ROBUST ESTIMATION OF MOTION AND STRUCTURE USING A DISCRETE H_∞ FILTER

G. Qian A. Kale and R. Chellappa

Center for Automation Research
Department of Electrical and Computer Engineering
University of Maryland
College Park, MD 20742-3275
{gqian,kale,rama}@cfar.umd.edu

ABSTRACT

In this paper, a robust Structure from Motion (SfM) algorithm using a discrete H_∞ filter is presented. Existing SfM algorithms do not work well when large motion or measurements modeling uncertainties are present. By using a H_∞ filter, these modeling uncertainties can be successfully handled. This algorithm has been tested on synthetic image sequences and results show the superiority of the H_∞ filtering approach.

1. INTRODUCTION

The Structure from Motion (SfM) algorithms are essentially model-based [1, 2, 3] and mainly an Extended Kalman Filter (EKF) or its extensions are used for the motion estimation. It is well known that the Kalman filter is optimal when the system and measurement models are linear and both dynamic and measurement noise is additive Gaussian with known covariances. However, due to the nonlinearities in the measurement (perspective projection) equations and the unavailability of the complete knowledge about platform dynamics, significant uncertainties and modeling errors are introduced during the linearization stage when implementing an EKF. Hence the EKF is not sufficient for the camera motion estimation problem.

Recent developments in optimal filtering [4, 5] have focused on the H_∞ estimators. In [6], H_∞ filtering has been applied to 2D visual motion estimation with known 2D scene structure. Application of H_∞ filtering for robust 3D SfM has not been addressed. In this paper we present an algorithm which is robust to modeling uncertainties. The strategy we used is two-pronged. Firstly we make use of the now well known H_∞ robust filtering technique and employ a simple strategy to keep the error covariance matrix large. This leads to superior performance in the presence of uncertainties in the motion and noise models employed. H_∞ filters minimize the infinity norm of the transfer function between the exogenous disturbances and the estimation error. This results in a filtering algorithm which is robust to uncertainties in the measurement and dynamic models used as evidenced by the experimental results obtained. The present approach used in this paper follows [5] wherein a solution to the filtering problem has been obtained by using a game theoretic approach. Approaches based on Krein spaces [4] are also being examined by us. The solution given in [5] reveals a systematic way of incorporating robustness by giving rise to an error covariance matrix which differs from that of the Kalman filter by a negative definite term. A larger error covariance matrix leads to an algorithm which is "alert" to discrepancies in the motion model and statistical properties of the noise.

Prepared through collaborative participation in the Advanced Sensors Consortium (ASC) sponsored by the U.S. Army Research Laboratory under the Federated Laboratory Program, Cooperative Agreement DAAL01-96-2-0001.

2. CAMERA AND MEASUREMENT MODELS

In our approach, we mainly follow [3] in camera and structure parameterization. Figure 1 illustrates the imaging procedure of a moving camera. Two 3D coordinate systems are used. I is an inertial world coordinate system, fixed on the ground and C is a camera-fixed coordinate system which uses the image plane as its XY plane. These two coordinate systems are coincident initially. When the camera moves, I remains on the ground and C moves with the camera. As shown in Figure 1, the camera motion between two image frames could be uniquely decomposed into a rotation about the focal point F and the translation of F. O is the center of the optical lens.

Because of the unavailability of the knowledge about translational dynamics, only T_k, the 3D camera position in I at time t_k, is used as the translational motion parameters. A random walk model is used to represent the translational dynamics. Hence the translation equation is given by

$$T_{k+1} = T_k + n_T \quad (1)$$

0-7803-6297-7/00/$10.00

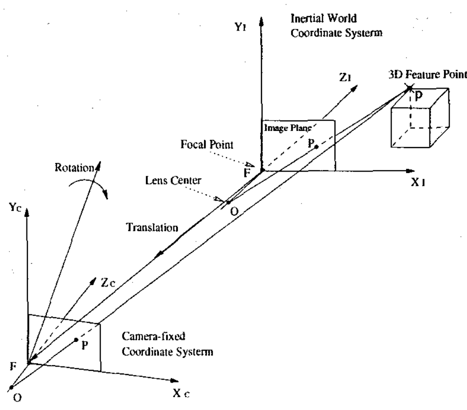

Fig. 1. Imaging model of moving camera

where n_T is the additive noise.

