

A Bayesian Sensor Fusion Scheme for Attitude Tracking

Junekey Jeon, Hwa-Suk Kim, Woo-Sug Jung, Sun-Joong Kim

ETRI (Electronics and Telecommunications Research Institute), Dajeon, Korea

{jkjeon,hwskim,wsjung,kimsj}@etri.re.kr

Abstract—Accurate attitude tracking is a vital process in AR/VR applications. To satisfy its stringent requirement, multiple types of sensors, such as IMU and vision sensors, must cooperate, as each type of sensors has its own set of advantages complementing each other. However, fusion of different types of sensors is not a trivial task. We have previously proposed an attitude tracking algorithm using an IMU, which belongs to the category of recursive Bayesian filtering. In this paper, we propose a both theoretically reasonable and practically useful sensor fusion scheme based on Bayesian statistics, to extend and complement our previous algorithm. Implementation of our scheme for an AR application is also successfully done.

Keywords—AR (Augmented Reality), VR (Virtual Reality), HMD (Head Mounted Display), IMU (Inertial Measurement Unit), Attitude Tracking

I. INTRODUCTION

Recently, interests on AR/VR technology is continuously increasing. One of the key ingredients in this technology is *pose tracking*, which is the process of real-time estimation of both the position and the attitude displacement of a device from/to an object of interest. For example, the AR solution made by Metaio produces the pose of a camera relative to a visual marker, while the VR solution made by Oculus produces the pose of their HMD relative to the position tracker placed on a room. The real-time estimation of the pose is then directly applied to virtual object rendering.

Precise pose estimation is crucial in AR/VR applications, because even small amount of estimation error is very easily observable, instantaneously leading to user's discomfort. Several types of sensors are used together, complementing each other, to accomplish this goal. Typical types of such sensors are vision sensors (e.g., RGB/Depth camera) and IMU. IMU is cheap, fast, and stable but the data produced, especially those by magnetometer, is sometimes not reliable at all, as it is largely affected by the environment. Without magnetometer, however, IMU inherently suffers from the yaw-drift problem. This is not an issue for vision sensors, but they are certainly much slower since they require a lot of heavy computations. Furthermore, pose estimation from vision sensors are usually not as stable as those from IMU, and may result considerable amount of “jitters” in virtual object rendering. While advantages and disadvantages of each type of sensors are clear, it is not an obvious task to establish a proper fusion of data produced from those sensors. The goal of this paper is to introduce a both theoretically reasonable and practically useful method of doing this.

In this paper, we focus on the *attitude tracking* part of the general pose tracking problem. Attitude tracking is in general much easier than the full pose tracking, and there are many AR/VR applications (e.g., 360° videos) only utilizing the attitude tracking but not the full pose tracking. Also, attitude tracking is in some sense more fundamental than position tracking, since it is the source of nonlinearity of the general pose tracking problem. It can be also utilized as an important subtask for the full pose tracking.

II. IMU FILTERING ALGORITHM

The set of possible attitudes, usually denoted as $SO(3)$ and called the *special orthogonal group of dimension 3*, is a well-studied example of Lie groups [2]. It is the set of 3×3 orthogonal matrices with positive determinant:

$$SO(3) := \{r \in \mathbb{R}^{3 \times 3} : r^T r = \mathbf{1}, \det r = 1\}$$

where $\mathbb{R}^{3 \times 3}$ is the set of 3×3 matrices of real entries, a^T denotes the transpose of a matrix a , $\mathbf{1}$ denotes the identity matrix of a given size, and $\det a$ denotes the matrix determinant of a matrix a .

In [1], we have derived an attitude tracking algorithm using IMU based on several families of probability distributions on $SO(3)$. The first one is described as follows. Let $\bar{r} \in SO(3)$ and Σ be a symmetric positive-definite homogeneous degree 2 element in the universal enveloping algebra [2] $U(\mathfrak{so}(3))$ of the Lie algebra [2] $\mathfrak{so}(3)$ associated to $SO(3)$. Then we consider the unique probability measure μ on $SO(3)$ having the Fourier-Stieltjes transform [3]

$$\hat{\mu}(\xi) := \int_{SO(3)} \xi(x)^{-1} d\mu(x) = \exp\left(\frac{1}{2}\xi_*\Sigma\right) \xi(\bar{r}^{-1}) \quad (1)$$

for each irreducible representation [2] ξ of $SO(3)$, where ξ_* is the derivative of ξ . We will denote this probability measure as $LD(\bar{r}, \Sigma)$ and call $LD(\bar{r}, \Sigma)$ it as the *left-translated diffusion distribution* with mean $\bar{r}$ and covariance Σ .

