2005－EC－1（7）
2005／6／4

人間とインタラクションを行う飛行ロボット

新田 亮　細井 一弘　屋比久 保史　杉本 雅則

東京大学大学院　新領域創成科学研究科

本稿では, 人間のジェスチャーを認識してアクションを起こす小型飛行船ロボットについて取り上げる. 飛行ロボットに関する研究はこれまで数多く行われてきたが, 屋内用の小型飛行ロボットと人間とのインタラクションについて論じられることはほとんど無かった. 本稿では, 初めに, HMM ベースのジェスチャー認識法を飛行船に適用した結果について述べる. 続いて, ジェスチャーに対する認識精度の向上を目指し, 対話型のジェスチャー認識法について提案を行う. ここでは, 人間とロボットがインタラクションを繰り返すことで, 認識したジェスチャーに対する確信度を高めていく枠組みについて定式化する. 最後にその有効性について考察する.

A Flying Robot That Interacts with Humans

Ryo Nitta　Kazuhiro Hosoi　Yasufumi Yabiku　Masanori Sugimoto

Graduate School of Frontier Sciences, The University of Tokyo

We have developed a small flying robot which reacts with human gesture. At the begining, we applied an HMM-based gesture recognition technique to a flying robot. After that, by innovating the theoretical framework in which a human and a robot interact with each other repeatedly, we have improved the accuracy of the recognition.

1　はじめに

近年, ロボットに関する様々な技術の向上に伴い, 掃除ロボットやペットロボットといった人間と同じ生活空間の中で活躍するロボットの研究・開発が盛んになってきている. これらのロボットにとって人間とのインタラクションは欠かせない. 人間の発話やジェスチャーに対応することはもちろん, より自然なインタラクションのためには, 人間や環境に適応した能力を獲得することがロボットに求められている. 中でも搬送やナビゲーションといった, 移動を伴いながら人間とコミュニケーションを行うロボットにとっては, 人間の歩行ペースや周囲の状況に適応しながら人間と同じ速度で移動することが望ましい. しかし, ロボットにとって障害物に衝突しないように移動することは容易ではなく, 段差などの不整地の移動にも時間がかかる. さらにオフィスや展示場のような周囲にたくさんの人間が存在する動的で不確定要素の多い環境ではその移動は益々困難である.

このように, 認知能力や身体能力に関してまだまだ差異のあるロボットと人間とが今後どのように共存していくかは大変重要な問題である. ロボットが活動しやすい環境を特別に設けることが考えられるが, それによって人間の生活空間が狭められてしまうことは望ましくない. だが, 屋外や屋内でも大型展示場のようなやや天井の高い環境では, 空中に人間の手が届かずあまり利用されていない空間が多く存在する. 特に屋内に関しては風などの影響を受けることも少なく, 比較的静的な空間が広がっていると考えることができる. そこで, ロボットが空中を自由に移動することができ, 人間とのコミュニケーションを取ることにも支障がなければ, 人間とロボットの共存に対する一つのモデルとして,「人間→地上, ロボット→空中」という形も十分現実的なものとなってくる.

自律飛行ロボットに関する研究はこれまで数多く行われてきた [2]. しかし人間にとって危険な大型のものが多く, 人間と直接インタラクションを行う目的で研究されてはこなかった. また, 小型のものについても, その搭載量に制約があることから十分なセンサーやハードウェアを載せることができずに制御が困難であるという問題があった. だが近年のハードウェアの急激な小型化・軽量化に伴い, 多くのハードウェアが小型飛行ロボットに搭載可能となってき

ている. 従って制御が容易になることはもちろん, 人間をサポートする上でも今後その可能性が大いに広がっていくと考えられる.

本稿では, 人間と飛行ロボットが共存していく上で重要となる, ロボットによる人間のジェスチャー認識について取り上げる. 最初に HMM を用いてその精度について考察した後、対話型のジェスチャー認識法の提案を行い, その有効性について検討する.

