

自律飛行ロボットを用いた屋内ナビゲーションシステム

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

本研究では、飛行船を自律的に移動させることでユーザを目的地まで誘導する、屋内型ナビゲーションシステムについて取り上げる。インタラクティブ性を持たせるため、本システムでは飛行船にユーザのジェスチャーを認識する機能を取り入れる。ジェスチャーには、静止ジェスチャーから時系列ジェスチャーまで様々な種類が存在するが、本稿では、周期性の伴った時系列ジェスチャーの認識を試みる。これは、本システムの利用を想定している混雑した展示場等において、周期性の伴ったジェスチャーの検出は比較的容易であると考えられるからである。しかし、飛行船は常に空中を移動しているため、周期的ジェスチャーの種類を判別することまでは容易ではない。そこで本稿では、その認識法について検討する。

An Indoor Navigation System Using Autonomus Flying Robots

Ryo Nitta[†] Kazuhiro Hosoi[†] Yasufumi Yabiku[†] Masanori Sugimoto[†]

[†] Graduate School of Frontier Sciences, The University of Tokyo

In this paper, we report an indoor navigation system using autonomous flying robots. We aim to make the system interactive, so we adapt a gesture recognition function. There are many kinds of gestures. We use cyclic gestures in this system, because cyclic gestures are very important in a crowded exhibition hall in which we want to use our system. But, the blimp is always moving in the air, so it is difficult for the blimp to discriminate that gestures. We think about some kinds of methods to recognize that gestures .

1 背景と目的

小型飛行ロボットには搭載量が非常に厳しいという制約があり、その利用はこれまで大変困難であった。しかし近年、ハードウェアの小型化・軽量化が著しく進んでおり、小型飛行ロボットの活躍する可能性が大きく広がろうとしている¹⁾。

小型飛行ロボットの特徴は、室内においても3次元空間を自由に移動出来ることにある。従って、混雑した通路を移動する時であっても、人間の頭上を

飛行することで容易にその移動を行うことが出来る。これは、障害物等を検出し、それを避ける経路を移動する必要のあった地上ロボットとは異なる、飛行ロボットの大きなメリットであると言える。また、飛行ロボットにカメラが積んである場合には、人間が直接見る事の出来ない視点からの映像を取得することも可能である。従って、モニタリング等の様々なアプリケーションへの応用も期待されているところである。

我々の目的は、小型飛行ロボットのメリットを活

かした高度なインタラクティブシステムを構築することである。現在は、展示場等でのナビゲーションシステムに飛行ロボットが有効利用出来ないかどうか検討中である。ナビゲーションシステムでは、1) ロボットに用件を伝えるためにユーザに近寄ってもらう、2) 案内中に速度を変更してもらう、3) 指差した場所の映像を取得してもらう、など、人間とロボットとのインタラクティブな機能を実現することが重要である。我々はこれらの機能を人間のジェスチャーを用いて実現したいと考えている。従って、飛行ロボットが人間のジェスチャーを検出し、その内容を理解する必要が生じる。

しかし、飛行ロボットにとって地上にいる人間のジェスチャーを認識することは容易ではない。これは、1) 不安定な動きによって飛行船の画面が常に揺れてしまう、2) 人間とは距離の離れたところから認識を行う必要がある、といったことが原因として挙げられる。また、本研究が想定している展示場等では、一般に複数の人が存在し、それぞれが動いたり止まったりしている。従って、映像の中で飛行ロボットにとって意味のあるジェスチャーだけを認識する必要が生じる。

本稿では、ナビゲーションシステムの中のジェスチャー認識の部分に焦点を置く。そして、常に不安定な動きを伴う飛行ロボットが、搭載したカメラの映像の中から、自分にとって意味のあるジェスチャーだけを検出し、その内容を判断するための方策について検討する。

