BAYESIAN STRUCTURE FROM MOTION USING INERTIAL INFORMATION

Gang Qian, Rama Chellappa and Qinfen Zheng

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

ABSTRACT

In this paper, a novel approach to Bayesian structure from motion (SfM) using inertial information and sequential importance sampling (SIS) is presented. The inertial information is obtained from camera-mounted inertial sensors and is used in the Bayesian SfM approach as prior knowledge of the camera motion in the sampling algorithm. Experimental results using both synthetic and real images show that more accurate results can be obtained when inertial information is used or same estimation accuracy can be obtained using inertial information at a lower cost.

1. INTRODUCTION

The structure from motion (SfM) problem refers to the reconstruction of 3-dimensional (3D) scene structure and camera motion from an image sequence captured by a moving camera. SfM is a very crucial problem in computer vision and has been investigated for more than two decades. Although the underlying geometry of the problem is well understood, SfM remains a challenging problem due to feature tracking errors, feature occlusions, inherent ambiguities, mixed-domain sequences and the presence of independently moving objects, etc. To address these difficulties, researcher have been developing robust SfM algorithms.

Inertial information obtained from camera-mounted inertial sensors provides noisy measurements of camera motion parameters. Since the mid-1990s, inertial information has been used along with image sequences to estimate camera motion as well as scene structure [1, 2, 3]. [3] supplies a brief survey of the integration of inertial and visual systems. In [3], a robust SfM algorithm was designed to combine rate data (camera rotation velocity measured by rate sensors) and monocular image sequences using an extended Kalman filter (EKF). It was shown in [3] that the resulting SfM algorithm using rate data is more robust to feature tracking errors, mismatched feature points and is able to process mixed-domain sequences. The inherent ambiguities can also be reduced. Although there have been some successes in applying inertial information to SfM, the limitation of EKF is the main barrier to a complete solution to SfM by using inertial information.

Recently, researchers have attempted to solve the SfM problem in a Bayesian framework and some promising algorithms have been developed to handle the above difficult issues in SfM [4, 5]. In [5], an SfM algorithm was developed using the sequential importance sampling (SIS) [6]. With feature correspondences as observations, the posterior distribution of motion and structure parameters can be approximated by a set of samples and related weights. By using SIS, the samples and weights are updated to describe the new posterior distribution when new observations at the next time instant are available. It was shown in [5] that the SfM algorithm using SIS is capable of combating feature tracking errors and handling mixed-domain sequences. When the input image sequence is ambiguous, the resulting empirical distribution is multi-modal. Each mode represents a possible solution to the given observations. Moreover, in [7], another algorithm using SIS was proposed to detect moving objects from image sequences recorded by a moving camera. Computational cost is a main concern in applying sampling- based method in practices. Although SIS as a recursive technique is much less computationally intense than some Markov Chain Monte Carlo borne batch algorithms, it is still very crucial to find ways to reduce the computational load of the algorithm.

In this paper, we propose a novel SfM approach that integrate inertial information in a Bayesian SfM framework similar to the one developed in [5]. Inertial information is used as prior knowledge to the dynamics of the camera motion. It is a new way to fuse the inertial information in visual system. Moreover, both experiments using synthetic and real image sequences show that once inertial information is applied, more accurate results can be obtained with the same computational cost in the sense of needing fewer samples. Conversely, to achieve the same estimation accuracy, the pro-

Partially supported by the U.S. Army Research Laboratory (ARL) Collaborative Technology Alliance contract DAAD19-01-2-0008.

posed approach using inertial information is more computationally efficient and can converge to the true value much faster.

2. THEORETICAL BACKGROUND

2.1. Acquisition of inertial information

In our approach, the inertial information was measured by camera-mounted inertial sensors. The inertial measurement was then quantized and recorded in the form of bar-codes on the image frames synchronously captured by a video camera. In this paper, we will focus on the use of the rate data in SfM. Basically, from inertial rate sensors, we can obtain

$$\tilde{\Omega} = \Omega + \mathbf{n}_\Omega \tag{1}$$

where Ω is the camera rotation velocity and $\mathbf{n}_\Omega$ is the measurement noise. In practice, the measurement noise in the rate data is biased since the rate sensor has a drift, typically 3×10^{-4} radian/second [3]. Hence, in our implementation, an additive white Gaussian noise with zero mean and standard deviation of 0.01 radian/second is used to model the noise in rate data.

