Human Motion Prediction Considering Environmental Context

Igi Ardiyanto Jun Miura
Toyohashi University of Technology
1-1 Hibarigaoka, Tenpaku-cho, Toyohashi, Aichi, 441-8580, Japan
iardiyanto@aisl.cs.tut.ac.jp

Abstract

This paper describes an approach to predict the human motion. Instead of using a simple motion model as widely used, we take advantages of the environmental context, including the shape and structure, for predicting the human movement. First, we characterize the environment using a graph representation. Subsequently, we acquire the human trajectory tendency on each environment and build a probabilistic sequence model of the human motion. A particle filter-based predictor is then integrated into the system for generating possible future paths of the person. Evaluations on a real campus environment show the advantages of the proposed approach.

1 Introduction

For many robotic applications, it is necessary for a robot to understand its surrounding environment, including the human around it. One of the important cognition for the robot is to perceive the human motion and behavior. By figuring out the motion of each person, it enables the robot to take any necessary action, depending on the task's demand. For instance, a good prediction of the person movement can help the robot to generate an effective motion plan which tackles any possible collision in the future.

In many past works, especially for the people tracking purposes, the human motion is often assumed to follow a simple model such as the *constant velocity model* (*e.g.* [1] and [2]). Realizing the weakness of the simple model, some recent works suggest a more advanced approach for modeling the human motion.

Bennewitz et al. [3] tried to collect the pattern of the human trajectories using an *Expectation-Maximization* clustering and infer the human motion using a *Hidden Markov Model*. Later, Vasquez et al. [4] proposed an incremental model, so-called *Growing Hidden Markov Models*, to learn and predict the motion patterns.

From another perspective, Kitani et al. [5] employed an approach originated from the *optimal control theory* to forecast the long-term destinations of a person using a semantic scene. An interesting work by Luber et al. [6] utilized a *social force model*-based method for predicting the short-term intention of a moving person.

By assuming a person tends to move following the shape of the environment, we believe that a deep comprehension to the environmental information is necessary. Meanwhile, most of the mentioned works do not consider how the environment will affect the person movement (e.g. [1], [2], [3], [4], and [6]). In case of [5], it exploits the physical attribute information of the environment (such as building, car, and pavement), but it is basically used for separating the *walkable* and *non-walkable* area.

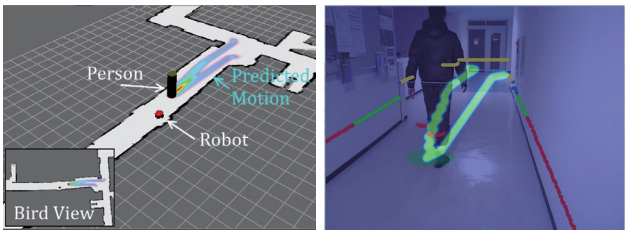

Figure 1: Human motion prediction: (left) perspective map view; (right) the robot camera view.

To close those gaps, we aim to incorporate the environmental information into our human motion prediction system, by analyzing the context of the environment. Here, the environmental context is defined as the leverage of a specific feature and attribute of the environment, including its shape and structure, to the outcome of the human motion. The movement of a person tends to follow the shape of a corridor, either it is a *T-junction*, a *cross-junction*, or a *straight way*. The person motion pattern on each junction is also *place-dependent*, i.e. the person motion preference will be governed by the functional entity on each edge of the junction. For example, we empirically found the students more likely choose the path towards the classroom rather than the one towards the toilet at a certain T-junction inside the building in our university. Utilizing such information will be useful for predicting where the person moves in the future.

We propose a novel framework for predicting the human motion. Initially, the environment is portrayed as a graph representation. We then extract the human trajectory trend and construct a probabilistic sequence model using *Hidden-state Conditional Random Field* (HCRF) [7], considering the person motion and environmental features. A particle filter-based predictor is then employed for yielding probable future paths and goals to where the person may proceed (see Fig. 1).

Subject to the description above, our main contribution lies on the contemplation of the environmental context to the human motion prediction. It is also worth to count the usage of the graph representation for describing the environment as another contribution.

