Mammalian Brain Inspired Localization Algorithms with von Mises Distributions

Tsang-Kai Chang and Ankur Mehta

Abstract—Biological agents still outperform the artificial counterparts in navigating the first-visited environments, even with the advance of deep neural networks nowadays. To bridge this gap, by taking the localization problem as the initial step, we investigate the localization principles in mammalian brains to establish the common localization framework in both biological and artificial systems. Furthermore, inspired by the grid cells discovered in mammalian brains, a localization algorithm with circular representation is proposed. Compatible with bearing-and-distance measurement, the proposed algorithms avoid the linearization inconsistency which remains a severe problem in conventional extended Kalman filter algorithms. As the effectiveness of the proposed algorithms shown in simulation results, this paper indicates a novel localization method that is promising to further tackle the simultaneous localization and mapping (SLAM) problem.

I. INTRODUCTION

The ability to self-localize in the environment is essential for all autonomous agents, ranging from biological rodents to artificial robots. Even with the advance of deep neural networks in artificial intelligence, biological agents still outperform artificial counterparts in spatial navigation, where biological agents do not require reinforcement learning over several trials [1]. To further improve the state-of-the-art navigation algorithms in artificial agents, insights from biological systems are undoubtedly useful [2]. Therefore, a unified framework for localization methods paves the road to import ideas from biological systems, and to further ameliorate the existing algorithms for robots.

In mammals, there are several cells whose firing patterns obtain spatial selectivity. Place cells in hippocampus fire when the position of the rodent is close to certain spatial point; head direction cells in postsubiculum are activated when the head of the rodent is facing certain direction; and grid cells in medial entorhinal cortex fire when the position of the rodent is in one of the vertices of the hexagonal pattern [3]. A comprehensive review on the associated neuroscience background can be found in [4], while the essential information is summarized in Section III. Among all cells, grid cells are especially crucial since the firing pattern is context-independent, which suggests that grid cells are embedded with universal spatial structure. Surprisingly, the grid pattern shows up in some of the artificial neurons of recurrent networks conducting navigation tasks [1]. Even though a single grid cell gives ambiguous positional information, an ensemble of grid cells does provide accurate position representation. Specifically, proximate grid cells have the same spatial scale and orientation but different spatial phases, as those grid cells are organized as a module. Several modules of grid cells can then provide an unique spatial representation with a relatively large coverage comparing to the scales of each module. As a result, circular representation of the spatial state is pervasive in mammalian brain, with head direction cells for orientation and grid cells for position. The circular representation itself is also believed to obtain arithmetic advantages over conventional fixed-base representation [5].

To unveil the underlying principles of both biological and artificial agents, we concentrate on localization problem in this paper. That is, we assume that the map of environment is already built, and leave the investigation of SLAM for further research. In mammalian brain, postsubiculum with head direction cells and medial entorhinal cortex with grid cells are believed to keep the estimates of orientation and of position, respectively. While the essence of localization is estimation problem, the utilization of both proprioceptive and exteroceptive information should be investigated in detail. As for path integration, also known as dead reckoning in engineering, head direction and grid cells are able to represent the latest estimate with proprioceptive signals. The exteroceptive information, possibly from lateral entorhinal cortex to hippocampus, also helps to refine the internal spatial estimate [4], [6], [7]. The dependency on proprioceptive and exteroceptive information in mammalian brains shares the same framework in robotic localization, summarized in Fig. 1 and further explained in Section III, which serves as the basis for applying biological principles on artificial agents. In [8], an architecture inspired by grid cells is proposed for SLAM to map suburb area. However, no proprioceptive information is used in this system, which not only deviates from the current understanding of the mechanism of grid cells, but also remains uninformative for

Fig. 1. The localization architecture of mammalian brains with engineering interpretation.

The authors are with the Department of Electrical and Computer Engineering, University of California, Los Angeles, CA, 90095 USA. e-mail: {tsangkaichang, mehtank}@ucla.edu
The engineering system relying on dead reckoning.

While the application of biological principles seems promising, the intrinsic differences between biological and artificial systems should be carefully acknowledged. In particular, besides estimation uncertainty from the stochastic modeling, we also have to consider the representation uncertainty. Especially in mammalian brains, while the spatial state is represented by the firing patterns of neurons, the representation of spatial state is stochastic, due to the intrinsic neuron firing randomness. On the contrary, the numerical representation in machine is static.

Therefore, what we are going to do in this paper is not a faithful replica of brain activity, but an engineering implementation of the localization algorithm with the principles extracted from mammalian brains.

We propose the algorithm with mixture representation, circular representation for orientation and Cartesian representation for position in particular, and also the one with fully circular representation. The former is more consistent with existing engineering systems, while the later resembles the one in mammalian brains. To broaden the engineering applicability, we incorporate bearing-and-distance measurements in the proposed algorithms. Conventional methods based on extended Kalman filtering (EKF) are notorious for the approximation will be used extensively later as the connection between biological and artificial distributions.