2.2. Sequential importance sampling

The SIS method has been proposed for approximating the posterior distribution of the state parameters of a dynamic system [6]. Denote the measurement as $\mathbf{y}_t$ and the state parameter as $\mathbf{x}_t$ and let $\mathcal{X}_t = \{\mathbf{x}_i\}_{i=1}^t$ and $\mathcal{Y}_t = \{\mathbf{y}_i\}_{i=1}^t$. Samples drawn from the posterior distribution of the states $(\mathbf{x}_1, \mathbf{x}_2, \cdots, \mathbf{x}_t)$ given all the available observations up to t, $\pi_t(\mathcal{X}_t) = P(\mathcal{X}_t|\mathcal{Y}_t)$ are needed to evaluate the ensemble statistics such as mean or modes. In SIS, the posterior distribution of the state parameters are approximated by a set of samples called *properly weighted samples* and their corresponding weights [6].

Suppose $\{\mathcal{X}_t^{(j)}\}_{j=1}^m$ is a set of random samples properly weighted by the set of weights $\{w_t^{(j)}\}_{j=1}^m$ with respect to π_t and let g_{t+1} be a trial distribution. Then the recursive SIS procedure to obtain the random samples and weights properly weighting π_{t+1} is as follows [6].

SIS steps: for $j = 1, \cdots, m$.

(A) Draw $X_{t+1} = \mathbf{x}_{t+1}^{(j)}$ from $g_{t+1}(\mathbf{x}_{t+1}|\mathcal{X}_t^{(j)})$. Attach $\mathbf{x}_{t+1}^{(j)}$ to form $\mathcal{X}_{t+1}^{(j)} = (\mathcal{X}_t^{(j)}, \mathbf{x}_{t+1}^{(j)})$.

(B) Compute the "incremental weight" u_{t+1} by

$$u_{t+1}^{(j)} = \frac{\pi_{t+1}(\mathcal{X}_{t+1}^{(j)})}{\pi_t(\mathcal{X}_t^{(j)}) g_{t+1}(\mathbf{x}_{t+1}|\mathcal{X}_t^{(j)})}$$

and let $w_{t+1}^{(j)} = u_{t+1}^{(j)} w_t^{(j)}$.

It can be shown [6] that $\{\mathcal{X}_{t+1}^{(j)}\}_{j=1}^m$ is properly weighted by $\{w_{t+1}^{(j)}\}_{j=1}^m$ with respect to π_{t+1}. Hence, the above SIS steps can be recursively applied to obtain the properly weighted set for any future time instant when the corresponding observations are available. The choice of the trial distribution g_{t+1} is very crucial in the SIS procedure since it directly affects the efficiency of the proposed SIS method. In our approach, we used the one step Markov transition probability distribution as the trial distribution

$$g_{t+1}(\mathbf{x}_{t+1}|\mathcal{X}_t) = q_{t+1}(\mathbf{x}_{t+1}|\mathbf{x}_t).$$

It can be shown that in this case $u_{t+1} \propto f(\mathbf{y}_{t+1}|\mathbf{x}_{t+1})$, which is the conditional probability density function of the observations at $t+1$ given the state sample $\mathbf{x}_{t+1}$ and is also known as the likelihood function of $\mathbf{x}_{t+1}$ since the observations are fixed.

3. BAYESIAN SFM USING INERTIAL INFORMATION

In this section, we will develop a novel approach to integrate the rate data in a Bayesian SfM framework using SIS. In [5], since no knowledge of the camera dynamics was available, a random walk model was adopted to describe the dynamics of the camera motion from one time instant to the next time instant. In our approach, we have direct measurements of camera rotation velocity. Although the measurements are noisy, we can use this information to predict the rotation angles of the camera at the next time instant instead of random guess. We will first review the dynamic system model of a moving camera and then describe the algorithm in details.