2 Proposed approach

Our objective is to utilize the environment information to aid the human motion prediction. We begin with characterizing the environment to obtain its meaningful context. For example in the real world, we can semantically categorize an indoor environment into a hallway and a junction type, from which an environment is basically a connected sequence of both types. Therefore, it is easy to predict the human motion on the hallway (e.g. getting close or going away), and be-

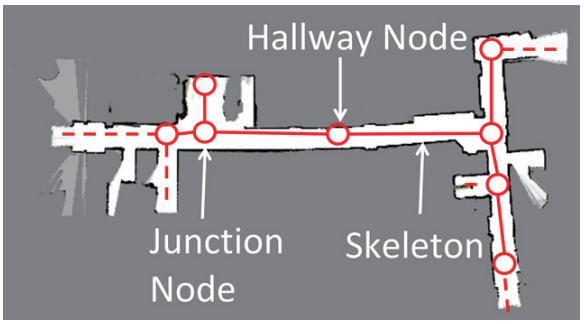

Figure 2: Graph representation of the map.

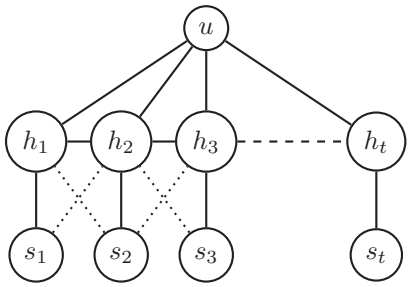

Figure 3: Trajectory model using the HCRF.

comes more difficult on the junction cases. Here we intend to imitate the above human reasoning.

2.1 Environment representation as a graph

Given two dimensional grid map $\mathbb{Q} \subset \mathbb{R}^2$ acquired by a SLAM algorithm [8], we then simplify the map $\mathbb{Q}$ such that $f_{simp} : \mathbb{Q} \mapsto \{\mathcal{P}, \mathcal{K}\}$. The mapping function f_{simp} follows the procedure mentioned in [9] to obtain a polygonal model $\mathcal{P}$, as well as the skeleton $\mathcal{K}$ of the map. From $\mathcal{K}$, we can determine the junction by applying a template matching over the map using the junction models [9].

Now, we are able to represent the map as a graph which connects the hallway and junction nodes, as shown in Fig. 2. We set the range area of each node to 10 meters, assuming the robot ability to detect and track the person is limited (i.e. the robot visibility, denoted by $\mathcal{V}$). Here, the environmental context reasoning are employed. We assume the human motion on the hallway nodes can be classified into two classes, getting close and going away. Regardless of the junction node, we also presume the human motion will follow the skeleton shape.

As the robot's "view" is limited, the person motion can be predicted to go towards *the frontiers*, i.e. the intersection of the robot visibility $\mathcal{V}$ and the skeleton $\mathcal{K}$. Thereafter, we define $\mathcal{G} = \{g_1, g_2, \ldots, g_n\}$ as the goal locations the person may lead to, as follows

$$\mathcal{G} = \{\forall q \in \mathbb{Q} | q = \mathcal{V} \cap \mathcal{K}\}. \tag{1}$$

2.2 Predicting the human motion

We formally define the trajectory of human motion as $\mathcal{S} = \{s_1, s_2, \ldots, s_t\}$ which is a sequence of the human position until the time t, where $\mathcal{S}$ are interchangeable with $\mathbb{Q}$ through a projection mapping. Let $\mathcal{U} = \{u_1, u_2, \ldots, u_n\}$ be n class trajectory labels denoting the person intention to go towards each goal in $\mathcal{G}$. Let $\phi_{\mathcal{S}}$ and $\phi_{\mathcal{G}}$ respectively represent the observation of the human position and the goals located on the frontiers from the current robot pose.

We aim to model the relationship between the person trajectory, the goal locations, the predicted motion towards the goals (labels), and the observations as $p(\mathcal{S}, \mathcal{G}, \mathcal{U} | \phi_{\mathcal{S}}, \phi_{\mathcal{G}})$ respectively. Under the independence assumption, the model can be written as

$$p(\mathcal{S}, \mathcal{G}, \mathcal{U} | \phi_{\mathcal{S}}, \phi_{\mathcal{G}}) = p(\mathcal{U} | \mathcal{S}, \mathcal{G}) p(\mathcal{S}, \mathcal{G} | \phi_{\mathcal{S}}, \phi_{\mathcal{G}}). \tag{2}$$

The first term of the right-hand side of eq. (2) is basically the label prediction involving a sequence structure, which is naturally solved using a sequence classifier. The second term models the person and goal locations, which in our case, a *gaussian distribution* is used.

