outlier features are distracted from the true object due to its small size, noise, similar objects in the background and reflections from the shiny floor. It is possible to improve the feature tracking results by using methodologies from [14], [21], [24], but good feature tracking should be complemented with a robust methodology in order to compensate for occasional mistakes. Although the deficiencies of the EKF approach are mentioned in [5], [6], [7], no examples are given and no remedies are suggested in the camera motion estimation literature. As anonymous reviewers have suggested, we

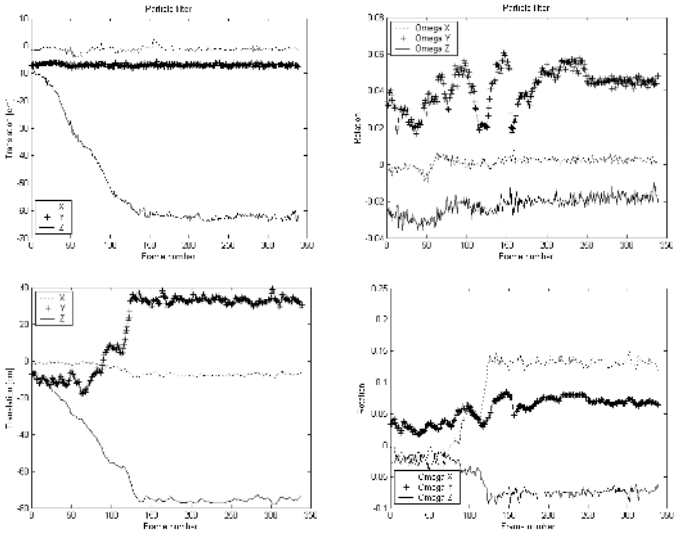


Fig. 3. Estimated camera 3D trajectory using the proposed approach (*upper row*) and the EKF approach (*lower row*)

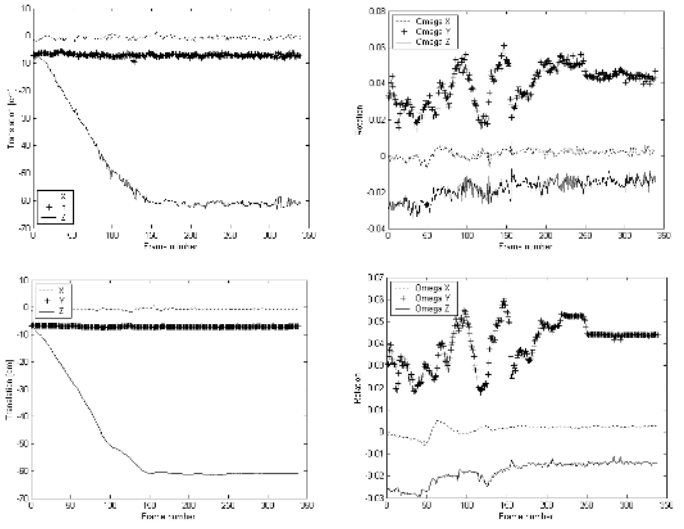


Fig. 4. Estimated camera 3D trajectory using only the inlier points, the proposed approach (*upper row*) and the EKF approach (*lower row*)

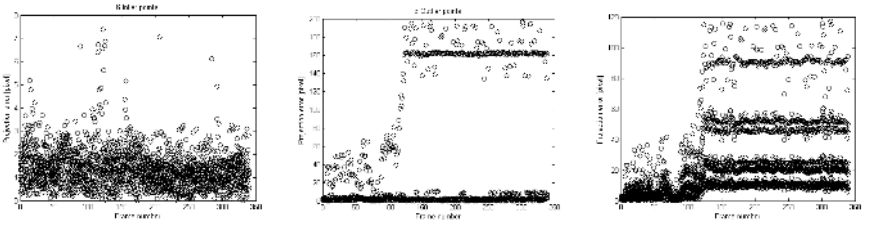


Fig. 5. Image plane errors. Robust approach showing the 6 inliers (*left*) and 3 outliers (*middle*). EKF approach with all 9 features (*right*)

examined two methods of making the EKF solution more robust: 1. By incorporating measurements only from features which have a geometric error norm below a threshold and 2. By applying the robust distance function on the norm of the geometric error of each feature. Both failed to improve the results of the EKF. Rejection of outliers in Kalman filtering may succeed if the outliers appear scarcely or when their proportion is small. In our example these conditions are clearly violated.

5.2.4 Structure Computation Example

Structure of unknown features in the scene can be recovered using the estimated camera motion obtained by the robust method and the image measurements in a partially decoupled scheme as explained in section 4.2. As an example, consider the middle of the letter B appearing on the air conditioner which was tracked from frame 0 to frame 50 (it is occluded shortly afterwards). The reconstructed location of this point in the world axes is: X=42.3cm; Y=-45.2cm; Z=159.6cm. The tape measure world axes location is: X=42.0cm; Y=-43.7cm; Z=155cm. The X, Y, Z differences

are: 0.3, 1.5 and 4.6 [cm] respectively. As expected, the estimation error is larger along the optical axis (approximately the world's Z axis). The accuracy is reasonable, taking into account the short baseline of 19cm produced during the two seconds of tracking this feature (the overall translation from frame 0 to frame 50). As discussed in [5], a long baseline improves the structure estimation accuracy when the information is properly integrated over time.

