Travel Time-dependent Maximum Entropy Inverse Reinforcement Learning for Seabird Trajectory Prediction

Tsubasa Hirakawa*, Takayoshi Yamashita*, Ken Yoda†, Toru Tamaki‡, Hironobu Fujiyoshi*
Chubu University* Nagoya University† Hiroshima University‡

***Abstract*—Trajectory prediction is a challenging problem in the fields of computer vision, robotics, and machine learning, and a number of methods for trajectory prediction have been proposed. Most methods generate trajectories that move toward a goal in a straight line (goal-directed) while avoiding obstacles. However, there are not only such goal-directed trajectories but also trajectories that taking detours to reach the goal (non-goal-directed). In this paper, we propose a method of predicting such non-goal-directed trajectories based on the maximum entropy inverse reinforcement learning framework. Our method introduces travel time as a state of the Markov decision process. As a practical example, we apply the proposed method to seabird trajectories measured using global positioning system loggers. Experimental results show that the proposed method can effectively predict non-goal-directed trajectories.**

***Keywords*-Trajectory Prediction; Maximum Entropy Inverse Reinforcement Learning; Markov Decision Process; Animal Behavior Analysis**

I. Introduction

Trajectory prediction (path prediction), in which the task is to predict a sequence of future actions or the remaining parts of a trajectory, has a large number of applications and has received much attention in the fields of computer vision, robotics, and machine learning [1], [2], [3], [4], [5], [6]. Due to the diversity of applicable fields, many studies [7], [8], [9], [10], [11] have been conducted in recent years.

Most prediction methods are focused on trajectories of individuals (pedestrians, cyclists, and even vehicles). Typically, these individuals tend to move toward their internal goal in a linear manner while avoiding obstacles. Furthermore, people tend to move according to a certain social rule, e.g., prefer walking on sidewalks or hesitate to walk on the grass. People use such information as latent prior knowledge to decide their future trajectories.

Most trajectory-prediction methods introduce such information to provide an accurate prediction result. With these methods, the trajectory becomes like a straight line toward the goal. We call such a trajectory *goal-directed* in this paper. However, trajectories are not only goal-directed but also *non-goal-directed*, i.e., taking a detour to reach the goal. Current trajectory-prediction methods cannot treat a non-goal-directed trajectory appropriately.

In this paper, we propose a method of predicting a non-goal-directed trajectory. Our method is based on the maximum entropy inverse reinforcement learning (MaxEnt IRL) framework. Some trajectory-prediction methods based on this framework have been proposed [1], [2], [11], [3] and have successfully predicted long-term trajectories by introducing the physical environment and movement of other pedestrians. However, these methods predict only goal-directed trajectories while avoiding obstacles. We introduce travel time as a state of the Markov decision process (MDP) to predict non-goal-directed trajectories. Explicitly given travel time from the start state to the goal state, we can predict a trajectory while taking detour actions into account.

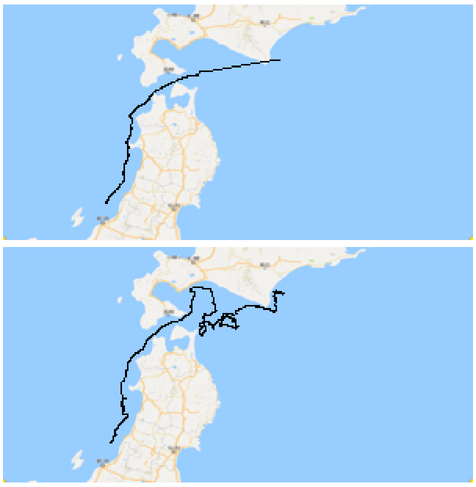

Figure 1. Two examples of shearwater trajectories [12]. Black lines show trajectories recorded using GPS loggers. Both trajectories start from same location at bottom-left and go to destination at top-right. In top image, shearwater moves directly toward goal by avoiding obstacles (land). In bottom image, shearwater does not move toward goal and takes indirect route.