On the other hand, another family of probability distributions is obtained from the following procedure: first, consider a PDF (Probability Density Function)

$$f(q) = \frac{1}{K(M)} \exp(q^T M q) \quad (2)$$

on the unit sphere $S^3 \subseteq \mathbb{R}^4$, for some 4×4 symmetric matrix M of real entries. Think of each $q = (\cos \frac{\theta}{2}, v \sin \frac{\theta}{2}) \in \mathbb{R}^{1+3}$ on S^3 as the quaternion representation of the rotation along the axis v with angle θ , then (2) induces a probability measure

on $SO(3)$. Since $f(q)$ has antipodal symmetry, that is $f(q) = f(-q)$ for all q , it follows that the induced probability measure on $SO(3)$ has the PDF that is exactly the double of $f(q)$. We call this induced measure *Bingham distribution* associated to M , and denote it as $BH(M)$. There is another description of $BH(M)$ as a *matrix Fisher distribution* [4]. After some computation, one can show that it is possible to obtain from M a unit quaternion $\bar{q}$ and a 3×3 symmetric positive-definite matrix Σ so that

$$f(q) = \frac{e^{\lambda_0}}{K(M)} \exp \left((\bar{q}^{-1}q)^T \begin{bmatrix} 0 & 0 \\ 0 & -\frac{1}{2}\Sigma^{-1} \end{bmatrix} (\bar{q}^{-1}q) \right) \quad (3)$$

where λ_0 is the largest eigenvalue of M , whenever λ_0 is strictly larger than other eigenvalues of M . Here, $\bar{q}$ (as well as $-\bar{q}$) is the “center” of the distribution, while Σ describes how the distribution “spreads” from $\bar{q}$. Eigenvectors of Σ describe the “principal directions” of spreading, while the corresponding eigenvalues describe the amount of spreading along that directions.

We have hypothesized the following approximation in order to derive the algorithm in [1]:

$$LD(\bar{r}, \Sigma) \approx BH(M), \quad (4)$$

where $\bar{r}$ is the rotation matrix corresponding to $\bar{q}$, which together with Σ appears in (3). This hypothesis is indeed true when Σ is a scalar multiple of the identity matrix $\mathbf{1}$ [5], but as far as we know, there is no known mathematical results about the validity of this approximation for the general case, although it seems intuitively very natural.

The algorithm we derived in [1] is presented here for convenience:

Here, $\bar{r}_0, \Sigma_0, \bar{\omega}_0, \sigma_{\Omega,0}^2$ denote the initial estimations of attitude, attitude covariance, angular velocity, and angular velocity variance, respectively. σ_P^2 denote the variance of the increment of angular velocity at each time instance, and $\sigma_U^2, \sigma_V^2, \sigma_W^2$ are the noise variance of the gyroscope, the accelerometer, and the magnetometer, respectively, and Δt is the time interval between two consecutive samples from IMU. $\mathbf{g}, \mathbf{m}$ are the gravity vector and the Earth’s magnetic vector, respectively. Also, g, a, m denote the outputs of the gyroscope, the accelerometer, and the magnetometer, respectively, and $\bar{r}, \Sigma, \bar{\omega}, \sigma_{\Omega}^2$ denote the current estimation of attitude, attitude covariance, angular velocity, and angular velocity variance, respectively. Finally, $Q_L(q)$ for a quaternion q is the matrix representation of left-multiplication by q .

III. FUSION OF EXTERNAL ATTITUDE MEASUREMENTS

As depicted in Section I, magnetometer is in general not very reliable in our daily environment. It is sometimes even safer to set the magnetometer variance σ_W^2 much higher than the actual variance, so the data from the magnetometer is almost “turned off”. However, this makes IMU inherently insufficient for the full attitude measurement, and together with the inevitable gyroscope bias this causes the estimation to drift along the yaw direction. To fix this problem, some other sensors free from yaw-drift, for example vision sensors, must be accompanied.

