学術・技術論文

人間と機械のインタラクションを通じた
ジェスチャの理解と学習

村 嶋 照 久* 久 野 義 徳* 島 田 伸 敬* 白 井 良 明*

Understanding and Learning of Gestures through Human-Machine Interaction

Teruhisa Murashima*, Yoshinori Kuno*, Nobutaka Shimada* and Yoshiaki Shirai*

Humans can communicate each other by gestures. Even if they cannot recognize them at first, they can understand each other through interaction. This paper presents a robot system with such capability. The robot detects its user by recognizing his/her face. It can accept his/her commands given by gestures. The user may use unknown gestures to the robot. If the robot does not respond to his/her gesture, the user usually iterates the same gesture. The robot detects this repetitive pattern as an intentional gesture by which the user wants to give it some order. Then it shows a little action according to the meaning of the gesture that the robot guesses. It observes the user's reaction to its action. If he/she continues the same pattern gesture, the robot considers that its understanding is right, completing the action. It also registers the gesture pattern with the meaning. Otherwise, it iterates the same procedure taking another action as a candidate of the meaning. We have implemented such interactive capability on an intelligent wheelchair. It is convenient that we can make it come or go by gestures when we are off. Experimental results confirm that the proposed interaction method is useful in actual complex environments where even registered gestures cannot always be recognized.

Key Words: Interaction, Face Recognition, Gesture Recognition, Learning, Intelligent Wheelchair

1. は じ め に

人と共存し人の生活や仕事を支援する人間共存型ロボット [1]~[3] は、将来人間が生活を営む上で欠かせない存在になるであろう。そのとき、ロボットは人間とコミュニケーションをとることができなければいけない。日常社会で人間が共存できるのは、相手を認識し、情報を伝達し合うことによって、互いに相手の意思を理解しているからである。このようなコミュニケーションには言語によるものと非言語によるものがある。音声や文字を用いた言語によるコミュニケーションについては他多くの研究があるが [4] [5]、ここではジェスチャによる非言語コミュニケーションを検討する。ジェスチャを認識してヒューマンインタフェースに用いようという研究も多いが、そのほとんどは、あらかじめ「こういうジェスチャは、こういう意味」と取り決めておき、それを認識するものである [6]。人間同志の場合にもこのような定形のジェスチャは用いられるが、多くはその場で即興的に行われる。例として、言葉の通じない外国でのコミュニケーションの場合を考えてみよう。この場合は、ジェスチャがコミュニケーションの中心となる。うなずいたり、首

を振ったりというような、世界で広く共通するジェスチャもあるが、多くはその場で即興的に生み出される。したがって、一度では意味が理解できない場合も多い。しかし、お互いに何度かジェスチャで繰り返しやりとりをしているうちに、多分こういう意味であろう、というコンセンサスが生まれる。人間同志ではこのようなジェスチャによるインタラクションを通して、初めて見るジェスチャを理解していくことができる。

本論文では、ロボットにも人間と同様のインタラクションをすることにより、未知のジェスチャを理解し、学習する能力を与えることを検討する。これまでのジェスチャ認識手法では、事前に登録された定形パタンのジェスチャでも複雑な環境下で確実に認識を行うことは難しく、人間は同一のジェスチャをしたつもりでも、ロボットにとっては別の未知のジェスチャに見える場合がある。従来手法では、このような場合にロボットに指示を送ることができなくなってしまうが、インタラクションを通じて未知のジェスチャを理解する能力があれば、複雑な環境でもロボットとコミュニケーションできる可能性がある。

本論文では、以上のようなインタラクションを通じたジェスチャ理解と学習の方式を提案する。具体的なロボットの例として自動走行する車椅子を取り上げ、車椅子から降りているとき、ジェスチャで車椅子を操作することのできるシステムを開発した。このシステムを使って本手法の有効性を示す。