Non-goal-directed trajectory prediction is valuable in the field of biology, ecology, and animal science *Bio-logging*. The use of data recorded using global positioning system (GPS) loggers and small cameras attached to an animals body or leg provides various information to understand the animals behavior, habits, and environments. Recent development of a small GPS logger [13], [14] enables us to collect trajectory data and apply learning-based methods. In particular, analyzing trajectories of seabirds benefits from such small GPS devices. Figure 1 shows examples of trajectories of streaked shearwaters (*Calonectris leucomelas*) [12]. The shearwaters start their flight from their breeding area on a small island (bottom-left of figures) and go to feed on fish over 1,000 km away from the breeding area (top-right of figures). It takes a few days to return the breeding area. However, two trajectories shown in Figure 1 pass through different paths although those head to the same goal. Because the GPS trajectories do not tell us their movements, predicting trajectories of shearwaters could be greatly helpful to reveal their behavior. Therefore, we applied the proposed method to predicting trajectories of shearwaters in our experiment.

Our contribution is two-fold. First, we introduce travel time in an MDP framework. The concept of time is implicitly included as an action step in the MDP. However, travel time has never explicitly been defined as a state. Second, to the best of our knowledge, this is the first attempt at applying a trajectory-prediction method to animal behavior.

II. Related work

Trajectory prediction is a challenging problem. To obtain reliable prediction results, various methods have been proposed based on the Kalman filter [7], dynamic Bayesian networks [9], optical flow [8], Markov Chain Monte Carlo (MCMC) [15], patch-based [16], [17], and social force model [18].

The most common methods use the recent developments in deep neural network frameworks. Alahi et al. [4] proposed a trajectory-prediction method based on long short-term memory (LSTM). They also proposed a pooling layer called the social pooling layer to represent the interaction between neighbor pedestrians. Fernando et al. introduced an attention-based LSTM encoder-decoder model [5] and proposed a network called the Tree Memory Network [19]. Yi et al. [6] proposed a convolutional neural network (CNN) framework to predict trajectories of multiple pedestrians. These methods have two drawbacks: the necessity of the past trajectory as an input for prediction and the short range of prediction.

Another common method is based on IRL, especially the MaxEnt IRL [20] framework. Kitani et al. [2] assume that a pedestrian's trajectories are decided due to the physical environment such as sidewalks, pavements, and vehicles. They model this concept as an MDP model and learned the optimal reward weight from demonstrated (training) trajectories. Ma et al. [3] extended this method to consider multiple-agent interactions. They introduced fictitious play to represent interactions between pedestrians and used attributes such as gender and age to consider walking speed. Other prediction methods using IRL have been proposed [1], [11], [21], [22]. Prediction methods based on the MaxEnt IRL framework has an advantage of predicting long-term trajectories. However, these methods cannot predict non-goal-directed trajectories because of the effect of negative rewards (details are given in Section III). Consequently, we also adopted the MaxEnt IRL framework in this study.

In contrast to the above methods, our method effectively predicts non-goal-directed trajectories. In this study, we compare the proposed method with that by Kitani et al. [2] using the shearwater trajectory dataset.

III. Travel-time-dependent MaxEnt IRL framework

We briefly introduce the basic properties of IRL and explain our method. An MDP [23], [24] can be defined as a tuple $(S, A, p(s_0), T, R)$, where $s \in S$ is a state, $a \in A$ is an action, $p(s_0)$ is an initial state distribution, $T = \{p(s'|s,a)\}$ is a set of state transition probabilities, and R is a reward function (or value). A trajectory ζ is defined as a sequence of state-action pairs, i.e., $\zeta = \{(s_0, a_0), (s_1, a_1), \dots\}$. Trajectories can be predicted by solving the MDP to maximize the reward value. In reinforcement learning, the reward values are known or should be defined.

The reward function is often not given or difficult to define manually. In this situation, it is more reasonable to estimate from training data. *Inverse reinforcement learning* [25] is a problem of estimate the optimal reward function from demonstrated (expert) data. There are several methods of estimating the reward function based on linear programming [26] and max-margin and projection methods [27]. Among those, we follow the MaxEnt IRL framework [20], as in [1], [2], [11], [3]. In the MaxEnt IRL framework, the reward function for a ζ is defined as

$$R(\zeta; \boldsymbol{\theta}) = \sum_t \boldsymbol{\theta}^{\mathrm{T}} \boldsymbol{f}(s_t), \tag{1}$$

where $\boldsymbol{\theta}$ is a weight vector. $\boldsymbol{f}(s_t)$ is a feature response vector observed at state s_t along ζ, that is given by feature maps (the example is shown in Figure 3). The MaxEnt IRL framework is aimed at estimating the optimal weight vector $\hat{\boldsymbol{\theta}}$ from demonstrated trajectories.