2 問題の設定

2.1 飛行船

本研究では、飛行ロボットとして小型飛行船を用いる(図1参照)。小型飛行船は、他の飛行ロボットと比較してその搭載量の制約が厳しく、また、動きが不安定といった問題がある。しかし、人間に接触しても危害を加えることが無く、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、その制御法について論じたものが多い。制御に用いるセンサーとしては、屋内型の飛行船の多くがビジョンのみを用いているのに対し、屋外型ではGPS等のあらゆるセンサーを利用している。これは、飛行船の大きさに対する搭載量の制約の違いから生じた結果である。Zwaanらは、ビジュアルサーボを用いて、屋内に設置された目印を見失わないための安定した制御を

実現している²⁾。また、Fukaoらは、飛行船の位置認識に利用する特徴点の追従をより確実に行うことで、位置推定の精度を向上させ、安定した制御を実現している⁶⁾。飛行船と人間とがインタラクティブを行う研究例としては、人の拍手に反応する機能を取り入れた川村らの研究⁷⁾が挙げられる。しかし、全体としては散見される程度であり、本研究のように人間のジェスチャーを認識させる例は非常に少ない。

本研究では、市販されている全長約1mのラジコン飛行船を用いる。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、それにバッテリーを統合した駆動部から構成される。さらに、外界の情報を得るための2つの小型ワイヤレスカメラを搭載している。

飛行船は3枚のプロペラを独立に回転させることで、前後移動、回転移動、上下移動を行う。その制御命令の決定は外部のPCで行い、無線を通じて駆動部に送信される(図2参照)。そのため、ラジコン飛行船付属のコントローラを改造した。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1: 飛行船

2.2 ナビゲーションシステム

我々が開発している屋内ナビゲーションシステムは、天井の高い大型の展示場などで利用することを想定している。複数台の小型飛行船が空中を行き交い、なるべく多くのユーザが、飛行船をパーソナルロボットとして利用出来るシステムの構築を目指している。現在実装しているシステムの機能は、1) 目的地に向かってユーザを案内する、2) 予め登録されたユーザのジェスチャーを検出し、それに

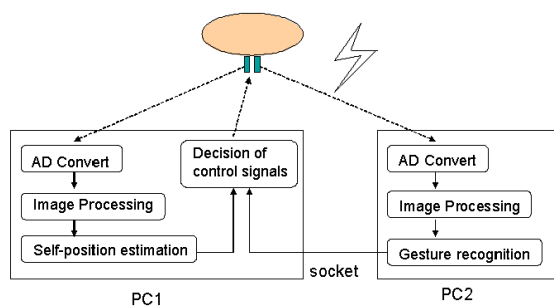


図 2: システム構成

対応した行動を行う、3) ジェスチャーに対する反応を終えると、再び目的地を目指して進む、といった機能である。

尚、本システムでは、大掛かりな機器は周囲に出来る限り設置せず、飛行船に搭載したセンサー情報を利用して飛行船を制御することを目指す。これは、展示場では内部の配置が頻繁に変わる場合が多いことから、環境の変化に柔軟に対応することを想定しているためである。

2.3 ジェスチャー認識の形態

本稿では、飛行船による人間のジェスチャー認識について論じる。ジェスチャーには、一般に静止ジェスチャーと時系列ジェスチャーが存在し、状況に応じて認識のしやすさが異なる。本研究で扱うナビゲーションシステムでは、歩いている人、止まって展示物を眺めている人、他の飛行船にジェスチャーを行っている人、など、あらゆる人が存在する。従って、静止ジェスチャーを用いても、止まって展示物を眺めている人との区別は容易では無い(図3参照)。また、一般的な時系列ジェスチャーを用いても、動いている人との区別が困難である。

そこで本研究では、両手を左右に振るなどの、周期性を伴う時系列ジェスチャーを用いる。これらのジェスチャーは、画面上を一定時間観測すると、ジェスチャーが行われている領域だけ、特定の周波数にピークを持つようなパワースペクトルが得られる(3.3節参照)。従って、ジェスチャーとジェスチャー以外の動きとの区別がしやすいという利点がある。ただし、本研究においては、以下の二つの問題が発生する。