2 飛行船ロボット

本稿では, 飛行ロボットとして飛行船を取り扱う. 飛行船は搭載量の制約が厳しかったり動きが不安定といった問題があるが, 人間に接触することがあっても危害を与えることはなく, 人間に安全な飛行ロボットであると言える. 飛行船に関する研究としては, 人間のいない環境においてその制御を扱ったものがほとんどである [5]. 飛行船には, 市販されている全長約 90cm の屋内用のラジコン飛行船 (タカラドリームフォース 02) を改造したものを用いる. この飛行船は, 浮力を得るためのエンベロープ部と, モータやプロペラ, 制御信号受信機, それにバッテリを統合した駆動部から構成される. また, 外界の情報を得るための小型のワイヤレスカメラを搭載している. システム構成としては, 飛行船に取り付けられたカメラより取得された画像情報を PC に送り, そこで画像処理をして次に飛行船が取るべき行動の決定を行い, 制御信号を飛行船の駆動部に送信するという仕組みになっている（図 2 参照）.

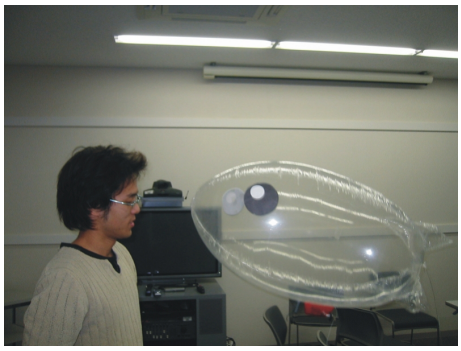

図 1: 人間とインタラクションする飛行ロボット

飛行船は 2 枚のプロペラを独立に回転させることで, 前進・後退・回転を行う. また, その 2 枚のプロペラをつなぐ主軸を回転させることで, 上下移動を行う.

飛行船を PC から直接制御するために, ここではラジコン飛行船付属のコントローラを改造したものを用いる. コントローラと PC の通信には PIC を用い, PC のシリアルポートからの信号をプロペラの正回転, 逆回転の ON, OFF を表す制御信号に変換している.

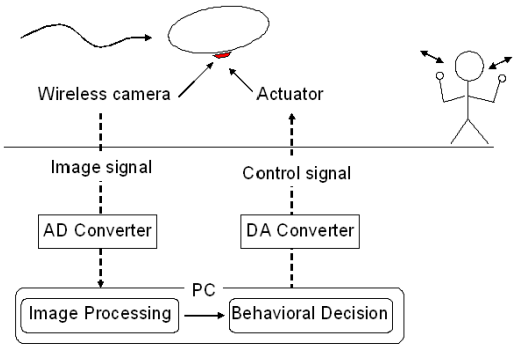

図 2: システム構成

3 ジェスチャー認識

3.1 HMM を用いたジェスチャー認識

ジェスチャー認識に関してはこれまで数多くの研究が行われてきた. 代表的な方法として隠れマルコフモデル (Hidden Markov Model, 以下 HMM)[4] や DP マッチング [3] を用いたものが挙げられる. 本稿では HMM を用いる. HMM は画像特徴量の揺らぎによる影響を緩和した時系列データの認識が可能であり, また, 高速計算のための様々なアルゴリズムが提案されている.

HMM は状態の有限集合 $\boldsymbol{Q} = \{q_1, ..., q_N\}$, 出力記号の有限集合 $\boldsymbol{S} = \{\boldsymbol{o}_1, ..., \boldsymbol{o}_N\}$, 状態遷移確率分布 $\boldsymbol{A} = \{a_{ij}\}$, 記号出力確率分布 $\boldsymbol{B} = \{b_j(k)\}$, 初期確率分布 $\boldsymbol{\pi} = \{\pi_i\}$ の, 以上 5 組組み $\lambda = (\boldsymbol{Q}, \boldsymbol{S}, \boldsymbol{A}, \boldsymbol{B}, \boldsymbol{\pi})$ で表される. ただし, a_{ij} は状態 q_i から q_j への遷移確率を示し, $b_j(k)$ は状態 q_j から出力記号 $\boldsymbol{o}_k$ を出力する確率を示す.

HMM で用いる画像特徴量については、人間の体に装着した 4 つの赤外線 LED の位置を, 可視光を遮断する IR フィルタで覆われた飛行船カメラで取得する。遠方に取り付けられたカメラから人間の部位を認識する手法としては, 肌色認識を行う手法一般的であるが、周囲の環境の変化によって認識率が大きく変動するという問題がある. そこで本稿では LED を用いた方法を採用して人間の部位の認識を容易に

する.