To reduce the impact of the nonlinearity in the rotation dynamic equation and make the linearization more accurate, only the angular velocity vector, $\Omega = (\omega_x, \omega_y, \omega_z)^T$, is used as the rotational state parameter. The dynamic equation of the angular velocity Ω is:

$$\Omega_{k+1} = \Omega_k + n_\Omega \tag{2}$$

where n_Ω is the random disturbance of rotation velocity. The global rotational angle vector, $\Psi = (\psi_x, \psi_y, \psi_z)^T$, is updated outside the filtering procedure as follows. Let $R(\Psi) = \Re(\Psi, t)\mid_{t=1}$[1] be the rotation matrix generated by Ψ. At first, $R(\Psi)$ is updated by

$$R(\Psi_{k+1}) = \Re(\Omega_k, t_{k+1} - t_k) \cdot R(\Psi_k) \tag{4}$$

and then Ψ_{k+1} is computed from $R(\Psi_{k+1})$ using

$$\Psi_{k+1} = \frac{\phi}{2\sin\phi}(r_{23}-r_{32}, r_{31}-r_{13}, r_{12}-r_{21})^T \tag{5}$$

where $\phi = cos^{-1}\{\frac{1}{2}(tr(R(\Psi_{k+1})) - 1)\}$ and r_{jk} is the element in the j^{th} row and k^{th} column of $R(\Psi_{k+1})$.

Assume that in a static scene, L feature points are detected and tracked through the image sequence. Denote the 3D coordinates of the a feature point p in the world system I by (X, Y, Z). At time t_k, due to the motion of the camera, feature point p has coordinates (X_k^c, Y_k^c, Z_k^c) in the camera system C, which is given by

$$\begin{pmatrix} X_k^c \\ Y_k^c \\ Z_k^c \end{pmatrix} = R(\Psi_k) \cdot \begin{pmatrix} X \\ Y \\ Z \end{pmatrix} - T_k \tag{6}$$

[1]
$$\Re(\Psi, \tau) = \begin{pmatrix} n_1^2 + (1-n_1^2)\eta & n_1n_2(1-\eta)+n_3\zeta & n_1n_3(1-\eta)-n_2\zeta \\ n_1n_2(1-\eta)-n_3\zeta & n_2^2+(1-n_2^2)\eta & n_2n_3(1-\eta)+n_1\zeta \\ n_1n_3(1-\eta)+n_2\zeta & n_2n_3(1-\eta)-n_1\zeta & n_3^2+(1-n_3^2)\eta \end{pmatrix} \tag{3}$$

where $n = (n_1, n_2, n_3)^T = \frac{\Psi}{|\Psi|}$, $\theta = \tau \cdot |\Psi|$, $\zeta = \sin\theta$, and $\eta = \cos\theta$.

Denote $(\tilde{u}_k, \tilde{v}_k)$ as the feature tracking result of this feature point in the k^{th} image frame. They are the noisy measurements resulting from the perspective projection of this feature point on the image plane.

$$\begin{cases} \tilde{u}_k = \frac{X_k^c}{1+\beta Z_k^c} + \eta_x \\ \tilde{v}_k = \frac{Y_k^c}{1+\beta Z_k^c} + \eta_y \end{cases} \tag{7}$$

where β is the inverse of the focal length of the camera. η_x and η_y represent errors in feature tracking. $R^{(i)}$ denotes the i^{th} row vector in rotational matrix R.

