6—5 Proposal of An Adaptive Vision-based Interactional Intention Inference System in Human/Robot Coexistence

Minh Anh T. Ho, Y. Yamada, T. Sakai, T. Morizono, Y. Umetani
Intelligent Systems Lab.,Toyota Technological Inst.
2-12-1 Hisakata, Tempaku-ku, Nagoya 468-8511, Japan
ho@s2.toyota-ti.ac.jp

Abstract

In the paper, we propose a vision-based system for adaptively inferring the interactional intention of a person coming close to a robot, which plays an important role in the succeeding stage of human/robot cooperative operations typically in production lines. It is naturally supposed that the interactional intention of a person induces her action of approaching to the robot and touching it. Therefore, in our Adaptive Vision-based Interactional Intention Inference System (AVIIIS), the human interactional intention is inferred from the direction in which the human body is moving and the trajectory of the human movements. In the process, human motion trajectory is generated by traces of the head movement and interpolated with a cubic Spline approximation technique. Hidden Markov Model (HMM) is applied as an adaptive noise reduction filter at the stage of inferring the human interactional intention. The HMM algorithm with a stochastic pattern matching capability is extended to supply whether or not a person has an intention toward the robot at any time. Experimental results demonstrate the adaptiveness of the inference system using the extended HMM algorithm that filters out motion deviation over the trajectory.

1 Introduction

In the field of studying human/robot coexistence systems, total risk analysis reveals that an outsider's entrance to the working volume of a robot possibly leads to some hazardous situations [1]. Therefore, a robot is expected to have at least the capability of detecting the presence of humans around, so that the robot can prevent such dangerous situations from occurring. As a little child can understand the objective of the person just by a short time of watching the trajectory of human movement [2], the use of visual information processing technology to capture a person entering the camera view potentially supplies more meaningful information. We consider having possibility of providing a robot with the capability to infer if a person getting close to the robot has intention to operate it or not. In the study, we define the function of inferring the desire of a person in close vicinity to a robot to interact with it as interactional intention inference. The clues in this case are the sequence of human walking directions discretized relatively to the orientation where the robot arm exists. We focus on inferring human interactional intention because such an inference function plays an important role that the robot can exhibit a smooth change in its mechanical impedance in the succeeding stage of human/robot cooperation [3].

From the viewpoint that intention inference is attained by evaluating the gap between the resultant action of the operator and the task goal given to the system [4], such an inference capability is considered to be acquired through machine learning processes. In the study, we regard that the human interactional intention inference problem is a pattern matching one from early incomplete data in a human movement trajectory, and that the reliability of the matching function is enhanced through a learning process of the robot which is capable of evaluating the above mentioned gap.

The general process of the system is shown in **Figure 1**. In the following three sections, we will give

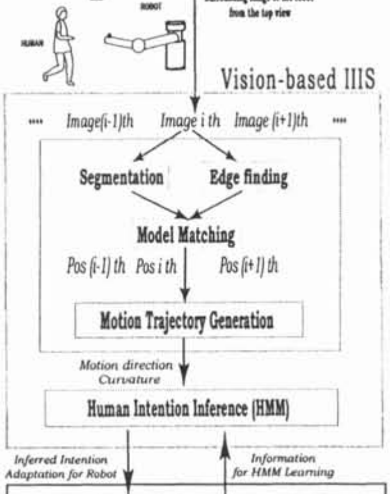

CAMERA
ROBOT
HUMAN
Surrounding image of the robot from the top view
Vision-based IIIS
.... Image(i-1)th Image i th Image (i+1)th ....
Segmentation
Edge finding
Model Matching
Pos (i-1) th Pos i th Pos (i+1) th
Motion Trajectory Generation
Motion direction Curvature
Human Intention Inference (HMM)
Inferred Intention Adaptation for Robot
Information for HMM Learning
ROBOT
Contact force sensing information

Figure 1: Overview of the AVIIIS

the details on the processes and the experimental results, excluding the contact force sensing process by the robot. The reward obtained through an interactive process of interacting with the robot can be judged if any of his/her actual operational force is detected by the robot. As a result, HMMs function as an adaptive noise reduction filter for gaining wide adaptability of the system.

2 The Human Tracking Process

First, human entrance is detected and is modeled by an ellipse to supply information about the head position to implement the tracking capability. Second, cubic Spline approximation technique is used intensively to approximate the trajectory with reduced control points in order that the system acquires information about the human motion direction and the curvature of the motion trajectory. This information will be used for intention inference in the next step.