3.2 行動要素列の決定

本稿では, 連続なパターンである時系列データを離散化して HMM に適用する. ここでは, 予め用意した M 個の行動要素 $\{\boldsymbol{u}_1, ..., \boldsymbol{u}_M\}$ の列に離散化することを考える. HMM の出力記号にはこの行動要素列を用いることにする. すなわち, $\boldsymbol{o}_i = \boldsymbol{u}_i (i = 1, ..., M)$ とする. 各行動要素 $\boldsymbol{u}_i$ には, 人間の体に装着した 4 つの LED の位置関係が表現される位相空間上の点 $\boldsymbol{\mu}_i$ と, その位相空間上の近傍の大きさを決定する共分散行列 $\sum_i$ を付加する.

$$\boldsymbol{u_i} \equiv \{\boldsymbol{\mu}_i, \sum\nolimits_i\} \tag{1}$$

観測された時系列データを行動要素列に変換するために, 微小時間単位で行動パターンをサンプリングし, その値 $\boldsymbol{x}$ に対して次の計算を行う.

$$j = arg\max_i \frac{1}{\sqrt{(2\pi)^D det \sum_i}} exp\Big\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu}_i)^T \sum\nolimits_i^{-1} (\boldsymbol{x}-\boldsymbol{\mu}_i)\Big\} \tag{2}$$

ただし, D は要素表現の次元数, det は行列式, T はベクトルの転置を表す. この計算により, 位相空間上で $\boldsymbol{x}$ に最も近い行動要素 $\boldsymbol{u}_j$ が求まる. これを全ての時刻に対して繰り返し, 行動要素列 $[\boldsymbol{u}_{k_1}, \boldsymbol{u}_{k_2}, ..., \boldsymbol{u}_{k_t}]$ を得る.

3.3 HMM のパラメータ学習

ジェスチャー $i (i = 1, ..., K)$ に対する行動要素列の時系列パターン $\boldsymbol{O} = \{\boldsymbol{u}_{k_1}, \boldsymbol{u}_{k_2}, ..., \boldsymbol{u}_{k_t}\}$ を複数用いて, これらを最もよく発生させるような HMM のパラメータ λ_i を学習し, 登録する. ここでのパラメータの推定には, EM アルゴリズムの一種である Baum-Welch アルゴリズムを用いる. これを K 個全てのジェスチャーについて行う.

3.4 HMM を用いた時系列パターンの認識

人間の行動を認識するために, 観測された行動要素列の時系列パターン $\boldsymbol{O}$ が HMM からどの程度の確率で生成されるかを示す $P(\boldsymbol{O}|\lambda)$ を用いる. この確率は λ を変数とする尤度であり, Viterbi アルゴリズムによって高速に計算することが出来る. 登録した全てのジェスチャーに対する HMM を用いて尤度 $P(\boldsymbol{O}|\lambda_i)$ を算出し, この尤度の最大値を検出することで認識を行う. 入力された時系列パターンに対応する HMM は高い尤度を示し, 関連性の低い HMM は低い尤度を示す.

3.5 実験

HMM を用いてジェスチャー認識を行った. ジェスチャーの種類は図 3 に挙げる 6 種類とし, HMM は Left-to-Right モデルを採用した. また, 量子化の際に用いる行動要素の数 M は 37 個とした. 学習の際は 3 人の被験者から採取した 1215 個の時系列データを用いた. 認識はそれぞれのジェスチャーについて 30 回ずつ計 180 回行った. 表 1 に, カメラを固定して行った場合と, 飛行船に取り付けて行った場合の結果を示す.