The connection between von Mises and Gaussian distributions also arises in conditioning, where the von Mises can be generated by conditioning on bivariate Gaussian distribution.

**Theorem 1.** Let \( v \) be bivariate random vector with mean \( d(\cos \phi, \sin \phi) \) and covariance \( \sigma^2 I_2 \), where \( I_n \) is a \( n \times n \) identity matrix. If \( v = r(\cos \theta, \sin \theta) \), then the conditional probability of \( \theta \) given \( r = r_0 \) is \( vM(\phi, r_0 d/\sigma^2) \).

**Proof.** The probability density function of \( f(r, \theta) \) is given by

\[
f(r, \theta) = \frac{1}{2\pi \sigma^2} \exp \left[ -\frac{1}{2\sigma^2} \left( r^2 d^2 - 2rd \cos(\theta - \phi) \right) \right].
\]

Therefore,

\[
f(\theta | r = r_0) \propto \exp \left[ \frac{r_0 d}{\sigma^2} \cos(\theta - \phi) \right],
\]

which completes the proof. \( \square \)

As von Mises random variable often appears in trigonometric functions, the properties of the trigonometric moment for von Mises distribution are thus essential. For zero-mean von Mises distribution, we have the following theorem from [13].

**Theorem 2.** For a random variable \( \theta_0 \) with distribution \( vM(0, \kappa) \), we have \( \mathbb{E}[\cos(p\theta_0)] = I_p(\kappa)/I_0(\kappa) \) for integer \( p \).

With Theorem 2 we have the following lemma regarding von Mises random variables with arbitrary parameters.

**Lemma 1.** For random variable \( \theta \) following \( vM(\mu, \kappa) \),

1. \( \mathbb{E}[\cos(\theta)] = \frac{I_1(\kappa)}{I_0(\kappa)} \cos(\mu) \),
2. \( \mathbb{E}[\cos^2(\theta)] = \frac{1}{2} + \frac{1}{2} \frac{I_1(\kappa)}{I_0(\kappa)} \cos(2\mu) \),
3. \( \text{Var}(\cos(\theta)) \leq 1 \).

**Proof.** The first two statements can be prove by Theorem 2 together with double-angle formulae and with the fact that \( \sin(\cdot) \) is odd. For the last statement,

\[
\text{Var}(\cos(\theta)) = \mathbb{E}[\cos^2(\theta)] - (\mathbb{E}[\cos(\theta)])^2
\]

\[
= \frac{1}{2} + \frac{1}{2} \frac{I_1(\kappa)}{I_0(\kappa)} \cos(2\mu) - \left( I_2(\kappa)/I_0(\kappa) \right)^2 \cos^2(\mu)
\]

\[
\leq 1 - \left( \frac{I_1(\kappa)}{I_0(\kappa)} \right)^2 \cos^2(\mu),
\]

where the inequality comes from Lemma 2 and \( \cos(\cdot) \leq 1 \). While \( \cos^2(\mu) \geq 0 \), we arrive the statement that \( \text{Var}(\cos(\theta)) \leq 1 \). \( \square \)

**III. Background in Neuroscience**

In this section, we summarize the constitution of grid cell modules, and the interaction of place and grid cells in circuitry level.
A. Spatial Representation by Grid Cells

To begin with, we consider $N$ one-dimensional grid cells within the same module whose spatial period is $\lambda$. As presented in [14], the tuning curve of cell $i$ in this module as a function of position $x \in \mathbb{R}$ can be modeled as

$$O_i(x) = f_{\text{max}} \tau \exp \left( \frac{1}{\sigma^2} \cos \left[ 2\pi \left( \frac{x - \phi_i}{\lambda} \right) \right] \right),$$

where $f_{\text{max}}$ is the peak firing rate, $\tau$ is the characteristic time interval over which spikes are counted, $\phi_i$ is the preferred phase, and $\sigma$ is the tuning width. Since the number of firings of this cell $i$ during the interval $\tau$ follows Poisson distribution with mean $O_i(x)$, the posterior distribution of $x$ can be constructed by collecting the firing patterns from all $N$ neurons, $O_1, \ldots, O_N$. While the tuning curves of those grid cells uniformly cover the interval by observation, our approximation of $O_i(x)$ can be approximated as constant for all $x$, the posterior density function is then given by

$$p(x|o_1, \ldots, o_N) \propto \exp \left( \sum_{i=1}^{N} o_i \cos(x - \phi_i) \right),$$

with $\kappa = \sigma^{-2}$. The module of grid cells behaves as a single grid cell by regarding the fact that the summation of cosine functions with the same period is still a cosine function, or

$$\bar{\kappa} \cos(x - \bar{\phi}) = \kappa \sum_{i=1}^{N} o_i \cos(x - \phi_i).$$

