Migrating Monarch Butterfly Localization Using Multi-Modal Sensor Fusion Neural Networks

Mingyu Yang*, Roger Hsiao*, Gordy Carichner*, Katherine Ernst*, Jaechan Lim†, Delbert A. Green II†, Inhee Lee†, David Blaauw*, and Hun-Seok Kim*

* University of Michigan, Ann Arbor, MI, USA & † University of Pittsburgh, Pittsburgh, PA, USA

Abstract—Details of Monarch butterfly migration from the U.S. to Mexico remain a mystery due to lack of a proper localization technology to accurately localize and track butterfly migration. In this paper, we propose a deep learning based butterfly localization algorithm that can estimate a butterfly’s daily location by analyzing a light and temperature sensor data log continuously obtained from an ultra-low power, millimeter (mm)-scale sensor attached to the butterfly. To train and test the proposed neural network based multi-modal sensor fusion localization algorithm, we collected over 1500 days of real world sensor measurement data by 82 volunteers all over the U.S. The proposed algorithm exhibits a mean absolute error of $< 1.7^\circ$ in latitude and $< 0.6^\circ$ in longitude Earth coordinate, satisfying our target goal for the Monarch butterfly migration study.

Index Terms—light-level geolocation, Monarch migration, neural networks, maximum likelihood estimation

I. INTRODUCTION

Each fall, millions of Monarch butterflies across central and eastern U.S. and southern Canada migrate up to 2,500 miles to overwinter in the same location in central Mexico. In spring these migrants mate and remigrate northwards to repopulate their northern breeding territory over 3 – 5 partially overlapping generations. Because no migrant Monarch lives long enough to make a return trip to the overwintering site, this navigational task cannot be learned and must be a genetically-encoded spatiotemporal program.

At present, only the largest animal migrants can be tracked continuously for significant portions of their migratory journey (e.g., [1]). Monarch butterflies, as small insects, cannot be tracked using the same strategy due to the weight and power constraints for mounted devices. A recent effort tracked Monarchs and green darner dragonflies up to hundreds of miles using the Motus Wildlife Tracking System [2]. While a substantial advance, this method has several limitations, such as unacceptable tracker weight, excessive power consumption, and very limited coverage of the Monarch migration territory.

While the global positioning system (GPS) is the most conventional method for determining locations, the smallest commercial GPS solution [3] has a total weight of 1 gram and size of 5cm, which is vastly too heavy and large for butterflies to carry. As an alternate to GPS, we propose to use daylight and temperature logging that are able to be integrated into the Michigan Micro Mote (M3) platform [4]–[6], which has a potential of weight less than 50 milligram and a size of 8×8×2.6 mm$^3$. Its duty-cycled operation is sustained by solar energy harvesting and it supports 50 meter distance wireless readout at the Monarch overwintering site. A new M3 platform customized for Monarch butterfly mounting is currently under manufacturing and evaluation.

This paper introduces a deep neural network based Monarch butterfly localization algorithm that utilizes the light intensity and temperature measurement data logged on the M3 platform. The proposed algorithm will be performed offline to analyze the migration trajectory when the log data is wirelessly retrieved from the butterfly at the overwintering site (it is possible because all butterflies migrate to the same site). To train and evaluate the proposed neural networks before deploying the final M3 system, we have conducted a data measurement campaign with 82 volunteers across the U.S. to record solar light intensity (in Lux) and temperature (in Celsius) using commercial HOBO sensors [7] as an emulator of the final M3 platform for the duration of Monarch fall migration in 2018.

II. RELATED WORK

Light intensity based localization has been applied to tracking marine animals that remain submerged and out of the reach of GPS [8]–[11]. Prior attempts tried to explicitly estimate the sunrise and sunset time from recorded light intensity curve. Then longitude and latitude coordinates are determined based on the estimated day center and day length respectively using standard astronomical equations [12]–[15]. However, it has the fundamental limitation of large latitude ambiguity around the equinox days (September 22 and March 20) when the day length is globally the same regardless of latitude. A second main challenge is the significant local light intensity variation due to weather and terrain factors that an ideal sunlight intensity model is unable to capture. In addition to light sensing modality, prior works [16], [17] augmented sea-surface temperature to improve the accuracy of tracking large sea animals for GPS-failing conditions (i.e., underseas). These approaches first construct a sea-surface temperature contour map based on satellite data and then localize the sensor by finding the position on the map where the temperature matches to the sensor reading. Although it is effective for sea animal tracking, the same method is not directly applicable to on-
above-ground butterfly localization where the temperature is significantly affected by the local terrain and weather.