2.2.1 Modeling the person trajectory as a sequence classification using the HCRF

To deal with the sequence prediction on eq. (2), we utilize the Hidden-state Conditional Random Field (HCRF) [7], as the structure is suitable for our trajectory motion model which consists of a sequence of the human pose with one label (see Fig. 3). Following the work of [7], our HCRF is modeled as

$$\begin{aligned} p(\mathcal{U} | \mathcal{S}, \mathcal{G}; \varphi) &\propto \sum_{\mathcal{H}} p(\mathcal{U}, \mathcal{H} | \mathcal{S}, \mathcal{G}; \varphi), \\ &\propto \frac{1}{Z(\mathcal{S}, \mathcal{G}; \varphi)} \sum_{\mathcal{H}} e^{f(\mathcal{U}, \mathcal{H}, \mathcal{S}, \mathcal{G}; \varphi)}, \end{aligned} \tag{3}$$

where φ is the parameter to be estimated, $f(\cdot)$ represents the feature function, and $Z(\mathcal{S}, \mathcal{G}; \varphi)$ denotes the normalization factor. Here, a vector of hidden-states $\mathcal{H} = \{h_1, h_2, \ldots, h_n\}$ are introduced as the possible hidden labels inside the model [7].

After the parameter φ is optimized using a gradient ascent method [10], we obtain the label score as follows

$$u_i = p(u_i | \mathcal{S}, \mathcal{G}; \varphi^*), \tag{4}$$

where φ^* is the learned parameter. For the classification purposes, we take the maximum score as the label for a trajectory.

2.2.2 Feature function

The feature function $f(\cdot)$ in eq. (3) is composed using following features to capture the trajectory traits:

1. **Position-based feature**. We describe $s_t = \{x_t, y_t\} \in \mathcal{S}$ as the human coordinate at the time t. Subsequently, we can extend it to derive the speed v and orientation θ of the human motion as follows

$$\begin{aligned} v_t &= \|s_t - s_{t-1}\|, \\ \theta_t &= \arctan(\frac{y_t - y_{t-1}}{x_t - x_{t-1}}). \end{aligned} \tag{5}$$

These features are then quantized into three bins and 16 bins histogram of the velocity and orientation respectively.

2. **Topology-based feature**. We want to figure out how the environment structure will affect the human motion. Hence, we utilize the skeleton map by calculating derivative of the distance function towards the skeleton $\mathcal{K}$ for each element $s_i \in \mathcal{S}$, as follows

$$r(s_i) = \frac{\partial(e^{\|s_i - s_\mathcal{K}\|})}{\partial s}, \quad (6)$$

where the numerator denotes the distance of s_i to the nearest point $s_\mathcal{K}$ in the skeleton. We expect to obtain a high magnitude of $r(s_i)$ when a person traverses the skeleton. This feature is then quantized into eight bins histogram.

2.2.3 Particle filter-based predictor

As the observation $\phi_\mathcal{S}$ and $\phi_\mathcal{G}$ are updated through the time, we recursively estimate the distribution in eq. 2 using a Bayesian framework, particularly a particle filter. The state model is composed by $\mathcal{X} = \{\mathcal{S}, \mathcal{G}, \mathcal{U}\}$, and the dynamical model is described as

$$p(\mathcal{X}_t|\mathcal{X}_{t-1}) = p(\mathcal{S}_t, \dot{\mathcal{S}}_t|\mathcal{S}_{t-1}, \dot{\mathcal{S}}_{t-1}) p(\mathcal{G}_t|\mathcal{G}_{t-1}) p(\mathcal{U}_t|\mathcal{U}_{t-1}), \quad (7)$$

where the first term is modeled using a *first-order dynamical model*, and the second and third terms are in the form of the *gaussian distribution*.

The observation is then modeled as

$$p(\phi_\mathcal{S}, \phi_\mathcal{G}|\mathcal{S}, \mathcal{G}) = p(\phi_\mathcal{S}|\mathcal{S}) p(\phi_\mathcal{G}|\mathcal{G}). \quad (8)$$

Again, we use the *gaussian distribution* model for the right-hand side of eq. (8).

We now have the confidence of the possible person intention to head up to each goal in $\mathcal{G}$ using the score of $\mathcal{U}$. The last step is to generate the possible trajectory of the person by connecting each goal to the current person pose using a *bezier curve* considering its confidence (e.g. Fig. 1). Please notice that the decision of choosing the goal can be determined when the confidence is above a threshold.

3 Experiments

The implementation of the described algorithm is done on a Windows PC (i7 2.4 GHz, 16 GB RAM) using C++ programming language.