原稿受付 1999年5月20日

*大阪大学大学院工学研究科

*Graduate School of Engineering, Osaka University

2. ジェスチャによるコミュニケーション実現のための要素

前章で述べた外国でのコミュニケーションの例のように、人間同士の場合はたとえ未知のものであってもジェスチャなどの行動を互に行うことで相手の意図を次第に理解できる。このような機能を一般的な環境でロボットに実現するためには以下のような要素が必要であると考ええる。なお、これを実現する際にロボットとしては知的車椅子[7]をとりあげ、その追加機能としてジェスチャによるコミュニケーションを検討した。すなわち、車椅子から降りたときに、ジェスチャにより車椅子に指示を送る機能を考えた。レストランやその他公共の場では、騒々しかったり、大きな声で会話をするのがはばかれる場合がある。そういうときに車椅子から降りたときに、車椅子を邪魔にならなように移動させたり、必要なときに車椅子を呼ぶということをジェスチャで指示できれば便利であると考えたからである。

1. 顔の検出・認識, 2. 手の検出・追跡

今回は主として手によるジェスチャを検討する。一部、顔によるジェスチャも考える。そのため、ロボットとしては、まず手を検出しなければならない。ロボットの視野にジェスチャをする人間が入っていると仮定できれば、これは簡単である。これまでの多くのジェスチャを利用したシステムはこれを仮定している。しかし、ここで考えている車椅子の場合はそういう仮定はできない。車椅子の回りでは多くの人間が手を動かしながら動いていることが考えられる。そのなかから車椅子に指示を送っている手の動きを検出しなければならない。そこで、今回は車椅子にはあらかじめ登録された人が指示を送ると限定した。具体的にはその車椅子のユーザが指示を送るものとする。顔は肌色でそのなかに目鼻などの特徴があり、複雑なシーンのなかでも検出しやすい。検出された顔を認識し[8]~[10]、それがユーザであれば、顔の位置により手の存在範囲が限定される。そのなかで肌色で動いている部分を手として検出し、追跡する。このように顔の認識は複雑なシーンのなかでのジェスチャを検出するのに有効な手段になる。また、登録された主人の言うことだけを聞くロボットが実現されていることにもなる。

3. 登録されたジェスチャの認識

外国でもすべてのジェスチャが未知ということはない。うなずき・首を振るなどはかなり共通のジェスチャであり、このような共通のジェスチャの存在がコミュニケーションを助けてくれる。ロボットの場合もいくつかの基本的なジェスチャはあらかじめ登録し認識できるようにしておく必要がある[11][12]。また、未知の手の動きの意味がわかったときには、以後はこの認識法により認識できるようにしておく必要がある。なお、この認識法はジェスチャがいつ始まってでも認識できるように、すなわちセグメンテーションと認識の両者が自動的に行われるものでなければならない。

4. 指示を意図しているジェスチャ部分の切り出し

上記の認識法で認識できるジェスチャだけならよいが、実際にはユーザは登録していないジェスチャをするかもしれない。また、ユーザは登録したジェスチャをしているつもりでも、実

際の環境ではロボットからの観察方向の違いや、照明条件や背景の条件の違いで手の検出・追跡が登録のときのようにできておらず、ロボットにとっては未知のジェスチャに見える場合がある(実際にはこのようなことが非常に多い)。このように広い意味での未知なジェスチャに対応できるようにしたい。そのためには、まず人間が指示を意図して手を動かしていることを認識し、その手の動きを切り出さなければいけない。人間は絶えずいろいろと手を動かしているだろうが、そのなかから指示の意図をもって動かしている部分を検出するという要素技術が必要になる。ここでは「H1: 人間は1度のジェスチャに相手が応答してくれなければ何回か同じジェスチャを繰り返すだろう」というヒューリスティックを利用する。すなわち、人間の手の動きが繰り返しパターンになっていることを検出したら、その繰り返しの1周期がある指示ジェスチャをしているのだろうと考えて切り出す。