III. PROPOSED METHOD

A. Overview

For the butterfly daily localization problem, the relationship between the measurements (light intensity and temperature), locations and time can be modelled by an observation model, which can be expressed as $[I_D, t_D]^T = g(x_D) + n_D$ where $D$ is the day index (with a unit of a day), $x_D$ denotes the state vector that represents the latitude and longitude coordinate of the butterfly on the day $D$, and $I_D$ and $t_D$ denotes the discrete sequence of light intensity and temperature sensor data measured on the day $D$, respectively. $n_D$ represents the sequence of process noise and observation noise, respectively and $g$ is the observation function relating the state vector to measurements.

In this paper, we mainly focus on the likelihood $p(I_D, t_D|x_D)$ to localize the butterfly’s daily position. As it is practically infeasible to find an exact expression of $g$ for Monarch butterfly migration, we rely on deep neural networks to learn the implicit observation model based on real world data. Then, we treat the output of the neural network as the estimation of likelihood, which can either be used for direct localization through maximum likelihood estimation (MLE), or can be combined with adaptive filtering/smoothing techniques (e.g., particle filtering/smoothing). Because light intensity and temperature measurements have different properties as discussed in Section III.C, we apply two distinct neural networks to learn their likelihoods separately. That is, $p(I_D|x) \approx \Phi_l(I_D, \tilde{x})$ and $p(t_D|x) \approx \Phi_D(t_D, \tilde{x})$ where $\Phi_l$ and $\Phi_t$ denote the two neural networks, $\tilde{x}$ denotes an arbitrary location. Then, with a simplifying assumption that light intensity and temperature measurement are conditionally independent given the state vector, we have $p(I_D, t_D|x) = p(I_D|x)p(t_D|x) \approx \Phi_l(I_D, \tilde{x})\Phi_t(t_D, \tilde{x})$. We call $\Phi_l$ the light intensity discriminator and $\Phi_t$ the temperature discriminator.

B. Light Intensity Discriminator

In this section, we propose a light discriminator network to estimate $p(I_D|x)$. For the network input, we first define a reshape function $r$: $I_D = r(I_D, \tilde{x}, D)$ where $I_D$ is the normalized light intensity data obtained by shifting and resampling the original $I_D$ based on the coordinate state $\tilde{x}$ and the date information $D$ so that the night center is located at the center and the length of the night is scaled to 12 hours as depicted in Fig. 1 (blue curve). The reason we normalize the light intensity curve based on the night center and length instead of the day center and length is because Monarch butterflies are known to rest without changing the location during the night. In our problem, the date information $D$ is known for the light intensity data measurement $I_D$ but the true coordinate state $x$ is unknown. Therefore, the normalized light intensity output $\hat{I}_D = r(I_D, \tilde{x}, D)$ becomes symmetric around the time reference (time of 24:00) with 12-hour night length only if a correct coordinate state $\tilde{x}$ is used in the normalization function as illustrated in Fig. 1. The objective of the light discriminator network $\Phi_l$ is to discriminate a correct vs. incorrect coordinate state $\tilde{x}$ given $\hat{I}_D$, $I_D$, and $D$.

The input to the light discriminator network is $I_D$ in log scale reshaped from the observation $I_D$ based on a state candidate $\tilde{x}$ given the measurement date information $D$. The neural network is trained to discriminate (i.e., binary classification) whether $\tilde{x}$ matches the true measurement location or not by observing the shape of $I_D$ normalized based on $\tilde{x}$. To generate the training dataset for this discriminator network, we use both matched and mismatched pairs of $(I_D, \tilde{x})$. The final activation of the discriminator network is the sigmoid function. Hence the output $\Phi_l(I_D, \tilde{x})$ can be interpreted as the likelihood probability $p(I_D|x)$. The length of $I_D$ as the input of the discriminator is set to 8 hours (±2 hours around the expected sunrise and sunset time) as shown in Fig. 1. Longer window length increases the complexity of the neural network unnecessarily without improving the discriminator accuracy as it mostly rely on the light intensity data shape near the sunrise and sunset time for binary classification.

Data around the night center are excluded from the discriminator input as the light measurement around midnight is mostly noise.