Algorithm 1 IMU Filtering Algorithm

Input: $\bar{r}_0, \Sigma_0, \bar{\omega}_0, \sigma_{\Omega,0}^2, \sigma_P^2, \sigma_U^2, \sigma_V^2, \sigma_W^2, \Delta t, \mathbf{g}, \mathbf{m}$
Output: $\bar{r}, \Sigma, \bar{\omega}, \sigma_{\Omega}^2$

```

1: // Initialize
2:  $\bar{r} \leftarrow \bar{r}_0, \Sigma \leftarrow \Sigma_0, \bar{\omega} \leftarrow \bar{\omega}_0, \sigma_{\Omega}^2 \leftarrow \sigma_{\Omega,0}^2$ 
3: for each time-instance do
4:   // Measurement
5:    $(g, a, m) \leftarrow getMeasurements()$ 
6:   // Update using gyroscope
7:    $\bar{r} \leftarrow \exp \left( - \left[ \frac{(\sigma_U^2 + \sigma_P^2)\bar{\omega}}{\sigma_U^2 + \sigma_P^2 + \sigma_{\Omega}^2} + \frac{\sigma_{\Omega}^2 g}{\sigma_U^2 + \sigma_P^2 + \sigma_{\Omega}^2} \right] \times \Delta t \right) \bar{r}$ 
8:    $\Sigma \leftarrow \Sigma + \frac{(\sigma_U^2 + \sigma_P^2)\sigma_{\Omega}^2 \Delta t^2}{\sigma_U^2 + \sigma_P^2 + \sigma_{\Omega}^2} \mathbf{1}$ 
9:    $\bar{\omega} \leftarrow \frac{\sigma_U^2 \bar{\omega}}{\sigma_U^2 + \sigma_P^2 + \sigma_{\Omega}^2} + \frac{(\sigma_P^2 + \sigma_{\Omega}^2)g}{\sigma_U^2 + \sigma_P^2 + \sigma_{\Omega}^2}$ 
10:   $\sigma_{\Omega}^2 \leftarrow \frac{\sigma_U^2(\sigma_P^2 + \sigma_{\Omega}^2)}{\sigma_U^2 + \sigma_P^2 + \sigma_{\Omega}^2}$ 
11:  // Update using accelerometer & magnetometer
12:   $t_1 \leftarrow -\frac{\|a - \mathbf{g}\|^2}{2\sigma_V^2} - \frac{\|m - \mathbf{m}\|^2}{2\sigma_W^2}, w_1 \leftarrow \frac{\mathbf{g} \times a}{\sigma_V^2} + \frac{\mathbf{m} \times m}{\sigma_W^2}$ 
13:   $B_1 \leftarrow \frac{a\mathbf{g}^T + \mathbf{g}a^T}{\sigma_V^2} + \frac{m\mathbf{m}^T + \mathbf{m}m^T}{\sigma_W^2} - \left( \frac{\|a + \mathbf{g}\|^2}{2\sigma_V^2} + \frac{\|m + \mathbf{m}\|^2}{2\sigma_W^2} \right) \mathbf{1}$ 
14:   $M_1 \leftarrow \begin{bmatrix} t_1 & w_1^T \\ w_1 & B_1 \end{bmatrix}$ 
15:   $q_0 \leftarrow toQuaternion(\bar{r})$ 
16:   $M_2 \leftarrow Q_L(q_0) \begin{bmatrix} 0 & 0 \\ 0 & -\frac{1}{2}\Sigma^{-1} \end{bmatrix} Q_L(q_0^{-1})$ 
17:   $(\bar{r}, \Sigma) \leftarrow Bingham\_decomposition(M_1 + M_2)$ 
18: end for
```

Let us assume that at some time instance the attitude of the system was R and a vision-based (or other kinds of) attitude measurement system estimated the attitude of the system as S . We may interpret this S as R disturbed by a random noise Z ; that is,

$$S = ZR \quad (5)$$

where R and Z are independent random variables taking their values in $SO(3)$. The order of multiplication matters here; the interpretation of $S = ZR$ and that of $S = RZ$ are different. This depends on how we defined the attitude: it could be either the coordinate transformation from the reference frame to the device’s local frame or vice versa. We derived **Algorithm 1** based on the first definition of attitude, so we stick to that. In this case, the noise coming *after* the true attitude (that is, Z in ZR) is due to the measurement device; for example, thermal noise is included in this category. On the other hand, the noise coming *before* the true attitude (that is, Z in RZ) is due to instability of the environment itself; for example, when using vision-based measurement, feature points may not be perfectly rigid. In reality, both forms of noise coexist, but we assume only one of them is significant. We will take look at the case $S = ZR$ here; the other case is almost the same.