In the MaxEnt IRL framework, the distribution over a ζ is defined using MaxEnt distribution, which is defined as

$$\begin{aligned} p(\zeta; \boldsymbol{\theta}) &= \frac{\exp\left(\sum_t \boldsymbol{\theta}^{\mathrm{T}} \boldsymbol{f}(s_t)\right)}{Z(\boldsymbol{\theta})} \\ &\propto \exp\left(\sum_t \boldsymbol{\theta}^{\mathrm{T}} \boldsymbol{f}(s_t)\right), \end{aligned} \tag{2}$$

where $Z(\boldsymbol{\theta})$ is a normalization function. The principle of MaxEnt enables us to handle imperfect demonstrated trajectories. This distribution means that a trajectory with a higher reward value is more often chosen than a lower value trajectory.

The MaxEnt IRL framework predicts a trajectory with a higher reward value and uses a negative reward value. The total reward value over a trajectory is computed as an accumulation of negative rewards of each state, as shown in Eq. 1. Because detour actions decrease the reward value, current methods predict only goal-directed trajectories while avoiding obstacles. To address non-goal-directed actions, we introduce *travel time* from the start state to the goal state. We define a state as $s = [x, y, z]$, where x and y are positions in a two-dimensional (2D) plane, z is travel time elapsed since the initial state, and $a = [v_x, v_y, v_z]$ is an action. Explicitly given travel time, a non-goal-directed action can also be considered.

To estimate the optimal weight vector $\hat{\boldsymbol{\theta}}$, we use a gradient decent algorithm, as did Ziebart et al. [20].

IV. Experimental results

This section demonstrates the effectiveness of the proposed method by using a trajectory dataset of shearwaters.

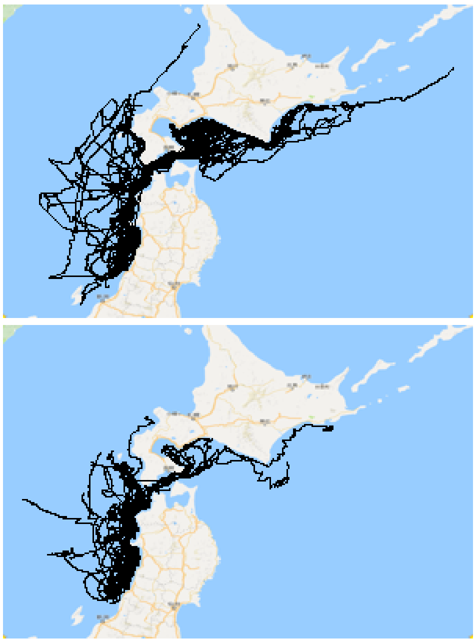

Figure 2. Trajectories of shearwaters [12]. Upper image shows trajectories of male shearwaters and bottom image shows trajectories of female shearwaters.

We give the details of the dataset and the prediction results.

A. Dataset

The shearwater-trajectory dataset consists of 106 trajectories (male: 53 and female: 53) [12]. Each trajectory was recorded using a GPS logger and had a series of longitude, latitude, and the corresponding travel time after leaving the nest. We defined the MDP state space as a 3D grid of $300 \times 200 \times 600$ to quantize the GPS trajectory data. Figure 2 shows all trajectories of the dataset. As we can see, male and female shearwaters seem to take different trajectories, that is, males go to distant goals and females go to nearby goals. This difference might be related to sex and/or body-weight differences [28], [29]. We used the male and female sub-datasets separately because mixing male and female trajectories would result in poor prediction performance.

We also generated feature maps for this dataset. As shown in Figure 2, shearwaters fly over the sea along coastlines. We therefore annotated physical attributes as *land* and *sea*. We generated additional features based on the physical attributes. We adopted exponentiated distance $d_{\text{exp}}(x, y)$ from each attribute, as done by Kitani et al. [2], which is defined as

$$d_{\text{exp}}(x, y) = \exp\left(\frac{-d_{\text{euc}}(x, y)}{\sigma^2}\right), \tag{3}$$

where $d_{\text{euc}}(x, y)$ is a Euclidean distance from an attribute to each state (x, y) in a 2D plane and σ^2 is a variance. In this experiment, we generated distance features with three variances $\sigma^2 = \{3, 5, 10\}$ with respect to three attributes: sea, land, and coastline. We used a total of 12 feature maps with an additional constant-value map. Note that each feature map is normalized in the range of $[-1, 0]$. The generated feature maps are shown in Figure 3.