C. Temperature Discriminator

The temperature discriminator network is designed to estimate $p(t|x)$. However, unlike the light intensity model that can be trained using a location dependent normalization function $r(I_D, \tilde{x}, D)$, the temperature data does not significantly depend on the longitude location coordinate while it is significantly affected by the local weather. Thus, we utilize the weather station data as a reference and train the discriminator to compare the two inputs; the temperature measurement data from the sensor and the weather station measurement data at a particular location on the same day. It produces the binary
classification result; matched or mismatched depending on whether the location of the weather station is closest to \( \tilde{x} \) or not. The final sigmoid function of the discriminator network quantifies the temperature data pattern similarity between the sensor and weather station data at a particular location \( \tilde{x} \) and date \( D \) estimating \( \Phi_t(t_D, \tilde{x}) \approx p(t_D | \tilde{x}) \). We restrict the length of the temperature discriminator input \( t \) to be ±8 hours around the night center because it is the time when butterflies rest without moving.

IV. EXPERIMENTS

A. Hardware & Data Collection

We used HOBO sensors [7] as an emulator of the final M3 platform to record the light intensity and temperature data. To collect the real world data, we disseminated HOBO sensors to 82 volunteers in the U.S. This volunteer data contain measurements of 1625 days with a time resolution of 10 sec for light intensity and 15 sec for temperature from 1st September to 19th December in 2018. The volunteer sensor placement locations are shown in the map (Fig. 2) of the U.S. We access the night temperature weather station data with time resolution of 1 hour using WeatherBit API [18]. The time offset and resampling factor for the light intensity reshaping function \( \hat{l}_D = r(l_D, \tilde{x}, D) \) are obtained by the astronomical equation MATLAB function [19], which calculates the sunrise and sunset time (which are converted to the night center time and night length) for a given coordinate \( \tilde{x} \) on the day \( D \).

B. Data Processing and Preparation

1625 days of sensor measurement data are divided into 1300 training and 325 testing data. The light intensity data are down-sampled to 1 minute interval and then the intensity is converted to log scale to magnify the change in low light level (near sunrise and sunset) while suppressing the temporal variation (due to shadow) in bright conditions. Since the WeatherBit weather station temperature data has time resolution of 1 hour, we also down-sample sensor temperature data to 1 hour interval to match the sampling rate.

For each of the 1300 training light intensity data \( l_D \), we prepare one matched pair \((\hat{l}_D, \tilde{x})\) and 24 mismatched pairs by applying random night center and night length offset in the range of 4 – 120 minutes. We end up with a training set of size 32500 in which 1300 entries are labeled Class 1 (match) and 31200 entries are Class 0 (mismatch) for binary classification. A similar process is applied to generate the training dataset for the temperature discriminator. For each of 1300 temperature measurements \( t_D \), we prepare one matched pair using the weather station data at the nearest location to label it Class 1 (matched). In addition, we create 15 mismatched pairs with the Class 0 label for each sensor measurement \( t_D \) using the weather station data randomly selected from 15 different locations within the range of ±20 degrees in both latitude and longitude around the ground-truth location. When there is no weather station data in the vicinity of a random position, we simply treat it as an outlier without adding it to the training dataset. This approach leads to 17198 Class 0 data and 1300 Class 1 data in total.

C. Network structure and training

The network structures for the proposed discriminators are shown in Fig. 3. The light intensity discriminator contains 4 convolution layers (conv - batch normalization - ReLU - max pooling) and 3 fully connected layers. The size of each layer is specified in Fig. 3. The temperature discriminator network shown on the right in Fig. 3 only contains three fully connected layer and a dropout layer with \( p = 0.25 \) placed after the first fully connected layer. These proposed networks were found after testing various network hyper-parameters to enhance the performance. Since the size of Class 0 dataset is much larger than that of Class 1 for both networks, we adopt a weighted sampling technique that samples the dataset unevenly so that the two classes are equally probable for each batch. Both discriminators are trained with the ADAM optimizer with betas = (0.9, 0.999) and a learning rate of \( 5 \times 10^{-4} \). The light discriminator is trained for 50 epochs while the temperature discriminator is trained for 200
epochs. We randomly selected 260 days of measurements to generate matched and mismatched pairs as our validation set. Considering the fact that our classification problem is highly imbalanced between Class 0 and 1, we searches for a network that produces the highest $F_{0.5}$ score instead of the highest validation accuracy, where $F_{\beta}$ score is defined as:

$$F_{\beta} = \frac{\text{recall} \cdot \text{precision}}{\beta^2 \cdot \text{precision} + \text{recall}} \quad (1)$$

Here, we choose $\beta = 0.5$ to prioritize precision because a higher precision score implies that our network eliminates false positives more aggressively producing high likelihood scores only for fewer candidate coordinate $\hat{x}$’s.