Since Bingham distributions are kinds of generalization of normal distributions, it seems reasonable to assume that the probability distributions of both R and Z are Bingham distributions; let Z follows $BH(M_Z)$ where $M_Z = \begin{bmatrix} 0 & 0 \\ 0 & -\frac{1}{2}\Sigma_Z^{-1} \end{bmatrix}$, and R follows $BH(M_R)$. Or, we may assume that they

follow left-translated diffusion distributions and relying on the hypothesis (4). Here, Σ_Z is the “covariance” of the noise. On the other hand, we have estimated the probability distribution of R as $LD(\bar{r}, \Sigma)$ using **Algorithm 1**, so using the hypothesis (4), we can find M_R from $\bar{r}$ and Σ , as described in the line 15 and the line 16 of **Algorithm 1**. Then, by Bayes’ rule, the posterior distribution of R after observing S is

$$p_{R|S}(r|s) = \frac{p_{S|R}(s|r)p_R(r)}{p_S(s)} = \frac{p_Z(sr^{-1})p_R(r)}{p_S(s)}. \quad (6)$$

Let q, q_s be quaternion representations of r, s . Note that

$$\begin{aligned} p_Z(sr^{-1}) &\propto \exp((q_s q^{-1})^T M_Z (q_s q^{-1})) \\ &= \exp((q^{-1})^T (Q_L(q_s)^T M_Z Q_L(q_s)) q^{-1}). \end{aligned} \quad (7)$$

Write $Q_L(q_s)^T M_Z Q_L(q_s) = \begin{bmatrix} t & w^T \\ w & B \end{bmatrix}$, and note that

$$(q^{-1})^T (Q_L(q_s)^T M_Z Q_L(q_s)) q^{-1} = q^T \begin{bmatrix} t & -w^T \\ -w & B \end{bmatrix} q, \quad (8)$$

so let $M_S := \begin{bmatrix} t & -w^T \\ -w & B \end{bmatrix}$, then

$$p_{R|S}(r|s) \propto \exp(q^T (M_R + M_S) q), \quad (9)$$

so we can conclude that the posterior distribution of R given S is the Bingham distribution $BH(M_R + M_S)$. Therefore, by decomposing $M_R + M_S$ we can update $\bar{r}$ and Σ .

It is practically not easy to determine the noise statistics; that is, it is a nontrivial task to find what Σ_Z is, which encodes the two things: the principal directions (the eigenvectors) and the corresponding uncertainties along that directions (the eigenvalues). Especially the principal directions are not easy to determine in practice. In this case, one may use only the overall uncertainty instead; that is, to assume $\Sigma_Z = \sigma_Z^2 \mathbf{1}$. This σ_Z^2 is relatively easy to find in practice: just measure the sample standard deviation of the measurement, then σ_Z is that standard deviation divided by $\sqrt{3}$ ¹. When $\Sigma_Z = \sigma_Z^2 \mathbf{1}$, the computation of M_S becomes simpler:

$$M_S = \frac{1}{2\sigma_Z^2} \begin{bmatrix} -\|v\|^2 & cv^T \\ cv & vv^T - \mathbf{1} \end{bmatrix} \quad (10)$$

when $q_s = (c, v) \in \mathbb{R}^{1+3}$. Interestingly, when $\Sigma_Z = \sigma_Z^2 \mathbf{1}$, the resulting M_S is same whatever we assumed between $S = RZ$ and $S = ZR$.

Note that our fusion scheme need not be restricted to vision-based measurement because we only need the assumption on distribution of the random noise Z , and nothing specific to vision is assumed. This fusion procedure can be done for any number of any type of attitude measurement systems. Another benefit is that timing synchronization between IMU and the external measurement system is not required; the procedure just can be done whenever the new measurement output is available. This is particularly appealing when fusing vision-based measurements, because usually vision-based measurements require high computational costs, so they have relatively slow measurement sampling speed.

¹This is of courses also only a rough measurement. For precise methods for determining Σ_Z , see [6]. It only explains the case when the distribution is on S^2 , but the case of S^3 can be done similarly.