Table I
MEAN NLL OF TRAJECTORY PREDICTION

Dataset	Baseline [2]	Proposed
Male	3.165 ± 0.581	1.916 ± 0.231
Female	3.699 ± 0.562	1.907 ± 0.118

As a baseline, we compared the method proposed by Kitani et al. [2]. In the experiment, we randomly selected 40 trajectories for training to estimate the optimal reward weight and the rest was used for testing.

B. Results

Figure 4 shows the predicted male trajectories. The baseline predicted trajectories that avoided crossing obstacles (land), while it failed to cover non-goal-directed trajectories. In particular, the trajectory on the left of Figure 4 seems to move randomly at the center of the figure. The proposed method could successfully cover such non-goal-directed trajectories.

Figure 5 shows the predicted female trajectories. We can see that the proposed method also provided reasonable prediction results. The baseline predicted trajectories that could not avoid crossing land As shown in Figure 2, female trajectories are relatively shorter than male trajectories. If we use such shorter trajectories for training, reward weight $\boldsymbol{\theta}$ cannot be estimated by considering physical attributes or feature maps. The suboptimal reward weight and property of goal-directed trajectory prediction would provide unreliable results. The proposed method successfully avoided crossing obstacles by adding travel time for increasing the possibility to take non-goal-directed trajectories.

As a quantitative evaluation of predicted trajectories, we used the negative log-loss (NLL), which is defined as

$$\text{NLL}(\zeta) = E_{\pi(a|s)}\left[-\log \prod_t \pi(a_t|s_t)\right]. \tag{4}$$

This is the expected log-likelihood of the demonstrated ζ under the predicted policy $\pi(a|s)$. Table I shows the mean NLL of both datasets. We can see that the proposed method performed better than the baseline.

V. Conclusion

We proposed a method of predicting non-goal-directed trajectories. Our method is based on the MaxEnt IRL framework with additional travel time as a goal state of the MDP, which enables us to handle non-goal-directed trajectories. In an experiment, we used a GPS trajectory dataset of shearwaters for evaluation. The experimental results indicate that the proposed method can effectively predict non-goal-directed trajectories.

The trajectory prediction of our method can be further improved based on the following aspects. The first aspect is considering a dynamic environment. Many studies were focused on the interaction between neighbor pedestrians and modeled it as a dynamic environment. However, a dynamic environment can be caused not only by the interaction but also in the physical scene. For instance,

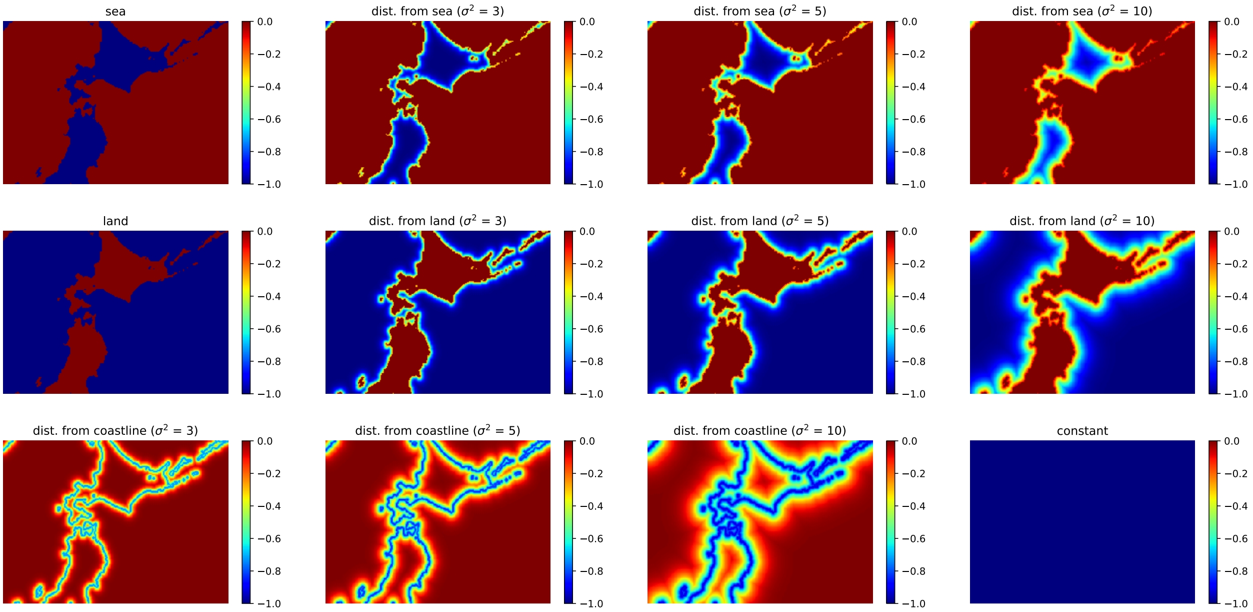
sea
dist. from sea (σ^2 = 3)
dist. from sea (σ^2 = 5)
dist. from sea (σ^2 = 10)
land
dist. from land (σ^2 = 3)
dist. from land (σ^2 = 5)
dist. from land (σ^2 = 10)
dist. from coastline (σ^2 = 3)
dist. from coastline (σ^2 = 5)
dist. from coastline (σ^2 = 10)
constant
0.0
−0.2
−0.4
−0.6
−0.8
−1.0