To reduce the size of the solution space of the 3D structure parameters and make the filtering procedure more stable, (l_u, l_v), the 2D coordinates of the feature points in the first image frame, are used to approximately describe the direction of location of the feature point. They are called direction of features (DOF) in [3]. The 3D structure parameter can be described as

$$\begin{cases} X = (1+\alpha\beta)(l_u + b_u) \\ Y = (1+\alpha\beta)(l_v + b_v) \\ Z = \alpha \end{cases} \tag{8}$$

where $B = (b_u, b_v)$ is the measurement bias of the direction of location of the feature point. Hence (b_u, b_v, α) can be used to represent the 3D structure of a feature point instead of (X, Y, Z). Because the dynamic range of the measurement bias is small, a small value can be used to initialize the associated terms in the estimated covariance matrix which means that the size of the solution space of the 3D parameters is reduced in the X and Y dimensions. After setting up the camera motion and imaging models, the state vector consists of the following parameters.

$$x_k = (T_k, \Omega_k, \beta, b_u^1, b_v^1, \alpha_1, \cdots \alpha_{L-1}, b_u^L, b_v^L)^T \tag{9}$$

3. DESIGN AND IMPLEMENTATION OF H_∞ FILTER

The design of discrete H_∞ filter is discussed in [5]. The following linear discrete system is considered.

$$\begin{aligned} x_{k+1} &= A_k x_k + B_k w_k \\ y_k &= C_k x_k + v_k \\ k &= 0, 1, \cdots, N-1 \end{aligned} \tag{10}$$

where x_k is a $n \times 1$ state vector, w_k is $n \times 1$ system dynamic noise vector. y_k is a $m \times 1$ measurement vector and v_k is the measurement noise with the same size as y_k. (A_k, B_k, C_k) are the system matrices. The estimation problem is to find the state estimate $\hat{x}_k$ using the noise-corrupted system output $\{y_k, k = 0, 1, \cdots, N-1\}$. Let the estimation performance measure be

$$J = \frac{\sum_{k=0}^{N-1} \|x_k - \hat{x}_k\|_I^2}{\|x_0 - \hat{x}_0\|_{P_0^{-1}}^2 + \sum_{k=0}^{N-1}\{\|w_k\|_{W_k^{-1}}^2 + \|v_k\|_{V_k^{-1}}^2\}} \tag{11}$$

where $||r||_Q^2 \equiv r^TQr$ and I is an identity matrix. W_k and V_k are the weight matrices for the system and measurement noise at t_k. In our case, they are simply the *a priori* covariance matrices of the system and measurement noise. Also, based on the camera motion model and imaging model presented in Section 2, the system matrices (A_k, B_k, C_k) in our camera motion estimation problem turn out to be

$$\begin{aligned} A_k &= I \\ B_k &= \begin{bmatrix} I & O \\ O & O \end{bmatrix} \\ C_k &= \frac{\partial h(x)}{\partial x}\Big|_{\hat{x}_k^-} \end{aligned} \tag{12}$$

where $h(x)$ is the image measurement equation described by (7). The Is and Os are identity and zero matrices with proper size, respectively. The H_∞ filter is required to find the optimal $\hat{x}_k$ in the sense that the supremum of the performance measure should be less than a positive prechosen noise attenuation factor $\frac{1}{\gamma}$, i.e.

$$\sup J < 1/\gamma \tag{13}$$

In [5], it was shown that for a fixed value of γ, a H_∞ filter for $\hat{x}_k$ exists if and only if there exists a stabilizing symmetric solution $P_k > 0$ to the Riccati equation:

$$\begin{aligned} P_{k+1} &= A_kP_k(I - \gamma P_k + C_k^TV_k^{-1}C_kP_k)^{-1}A_k^T \\ &\quad + B_kW_kB_k^T \\ P_0 &= p_0 \end{aligned} \tag{14}$$

and the H_∞ filter is given by

$$\hat{x}_{k+1} = A_k\hat{x}_k + H_k(y_k - C_k\hat{x}_k) \tag{15}$$

where H_k is the gain of the H_∞ filter at t_k given by

$$H_k = A_kP_k(I - \gamma P_k + C_k^TV_k^{-1}C_kP_k)^{-1}C_k^TV_k^{-1} \tag{16}$$

It is possible to rewrite the single covariance update equation (14) as composed of two parts viz a propagation part

$$P_k^- = A_kP_{k-1}^+A_k^T + B_kW_kB_k^T \tag{17}$$

and an update part

$$P_k^+ = P_k^- - H_kC_kP_k^- \tag{18}$$

In our implementation of the filter we modify the update part viz (18). As explained earlier the robustness can be improved by having the error covariance matrix larger. To achieve this we introduce a positive factor κ in (18) to get

$$P_k^+ = P_k^- - \frac{1}{\kappa}H_kC_kP_k^- \tag{19}$$

The κ is a tuning factor which is fixed based on the faith we have in the model chosen. Larger the uncertainty, larger should be the value of κ.