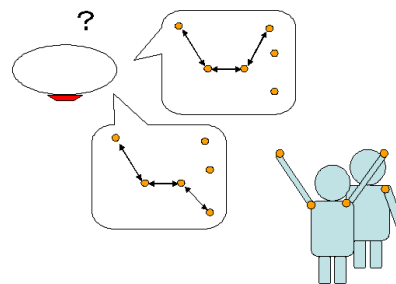


図 3: 複数人がいる場合

- 飛行船は常に動いているため、背景画像が変化する。
- 他の飛行船に対して行われているジェスチャーとの区別がつかない。

時系列ジェスチャーの周期性を検出し、その内容を理解するためには、ある固定された背景画像を基準とした上で、手の位置の変化を追跡することが望ましい。しかし、飛行船は常に動いているため、背景画像を固定することは容易では無い(図4上段参照)。これが前者の問題である。本稿では、ある時刻 t_0 のカメラ画像 (I_{t_0}) を基準とした場合、時刻 $t(t = t_1, t_2, \dots)$ の画像 (I_t) 上で確認された人間の部位が、 I_{t_0} 上では何処にあるのかを対応付けることを考える(図4下段参照)。そして、対応する画像領域を取得するために、飛行船の位置認識の結果を利用する手段を取る(3節)。また、後者の問題に対しては、ユーザのジェスチャーと飛行船の動きには何らかの相関があることを仮定して認識を行うことにする(4節)。

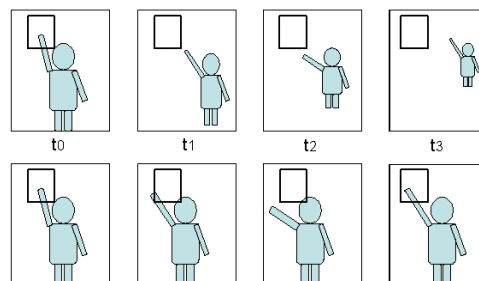


図 4: 背景画像の変化

尚、一般に人間のジェスチャー認識を行う場合は、手や顔の領域を肌色認識によって抽出して行う場合が多い。しかし、光の状況等によって抽出が安定しない場合がある。そこで本稿では、認識を容易にするために人間の手や肩に赤外線 LED を装着し、その光の位置を人間の部位として認識する。従って、飛行船に搭載した 2 台のカメラのうち、1 台は赤外線 LED の光を捉えるためのフィルタをかける。そして、飛行船の位置認識はもう 1 台のカメラの映像をもとに行う。

3 周期性のある時系列ジェスチャーの検出

3.1 画像領域の対応

飛行船は静止することが容易ではない。従って、静止物を見ていても、飛行船に搭載されたカメラの画面上では写り方が常に変化する。

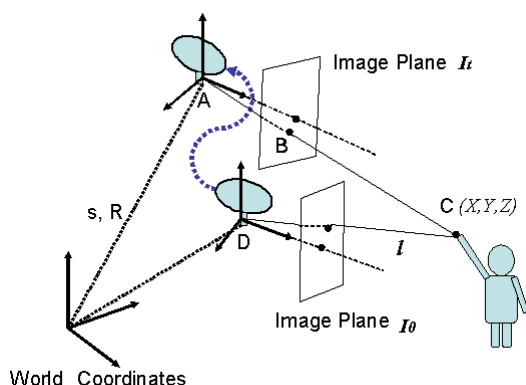


図 5: 画像領域の対応

前節でも触れた通り、時系列ジェスチャーの周期性を検出し、その内容を理解するためには、ある固定された背景画像を基準とした上で、手の位置の変化を追跡することが望ましい。今、世界座標系に対するカメラ座標系の位置と姿勢が、全ての時刻において推定出来ているとする。すると、以下のようにして画像領域の対応を行うことが出来る。

時刻 $t(t > t_0)$ での、世界座標系に対するカメラ座標系の位置と姿勢を s_t, R_t とする。ただし、 s_t は 3×1 の縦ベクトル、 R_t は 3×3 の行列である。ま

た、そのカメラ座標系の原点を点 A とする (図 5 参照)。このとき、画像平面に投影された人間の部位 (点 B) の世界座標系での位置 $v_t(3 \times 1$ の縦ベクトル) は、以下のようにして表すことが出来る。

$$\begin{pmatrix} v_t \\ 1 \end{pmatrix} = \begin{pmatrix} R_t & s_t \\ 0 & 1 \end{pmatrix} \begin{pmatrix} x_t \\ y_t \\ f \\ 1 \end{pmatrix} \quad (1)$$