We have implemented our fusion scheme for an AR application. When using only vision-based measurements (Metaio SDK with Intel® RealSense™ camera), there were lots of disturbing jitters, though dependence on the number of visual feature points was large. When using only IMU, either the magnetometer ran into trouble or the estimation suffered from the yaw-drift problem, depending on how much the magnetometer is actively used. By applying the proposed fusion scheme, we were able to complement the problems of both types of sensors.

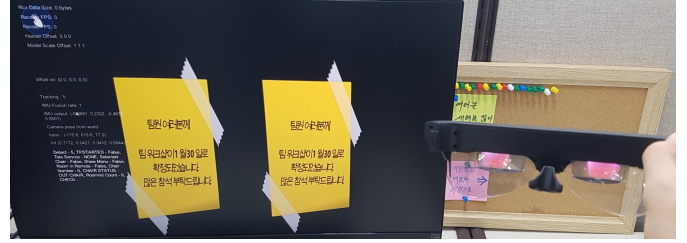


Fig. 1. The AR application which the fusion scheme is implemented

IV. CONCLUSION

In order to meet the stringent precision requirement in pose tracking in AR/VR, it is significant to use several different types of sensors in a way that their own advantages are well complementing each other. Based on the algorithm presented in [1], we proposed a Bayesian update scheme to complement the insufficiency prevalent in attitude measurement systems only using IMU. The proposed scheme uses Bingham distributions as both prior and noise distributions, and this enables a simple calculation of the posterior distribution. The implementation result was very successful, so we could conclude that the scheme is not only theoretically plausible but also practically useful.

ACKNOWLEDGMENT

This work was supported by Institute for Information & communications Technology Promotion (IITP) grant funded by the Korea government (MSIP) (No.B0101-16-1292, Development of Smart Space to promote the Immersive Screen Media Service).

REFERENCES

- [1] J. Jeon, H.-S. Kim, W.-S. Jung, and S.-J. Kim, “An attitude tracking algorithm using imu,” in *KIISE 2015 Winter Conference*. KIISE, 2015, pp. 1119–1121.
- [2] W. Fulton and J. Harris, *Representation Theory: A First Course*, corrected ed., ser. Graduate Texts in Mathematics. Springer, 1999.
- [3] G. B. Folland, *Real Analysis: Modern Techniques and Their Applications*, 2nd ed. Wiley, 2013.
- [4] C. Bingham, T. Chang, and D. Richards, “Approximating the matrix fisher and bingham distributions: Applications to spherical regression and procrustes analysis,” *Journal of Multivariate Analysis*, vol. 41, no. 2, pp. 314–337, 1992.
- [5] S. Matthies, J. Muller, and G. W. Vinel, “On the normal distribution in the orientation space,” *Textures & Microstructures*, vol. 10, pp. 77–96, 1988.
- [6] C. Bingham, “An antipodally symmetric distribution on the sphere,” *The Annals of Statistics*, vol. 2, no. 6, pp. 1201–1225, 1974.



Junekey Jeon received his M.S. degree in the field of electrical engineering from KAIST, Daejeon, Korea, in 2015. Since 2015, he has been a researcher with the Electronics and Telecommunications Research Institute (ETRI). His research interests include statistics, estimation theory, computer vision, machine learning, and information theory.



Hwa-Suk Kim received her M.S. degree from Chung-Nam National University, Daejeon, Korea, in 2002. Since 1991, she is currently a principal researcher on Broadcasting & Telecommunications Media Research Laboratory in Electronics Telecommunication Research Institute (ETRI). Her current research interests are the augmented/mixed reality, service control, and IoT (Internet of Things).



Woo-Sug Jung received his Ph.D. degree from Chung-Nam Univ, Daejeon, Korea, in 2010. Since 1994, he is currently a principal researcher on Broadcasting & Telecommunications Media Research Laboratory in Electronics Telecommunication Research Institute (ETRI). His current research interests are the augmented reality, wired network, and IoT (Internet of Things).



Sun-Jung Kim received her B.S. degree in computational statistics and her M.S. degree in computer science from Chungnam National University, Daejeon, Korea, in 1989 and 2000 respectively. In February 1989, she joined Electronics Telecommunication Research Institute (ETRI) where she is currently principal researcher and director. Her research interests includes convergence service control, smart TV, content knowledge mining.