2.1 2D Contour Human Modeling using Elliptic Matching Technique

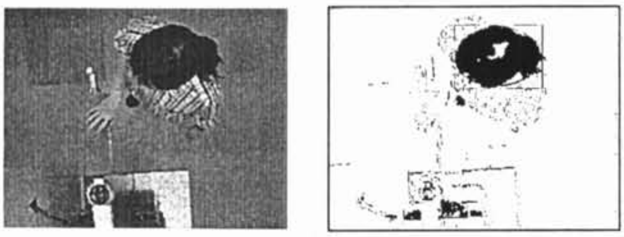


Figure 2: Real image and Result of Segmentation and Elliptic Matching technique

To detect the initial position of the person, the shape of the head is matched by an ellipse to supply precise information concerning the position and direction after the head part image is segmented from the dynamic scene by the black color. We adopted the Direct Least Squared Ellipse Matching technique [5]. An experimental result of applying the technique for the segmented head part is shown in **Figure 2**.

2.2 Motion Trajectory Interpolation Using Cubic Spline Approximation Technique

We use the cubic Spline approximation technique because of the local effect of control points on the curve. The succeeding detected positions help to lengthen the motion trajectory without affecting the whole curve. Furthermore, this technique maintains the interpolated curve to be C^2 continuous which still confirms us to compute motion direction vectors. The cubic parametric equation of the curve $\mathbf{p}(\tau) \equiv (p_x(\tau), p_y(\tau)) = \mathbf{a}\tau^3 + \mathbf{b}\tau^2 + \mathbf{c}\tau + \mathbf{d}$, with $0 \leq \tau \leq 1$, and parameters $\mathbf{a}$, $\mathbf{b}$, $\mathbf{c}$, and $\mathbf{d}$ can be calculated by

$$\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \\ \mathbf{c} \\ \mathbf{d} \end{bmatrix} = \frac{1}{6} \begin{bmatrix} -1 & 3 & -3 & 1 \\ 3 & -6 & 3 & 0 \\ -3 & 0 & 3 & 0 \\ 1 & 4 & 1 & 0 \end{bmatrix} \begin{bmatrix} \mathbf{p}_0 \\ \mathbf{p}_1 \\ \mathbf{p}_2 \\ \mathbf{p}_3 \end{bmatrix} \tag{1}$$

where $\mathbf{p}_0$ through $\mathbf{p}_3 (\in \mathbf{R}^2)$ are the control points.

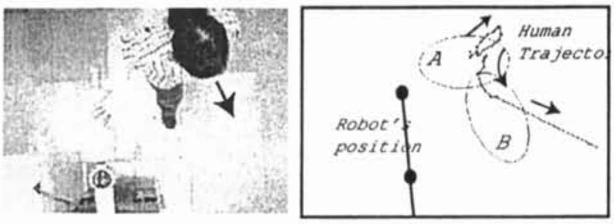


Figure 3: The trajectory by cubic Spline approximation with reduced control points

Figure 3 shows the result of an interpolated trajectory with reduced control points. At the head posture A, the person has the intention of operating the robot. Then, at the head posture B, the curve indicates that she changes her intention and goes outward.

2.3 Calculation of Motion Direction and Curvature

The motion direction is found by calculating the motion velocity vector $\mathbf{v} = (v_x, v_y)^T$ of the first derivatives of $p_x(t)$ and $p_y(t)$ with respect to time t. The curvature of the trajectory is calculated as

$$\kappa = \frac{d\alpha}{ds} \tag{2}$$

where $d\alpha$ is the angle (rad.) between the two binormal vectors $\mathbf{v}(s)$ and $\mathbf{v}(s + ds)$, and ds is the length of the curve from the initial point. In the study we assume that ds in the study is small enough to be approximated by the distance of a pair of two points in a sequence, and therefore $d\alpha$ is calculated as the angle between the motion vectors of these two points.

3 Application of HMM to the AVIIIS

Recently HMM has become widely used for pattern recognition problems, such as for speech recognition [6] or human gesture understanding[7] whose patterns are represented in sequences of events. HMM is one type of stochastic signal models to describe the observed sequence of probabilistic events and is characterized by states and transitions between states. One of the powerful properties of HMM is that the states are hidden, and the model infers those states by a sequence of observation symbols. In the system, we regard that the intention is hidden under the motion trajectory expressed as a sequence of motion direction observations. HMM needs two processes: the HMM learning process (or the reestimation process), and the succeeding process of using the trained HMM to estimate the appropriate model from the observation sequence.