5. 呈示する行動の選択, 6. 人間の反応の認識

前述の3でジェスチャの認識が信頼性高く行えた場合は、そのジェスチャで指示される行動をロボットは行う。そうでない場合は4の処理が行われるが、そこで指示のジェスチャをしているとわかっても、その意味はまだわからない。音声が使えそうな状況なら、ロボットが音声で尋ねてもよいが、ここでは知らない外国語の場合のように、ロボットが人間が指示していると推測される動作を少し示し(ロボットのジェスチャ)、それに対する人間の反応をロボットが見て、推測したジェスチャの意味が正しいかどうかを理解することを検討する。人間の場合は、人間同士の共通な知識を活用してこの推測をかなりうまく行うことができるが、ロボットには難しい。ここでは、登録されたジェスチャボタンにすべての特徴が一致するとは言えないが、一部一致するものを推定候補とする。そのようなものがない場合は、ロボットがその場でできる行動はある程度限られている(車椅子の場合は非常に限定されている)ので、そのなかからあらかじめ定めておいた順に選択することにする。

ロボットは推定した行動を示した後、人間の反応を観察する。もし、人間が、例えばうなずいて肯定を示すなど、ロボットが3の機能で認識できるジェスチャを示したらそれに従う。そうでない場合は「H2: 指示に合致した行動を示したときは同じジェスチャかそれを少し小さくしたジェスチャ(もう少しという意味で)を行う、指示と異なるときは、他のジェスチャを行う。また、一般に意思が伝わらないとジェスチャが大きくなる」というヒューリスティックを用いて判定する。

以下、上記の要素をどのように実現したかを示す。顔とジェスチャの認識については基本的には既存の手法を利用した。上にあげたヒューリスティックH1, 2に基づき、ジェスチャによる対話を通じてジェスチャを理解・学習していく部分が本論文の中心である。そして、最後に知的車椅子を用いた実験システムとそれによる実験結果を示す。

3. 顔の検出・認識

登録されているユーザを検出・認識する方法として、今回は固有空間法[13]を用いた顔の認識を利用する。

3.1 学習段階

学習段階として、さまざまな位置や方向から撮影したユーザの顔画像を準備し、すべての画像に対して、大きさの正規化、人間の顔型に沿うよう決まった雛型での背景部分の除去、輝度値の平均と分散に基づく階調値の正規化の三つの処理を行う。そして、これらを学習サンプル画像として固有空間（特徴空間）を作成しておく。

3.2 認識段階

初めに広角で撮影した観測画像のなかから、肌色で、フレーム間差分で差の大きい動いている領域を抽出する。このとき、顔の全領域が抽出されずに顔の中心部が空洞部分として残ることがあるので、小さな空洞部分を領域として埋める処理を行う。そして、現フレームから過去 12 フレームまでの履歴を見て、同じ位置に領域として抽出されたことが、ある回数以上あれば、その位置は顔の候補領域の一部であると考え、このようにして顔らしい物体の候補領域を抽出する。

次に、カメラを動かしズームインすることにより顔らしい候補領域をそれぞれ注視し、先ほどと同様に肌色と動きから探索領域を抽出する。そして、注視した探索画像内から、いろいろな大きさ、いろいろな場所で矩形の対象画像を切り出し、それぞれの対象画像内に肌色が存在するか、上に髪に相当する黒い領域が存在するかを見て、その対象画像に顔が写っているか判断する。顔が写っていると判断された対象画像は、学習段階と同様の正規化を行い、特徴空間に投影する。

対象画像を特徴空間に投影してからは、まず対象画像の投影点と最も近い位置に存在する学習画像との距離 'distance in feature space' (DIFS) [14] を求める (Fig. 1)。このとき、対象画像の投影点と特徴空間 F 上のある学習画像の DIFS が小さな値をとったとしても、余次元の空間 $\bar{F}$ 上での距離 'distance from feature space' (DFFS) が近いかわからないため、対象画像に特定のユーザの顔が実際に写っているとは断定できない。そこで、投影点に最も近い学習画像データを線形逆変換により復元した学習画像と対象画像との差の二乗和 (sum of squared distance, SSD) を計算する。そして、得られた DIFS と SSD が共に所定の値より小さな値であれば、その対象画像に特定ユーザの顔が写っていると考え、実際に、すべての対象画像に対して DIFS と SSD を計算しては、計算量が膨大になるため、DIFS の値が小さい対象画像を上位数十枚を選