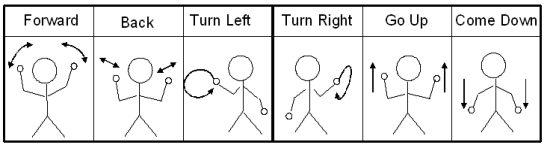

図 3: 6 種類のジェスチャー

	fixed camera	moving camera
rate	79.3%	43.4%

表 1: 固定カメラと飛行船カメラとの比較

実験結果より, カメラを固定した場合と比較して, 飛行船に取り付けられたカメラからのジェスチャー認識ではその識別率が著しく低下していることが確認出来る. これは, 飛行船の不安定な動きや, それに伴って人間の一部を視界から見失うことがあることに起因している. このような問題に対して, 飛行ロボットにも認識しやすいジェスチャーの種類を検討することも考えられる. しかし, 我々の目的は, 人間同士が行う自然なジェスチャーを用いて飛行ロボットとのインタラクションを行うことにある. そこで, ジェスチャーの種類に関する考察ではなく, 認識法についての提案を行い, 認識率の改善を目指す.

4 対話型ジェスチャー認識

人間同士がジェスチャーを用いてコミュニケーションを行う場合, ジェスチャーの理解が曖昧なときには, なんらかのアクションを起こして相手の反応を確認

してから, 先のジェスチャーの意味を解釈することがある. 飛行船のように, ジェスチャーの理解が容易ではないロボットが人間と自然なインタラクションを行っていくためには, このようなジェスチャー認識法が効果を発揮するのではないかと考えられる（図 4 参照）. 本節では数理モデルを用いてこの考え方を定式化する. そこで, それぞれのジェスチャーに対応

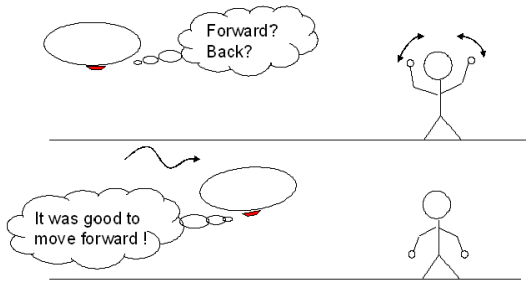

図 4: 対話型ジェスチャー認識

する HMM の操作性の向上や, HMM 同士の関係性の把握を容易にするために, 本稿では一つの HMM を一つのシンボルとして抽象化して考える. HMM がそのパラメータによって時系列データに影響を及ぼす確率過程の特性が決定されることを考慮してシンボル z を決定する [1].

4.1 シンボル空間の形成

シンボル同士の比較を行うために, 各シンボルが存在するシンボル空間を用意する. 空間の形成には距離の概念が必要であるが, 確率分布間関数間の隔たりを定量的に表す尺度として, Kullback-Leibler 情報量がある. 2 つの確率密度関数 p_1, p_2 間の Kullback-Leibler 情報量 $D(p_1, p_2)$ は次式で定義される.

$$D(p_1, p_2) = \int_{-\infty}^{\infty} \left(p_1(x) \log \frac{p_1(x)}{p_2(x)} \right) dx \qquad (3)$$

これを HMM に適用する場合は, 対象となる 2 つの HMM のパラメータを λ_1, λ_2 として次式で表される.

$$D(\lambda_1,\lambda_2)=\frac{1}{n}\sum_i \frac{1}{T_i}\left[\log p(\boldsymbol{y}_1^{T_i}|\lambda_1)-\log p(\boldsymbol{y}_1^{T_i}|\lambda_2)\right] \qquad (4)$$

ただし, $\boldsymbol{y}_1^{T_i}$ は λ_1 の学習の際に用いた長さ T_i の時系列データ, n は観測された時系列データの数である.

一般に $D(\lambda_1, \lambda_2) \neq D(\lambda_2, \lambda_1)$ であり, λ_1 と λ_2 に関して対称性が成り立たない. 従って, 上式を HMM 間同士の距離として用いるのは不適切である. そこで, 本稿では次式をシンボル間の距離として用いることにする.

$$D_s(\lambda_1, \lambda_2) = \frac{1}{2} \left(D(\lambda_1, \lambda_2) + D(\lambda_2, \lambda_1) \right) \qquad (5)$$