3.1 HMM Parametric Definitions

We set 2 HMMs: $\lambda_i = \{\mathbf{A}_i, \mathbf{B}_i, \pi_i\}$ with $i = 1, 2$. λ_1 is the model with interactional intention, and λ_2 is the model without interactional intention. Each model has 2 states, "Has Intention" and "No Intention". The observation data are composed of the direction of motion velocity O_d and the curvature of the motion trajectory O_c. The data are symbolized as follows. If the movement direction of the person turns toward the robot arm, O_d is set to 0, otherwise, it is set to 1. When the curvature of the trajectory is smaller than a threshold, O_c is 1, and otherwise, O_c is 0. Observation symbol $O_c = 1$ indicates a high chance of changing her movement direction, and consequently, O_d is likely to change the value, from 0 to 1, or vice versa. To make a general symbol by which both O_d and O_c are taken into account, the observation data vector $\mathbf{O}_t = \{O_{d_t}, O_{c_t}\}$ (t is any time in a sequence) is symbolized as

$$O_t = O_{d_t} * 2 + O_{c_t}, \tag{3}$$

which we call an integrated observation, or simply observation O_t in the study.

3.2 The Learning Stage (Reestimation)

With the learning capability of HMM, AVIIIS computes appropriate inference models through K sequences of observation $\mathbf{O}^1 ... \mathbf{O}^K$ for the 2 HMMs λ_1 and λ_2, according to the judgment of the system whether the person finally exerts an operational force to the robot arm endtip, $\mathbf{O}^k = \{O_1^k, O_2^k, ..., O_T^k\}$, with ($k = 1, ..., K$; T : *termination*) and O_t^k ($t = 1 \rightarrow T$) obtained from (3).

In the learning stage, the original Baum-Welch's algorithm [6] is extended to reestimate the models by calculating matrices $\mathbf{A} = \{a_{ij}\}$, $\mathbf{B} = \{b_j(l)\}$, and initial state vector π for each model.

3.3 The Inferring Stage (Estimation)

It is required that AVIIIS generate an output of the intention inference even at t in the middle of the time sequence $(1, ..., T)$. Therefore, we need to compute $P(\mathbf{O}_t|\lambda)$, probability of the partial observation sequence $\mathbf{O}_t = \{O_1, ..., O_t\}$, given the model λ. The inferred model is the one that has the highest probability for the observation sequence up to time t. This probability of the model is calculated by

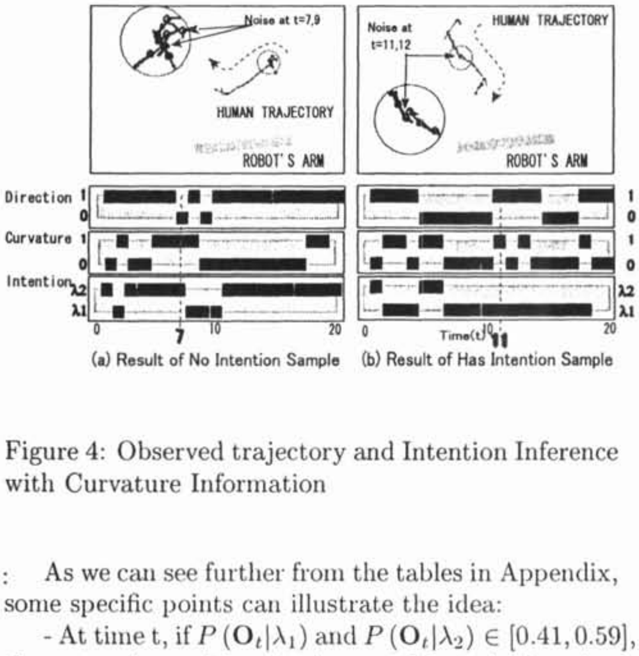
$$P(\mathbf{O}_t|\lambda) = \sum_{i=1}^{N} \alpha_t(i) \tag{4}$$

where N is the number of states in the model (N=2, in the study), and $\alpha_t(i)$ is the forward variable.

4 Experimental results

In the study, HMM is applied as a noise reduction filter for gaining wide adaptability of the system, and in our experiments, it is trained through observation sequences of the human movement direction and its curvature information.

Figure 4 shows two of these sample outputs. The upper figures depict the trajectories of the person, and the lower graphs are the resultant observation sequences O_d, O_c and the inferred interactional intention estimated by computation of $P(\mathbf{O}_t|\lambda_i)$. The left side is a No-Intention sample, and the right side is a Has-Intention one.



(a) Result of No Intention Sample (b) Result of Has Intention Sample

Figure 4: Observed trajectory and Intention Inference with Curvature Information

As we can see further from the tables in Appendix, some specific points can illustrate the idea:

- At time t, if $P(\mathbf{O}_t|\lambda_1)$ and $P(\mathbf{O}_t|\lambda_2) \in [0.41, 0.59]$, the system is considered to be *unstable*, and otherwise, it is *stable*. The system in *stable* phases can deal with some small noises from both observations.