D. Results

The proposed neural network based likelihood estimation ($\Phi_l(D, \hat{x})$ and $\Phi_t(D, \hat{x})$) is performed by collecting the neural network output for each test data $I_D$ and $t_D$ evaluated at various coordinates $\hat{x}$ in a grid surrounding the ground-truth sensor location with a range of $[-10, 10]$ degrees in latitude and longitude. The $\hat{x}$ grid resolution for the initial (coarse) likelihood evaluation of $\Phi_l(D, \hat{x})$ and $\Phi_t(t_D, \hat{x})$ is 1 degree for both longitude and latitude. The spatial resolution of the likelihood estimation is refined to 0.1 degree by upsampling and interpolating the coarse evaluation results.

Three example likelihood estimations on three different days are shown in Fig. 4 where red and blue color corresponds to high and low likelihood, respectively. The plots on the left, center, and right column show the (interpolated) neural network output $\Phi_l(I_D, \hat{x})$, $\Phi_t(t_D, \hat{x})$, and the product $\Phi_l(I_D, \hat{x})\Phi_t(t_D, \hat{x})$, respectively, for randomly selected sensor data instances $(I_D, t_D, \hat{x})$ while the ground-truth sensor locations $x_D$ are shifted to (0,0) for plotting. The sensor measurement data on the row (a), (b), and (c) were collected on the date of Sep. 28, Oct. 15, and Dec. 4, respectively. We observed that the light discriminator neural network mostly relies on the night length information to estimate the latitude while it uses the night center time information to estimate the longitude. Thus, when the night length is globally the same regardless of the coordinate around the equinox day (row (a) on Sep. 28), the light discriminator network fails to estimate the latitude and it produces a pattern of $\Phi_l(I_D, \hat{x})$ spread out along the latitude as shown on the top left of Fig. 4.

While the light discriminator has high ambiguity in latitude around the equinox and maintains low ambiguity in longitude, the opposite is true for the temperature discriminator as temperature varies significantly along latitude but less so along longitude as shown in Fig. 4 middle column. Therefore, light and temperature discriminator networks uniquely complement each other, resulting in significant accuracy improvement. When two neural network outputs are multiplied, it provides more reliable results to estimate the likelihood $\Phi_l(I_D, \hat{x})\Phi_t(t_D, \hat{x}) \approx p(I_D, t_D|\hat{x})$ as shown in Fig. 4 right column. In general, the output has smaller error in December (Fig. 4 row (c)) when both the light and temperature discriminators work reliably due to significant variations in night length, night center, and temperature across latitude and/or longitude.

Finally, we perform maximum likelihood estimation based localization to evaluate accuracy of the proposed method. We also measure false negative rates to quantify the reliability of the proposed likelihood estimator. Here, we compare the proposed approach to a baseline where the sunrise and sunset time...
are estimated by comparing the light intensity to a threshold calibrated for the minimum error. An optimal threshold value for the baseline method was found by exhaustively searching for the value that minimizes the night length and night center error.

The average longitude and latitude localization errors for different time intervals are shown in Fig. 5. All methods (threshold based baseline, light discriminator only, light and temperature discriminator combined) have similar performance in longitude estimation exhibiting less than 0.6° average absolute error for all periods. For latitude, our light discriminator significantly outperforms the baseline method for all periods. By combining the likelihood estimation from the light and temperature networks, the average error in latitude reduces dramatically from 11° to 1.5° around the fall equinox and remains under 0.7° over November and December.

Fig. 6 shows the false negative rate which is defined as the proportion of the absolute latitude localization error greater than 2°. In September (around the fall equinox), both the baseline method and our light discriminator produce > 80% false negatives while the combined estimation reduces it to around 20%. The light discriminator starts to significantly outperform the baseline from October and the proposed combined method exhibits only 1.23% false negative rate in November.

Fig. 7 shows the localization results in the Southwest U.S, in November and December using the proposed method, which demonstrates accuracy of 55.11km error on average. Most estimations are localized within the same city (Kansas City, Dallas, etc.) where actual measurements were made.

V. CONCLUSION

We present a neural network based butterfly localization algorithm that learns the observation model implicitly. The proposed method is applicable to ultra-low power ultra-small light and temperature sensors that can be attached to Monarch butterflies without impeding their migration. The maximum likelihood localization confirms that neural networks can learn implicit observation models to outperform traditional thresholding method. Testing results exhibit an average error under 1.7° in latitude and 0.6° in longitude, which is sufficient to study Monarch migration. We will continue collecting more volunteer measurements to improve the robustness of the neural networks.

REFERENCES