The method of finding the optimal γ is not discussed in [5]. In our camera motion estimation algorithm, we avoid the problem of finding a uniform γ for $\{\hat{x}_k, k = 0, 1, \cdots N-1\}$ by finding an optimal γ_k for $\hat{x}_k$ in each recursive step. As the estimation performance measure at each step is

$$J_k = \frac{||x_k - \hat{x}_k||_I^2}{||x_{k-1} - \hat{x}_{k-1}||_{P_{k-1}^{-1}}^2 + ||w_k||_{W_k^{-1}}^2 + ||v_k||_{V_k^{-1}}^2} \tag{20}$$

this will ensure that

$$\sup J_k < 1/\gamma_k, k = 0, 1, \cdots, N-1 \tag{21}$$

Propagation of P_k and the computation of H_k are governed by (17,19) and (16) except that γ is replaced by γ_k. Obviously, the Kalman filter is an extreme case of H_∞ filter when $\gamma = 0$ which has a trivial estimation performance condition ($\sup J_k < \infty$). In a Kalman filter, the performance measure is the cumulative residual error given by

$$E = \sum_{k=0}^{N-1} E_k = \sum_{k=0}^{N-1} ||y_k - h(\hat{x}_k)||^2 \tag{22}$$

where $h(x_k)$ is the perspective project measurement equation.

This output error E_k is chosen as the stopping criterion for update of γ. When further increment in γ saturates the output error, the γ iteration is terminated.

The following search scheme is used to compute the optimal local γ.

1. Set $\gamma_k = 0$ and compute the Kalman filter residual error E_k
2. Increase γ_k by a step value $\Delta\gamma$
3. Compute $\hat{x}_k$ using (15) and (16)
4. Compute $E_k = ||y_k - h(\hat{x}_k)||^2$
5. If reduction in E_k is "sufficiently large", go to step 2 and increase γ_k, else stop and use the value of γ_k at last search loop as the optimal γ_k.

The implementation of the H_∞ filter is very similar to the implementation of the standard EKF except that the propagation of the state estimate covariance matrix P_k and the computation of gain H_k is done using (19) and (16), respectively. The optimal γ_k for each recursive loop is found using the above mentioned search scheme.

4. EXPERIMENTAL RESULTS

The set-up of the simulation was as follows. A number of feature points were randomly distributed in the field of view of the camera. The depth of those feature points ranged from 0.5 to 3.8. Both H_∞ and Kalman filters were initialized with system dynamic noise covariance matrix $W_k = 1$.

To test the performance of our algorithm we simulated the motion with different uncertainties in motion and noise models. Uniformly distributed noise (over ±4 pixels) was added to the image measurements. The robustness tests we carried out included simulating motion with a model which

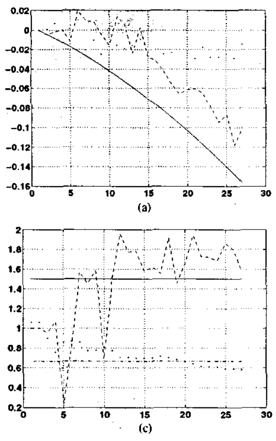

Fig. 2. (a) is the estimation of translation along optical axis and (b) the estimation of rotation about Y axis. In (a), it could be observed that due to the acceleration in the translation along the optical axis, estimates obtained from Kalman filter become poor while H_∞ performs well. (c) and (d) show the depths of two feature points. As is clear the error in estimating depth of the H_∞ filter is lower than that of the Kalman filter.