* According to No-Intention sample (Figure 5a):

  - We can see the effect of the curvature information O_c which is counted in O_t at $t = 2$ and $t = 18, 19$. The output of the interactional intention inference is so *unstable* at $t = 2$, that the result changes to a wrong model λ_1. But when the system is *stable* at $t = 18, 19$, observation $O_c = 1$ increases $P(\mathbf{O}_t|\lambda_1)$, but the result λ_2 is still maintained. The curvature information indicated at $t = 5 \rightarrow 8$ ($O_c = 1$) results in a change from model λ_2 (*unstable* status) to model λ_1 (*unstable* status).

  - The observation sequence from $t = 10 \rightarrow 20$ with O_d continuously equals 1 makes the system *stable* with the model λ_2.

- According to Has-Intention sample (Figure 5b):
  - The effect of curvature observation can be seen at $t = 2, 3$ and $t = 5, 6$, where $O_c = 1$ increases $P(\mathbf{O}_t|\lambda_1)$ while the computed result of O_d is 1 at $t = 2, 3$, and decreases $P(\mathbf{O}_t|\lambda_1)$ even $O_d = 0$ at $t = 5, 6$.
  - From $t = 7$ till the termination, the performance of the system is in a stable condition, even when the observation sequence contains some noises at $t = 11 \rightarrow 14$ and $t = 18 \rightarrow 20$. It is referred that the person still has an intention even his movement direction has changed to move outward for a while, and until he moves back to the robot arm (at $t = 15 \rightarrow 17$), the hidden state is still "Has-Intention".

It is evident that this property of the stable system is effective to deal with taggering noise on the human trajectories when a person is walking step by step.

5 Conclusion

We have discussed our work on the use of visual information and the combination of both extended HMM and cubic Spline approximation techniques to infer the human interactional intention. This combination has privilege as the requirement for real-time processing.

We described the visual information processing methods of both extracting the human entrance and modeling her head using a direct elliptic matching technique. The cubic Spline approximation technique applied to the trajectory of human head movement reduces the control points of the curve according to the relative change in the positions to the interpolated curve. After that, movement direction and curvature information over the trajectory were taken as observations for the system. Experimental results showed the advantage of using the observation including motion direction and curvature information of the human trajectory. By using the HMM with a modified reestimation process for the complete parameter set of the model, we finally proposed to infer human interactional intention as an adaptive pattern matching problem with incomplete sequences of observation. Experimental results showed the adaptiveness of reducing the staggering noise over the trajectory of the human movement.

Nevertheless, the camera in our system is settled considerably low, which results in a local sight of the trajectory. The HMM used is a discrete one which requires that observation data be symbolized. This one deteriorates the continuous information from observed data. Our ongoing work includes improvements for extending the human modeling operation to human arms and other body parts as well as for detecting the human eye direction, and further work will be processed with the continuous HMM in the same frame work of our proposal.

6 Acknowledgement

We thank Mr. Nagata Shigemi and his colleagues at Fijitsu Research Center for supplying the Fujitsu Color Tracking Vision Board and their various helpful information.

References

[1] Yamada, Yoji *et. al.*, *"FTA-Based Issues on Securing Human Safety in a Human/Robot Coexistence System", Proc of IEEE International Conf. on Systems, Man, and Cybernetics*, Tokyo, pp. II 1058-1063, 1999.

[2] Simon Baron-Cohen, Chapter 4: Developing Mindreading: The Four Steps, *"Mindblindness", The MIT Press*, 1995.

[3] Yamada, Yoji *et. al.*, *"Construction of a Human/Robot Coexistence System Based on A Model of Human Will-Intention and Desire, Proc. of 1999 IEEE International Conference on Robotics and Automation*, Detroit, pp. 2861-2867, 1999.

[4] D.A. Norman," Cognitive engineering. In Normal, D.A. & Draper, S.W.(Eds.)", *User Centered System Design.* Hillsdale, NJ:Lawrence Earlbaum, 1986.

[5] Andrew Fitzgibbon, Maurizio Pilu, and Robert B.Fisher, *"Direct Least Square Fitting of Ellipses", IEEE Trans on PAMI* Vol. 21, No. 5, May 1999.

[6] Lawrence R. Rabiner, *"A tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", Proceedings of the IEEE* Vol. 77, No. 2, February 1989.

[7] T. Startner and A. Pentland, *"Visual recognition of American Sign Language using Hidden Markov Models", Proceedings of Int. Conf. on Automatic Face and Gesture Recognition '95*, pp. 189-194, 1995.