To obtain $\bar{\phi}$ and $\bar{\kappa}$, we first rewrite the summation of cosine functions as

$$\bar{\kappa} \cos(x - \bar{\phi}) = \kappa \sum_{i=1}^{N} o_i \cos(x - \phi_i) = \text{Re} \left[ \kappa e^{\bar{x}i} \sum_{i=1}^{N} o_i e^{-i\phi_i} \right],$$

which gives that

$$\bar{\phi} = - \arg \left( \sum_{i=1}^{N} o_i e^{-i\phi_i} \right) = \arg \left( \sum_{i=1}^{N} o_i e^{i\phi_i} \right).$$

By letting $x = \phi$, we have

$$\bar{\kappa} = \kappa \sum_{i=1}^{N} o_i \cos(\bar{\phi} - \phi_i).$$

Therefore, the posterior function in (3) can be expressed as

$$p(x|o_1, \ldots, o_N) = \frac{1}{2\pi I_0(\kappa)} \exp \left( \bar{\kappa} \cos(x - \bar{\phi}) \right),$$

which follows von Mises distribution with mean $\bar{\phi}$ and concentration parameter $\bar{\kappa}$ exactly. That is, the maximum likelihood estimation from the posterior distribution is equivalent to detect the position from the firing pattern of a single neuron with preferred phase $\bar{\phi}$. As an example of population vector decoding [15], the collective representation of grid cells in the same module combats the representation uncertainty of each neuron, and signifies a single circular position with period $\lambda$.

While single circular representation leads to ambiguity over every period $\lambda$, the combination of $M$ orientation-aligned modules leads to uniquely represent space over a much larger coverage, since a signal with lower period emerges by summing two pure periodic signals [4]. While the orientations of different grid cell modules are aligned, the periods in each module, denoted by $\lambda_1, \ldots, \lambda_M$, form a discrete set [5]. In [16], [17], the optimal ratio of periods between adjacent modules is analyzed to lie around $3/2$, where the experiment on artificial agents shows similar result [1].

The extension to two-dimensional circular representation is straightforward with the setup of one-dimensional representation. However, direct assigning two independent one-dimensional circular representation to $x$- and $y$-axes in Cartesian coordinate system leads to a rectangular grid, rather than a hexagonal grid, which appears when two axes evolve along a twisted torus [6].

B. Hippocampus and Entorhinal Cortex Circuitry

With the orientation presented by the ensemble of head direction cells and the position by the ensemble of grid cells, we move on to the computation of estimation update. Based on [4], [6], [7], we summarize the current understanding of the estimate update in neuroscience as follows to the best of our knowledge. For path integration or dead reckoning, the proprioceptive information regarding angular and translational velocities, including vestibular information, sensorimotor information, and optic flow, is integrated to track the state estimate. While the exact location where the computation occurs is under debate, the firing patterns of head direction and grid cells are believe to represent the latest state estimates.

The external sensory information is also crucial to calibrate state estimates. In particular, lateral entorhinal cortex, connected with several areas in cortex bidirectionally, provides sensory information about the current position to hippocampus, which further improves the state estimation of head direction and grid cells. Therefore, the place cells can be regarded as a landmark map that locate the agent by the environment. In fact, hippocampus and entorhinal cortex form a loop to realize map building in novel environments. However, as the focus of this paper is localization rather than SLAM, we leave the discussion on the information flow from entorhinal cortex to hippocampus for the further work, and summarize the unified localization architecture in Fig. [1].

IV. KALMAN FILTERING WITH VON MISES DISTRIBUTIONS

We consider a discrete-time system with time index $t$ to describe the evolution of circular state $\theta_t \in [0, 2\pi)$ as

$$\theta_{t+1} = \theta_t + u_t + w_t,$$

where $u_t$ is the input at time $t$ and $w_t$ is the process noise, which is independent of the state and is modeled by $vM(0, \nu_w)$. To estimate the state, the state $\hat{\theta}_t$ can be observed by the observation model

$$o_t = \theta_t + \nu_t,$$

where $\nu_t$ is the observation noise modeled by $vM(0, \nu_v)$. Based on the principle in Kalman filtering, we can use $\hat{\theta}_t$, modeled by $vM(\mu_t, \kappa_t)$, to estimate $\hat{\theta}_t$, and update $\hat{\theta}_t$.
Recursive by the input $u_t$ and observation $o_t$. For time update, given the current estimator $\hat{\theta}_t$, the updated estimator $\hat{\theta}_{t+1}$ is the sum of two independent von Mises random variables [18]. Let $\theta_1$ and $\theta_2$ be independently distributed as $vM(\mu_1, \kappa_1)$ and $vM(\mu_2, \kappa_2)$, respectively. The probability density function of $\theta = \theta_1 + \theta_2$ is given by

$$h(\theta) = \frac{1}{4\pi^2 I_0(\kappa_1) I_0(\kappa_2)} \int_0^{2\pi} e^{\kappa_1 \cos(\xi-\mu_1) + \kappa_2 \cos(\theta-\xi-\mu_2)} d\xi,$$

which is not a von Mises distribution. However, $h(\theta)$ can be approximated by a von Mises distribution $vM(\mu_1 + \mu_2, A^{-1}(A(\kappa_1)A(\kappa_2)))$, where

$$A(\kappa) = \frac{I_1(\kappa)}{I_0(\kappa)}.$$

Figure 2 depicts the numerical value of $A(\kappa)$, while the following theorem states the effect of time update due to the properties of $A(\kappa)$.