6 Conclusion

A robust framework for camera motion estimation has been presented with extensions to the solution of the SLAM problem. The proposed algorithm can tolerate about 33% of outliers and it is superior in robustness relative to the commonly used EKF approach. It has been shown that a small number of visible features with known 3D structure are enough to determine the 3D pose of the camera. It may be implied from this work that some degree of decoupling between the motion estimation and structure recovery is a desirable property of SLAM algorithms which trades some accuracy loss for increased robustness. The robust distance function used in this work is symmetric for all the features with the underlying assumption that the probability of a feature to be an inlier or an outlier is independent of time. However, in most cases, a feature is expected to exhibit a more consistent behavior as an outlier or an inlier. This property may be exploited for further improvement of the algorithm's robustness and accuracy. Also, an interesting question for future work is: How to construct fiducials which can be quickly and accurately identified in the scene for camera localization purposes.

References

1. A. Azarbayejani and A. Pentland. Recursive Estimation of Motion, Structure and Focal Length. IEEE Trans. PAMI, Vol. 17, no. 6, pp. 562-575, 1995.
2. M. J. Black and P. Anandan. The Robust Estimation of Multiple Motions: Parametric and Piece-wise Smooth Flow Fields. CVIU, Vol. 63, No. 1, 1996.
3. M. J. Black and A. Rangarajan. On the Unification of Line Processes, Outlier Rejection, and Robust Statistics with Applications in Early Vision. IJCV 19(1), 57-91, 1996.
4. T.J. Broida, S. Chandrashekhar, and R. Chellappa. Recursive 3-d motion estimation from a monocular image sequence. IEEE Trans. on AES., 26(4):639- 656, 1990.
5. A. Chiuso, P. Favaro, H. Gin. and S. Soatto. Structure from Motion Causally Integrated Over Time, IEEE. Trans. on PAMI, Vol. 24, No. 4, 2002.
6. A. J. Davison. Real-Time Simultaneous Localization and Mapping with a Single Camera. ICCV 2003.
7. A. J. Davison and D. W. Murray. Simultaneous Localization and Map-Building using Active Vision. PAMI, Vol. 24, No. 7, July 2002.
8. O. Faugeras. Three Dimensional Computer Vision, a Geometric Viewpoint. MIT Press, 1993.
9. R. Hartley and A. Zisserman. Multiple View Geometry in Computer Vision, Cambridge press 2000.
10. T.S. Huang and A.N. Netravali. Motion and Structure from Feature Correspondences: a review. Proceeding of The IEEE Communications of the ACM, 82(2): 252-268, 1994.

11. X. Hu and N. Ahuja. Motion and Structure Estimation Using Long Sequence Motion Models. Image and Vision Computing, Vol. 11, no. 9, pp. 549-569, 1993.
12. M. Isard and A. Blake. Visual Tracking by Stochastic Propagation of Conditional Density. Proc. 4th ECCV, Pages 343-356.
13. M. Isard and A. Blake. CONDENSATION - Conditional Density Propagation for Visual Tracking, Int. J. Computer Vision, 29, 1, 5-28, 1998.
14. H. Jin, P. Favaro and S. Soatto. Real-time Feature Tracking and Outlier Rejection with Changes in Illumination, ICCV, July 2001.
15. H. Jin. Code from the web site: http://ee.wustl.edu/~hljin/research/
16. B.D. Lucas and T. Kanade, An iterative Image Registration Technique with an Application to Stereo Vision, In IJCAI81, pages 674-679, 1981.
17. J. MacCormick and M. Isard, Partitioned Sampling, Articulated Objects and Interface-Quality Hand Tracking. Proc. Sixth European Conf. Computer Vision, 2000.
18. M. Montemerlo and S. Thrun. Simultaneous Localization and Mapping with Unknown Data Association using FastSLAM. Proc. ICRA, 2003, to appear
19. M. Montemerlo, S. Thrun, D. Koller, and B. Wegbreit. FastSLAM: A Factored Solution to the Simultaneous Localization and Mapping Problem. In AAAI-2002.
20. P.J. Huber. Robust Statistics. Wiley 1981.
21. J. Shi, C. Tomasi. Good Features to Track. CVPR '94, June 1994, pub. IEEE, pp. 593-600.
22. M. Spetsakis and J. Aloimonos. A Multi-Frame Approach to Visual Motion Perception. Int. J. Computer Vision, Vol. 6, No. 3, pp. 245-255, 1991.
23. S. Thrun, W. Burgard and D. Fox. A Real-Time Algorithm for Mobile Robot Mapping With Applications to Multi-Robot and 3D Mapping. IEEE. International Conference on Robotics and Automation. April 2000.
24. T. Tommasini, A. Fusiello, E. Trucco, and V. Roberto. Making Good Features Track Better. Proc. IEEE Conf. on Computer Vision and Pattern Recognition, 178-183, 1998.
25. Web site: http://www.cs.technion.ac.il/~taln/