ただし、 x_t, y_t は、時刻 t での画像平面における点 B の座標である。また、 f はカメラの焦点距離である。人間に装着している LED の高さを全て一定に近似すると、AB を結ぶ直線の式を用いることで、人間の部位 (点 C) の世界座標 (X, Y, Z) を一意に求めることが出来る。さらに、点 C の位置が求まると、今度は時刻 t_0 でのカメラ中心 (点 D) とを結ぶ直線 (l) を決定することが出来、その直線と時刻 t_0 でのカメラ平面 (I_{t_0}) との交点を求めることで、対応する画像領域を求めることが出来る。

ここでは、全時刻におけるカメラの位置認識結果が求まっていることが前提となっている。そこで以下では、本研究で用いる位置認識の方法について説明する。

3.2 位置認識

飛行船は搭載量の制約が厳しいことから、多くのセンサーを積むことが出来ないという問題がある。そこで我々は、搭載しているセンサーとしてカメラのみを用い、カメラ画像から自己位置を推定する。

画像から自己位置を推定する手法として、ランドマークを利用する方法がある。各ランドマークに環境内での絶対座標を登録しておくことで、ランドマークの見え方から比較的容易に自己位置を推定することが出来る。また、一般にランドマークの設置は容易であるため、環境の変化にも対応しやすいというメリットがある。しかし、ランドマークの数が増えるに従って誤認識を行う確率も増大する。従って、ランドマークを多様するという手段にも限界がある。さらに、本稿のような比較的大型の展示場で利用するシステムを想定した場合には、常に飛行船の視野に収まる位置にランドマークを配置することが出来るとは限らないという問題もある。

カメラ画像に写る特徴点の軌跡から、カメラの移動した軌跡を推定する手法 (SFM : Structure

from Motion) も数多く研究されている。しかし、ノイズに弱いといった問題や、対象物とカメラとの間の絶対的な長さを与える必要があるといったことから、それだけで自己位置を推定することは容易では無い。そこで本研究では、ランドマークの認識とSFMの手法を併用することで、自己位置の推定を行う。

3.2.1 因子分解法

SFMの手法は数多く提案されている。本稿では、代表的な手法の一つでもある因子分解法⁵⁾を利用する。因子分解法は、他のSFMの手法と比較してノイズの影響を受けにくい手法として知られている。また、カメラの運動軌跡と対象物の形状を同時に推定することが出来る。ただし、カメラと対象物との間の絶対的な距離(以下、スケール情報と呼ぶ)までは復元出来ない。従って、事前に与える必要がある。

基本的な因子分解法では、数フレーム間で見失わなかった特徴点の座標から、各フレームでのカメラの位置と姿勢をバッチ処理的に算出する(図6参照)。因子分解法は元々、カメラモデルとして平行投影モデルを仮定した方法であった。しかしその後、様々なカメラモデルに対する拡張が行われた。本研究では、透視投影モデルを良く近似するparaperspectiveモデルを仮定した因子分解法³⁾を利用する。

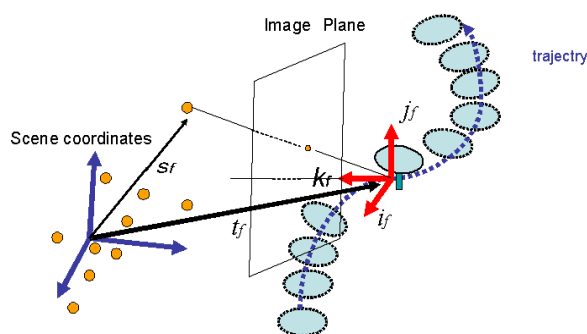


図 6: 因子分解法

3.2.2 ランドマーク認識と因子分解法の併用

本研究では、前節で紹介した因子分解法とランドマーク認識を併用することで位置認識を行う。ランドマークの認識を行うことで、因子分解法で利用

するためのスケール情報の獲得が可能になる。また、因子分解法を使い続けることによって積算される推定誤差をリセットする効果もある。一方、因子分解法を利用することで、ランドマークの数を抑えることが出来、ランドマークが見当たらない場所においてもある程度正確に位置を推定することが出来る。