**Theorem 3.** For $\kappa_1, \kappa_2 > 0$,

$$A^{-1}(A(\kappa_1)A(\kappa_2)) < \min(\kappa_1, \kappa_2).$$

The proof of Theorem 3 is presented in Appendix I. Theorem 3 states that the approximated von Mises distribution after time update has smaller concentration parameter, which conforms to the fact that time update introduces uncertainty in estimation. Consequently, in time update, $\hat{\theta}_{t+1}$ can be approximated by the distribution $vM(\mu_{t+1}, \kappa_{t+1})$, with $\mu_{t+1} = \mu_t + u$ and $\kappa_{t+1} = A^{-1}(A(\kappa_t)A(\kappa_w))$.

In observation update, the posterior probability density of the state estimate given the observation $o_t$ is given by

$$p(\theta_t = \theta | o_t) = \frac{p(o_t | \theta_t = \theta)p(\theta_t = \theta)}{p(o_t | \theta_t)},$$

where $p(o_t | \theta_t = \theta)$ follows $vM(\sigma - \theta, \kappa)$ and $p(\theta_t = \theta)$ follows $vM(\mu_t, \kappa_t)$. The time index $t^-$ represents the time instance before the observation. With the multiplication of von Mises distribution in the nominator, we have

$$p(\theta_t = \theta | o_t) \propto e^{\kappa_\nu \cos(\theta - \nu) + \kappa_- \cos(\theta - \mu_-)}.$$
respectively. The agent keeps the estimates \( \hat{\theta}_t, \hat{x}_t, \text{ and } \hat{y}_t \) to track the corresponding terms, where \( \hat{\theta}_t \) is a von Mises random variable with distribution \( vM(\hat{\theta}_t, \kappa_t) \) and \( \hat{x}_t, \hat{y}_t \) follow \( N(\bar{x}_t, \sigma_{\bar{x},t}^2) \) and \( N(\bar{y}_t, \sigma_{\bar{y},t}^2) \), respectively. The time update for \( \hat{\theta}_t \) is relatively easy, since we can approximate \( \hat{\omega} \Delta t \) by \( vM(0, \Delta t \kappa_t) \) and Algorithm 1 thus follows. However, in the time update of \( \hat{x}_t \), the exact distribution of \( v_1 + \hat{v} \) cos \( \hat{\theta}_t \) lays a burden on any practical application. Therefore, we characterize its first two moments of \( (v_1 + \hat{v}) \cos \theta_t \) for reasonable approximation instead. With the distribution of \( \hat{\theta}_t \) denoted by \( vM(\hat{\theta}_t, \kappa_t) \) and the properties of trigonometric moments of von Mises distributions listed in Section II, we have

\[
E[(v_1 + \hat{v}) \cos \hat{\theta}_t] = v_1 E[\cos \hat{\theta}_t] + E[\hat{v}] E[\cos \hat{\theta}_t] = v_1 A(\kappa_t) \cos \theta_t,
\]

and together with the law of total variance

\[
\text{Var}[(v_1 + \hat{v}) \cos \hat{\theta}_t] = E[\sigma_v^2 \cos \hat{\theta}_t] + \text{Var}(v_2 \cos \hat{\theta}_t) \\
\leq \sigma_v^2 + \sigma^2_v.
\]

Although the random variable \((v_1 + \hat{v}) \cos \hat{\theta}_t\) is not Gaussian, we proceed to approximate it as a Gaussian random variable with mean \( v_1 A(\kappa_t) \cos \hat{\theta}_t \) and variance assigned as the upper bound \( \sigma_v^2 + \sigma_v^2 \) to compensate the approximation discrepancy. As a consequence, we rewrite the time update model as

\[
x_{t+1} \approx x_t + v_1 \Delta t A(\kappa_t) \cos \theta_t + w_x,t, \\
y_{t+1} \approx y_t + v_1 \Delta t A(\kappa_t) \sin \theta_t + w_y,t,
\]

where both \( w_x,t \) and \( w_y,t \) follows \( N(0, \sigma^2_v + \sigma_v^2) \). The position estimates can now be updated by traditional Kalman filtering.