3.1. Parameterization of the camera motion

In our approach, the parameterization of the camera motion follows [5]. Two 3D Euclidean coordinate systems are used. One coordinate system is attached to the camera and uses the center of projection of the camera as its origin. It is denoted by C. The Z axis of C is along the optical axis of the camera, with the positive half axis pointing to the scene being observed. The $X - Y$ plane of C is perpendicular to Z axis with X and Y axes parallel to the borders of the image plane. Also, $X - Y - Z$ axes of C agree with the right-hand rule. The other coordinate system, denoted by I, is a world inertial frame. I is fixed on the ground and the axes of I are configured in such a way that initially, I and C are coincident. When the camera moves, C travels with the camera and I stays at the initial position.

Five parameters are employed to describe the camera motion at time t relative to the world inertial frame I.

$$\mathbf{x}_t = (\psi_x, \psi_y, \psi_z, \alpha, \beta)$$

$\psi = (\psi_x, \psi_y, \psi_z)$ are the rotation angles of the camera about the coordinate axes of the world frame I. (α, β) are the elevation and azimuth angles of the camera translation direction, measured in the world system I. Let $T(\alpha, \beta)$ be the unit vector in the translation direction given by $T(\alpha, \beta) = (\sin(\alpha)\cos(\beta), \sin(\alpha)\sin(\beta), \cos(\alpha))$ Moreover, the rotation angle $\psi(t)$ is given by

$$\psi(t) = \int_0^t \Omega(\tau) d\tau \tag{2}$$

Given the above motion parameterization, a state space model can be used to describe the behavior of a moving camera.

$$\mathbf{x}_{t+1} = \mathbf{x}_t + \mathbf{n}_x \quad (3)$$

$$\mathbf{y}_t = Proj(\mathbf{x}_t, \mathcal{S}_t) + \mathbf{n}_y \quad (4)$$

where $\mathbf{x}_t$ is the state vector and $\mathbf{y}_t$ is the observation at time t. $Proj(\cdot)$ denotes the perspective projection, a function of camera motion $\mathbf{x}_t$ and scene structure $\mathcal{S}_t$. $\mathbf{n}_x$ denotes the dynamic noise in the system, describing the time varying property of the state vector. Since rate data $\tilde{\Omega}$ is available, the dynamic noise $\mathbf{n}_\psi$ in the rotation angle can be described by

$$\mathbf{n}_\psi = \tilde{\Omega}\Delta t + \mathbf{n}_{n,\psi} \quad (5)$$

where Δt is the time between two time instants and $\mathbf{n}_{n,\psi}$ represents the rotation prediction errors using the rate data. As no prior knowledge about translation direction is available, a random walk will still be used to model the dynamics of camera translation.

SIS procedure for SfM using inertial information

1. **Initialization.** Draw samples of the motion parameters $\{\mathbf{x}_0^{(j)}\}_{j=1}^m$ from the initial distribution π_0. In $\{\mathbf{x}_0^{(j)}\}$, the components of the rotation angles are all set to zero and the samples of α and β are drawn from the uniform distribution in $[0, \pi]$ and $[0, 2\pi]$, respectively. Since all the samples are drawn from the exact distribution, equal weights are assigned to these samples.

   For $t = 1, \cdots, \tau$:

2. **Samples generation.** Draw $\{\mathbf{x}_t^{(j)}\}_{j=1}^m$ from the distributions of $\{\mathbf{x}_{t-1}^{(j)}\}_{j=1}^m + \mathbf{n}_x$. Since rate data is available, the new samples of rotation angles can be drawn as

   $$\psi_t = \psi_{(t-1)} + \tilde{\Omega}\Delta t + \mathbf{n}_{n,\psi} \quad (6)$$

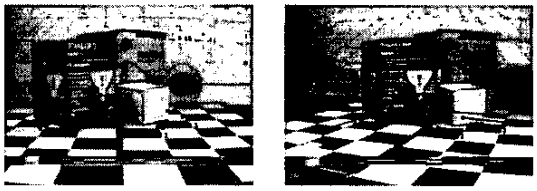
   For the trial distribution of the translation direction angles and the prediction errors $\mathbf{n}_{n_\Omega}$, the following distribution can be used.

   $$\begin{cases} n_{n,\psi_\iota} \sim \mathcal{N}(0, \sigma_\iota), \iota \in \{x, y, z\} \\ n_\kappa \sim U(-\delta_\kappa, \delta_\kappa), \kappa \in \{\alpha, \beta\} \end{cases} \quad (7)$$

   where σ_ι,δ_α and δ_β can be chosen as some positive numbers.