4.2 対話型ジェスチャー認識モデル

各々のジェスチャー認識に対する確信度は, シンボル空間上の確率分布関数で表現される. 本稿では, 予め登録されたジェスチャーとそれに対応する飛行船の行動のみを扱う. 従って分布は離散分布となる. 登録されたジェスチャーのシンボル集合を $\boldsymbol{Z} = \{z_1, ..., z_K\}$, 対応する飛行船行動集合を $\boldsymbol{U} = \{u_1, ..., u_K\}$ とする. 時刻 t の時点での飛行ロボットの行動を $u^t \in \boldsymbol{U}$, 観測された人間の反応を $z^t \in \boldsymbol{Z}$ とすると, 飛行ロボットが人間のジェスチャーをシンボル空間上の $x^t \in \boldsymbol{Z}$ のジェスチャーだと解釈している確率は次式で表すことが出来る.

$$Bel(x^t) = p(x^t | z^t, ..., z^1) \qquad (6)$$

ここで, 未来の状態は現在の状態のみに依存するといういうマルコフ性を仮定すると,

$$p(x^t | z^{t-1}, ..., z^1) = \sum_{x^{t-1}} p(x^t | x^{t-1}, u^{t-1}) Bel(x^{t-1}) \qquad (7)$$

となる. この式とベイズの定理を利用することで, 次式のように確信度 Bel を更新することが出来る.

$$Bel(x^t) = \eta p(z^t | x^t) \sum_{x^{t-1}} p(x^t | x^{t-1}, u^{t-1}) Bel(x^{t-1}) \qquad (8)$$

尚, η は正規化のための定数である. ここでの計算を進めるにあたり, $p(z^t|x^t)$ と $p(x^t|x^{t-1}, u^{t-1})$ を求める必要がある. 前者に関しては, 3.4 節で登場した Viterbi アルゴリズムによって求めることが出来る. 後者に関しては次節で定義する.

4.3 行動モデル

前節で登場した $p(x^t|x^{t-1}, u^{t-1})$ を本稿では行動モデルと呼ぶことにする. 行動モデルを決定するにあたり, ここでは人間がジェスチャーを行う際によく見られる次の 3 つの特徴を盛り込むことにする.

1. 人間は相手が自分の意図した動作を行っていた場合であっても, それが終了するまでは絶え間なく同じジェスチャーを続ける.

2. 人間は相手に自分の意図が伝わったと判断した場合は, それ以上ジェスチャーを続けずに相手の動作が終了するのを待つ.

3. 人間は相手が自分の意図していない行動を行った場合は, 相手に自分の意図がきちんと伝わることを目的に, 同じジェスチャーを以前よりも大げさに繰り返す.

例えばシンボル空間上の位置 x が, 飛行船に前進の行動を要求するジェスチャー (両手を振る) を抽象化したシンボルである場合を例に取る. このとき, 飛行船が前進という動作を起こすと, 飛行船の観測するジェスチャーとしては人間が両手を振り続ける場合 (ルール 1) と, 全くジェスチャーをしない場合 (ルール 2) の二つがあると考える. また, 飛行船が後退などといった全くその状態では要求されていない行動を行った場合は, 人間は自分の意図が飛行船に前進してほしいことだということを強く伝えるために, 両手を振るというジェスチャーをやや大げさに繰り返す (ルール 3) と考える（図 5 参照）. 以下, ルール 2 での人間のジェスチャーを”見守りポーズ”, ルール 3 でのジェスチャーを”オーバージェスチャー”と呼ぶ.

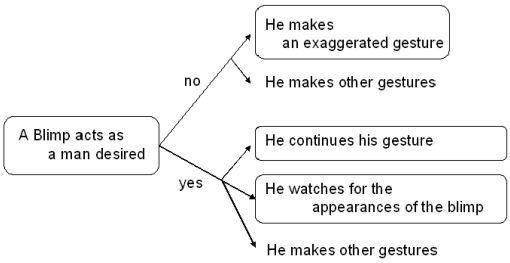

図 5: ルール

今, シンボル空間上において, ジェスチャー i に対応するシンボルを z_i とする. また, ジェスチャー i のオーバージェスチャー i' と見守りポーズ i_s に対応するジェスチャーをそれぞれ z_i', z_s とする.