As for observation update, we consider the direct measurement model first:

\[
o_{\theta,t} = \theta_t + \nu_\theta, \\
o_{x,t} = x_t + \nu_x, \\
o_{y,t} = y_t + \nu_y,
\]

where the orientation observation noises \( \nu_\theta \) is modeled by \( vM(0, \kappa_{\nu_\theta}) \) and the position observation noise \( \nu_x \) by \( N(0, \sigma^2_x) \) and \( \nu_y \) by \( N(0, \sigma^2_y) \). The observation update for \( \hat{\theta}_t \) then simply follows the corresponding equation in Algorithm 1, while that for \( \hat{x}_t \) and \( \hat{y}_t \) can be done by conventional Kalman filtering.

### B. Circular Representation

While the orientation is naturally circular, the key of circular representation of position relies on a module containing several circular estimates as explained in Section III. Particularly, the circular representation of \( x \)-axis can be achieved by multiple phases \( \phi_{1,t}, ..., \phi_{M,t} \), where each phase is associated with certain spatial period \( \lambda_i \). The \( y \)-axis position can be represented by \( \psi_{1,t}, ..., \psi_{M,t} \). With the state fully represented in circular form, the estimates of the state, including \( \hat{\theta}_t, \phi_{1,t}, ..., \phi_{M,t}, \psi_{1,t}, ..., \psi_{M,t} \), now all follow von Mises distribution, whose update relies on Algorithm 1. In addition to the estimate update, the conversion from circular representation to Cartesian representation is also essential for further application that needs explicit position information.

As for the localization algorithm with circular representation, since the orientation estimate \( \hat{\theta}_t \) is identical to the mixture representation case, we only discuss the position estimates in the following. With angular velocity \( \omega_t \) and translational velocity \( v_1 \) as input in time update, we have the update equations for positions

\[
\phi_{i,t+1} = \phi_{i,t} + 2\pi \left( \frac{v_1 + \hat{v}}{\lambda_i} \right) \sin \theta_t \Delta t, \\
\psi_{i,t+1} = \psi_{i,t} + 2\pi \left( \frac{v_1 + \hat{v}}{\lambda_i} \right) \cos \theta_t \Delta t,
\]

As developed in the mixture representation case, with the characterization of first two moments of \( \hat{v} \cos \theta_t \), the time update equation can be rewritten in the form compatible with [8] as

\[
\phi_{i,t+1} = \phi_{i,t} + \nu_{\phi,t} + w_{\phi,t},
\]

where

\[
u_{\phi,t} = 2\pi \left( \frac{v_1 A(\kappa_t)}{\lambda_i} \cos \hat{\theta}_t \right) \Delta t
\]

and \( w_{\phi,t} \) follows \( vM(0, \kappa_{w_{\phi,t}}) \) with

\[
\kappa_{w_{\phi,t}} = \frac{\lambda_i^2}{4\pi^2 (\Delta t)^2 (\sigma^2_v + \sigma_v^2)}.
\]

The procedure can be applied for \( y \)-axis update of \( \hat{\psi}_{i,t} \) as well. Consequently, the update for the state estimates \( \theta_t, \phi_{1,t}, ..., \phi_{M,t}, \psi_{1,t}, ..., \psi_{M,t} \) in time update simply follow Algorithm 1 with respective parameters.

With the observation model in [14], the observation update for \( \hat{\theta}_t \) is direct. Those for \( \phi_{1,t}, ..., \phi_{M,t}, \psi_{1,t}, ..., \psi_{M,t} \) can also be realized with appropriate von Mises approximation. By expressing [14] in circular representation as

\[
o_{\phi_{i,t}} = \phi_{i,t} + \nu_{\phi_{i,t}},
\]

where \( \nu_{\phi_{i,t}} \) follows \( vM(0, \lambda_i^2/4\pi^2 \sigma_{\phi_{i,t}}^2) \), the updates of circular represented position updates simply follow. The exact procedure can be applied for \( y \)-axis position phases for certain.

We now turn to the conversion from the fully circular representation to conventional Cartesian coordinate, which is known as readout algorithm and discussed extensively in [19]. To begin with, we take \( x \)-axis for example. Since each estimate gives a most likely position with \( \lambda_i \) spatial period, we then pick the \( x \) estimate as the position that maximize the overall likelihood, as

\[
\hat{x}_t = \arg \max_{x \in C_x} \sum_{i=1}^{M} \kappa_i \cos \left( \frac{2\pi}{\lambda_i} x - \phi_{i,t} \right),
\]

where \( C_x \) indicates the coverage that the circular representation is valid. The accuracy of this translation not only depends on the number of estimates in a module \( M \) and the spatial periods \( \lambda_1, ..., \lambda_M \), but also on the representation accuracy of \( \phi_{i,t} \) and the algorithm to obtain the maximal argument in implementation, which is not complete yet to the best of our knowledge. Therefore, we leave the complete characterization of readout algorithm for further research but use only the straightforward implementation in this paper.
C. Bearing-and-Distance Measurement