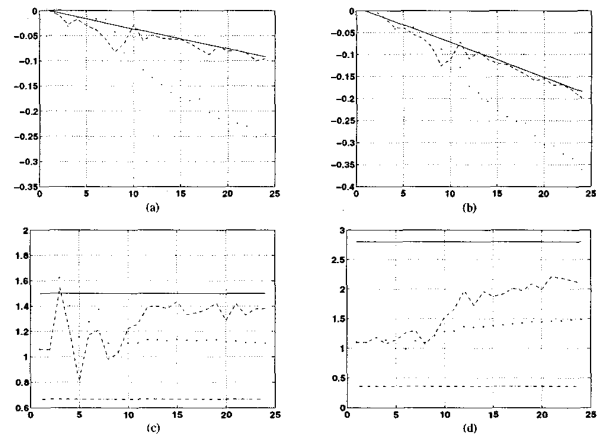

Fig. 3. The solid line represents ground truth, dotted line represents Kalman estimates and dashed line represents H_∞ estimates. (a) is the estimation of translation along X axis and (b) the estimation of rotation about Y axis. In (a), it could be observed that due to the bias in the noise estimates obtained from Kalman filter drifted away from the ground-truth gradually while H_∞ performs well. (c) and (d) show the depths of two feature points. As is clear the error in estimating depth of the H_∞ filter is lower than that of the Kalman filter.

is at variance with the assumed one. Examples of motion uncertainties included using constant acceleration in the translational and rotational motion. In such a case, it is evident that the value used for the dynamic noise covariance would not remain valid for large time (since the velocity predicted by the assumed model will be much lower than the true velocity when a constant acceleration is present). We may expect that the performance of the Kalman filter will be considerably worse than that in the case without constant acceleration. Figure 2 shows the case when there is an acceleration of 0.1 in the translation along the optical axis. Figure 2(a) shows the estimates of translation along optical axis. It is seen that the Kalman filter estimates drift away from the ground truth. However, the performance of the H_∞ filter does not degrade as much as the Kalman filter.

Examples of uncertainties in the assumed noise models included adding a bias in the measurement noise, using a noise covariance different from that used in the covariance matrices, etc. We added a bias of 0.15 in the measurement noise. The tracking of the translation along X axis, rotation about Y axis and the depths of two feature points are shown in figure 3. It can be seen that the H_∞ filter produces more robust estimates of motion and depth of feature points. We also tested the filters for the estimation of focal length in the presence of uncertainty. Here too the estimates using H_∞ filter converged to the true value of the focal length whereas the Kalman filter estimate of focal length diverged when there was significant uncertainty in motion and noise models.

5. CONCLUSIONS

The design and implementation of a discrete H_∞ filter for the structure from motion problem was discussed. The design criteria were based on the worst-case disturbance [5] and residual square errors. A search scheme was used to find the optimal noise attenuation γ_k for each recursive loop. Simulation results show the superiority of the H_∞ filter in the presence of significant modeling uncertainties.

6. REFERENCES

[1] T. J. Broida, S. Chandrashekhar, and R. Chellappa, "Recursive estimation of 3-d kinematics and structure from a noisy monocular image sequence," *IEEE Trans. on Aerospace and Electronic Systems* **vol. 26**, pp. 639–656, 1990.

[2] T. Wu, R. Chellappa, and Q. Zheng, "Experiments on estimating egomotion and structure parameters using long monocular image sequences," *International Journal of Computer Vision* **vol. 15**, pp. 77–103, 1995.

[3] A. Azarbayejani and A. Pentland, "Recursive estimation of motoin, structure, and focal length," *IEEE Trans. on Pattern Analysis and Machine Intelligence* **17**, pp. 562–575, 1995.

[4] B. Hassibi and T. Kailath, "h_∞ adaptive filtering," in *IEEE International Conference on Acoustics, Speech and Signal Processing, Detroit, MI*, pp. 949–952, 1995.

[5] X. Shen and L. Deng, "Game theory approach to discrete h_∞ filter design," *IEEE Trans. Signal Processing* **45**, pp. 1092–1095, 1997.

[6] A. Kale and R. Banavar, "Robust techniques for visual motion estimation," in *International Symposium on Intelligent Robotic Systems*, 1998.