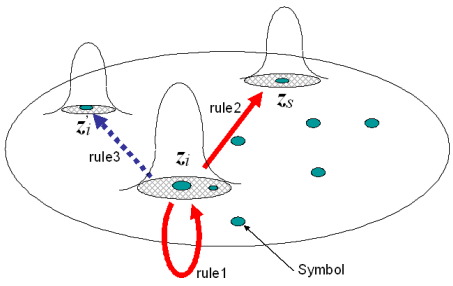

図 6: シンボル z_i からの確率伝播

- $x^{t-1} = z_i$ のとき

$u^{t-1} \neq u_i$ の場合, 人間はルール 3 に基づいてオーバージェスチャー i' を行う可能性が高い. また, $u^{t-1} = u_i$ の場合, 人間はルール 1 に基づいてジェスチャー i を行うか, ルール 2 に基づいて見守りポーズ z_s を行う可能性が高い. そこで以下の分布関数に従って確率を割り振る（図 6 参照）.

$$p(x^t|x^{t-1},u^{t-1}) \sim \begin{cases} \mathcal{N}(z_i',\sigma) & (u^{t-1} \neq u_i) \\ \{\mathcal{N}(z_i',\sigma)+\mathcal{N}(z_s,\sigma)\}/2 & (u^{t-1} = u_i) \end{cases} \quad (9)$$

尚, $\mathcal{N}$ は正規分布であり, 4.1 節で定義した D_s をそのノルムとして利用する. また, σ については, 簡単のため以下の全ての正規分布で同一とする.

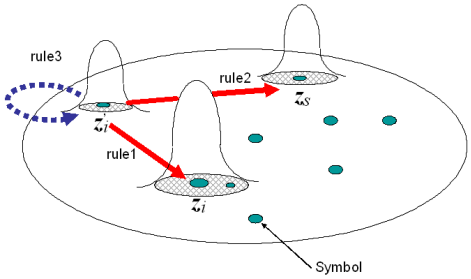

図 7: シンボル z_i' からの確率伝播

- $x^{t-1} = z_i'$ のとき

$u^{t-1} \neq u_i$ の場合, 人間はルール 3 に基づいてオーバージェスチャー i' を行う可能性が高い. また, $u^{t-1} = u_i$ の場合, 人間はルール 1 に基づいてジェスチャー i を行うか, ルール 2 に基づいて見守りポーズ z_s を行う可能性が高い. そこで以下の分布関数に従って確率を割り振る（図 7 参照）.

$$p(x^t|x^{t-1},u^{t-1}) \sim \begin{cases} \mathcal{N}(z_i',\sigma) & (u^{t-1} \neq u_i) \\ \{\mathcal{N}(z_i,\sigma)+\mathcal{N}(z_s,\sigma)\}/2 & (u^{t-1} = u_i) \end{cases} \quad (10)$$

- $x^{t-1} = z_s$ のとき

$u^{t-1} = u^{t-2}$ の場合, 人間は再び同じ見守りポーズを行う可能性が高いと. また, $u^{t-1} \neq u^{t-2}$ の場合, 行動 u^{t-2} に対応するシンボルを z_j とすると, ルール 3 に基づいてジェスチャー j を行うかオーバージェスチャー j' を行う可能性が高い. そこで以下の分布関数に従って確率を割り振る.

$$p(x^t|x^{t-1},u^{t-1}) \sim \begin{cases} \mathcal{N}(z_s,\sigma) & (u^{t-1} = u^{t-2}) \\ \{\mathcal{N}(z_j,\sigma)+\mathcal{N}(z_j',\sigma)\}/2 & (u^{t-1} \neq u^{t-2}) \end{cases} \quad (11)$$

4.4 行動決定方法

人間がジェスチャー i を行ったと考えられるのは, シンボル空間における z_i, $z_i^{'}$, z_s の確信度が高い場合である（ただし,$u^{t-1} \neq u_i$ の場合は z_s は含まれない）. そこで, 時刻 t におけるジェスチャー i の確信度を

$$Bel_g(i)=\begin{cases} Bel(z_i)+Bel(z_s)+Bel(z_i^{'}) & (u^{t-1}=u_i) \\ Bel(z_i)+Bel(z_i^{'}) & (u^{t-1}\neq u_i) \end{cases} \quad (12)$$

と定義する.