Figure 3. Feature maps for shearwater-trajectory dataset

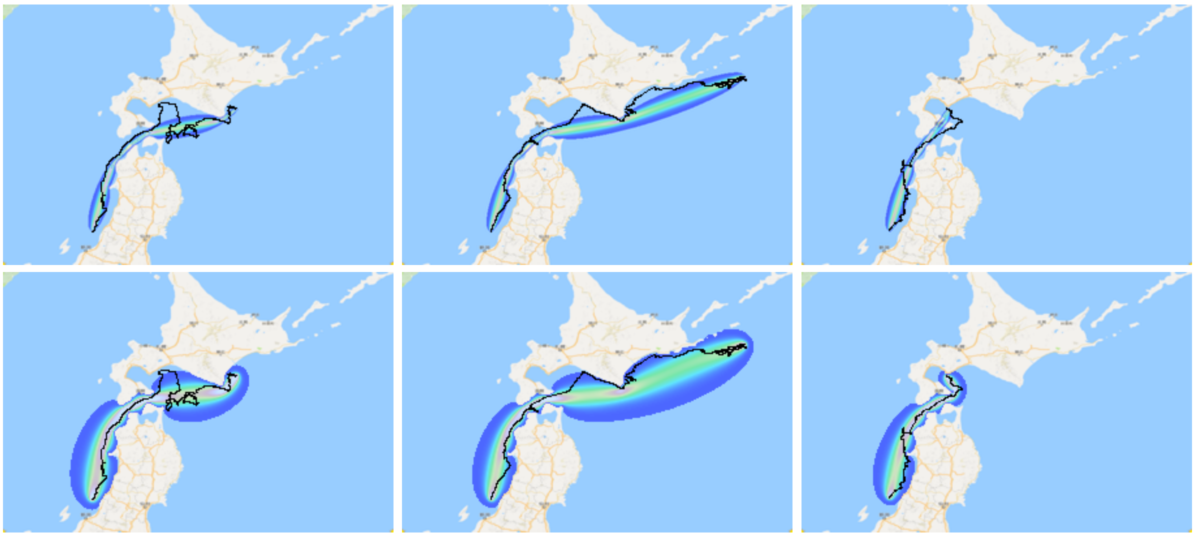

Figure 4. Predicted trajectories of male dataset. Upper row shows trajectories predicted with baseline [2] and bottom row shows those with proposed method. Each column shows results of same trajectory. Black lines show trajectories recorded using GPS logger. Predicted distributions are shown as heat map: higher probability is shown as warmer colors and lower probability is shown as cooler colors.

trajectories might be affected by weather and wind patterns in the case of shearwaters. Hence, introducing factors temporally changing over a scene is one of our future work. The second aspect is considering the attributes of an individual. Trajectories of shearwaters are affected by certain attributes, e.g., sex and/or age. Introducing attributes and predicting trajectories in a single framework is also our future work.

Acknowledgement

This work was supported in part by JSPS KAKENHI grant numbers JP16H06540, 16H06541, and 16K21735.