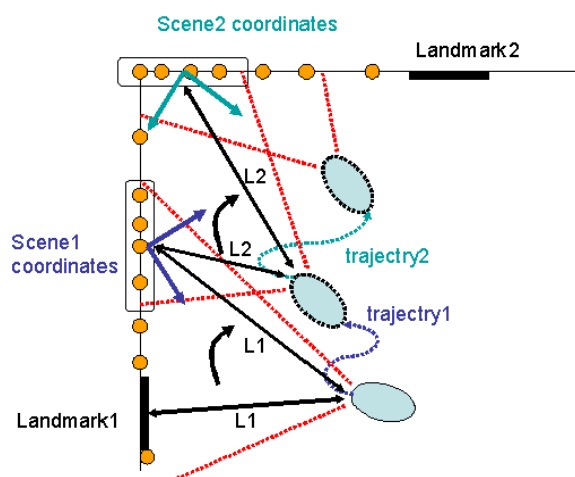


図 7: ランドマーク認識と因子分解法の併用

ランドマークが見えた場合は、その見え方から飛行船の位置を推定する。ランドマークを見失った場合は、特徴点の軌跡から因子分解法を用いて飛行船の軌跡を推定する。本研究では、15フレームに1回因子分解法を利用する。因子分解法で利用するスケール情報に関しては、完璧な環境地図を飛行船が予め持っていれば、現在の飛行船の位置と姿勢から対象物までのスケール情報を求めることが出来る。しかし、3次元の完全な環境地図を作成するためには多大なコストを要する。そこで本研究では、環境地図を用いずに以下のようにして近似的にスケール情報を与える(図7参照)。

(1) 因子分解法を使う直前(15フレーム前)までにランドマークが見えていた場合はランドマークと飛行船との距離(L1)をスケール情報として用いる。(2) 直前にランドマークが見えていなかった場合は、因子分解法で求めた15フレーム前の飛行船と対象物との距離(L2)をスケール情報として用いる。

3.2.3 特徴点の取得

特徴点の取得はコーナー検出を用いて行い、追跡は Lucas-Kanade 法を用いる。追跡に失敗する等で特徴点を見失った場合は、再びコーナー検出を行って特徴点を生成し、画面には常に 30 個の特徴点があるようにしている。尚、これらの処理は画像処理ライブラリである OpenCV を用いて行っている。

3.2.4 ノイズへの対応

追跡した特徴点の中には、外れ値が含まれることも多く、それらが位置推定結果に大きく悪影響を及ぼすことがある。そこで本研究では、Huynh の方法⁴⁾を用いて外れ値の除去を行っている。

また、無線カメラの電波にノイズが多く載った場合や、飛行船が激しく動いた場合等には、多数の特徴点の追跡に失敗してしまっ因子分解法を利用出来ないことがある。そのような場合には、直前までの飛行船の動きから現在の位置と姿勢を推定している。さらに、カメラの影響で、飛行船が回転運動しかしていない場合であっても推定結果に横移動成分が含まれてしまうことがある。しかし、我々が利用している飛行船は実際にはほとんど横移動をすることがない。そこで、推定結果から横移動成分のみを除去するように補正をかけている（図 8 参照）。

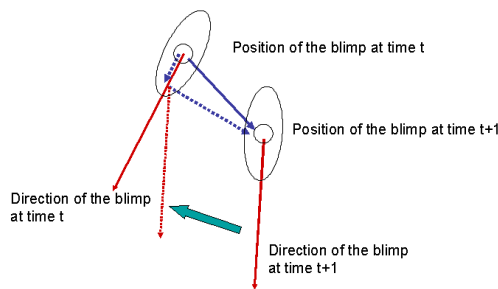


図 8: 補正

3.3 周期性のあるジェスチャー領域の検出

前節までの手法で、時刻 $t(t = t_1, t_2, \dots)$ における、人間の部位の背景画像 (I_{t_0}) への対応点を得た。これらの点の座標の変化から、FFT(高速フーリエ変換)を用いることで周期性を検出する。 I_{t_0} 上を小さな領域に分割し、それぞれの領域において FFT を利用する。周期性のジェスチャーが行われている