飛行船が人間と対話を重ねていく中で, 人間の意図の解釈を一つに絞り込んでいくためには, それぞれの時刻 t で行動 u_t を如何に選択するのかが重要な問題となる. 本稿では, 時刻 t の時点で Bel_g の値が最も高いジェスチャーに対応する行動を飛行船が選択する, という戦略を採用する.

4.5 実験

提案手法の有効性を確認するために実験を行った. 左回転命令のジェスチャー 1 と, 前進命令のジェスチャー 2, またそれぞれのオーバージェスチャーと, 見守りポーズの計 5 種類のジェスチャーに対する HMM のパラメータを登録したシンボル空間を用意した. 実験中, 人間は常に飛行船が前進することを望んでいる. 図 8 は飛行船と人間との一連のインタラクションの過程において, 飛行船のジェスチャー認識に対する確信度 Bel_g の推移を表したものである. 飛行船はジェスチャー 1 とジェスチャー 2 の識別が曖昧になり, 途中で確信度の優先順位が逆転し, 4.4 節の戦略に従って人間の期待しない左回転運動を始めようとする（フレーム数 160 付近）. それを見た人間が, 前進命令のオーバージェスチャーを行うことで（フレーム数 200 付近）, ジェスチャー 2 に人間の意図があるという確信度を再び飛行船が高めたをことを表している.

これから、ジェスチャー認識に対する判断に曖昧性が生じた場合でも、提案手法を用いることで曖昧性を解消して, 正しい認識を飛行船が行っていることが確認できる。

5 まとめと今後の課題

本稿では, 飛行船に人間のジェスチャーを認識させて人間の意図する行動を行わせるための方法について考察した. 飛行船の不安定な動き等の影響によって, 従来広く用いられてきた HMM を単に利用するだけではジェスチャーの識別率の低下が避けられないことが分った. これに対して, 我々は対話型のジェスチャー認識法を提案し, 実験を通じて曖昧で識別判断に迷う状況になっても, 人間とのインタラクションを繰り返すことで人間の意図を最後には正確に飛行船が判断していることを確認した. 今後はジェスチャーの種類をさらに増やした状態においても本手法が有効かどうかについて検討したい.

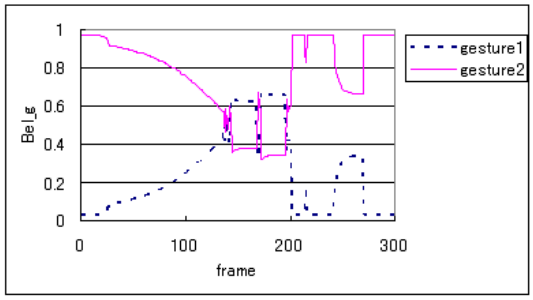

図 8: 提案手法の実験結果

謝辞

飛行船を提供された (株) タカラに感謝の意を表明する.

文献

[1] T. Inamura, H. Tanie and Y. Nakamura, ”From Stochastic Motion Generation and Recognition to Geometric Symbol Development and Manipulation” In Proceeding of Int’l Conference on Humanoid Robots (Humanoids 2003), 2003.

[2] S. Saripalli, J. F. Montgomery, and G. S. Sukhatme, ”Visually-Guided Landing of an Unmanned Aerial Vehicle,” In IEEE Transactions on Robotics and Automation, Vol. 19, No. 3, pp. 371-381, Jun 2003.

[3] S. Seki, K. Takahashi, and R. Oka, ”Gesture recognition from motion images by spotting algorithm”, In Proceeding of Asian Conference on Computer Vision, pp.759-762. Osaka, Japan, 1993.

[4] J. Yamato, S. Kurakake, A. Tomono and K. Ishii, ”Human Action Recognition Using HMM with Category-Separated vector Quantization”, Transactions of the Institute of Electronics, Information and Communication Engineers D-II, Vol.J77D-II, No.7, pp.1311-1318, 1994.

[5] S. Zwaan, A. Bernardino, Jose’ Santos-Victor, ”Vision based station keeping and docking for an Aerial blimp”, VisLab-TR 11/2000 - Intl Conference on Intelligent Robots and Systems IROS’2000, Kagawa University, Takamatsu, Japan, Oct 30 - Nov 5, 2000.