3. **Weight computation and re-sampling.** Compute the weights of the samples, $\{w_t^{(j)}\}$, using the observed feature correspondence according to the weight computation equation (5) in [5]. The resulting samples and their corresponding weights $(\mathcal{X}_t^{(j)}, w_t^{(j)})$ are properly weighted with respect to $\pi_t(\mathcal{X}_t)$. Re-sample the above samples.

By using the above properly weighted sample-weight sets for each time instant, the mean of motion parameters can be computed directly. Also since the sample-weight sequences after re-sampling approximates $\mathcal{X}_t$ in distribution, the MAP estimates of $\mathcal{X}_t$ can also be obtained by locating the modes of $\pi_t(\mathcal{X}_t)$. Once the motion distribution is obtained, the depth inference method proposed in [5] can be used directly to find the MAP or MMSE estimates of the depths at feature points.



(a) (b)

Fig. 1. Feature points and their trajectories tracked through an image sequence with inertial data. The inertial data were recored as bar-codes at the bottom of the image frames.

4. EXPERIMENTAL RESULTS

The proposed algorithm has been tested using both synthetic and real image sequences. Due to space limitations, only one example with real images is included here.

4.1. An example using a real image sequence

A real image sequence with inertial rate data was used in this experiment. Figure 1 (a) shows the first frame of the sequence with detected features and Figure 1 (b) shows the last frame with trajectories of the features. Results of the posterior distributions of the camera motion parameters obtained with and without using the inertial rate data with different numbers of motion samples are shown in Figure 2. The figures of each column correspond to the five motion parameters in the state space. In each figure, time goes from top to bottom. The figures in the first column from the left are the marginal posterior distributions of the motion parameters when 1000 motion samples were used without using inertial rate data. Since the number of samples was not large enough, the SfM estimates were not close to the ground-truths, shown by the solid lines crossing the plot from top to bottom. The figures in the second column show the motion distributions without using inertial information when the number of samples increased to 7000. Due to the increase of samples, the results were more accurate than that when only 1000 samples were used. In the third column, the figures show the estimation results of using inertial rate data with only 1000 motion samples. It can be seen that in the case of using inertial rate data, although only 1000 motion samples were used, the results are comparable to that obtained using 7000 samples without using inertial information.

4.2. Performance analysis

Various numbers of motion samples from 200 to about 10,000 were applied to these observations. The average epipolar distance was used to indicate the goodness of the estimates. Figure 3 shows the average epipolar distance with and without using inertial rate data when different numbers of mo-

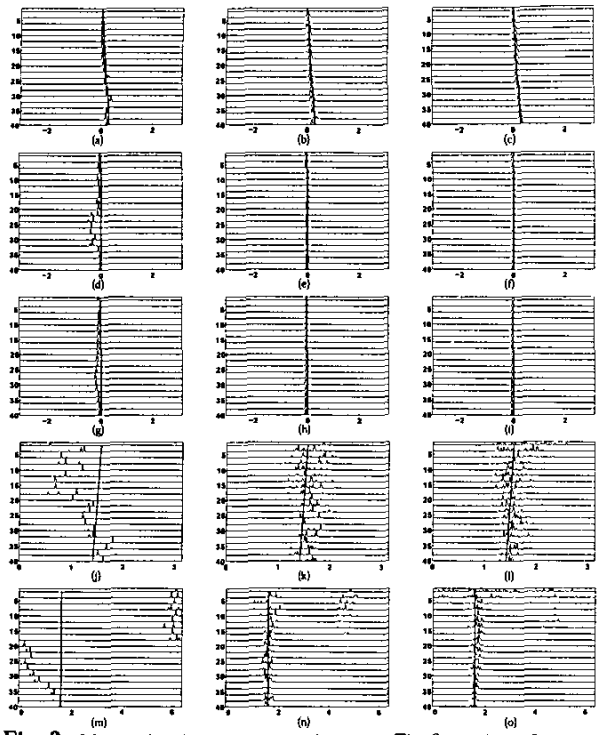


Fig. 2. Motion distributions using real images. The first column from the left shows the marginal posterior distributions of the motion parameters when 1000 motion samples are used without using inertial rate data. The figures in the second column show the motion distributions without using inertial information when the number of samples increases to 7000. In the third column, the figure shows the estimation results of using inertial rate data with only 1000 motion samples.