For most localization systems in practice, agents hardly can observe their spatial states directly. To ensure the applicability, we develop the localization algorithm relying on bearing-and-distance measurement. We assume that a landmark with known position \((x_l, y_l)\) is present in the environment, and the agent can observe the relative bearing \(b_t\) and distance \(r_t\) of the landmark at time \(t\) as

\[
\begin{align*}
    b_t &= \tan^{-1}\left(\frac{y_l - y_t}{x_l - x_t}\right) - \bar{\theta}_t, \\
    r_t &= \sqrt{(x_l - x_t)^2 + (y_l - y_t)^2}.
\end{align*}
\]

(19)

In real observation, both quantities are perturbed by noises as

\[
\begin{align*}
    o_{b,t} &= b_t + \nu_b, \\
    o_{r,t} &= r_t + \nu_r.
\end{align*}
\]

(20)

For convenience, we model \(\nu_b\) as \(vM(0, \kappa_b)\) and \(\nu_r\) as \(N(0, \sigma_r^2)\), respectively.

The spatial states \(\theta, x, y\) can be reconstructed with the observation terms \(b, r\) and be related to direct observation measurement by

\[
\begin{align*}
    o_{\theta,t} &= \tan^{-1}\left(\frac{y_l - y_t}{x_l - x_t}\right) - o_{b,t} = \bar{\theta}_t + \nu_{\theta,t}, \\
    o_{x,t} &= x_t - o_{r,t} \cos(\bar{\theta}_t + o_{b,t}) = x_t + \nu_{x,t}, \\
    o_{y,t} &= y_t - o_{r,t} \sin(\bar{\theta}_t + o_{b,t}) = y_t + \nu_{y,t}.
\end{align*}
\]

(21)

Equivalently, we have an direction observation of \(\theta\) constructed from \(\hat{x}_t, \hat{y}_t\), and \(o_{\theta,t}\). By quantifying the noise profile of this equivalent direct observation, we can apply the observation update developed previously for direct measurement. While \(\left[x_l - \hat{x}_t, y_l - \hat{y}_t\right]^T\) is a Gaussian random vector, \(\tan^{-1}\left(\frac{y_l - \hat{y}_t}{x_l - \hat{x}_t}\right)\) is thus a von Mises random variable by Theorem III. However, the exact covariance matrix of \(\left[x_l - \hat{x}_t, y_l - \hat{y}_t\right]^T\) is not attainable since the covariance between \(\hat{x}_t\) and \(\hat{y}_t\) is not tracked in the algorithm. To overcome this problem, we first note that \(\text{Var}(x_l - \hat{x}_t) = \sigma_x^2\) and \(\text{Var}(y_l - \hat{y}_t) = \sigma_y^2\). Furthermore, we have \(\sigma_x^2 = \sigma_y^2\) since they both follow the same update equation with identical parameters in the algorithm. As a consequence, we can choose \(2\sigma_x^2 I_2\) as the nominal covariance of \(\left[x_l - \hat{x}_t, y_l - \hat{y}_t\right]^T\), which is guaranteed to be no smaller than the real covariance matrix of \(\left[x_l - \hat{x}_t, y_l - \hat{y}_t\right]^T\) according to Theorem II. Therefore, we can therefore model \(\tan^{-1}\left(\frac{y_l - \hat{y}_t}{x_l - \hat{x}_t}\right)\) as \(vM(0, \kappa)\), \(\text{Var}(\bar{\theta}_t + o_{b,t}) = \nu_{\theta,t}\), where \(\bar{\theta}_t = \sqrt{(x_l - \hat{x}_t)^2 + (y_l - \hat{y}_t)^2}\), by Theorem III. Combining with \(o_{b,t}\), we have the condition from equivalent direct observation that

\[
\begin{align*}
    o_{\theta,t} &= \tan^{-1}\left(\frac{y_l - \hat{y}_t}{x_l - \hat{x}_t}\right) - o_{b,t},
\end{align*}
\]

with \(\nu_{\theta,t}\) characterized by \(vM(0, \kappa)\).

Similarly, we have equivalent direct position observation constructed from \(\bar{\theta}_t, o_{r,t}\) and \(o_{b,t}\). As \(\bar{\theta}_t + o_{b,t}\) approximated by \(vM(\bar{\theta}_t + o_{b,t}, A^{-1}(\kappa_1 A(\kappa_1)))\), we know \(\text{Var}(\bar{\theta}_t + o_{b,t})\) and that \(\text{Var}(\bar{\theta}_t + o_{b,t}) \leq \sigma^2 + \sigma^2_{r,t}\). Consequently, the observation gives the condition that

\[
\begin{align*}
    o_{x,t} &= x_t - o_r A(\kappa_1) A(\kappa_2) \cos(\bar{\theta}_t + o_{b,t})
\end{align*}
\]

accompanied with the noise \(\nu_{x,t}\) characterized by \(N(0, (\sigma_r^2 + \sigma^2_{r,t})^{-1})\), which can update \(\hat{x}_t\) by the direct observation procedure. The estimate \(\hat{y}_t\) can be updated with similar arguments. Also, the application to the algorithm based on fully circular representation is direct, which is omitted here.