3.1 Dataset evaluations

At first, we collect a set of person trajectories on five different locations/junctions at our campus (see Fig. 4), using a laser-based person tracker [11]. In total, 983 trajectory sequences were captured, yielding three to six trajectory classes per location. For each location, we then randomly divide the data into two different sets, i.e. for training and testing purposes.

We evaluate performance of the proposed method utilizing the HCRF [7] for discriminating the person trajectories on each location. We compare it with two baseline methods, Conditional Random Field (CRF) [12] and Hidden Markov Model (HMM) [13], due to the nature of the problem as the sequence classification. In the CRF experiments, each state in one trajectory is respectively labeled using a same class, as the opposite

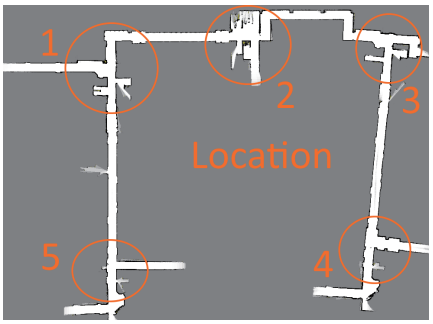

Figure 4: Environment map showing locations/junctions.

Table 1: Comparison of trajectory classification

Method	Accuracy (%) on Location				
	1	2	3	4	5
CRF	52.66	73.45	46.45	43.83	42.64
HMM	58.90	77.25	54.67	50.24	51.30
HCRF	55.83	77.76	51.23	46.87	46.28
CRF + context	52.90	74.23	49.00	46.47	44.64
HMM + context	60.67	75.23	58.96	52.45	48.20
HCRF + context	63.34	80.45	61.62	57.90	53.44

to the HCRF which uses only one class label per trajectory. Both CRF and HCRF are accordingly trained for each location as a multi-class classifier. On the other hand, we generate the model for each trajectory class for the HMM.

On each mentioned method, we engage two different types of the feature usage; using only the positional information, and combining the positional and context (i.e. topological features).

Table 1 shows the accuracy of the trajectory classification. Please note that at "location 2", we have only three classes of the trajectory, make it easier to do the classification here rather than the one on the other locations and achieve a high accuracy. We can clearly see that taking into account the environmental context enhances the trajectory class recognition rate. Moreover, The usage of the HCRF has a benefit over the other methods. It can be explained by the ability of the HCRF to model the hidden structures of the trajectory sequence and its relationship toward one single label, which is lack in the CRF and HMM.

3.2 Predicting the human motion on a robot

We employ a mobile robot equipped by a laser range finder and a camera to verify the performance of our human motion prediction. The same laser-based person tracker mentioned in section 3.1 is utilized. We carry out the experiments on "location 1".

Figure 5 shows the prediction performance of our system. Initially, each possible trajectory of the person towards the predicted goal has an equal distribution. The predicted goals are determined by the current frontiers of the robot, explained in section 2.1.

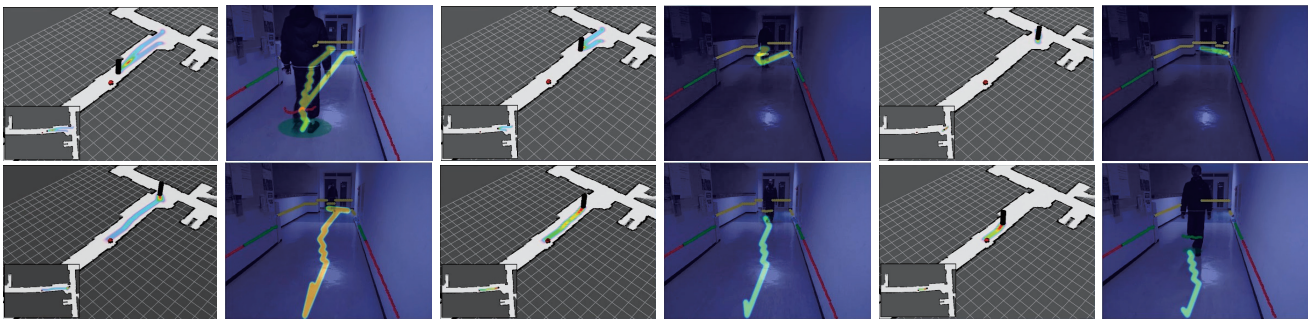

Figure 5: Human motion prediction on a robot (left-to-right): (top) the person moves away from the robot; (bottom) the person comes closer to the robot. For each set, the left figure is the perspective map view, the right one shows the camera view.