tion samples were used. In Figure 3, the horizontal axis is the number of motion samples used in the SIS procedure and the vertical axis indicates the average epipolar distance. The dotted line shows the average epipolar distances without using inertial rate data and the solid line shows the average epipolar distances when inertial data was used. As the number of motion samples increases, both lines with and without using inertial rate data decrease and converge to a steady state so that the performance of the algorithm does not improve by simply using more motion samples. It can be seen that with the same number of motion samples, the proposed algorithm by using inertial rate data has lower average epipolar distances and converges to the steady status much faster than the algorithm without using inertial data. Moreover, from this figure, we can see that the residue errors obtained using inertial rate data are always smaller than that without using rate data, even when sufficient large number of samples have been used. This tells us that with the help of rate data, the proposed algorithm can reach steady states with lower residue errors.

5. CONCLUSIONS

A novel approach to Bayesian SfM using inertial information is presented in this paper. By using inertial information as prior knowledge of the camera motion dynamics, more

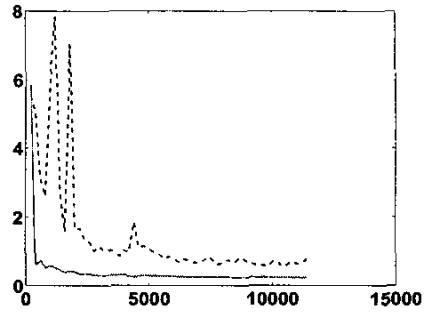


Fig. 3. The average epipolar distance with (solid lines) and without (dotted lines) using inertial rate data with different number of motion samples.

accurate SfM estimates can be obtained than that from the algorithm without using inertial information when the numbers of samples used in the SIS procedure are the same. It was also found that by using inertial information, the estimates can converge to the ground-truths much faster and less samples are needed to reach the steady status. Moreover, with the aid of inertial information, the proposed algorithm can reach a steady state with lower residue errors which indicates that the estimates are closer to the ground-truths. This observation can be explained by the lower CRLB due to the use of inertial information reported in [3].

6. REFERENCES

[1] F. Viéville, T.and Romann, B. Hotz, H. Mathieu, M. Buffa, L. Robert, P. Facao, O. Faugeras, and J. Audren, "Autonomous navigation of a mobile robot using inertial and visual cues," in *Intelligent Robots and Systems*, M. Kikode, T. Sato, and K. Tatsuno, eds., (Yokohama), 1993.

[2] T. Mukai and N. Ohnishi, "The recovery of object shape and camera motion using a sensing system with a video camera and a gyro sensor," in *International Conference on Computer Vision, Corfu, Greece*, pp. 411–417, Septempber 21-24, 1999.

[3] G. Qian, R. Chellappa, and Q. Zheng, "Robust structure from motion estimation using inertial data," *Journal of the Optical Society of America A* **18**, pp. 2982–2997, 2001.

[4] D. Forsyth, S. Ioffe, and J. Haddon, "Bayesian structure from motion," in *International Conference on Computer Vision, Corfu, Greece*, pp. 660–665, 1999.

[5] G. Qian and R. Chellappa, "Structure from motion using sequential monte carlo methods," in *International Conference on Computer Vision, Vancouver, Canada*, pp. II: 614–621, 2001.

[6] J. S. Liu and R. Chen, "Sequential monte carlo methods for dynamic systems," *J. Amer. Statist. Assoc.* **93**, pp. 1032–1044, 1998.

[7] G. Qian and R. Chellappa, "Moving targets detection using sequential importance sampling," in *IEEE International Conference on Acoustics, Speech and Signal Processing, Salt Lake City, UT*, 2001.