領域では、ある周波数でピークを持つようなパワースペクトルが観測されることになる（図 9、図 10 参照）。

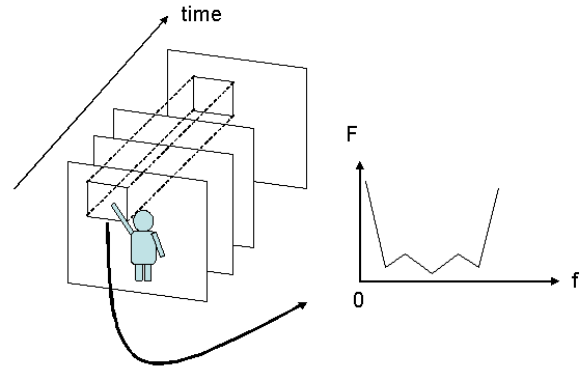


図 9: FFT を用いたパワースペクトルの導出

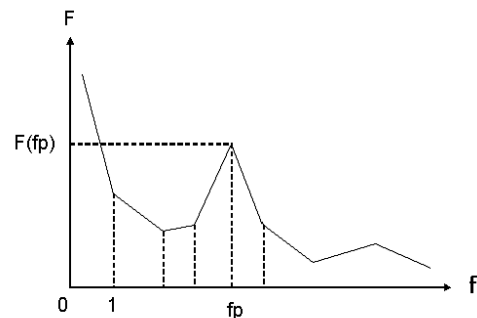


図 10: FFT によるピークの検出

4 議論

3 節で述べた手法を用いることで、理想的には、周期性を伴うジェスチャーの領域を抽出することが出来る。しかし、飛行船の位置推定誤差の影響などから、実際にはジェスチャーを行っていない領域を誤って抽出する場合がある。また、本研究が想定しているナビゲーションシステムでは、別の飛行船に対してジェスチャーを行っているユーザがいることも想定しており、自分には関係の無いジェスチャーを飛行船が観測することもある。そこで、周期性を伴うジェスチャーとして抽出された候補の中から、自分にとって意味のあるジェスチャーを検出し、その内容を理解することが飛行船にとって必要になってくる。

飛行船に対してメッセージを送ろうとしているユーザのジェスチャーは、その飛行船の動きに何らかの相関があるのではないかと我々は考える。そこで、この考えに基づき、以下の2通りの方法を用いて、意味のあるジェスチャーの検出・理解を行うことを検討している。

4.1 方法1

ユーザは、飛行船の動作を見て自分の行うジェスチャーを決める。従って、飛行船の行っている動作の変化(前進 左回転, 等)と、ユーザの行うジェスチャーの変化(ジェスチャーの種類の変更, ジェスチャーをする しない, しない する)には、他に周期的ジェスチャーの候補として挙げられたものと比較して、より強い関連性があるのではないかと考えられる。そこで、周期的ジェスチャーとして候補に挙げられているものの中で、飛行船が動作を変更した後の時間 S 内にジェスチャーの様子が変化した回数を数え上げる。これを、一定時間 $T(T > S)$ 繰り返し、その回数がある閾値以上の場合、意味のあるジェスチャーとする(図11参照)。

尚、飛行船は、位置認識の結果を用いることで、自分の動作が変化しているかどうかを確認する。ジェスチャーの種類は有限とし、それぞれに対応する飛行船の行動を登録しておく。飛行船の動作の変化は、これら有限個の中でクラスタリングして認識する。