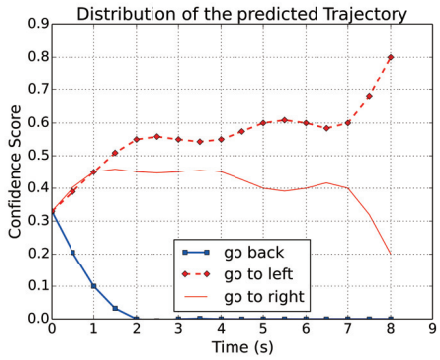

Figure 6: Distribution of the predicted trajectory over the time, according to the top figures of Fig. 5.

As the person data sequence grows, the information about the person speed, orientation, and the environmental context becomes more certain and will be fed to our system. Hereupon, the predicted trajectory will be condensed towards the predicted goals which have a higher likelihood according to the classifier.

The human motion prediction performance is qualitatively satisfying according to Fig. 5. It is supported by the tendency graph on Fig. 6, which shows the distribution of each predicted trajectory over time.

4 Conclusions

We have established an algorithm for predicting the human motion, considering the context of the environment. By composing the probabilistic sequence model of the human motion, we capture the human trajectory tendency on each environment structure. Afterwards, we predict the human intention by incorporating the model on a particle filter-based system. Experimental results support the benefit of our approach over the other methods.

Although the evaluations were done for predicting the human movement on an indoor campus environment, our algorithm may potentially be generalized to the cases on any structured environment with any moving object (e.g. vehicle movement on the road). While the current work is restricted to the single-person context, in the future it will be interesting to consider a multi-person motion prediction with its mutual trajectory and social behavior. Another possible future direction is to carefully examine the feature nonlinearity for increasing the prediction ability.

References

[1] J. Cui, H. Zha, H. Zhao, and R. Shibasaki. "Tracking multiple people using laser and vision". In Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems, pp. 1301-1306, 2005.

[2] A. Fod, A. Howard, and M.J. Mataric. "Laser-based people tracking". In Proc. of the IEEE Int. Conf. on Robotics and Automation, pp. 3024-3029, 2002.

[3] M. Bennewitz, W. Burgard, G. Cielniak, and S. Thrun. "Learning motion patterns of people for compliant robot motion". In Int. Journal of Robotics Research, Vol. 24 no. 1, pp. 31-48, 2005.

[4] D. Vasquez, T. Fraichard, and C. Laugier. "Incremental learning of statistical motion patterns with growing hidden Markov models". In IEEE Trans. on Intelligent Transportation Systems, vol. 10, no. 3, pp. 403-416, 2009.

[5] K. Kitani, B.D. Ziebart, J.A. Bagnell, and M. Hebert. "Activity forecasting". In Computer Vision - ECCV 2012, pp. 201-214, 2012.

[6] M. Luber, J.A. Stork, G.D. Tipaldi, and K.O. Arras. "People tracking with human motion predictions from social forces". In Proc. of the IEEE Int. Conf. on Robotics and Automation, pp 464-469, 2010.

[7] A. Quattoni, S. Wang, L.P. Morency, M. Collins, and T. Darrell. "Hidden conditional random fields". In IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 29, no. 10, pp. 1848-1853, 2007.

[8] S. Thrun, W. Burgard, and D. Fox. "Probabilistic Robotics". The MIT Press, 2005.

[9] I. Ardiyanto and J. Miura. "Visibility-based viewpoint planning for guard robot using skeletonization and geodesic motion model". In Proc. of the IEEE Int. Conf. on Robotics and Automation, pp. 652-658, 2013.

[10] C. Sminchisescu, A. Kanaujia, Z. Li, and D. Metaxas. "Conditional models for contextual human motion recognition". In Proc. of the IEEE Int. Conf. on Computer Vision, pp. 1808-1815, 2005.

[11] K. Koide, I. Ardiyanto, and J. Miura. "Person detection and tracking using camera and laser range finder for attendant robots". In SI2013, 2013. (In Japanese)

[12] J. Lafferty, A. McCallum, and F. Pereira. "Conditional random fields: Probabilistic models for segmenting and labeling sequence data". In Proc. of the 18th Int. Conf. on Machine Learning, pp. 282-289, 2001.

[13] A. Sand, C. Pedersen, T. Mailund, and A. Brask. "HMMlib: A C++ library for general hidden Markov models exploiting modern CPUs". In Proc. of the 2nd Int. Workshop on High Performance Computational Systems Biology, pp. 126-134, 2010.