References

[1] B. D. Ziebart, N. Ratliff, G. Gallagher, C. Mertz, K. Peterson, J. A. Bagnell, M. Hebert, A. K. Dey, and S. Srinivasa, “Planning-based prediction for pedestrians,” in *The IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 3931–3936, Oct 2009.

[2] K. M. Kitani, B. D. Ziebart, J. A. Bagnell, and M. Hebert, “Activity forecasting,” in *European Conference on Computer Vision (ECCV)*, pp. 201–214, 2012.

[3] W.-C. Ma, D.-A. Huang, N. Lee, and K. M. Kitani, “Forecasting interactive dynamics of pedestrians with fictitious play,” in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

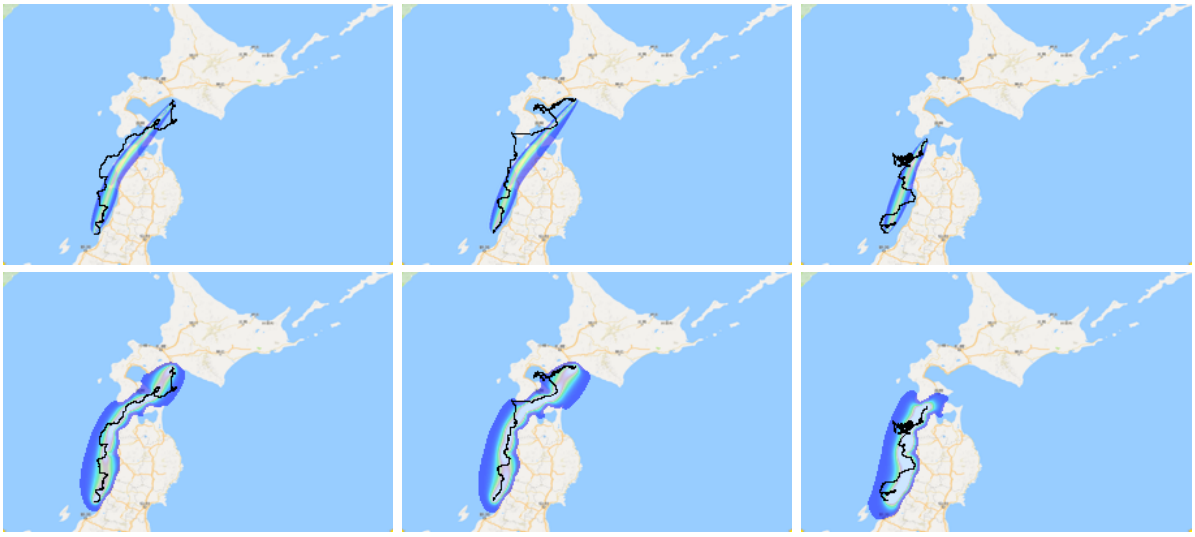

Figure 5. Predicted trajectories of female dataset. Upper row shows trajectories predicted with baseline [2] and bottom row shows those with proposed method. Each column shows results of same trajectory. Predicted distributions are shown as heat map: higher probability is shown as warmer colors and lower probability is shown as cooler colors.

[4] A. Alahi, K. Goel, V. Ramanathan, A. Robicquet, L. Fei-Fei, and S. Savarese, "Social lstm: Human trajectory prediction in crowded spaces," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[5] T. Fernando, S. Denman, S. Sridharan, and C. Fookes, "Soft + hardwired attention: An LSTM framework for human trajectory prediction and abnormal event detection," *CoRR*, vol. abs/1702.05552, 2017.

[6] S. Yi, H. Li, and X. Wang, "Pedestrian behavior understanding and prediction with deep neural networks," in *European Conference on Computer Vision (ECCV)*, pp. 263–279, 2016.

[7] N. Schneider and D. M. Gavrila, "Pedestrian path prediction with recursive bayesian filters: A comparative study," in *The 35th German Conference on Pattern Recognition (GCPR)*, pp. 174–183, 2013.

[8] C. G. Keller and D. M. Gavrila, "Will the pedestrian cross? a study on pedestrian path prediction," *IEEE Transactions on Intelligent Transportation Systems*, vol. 15, pp. 494–506, April 2014.

[9] J. F. P. Kooij, N. Schneider, F. Flohr, and D. M. Gavrila, "Context-based pedestrian path prediction," in *European Conference on Computer Vision (ECCV)*, pp. 618–633, 2014.