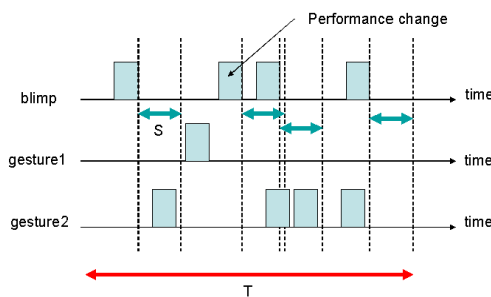


図 11: 方法 1

4.2 方法2

方法1では、飛行船にとって意味のあるジェスチャーの検出が可能であっても、ジェスチャーの種類を判別することまでは出来ない。そこで、ベイズ

推定の考え方をもとに、飛行船の動作に対するユーザのジェスチャーの変化を観測することで、ジェスチャーが表すユーザの意図について推定する方法を考える(図12参照)。

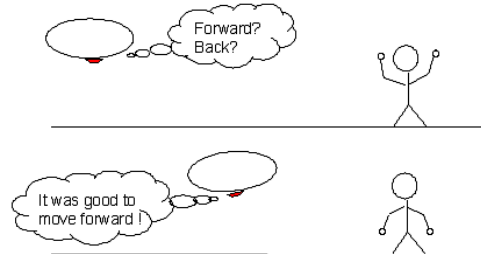


図 12: 方法 2

予め登録されたジェスチャーのシンボル集合を $X = \{x_1, \dots, x_K\}$, 対応する飛行船行動集合を $U = \{u_1, \dots, u_K\}$ とする。時刻 t の時点での飛行ロボットの行動を $u^t \in U$, 観測された人間の反応を z^t とする。すると、時刻 t の時点で、飛行ロボットが人間のジェスチャーをシンボル集合中の $x^t \in X$ のジェスチャーだと解釈している確率 $Bel(x^t)$ は次式で表すことが出来る。

$$Bel(x^t) = p(x^t | z^t, \dots, z^1) \quad (2)$$

ここで、マルコフ性を仮定することにより、次式が得られる。

$$p(x^t | z^t, \dots, z^1) = \sum_{x^{t-1}} p(x^t | x^{t-1}, u^{t-1}) Bel(x^{t-1}) \quad (3)$$

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。

$$Bel(x^t) = \eta p(z^t | x^t) \sum_{x^{t-1}} p(x^t | x^{t-1}, u^{t-1}) Bel(x^{t-1}) \quad (4)$$

ただし、 η は正規化のための係数である。

ここで、 $p(z^t | x^t)$ に関しては、HMM の viterbi アルゴリズム等を用いることで導出することが出来る。一方、 $p(x^t | x^{t-1}, u^{t-1})$ に関しては、複数の人間に対して実際に何度か飛行船へのジェスチャーを行ってもらい、その結果をもとに作成する。このような方法を導入することで、ジェスチャー認識に曖昧性が生じている場合であっても、ユーザの意図を

正しく理解することが出来るのではないかと考えられる。

5 まとめ

本稿では、飛行ロボットを用いた屋内ナビゲーションシステムにおける、ユーザのジェスチャー認識について報告した。始めに、周期性を伴うジェスチャーの、本システムにおける重要性について説明し、その認識のために重要となる、1) 背景の変化への対応法と、2) 飛行船にとって意味のあるジェスチャーの認識法、について検討した。現在は、上記の検討をもとに実験を進めているところである。今後は実験結果をもとに、ジェスチャーの認識精度を向上させていく方針について検討する予定である。

参考文献

- 1) 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
- 2) 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.
- 3) C. J. Poelman, T. Kanade, "A paraperspective factorization method for shape and motion recovery", In IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 19, No. 3, pp. 206-218, March 1997
- 4) D. Q. Huynh, A. Heyden, "Outlier Detection in Video Sequences under Affine Projection", Proc. IEEE Conf. on Computer Vision and Pattern Recognition, vol. 1, pp. 695-701, December 2001.
- 5) C. Tomasi, T. Kanade, "Shape and motion from image streams under orthography: a factorization method", International Journal of Computer Vision archive, Vol. 9, No.2, November 1992.
- 6) T. Fukao, K. Fujitani and T. Kanade: Image-based tracking control of a blimp, In Proc. 42nd IEEE Conference on Decision and Control, pp. 5414-5419 December 2004.
- 7) 川村 秀憲, 角田 久雄, 山本 雅人, 高谷 敏彦, 大内 東: "ホバリング制御に基づくエンターテインメントバルーンロボットの開発" エンターテインメントコンピューティング 2004 (函館), pp. 69-74 August 2004.