While most of the localization algorithms with bearing-and-distance measurement are based on EKF implementation, the Jacobian matrix is calculated in updates. However, the performance of these algorithms is undermined by the linearization inconsistency, as discussed in [9–11]. On the contrary, the proposed algorithms here rely on the approximation between von Mises and Gaussian distributions, which is analyzable and accurate as presented in the following section.

VI. SIMULATION

A. Direct Measurement

We consider an artificial agent that performs time update at 50 Hz and observation update at 2.5 Hz. For time update, constant odometry inputs are applied with \(\omega_l = 0.2\) rad/s and \(v_l = 0.1\) m/s, and the noise parameters \(\kappa_{\omega \Delta t} = 500\) and \(\sigma_\omega = 0.01\) m/s. As for observation update, the parameters of noise are set to \(\kappa_{\nu x} = 500\) and \(\sigma_{o,x} = \sigma_{o,y} = 0.01\) m. The agent uses both mixture and circular representations to estimate its spatial states concurrently. For circular representation, we consider
Inspired by the grid representation in mammalian brains, we establish the estimation foundation for circular estimates for orientation estimation error in particular. However, the property of the proposed algorithms is demonstrated in this paper as the foundation to borrow the biological principles to improve navigation algorithms. With the accomplishment of localization as the first step, we believe that it is promising to revisit other navigation tasks in robotics, especially SLAM, with inspiration from mammalian brains.

B. Bearing-and-Range Measurement

We simulate the proposed algorithms with bearing-and-distance measurement against the conventional EKF localization algorithms. From simulation result, the property of the proposed algorithms, which is especially obvious in Fig. 3 not only shows the effectiveness of the proposed algorithm, it also shows the conversion between circular and Cartesian representation introduces negligible error.

APPENDIX I: PROPERTIES OF $A(\kappa)$

Lemma 2 (Modified Bessel functions [20]). For real $p \geq 0$,

$$-p + \sqrt{p^2 + \kappa^2} \leq I_p(\kappa) < \frac{I_p(\kappa)}{I_{p-1}(\kappa)}.$$  

(22)

Also, for $p \geq 1/2$, the inequality $I_p(\kappa)/I_{p-1}(\kappa) < 1$ holds.

Theorem 3 is a direct result of the properties of $A(\kappa)$, which are presented in the following lemma.

Lemma 3. For $\kappa > 0$,

1) $0 < A(\kappa) < 1$,
2) $0 < A'(\kappa)$.

Proof. By Lemma 2 we have $A(\kappa) < 1$ and for $\kappa > 0$,

$$A(\kappa) > \frac{\sqrt{\kappa^2 + 1} - 1}{\kappa} > 0,$$  

(23)

which gives the lower bound of $A(\kappa)$. In [12], we have the recurrent equation

$$A'(\kappa) = 1 - A(\kappa) \left( A(\kappa) + \frac{1}{\kappa} \right).$$  

(24)

Together with (23), $A'(\kappa) > 0$ is proved. $\Box$

The results from Lemma 3 state that $A(\kappa_1)A(\kappa_2) < A(\min(\kappa_1, \kappa_2))$, which leads to Theorem 3 directly.

APPENDIX II: PROPERTIES OF POSITIVE DEFINITE MATRICES

Lemma 4. Suppose that $P > 0$ and

$$P = \begin{bmatrix} A_{m \times m} & B \\ B^T & C_{n \times n} \end{bmatrix},$$

then for $c \in (0, 1)$

$$P < P' = \begin{bmatrix} \frac{1-c}{c} A & 0 \\ 0 & \frac{1-c}{c} C \end{bmatrix}.$$  

Proof. Consider a nonzero vector $v^T = [x^T, y^T]$ where $x \in \mathbb{R}^n$ and $y \in \mathbb{R}^m$. We can construct $v^T = [(1-c)x^T, cy^T]$. Since $P > 0$, $v^TPv > 0$, or $u^TPu > 0$, where

$$Q = \begin{bmatrix} (1-c)^2A & (1-c)B \\ c(1-c)B^T & c^2C \end{bmatrix} > 0.$$  

The result then follows by

$$P' - P = \begin{bmatrix} \frac{1-c}{c^2} A & -B \\ -B^T & \frac{1-c}{c^2} C \end{bmatrix}$$

$$= \frac{1}{c(1-c)} \begin{bmatrix} I_m & 0 \\ 0 & -I_n \end{bmatrix} Q \begin{bmatrix} I_m & 0 \\ 0 & -I_n \end{bmatrix} > 0.$$  