[10] E. Rehder and H. Kloeden, "Goal-directed pedestrian prediction," in *IEEE International Conference on Computer Vision (ICCV) Workshop*, pp. 139–147, Dec 2015.

[11] N. Lee and K. M. Kitani, "Predicting wide receiver trajectories in american football," in *The IEEE Winter Conference on Applications of Computer Vision (WACV)*, pp. 1–9, March 2016.

[12] S. Matsumoto, T. Yamamoto, M. Yamamoto, C. B. Zavalaga, and K. Yoda, "Sex-related differences in the foraging movement of streaked shearwaters calonectris leucomelas breeding on awashima island in the sea of japan," *Ornithological Science*, vol. 16, pp. 23–32, 2017/06/04 2017.

[13] H.-G. Hao, H.-X. Lu, W. Chen, and C. An, *A Novel Miniature Microstrip Antenna for GPS Applications*, pp. 139–147. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012.

[14] L. Wang, L. Deng, X. Xi, and Y. Du, "A miniature gps microstrip antenna," in *ISAPE2012*, pp. 250–252, Oct 2012.

[15] D. Xie, S. Todorovic, and S. C. Zhu, "Inferring dark matter and dark energy from videos," in *The IEEE International Conference on Computer Vision (ICCV)*, pp. 2224–2231, Dec 2013.

[16] L. Ballan, F. Castaldo, A. Alahi, F. Palmieri, and S. Savarese, "Knowledge transfer for scene-specific motion prediction," in *European Conference on Computer Vision (ECCV)*, pp. 697–713, 2016.

[17] J. Walker, A. Gupta, and M. Hebert, "Patch to the future: Unsupervised visual prediction," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 3302–3309, June 2014.

[18] A. Robicquet, A. Sadeghian, A. Alahi, and S. Savarese, "Learning social etiquette: Human trajectory understanding in crowded scenes," in *European Conference on Computer Vision (ECCV)*, pp. 549–565, 2016.

[19] T. Fernando, S. Denman, A. McFadyen, S. Sridharan, and C. Fookes, "Tree memory networks for modelling long-term temporal dependencies," *CoRR*, vol. abs/1703.04706, 2017.

[20] B. D. Ziebart, A. L. Maas, J. A. Bagnell, and A. K. Dey, "Maximum entropy inverse reinforcement learning.," in *the Advancement of Artificial Intelligence (AAAI)*, vol. 8, pp. 1433–1438, Chicago, IL, USA, 2008.

[21] V. Karasev, A. Ayvaci, B. Heisele, and S. Soatto, "Intent-aware long-term prediction of pedestrian motion," in *The IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2543–2549, May 2016.

[22] N. Rhinehart and K. M. Kitani, "Online semantic activity forecasting with DARKO," *CoRR*, vol. abs/1612.07796, 2016.

[23] R. Bellman, “A markovian decision process,” *Journal of Mathematics and Mechanics*, vol. 6, no. 5, pp. 679–684, 1957.

[24] R. S. Sutton and A. G. Barto, *Reinforcement learning: An introduction*, vol. 1. MIT press Cambridge, 1998.

[25] S. Russel, “Learning agents for uncertain environments,” in *The Fifteenth International Conference on Machine Learning (ICML)*, pp. 278–287, 1998.

[26] A. Y. Ng, S. J. Russell, *et al.*, “Algorithms for inverse reinforcement learning,” in *The Seventeenth International Conference on Machine Learning (ICML)*, pp. 663–670, 2000.

[27] P. Abbeel and A. Y. Ng, “Apprenticeship learning via inverse reinforcement learning,” in *The Twenty-first International Conference on Machine Learning (ICML)*, pp. 1–8, ACM, 2004.

[28] H. Weimerskirch, M. Louzao, S. de Grissac, and K. Delord, “Changes in wind pattern alter albatross distribution and life-history traits,” *Science*, vol. 335, no. 6065, pp. 211–214, 2012.

[29] T. Yamamoto, H. Kohno, A. Mizutani, K. Yoda, S. Matsumoto, R. Kawabe, S. Watanabe, N. Oka, K. Sato, M. Yamamoto, *et al.*, “Geographical variation in body size of a pelagic seabird, the streaked shearwater calonectris leucomelas,” *Journal of Biogeography*, vol. 43, no. 4, pp. 801–808, 2016.