(25)

$\Box$
REFERENCES

[1] A. Banino, C. Barry, B. Uria, C. Blundell, T. Lillicrap, P. Mirowski, A. Pritzel, M. J. Chadwick, T. Degris, J. Modayil, G. Wayne, H. Soyer, F. Viola, B. Zhang, R. Goroshin, N. Rabinowitz, R. Pascanu, C. Beattie, S. Petersen, A. Sadik, S. Gaffney, H. King, K. Kavukcuoglu, D. Hassabis, R. Hadsell, and D. Kumaran, “Vector-based navigation using grid-like representations in artificial agents,” Nature, vol. 557, no. 7705, pp. 429–433, May 2018.

[2] D. Hassabis, D. Kumaran, C. Summerfield, and M. Botvinick, “Neuroscience-inspired artificial intelligence,” Neuron, vol. 95, no. 2, pp. 245–258, Jul. 2017.

[3] T. Hafting, M. Fyhn, S. Molden, M.-B. Moser, and E. I. Moser, “Microstructure of a spatial map in the entorhinal cortex,” Nature, vol. 436, no. 7052, pp. 801–806, Aug. 2005.

[4] B. L. McNaughton, F. P. Battaglia, O. Jensen, E. I. Moser, and M.-B. Moser, “Path integration and the neural basis of the ‘cognitive map’,” Nature Reviews Neuroscience, vol. 7, no. 8, pp. 663–678, Aug. 2006.

[5] I. R. Fiete, Y. Burak, and T. Brookings, “What grid cells convey about rat location,” Journal of Neuroscience, vol. 28, no. 27, pp. 6858–6871, Jul. 2008.

[6] A. Guanella, D. Kiper, and P. Verschure, “A model of grid cells based on a twisted torus topology,” International Journal of Neural Systems, vol. 17, no. 04, pp. 231–240, Aug. 2007.

[7] H. Sanders, C. Rennó-Costa, M. Idiart, and J. Lisman, “Grid cells and place cells: An integrated view of their navigational and memory function,” Trends in Neurosciences, vol. 38, no. 12, pp. 763–775, Dec. 2015.

[8] M. J. Milford and G. F. Wyeth, “Mapping a suburb with a single camera using a biologically inspired SLAM system,” IEEE Transactions on Robotics, vol. 24, no. 5, pp. 1038–1053, Oct. 2008.

[9] S. J. Julier and J. K. Uhlmann, “A counter example to the theory of simultaneous localization and map building,” in Proceedings of the IEEE International Conference on Robotics and Automation, vol. 4, May 2001, pp. 4238–4243.

[10] T. Bailey, J. Nieto, J. Guivant, M. Stevens, and E. Nebot, “Consistency of the EKF-SLAM algorithm,” in 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems, Oct. 2006, pp. 3562–3568.

[11] S. Huang and G. Dissanayake, “Convergence and consistency analysis for extended Kalman filter based SLAM,” IEEE Transactions on Robotics, vol. 23, no. 5, pp. 1036–1049, Oct. 2007.

[12] K. V. Mardia and P. E. Jupp, Directional Statistics. New York: Wiley, 1999.

[13] M. C. Jones and A. Pewsey, “A family of symmetric distributions on the circle,” Journal of the American Statistical Association, vol. 100, no. 472, pp. 1422–1428, 2005.

[14] A. Mathis, A. V. M. Herz, and M. B. Stemmler, “Multiscale codes in the nervous system: The problem of noise correlations and the ambiguity of periodic scales,” Physical Review E, vol. 88, no. 2, p. 022713, Aug. 2013.

[15] T. P. Trappenberg, Fundamentals of Computational Neuroscience. New York: Oxford University Press, 2002.

[16] X.-X. Wei, J. Prentice, and V. Balasubramanian, “A principle of economy predicts the functional architecture of grid cells,” eLife, vol. 4, p. e08362, Sep. 2015.

[17] M. Stemmler, A. Mathis, and A. V. M. Herz, “Connecting multiple spatial scales to decode the population activity of grid cells,” Science Advances, vol. 1, no. 11, p. e1500816, Dec. 2015.

[18] I. Markovi and I. Petrovi, “Bearing-only tracking with a mixture of von Mises distributions,” in 2012 IEEE/RSJ International Conference on Intelligent Robots and Systems, Oct. 2012, pp. 707–712.

[19] D. Bush, C. Barry, D. Manson, and N. Burgess, “Using grid cells for navigation,” Neuron, vol. 87, no. 3, pp. 507–520, Aug. 2015.

[20] A. Laforgia and P. Natalini, “Some inequalities for modified Bessel functions,” Journal of Inequalities and Applications, vol. 2010, no. 1, p. 253035, Dec. 2010.