出し、その選出画像のみの SSD を計算することにより計算量を軽減する。

4. 顔と手の検出・追跡

人間は顔や手といった部分を動かしてジェスチャを行う。そこで、ジェスチャ認識のために、顔や手の検出・追跡を行う。

顔の追跡は以下の方法で行う。顔の認識が終了した直後は、認識された顔の位置をもとに設定した探索領域内で、認識結果より得られた結果画像をテンプレートとして、パターンマッチングを行うことにより実現する。それ以降は、前フレームの顔の位置をもとに設定した探索領域内で同様にマッチングを行い実現する。

手の追跡では、初めに 3.2 節と同様に肌色と動き情報から手の領域を抽出する。このとき、顔の認識直後は、ユーザを注視している画像領域から認識結果の顔領域を除いた領域より手の領域を抽出する。ただし、顔の位置が Fig. 2 (a) のように画像上方に大きく写るようにあらかじめカメラを制御しておく。その次からは、Fig. 2 (b) のように、顔の追跡により求められた顔領域と前のそれぞれの手の位置を中心とした半径 R の円の範囲外の領域を除去した領域内より手の領域を抽出する。

次に、ユーザが動かしている手は、両手か片手か判断する。抽出された手の領域に対して横軸方向の射影を求め、二つの領域に分割したときの距離があるしきい値以上であれば両手として分割する。もし横軸方向に対して領域分割できなければ、同様に縦軸方向に対して分布を求めて分割する。そして、横軸・縦軸どちらに対しても二つに分割することができなければ、片手のみが検出されたとする。Fig. 3 に、ユーザが手を左右で動か

Fig. 2 Hand search regions

Fig. 1 Eigenspace

Fig. 3 Hand region extraction

Fig. 4 Features for gesture recognition

Table 1 Known gestures

Face gesture	Right hand gesture	Both hand gesture
Yes, No	No, Bye-bye, Come, Back, Right, Left	Circle

している場合と手を上下で動かしている場合の領域の分割結果を示す。図中において黒領域が抽出された手の領域を表し、点線が領域の端を表し、破線が分割された位置を表している。

そして、両手と判断された場合はそれぞれの領域、片手と判断された場合はその領域の重心を手の位置とする。手は前フレームの左右の手の近い方から移動してきたと考えて前の手の位置からの距離を計算し、左手と右手の判定を行う。もし得られた手が片手のみの場合は、もう一方の手は移動しなかったと考え、現フレームでの手の位置は前フレームでの位置とする。

5. ジェスチャの認識

ジェスチャの認識は、ジェスチャ区間をセンシングメントンする必要がなく、認識がリアルタイムに行える連続 DP を用いたマッチング法 [15] で行う。

認識の際には、追跡によって得られた顔と手の画像中での位置に基づく特徴量を用いる。対象人物の画像中での位置ずれを考慮して、顔によるジェスチャの場合は、Fig. 4 (a) のように前の画像内での顔の位置から現フレームでの位置までの移動量の特徴量として用いる。手によるジェスチャの場合は、Fig. 4 (b) のよう現フレームで得られた顔の位置を原点とした相対座標値の特徴量として用いる。顔を注視するときにある一定の大きさに写るようにカメラを制御して大きさの正規化を行っているため、人物の大きさによる影響はないと考えられる。

今回は、Table 1 にある 9 種類のジェスチャをあらかじめ登録して、認識できるようにした。

6. 未知ジェスチャの切り出し

人間の返答ジェスチャは多彩であるため、ロボットにとって未知なジェスチャに遭遇することが起こる。そのような場合でも、ロボットが対応できるようにするために、ヒューリスティック H1 を基に、その未知ジェスチャを切り出し、以後そのジェスチャを認識できるように学習する方法について述べる。

Fig. 5 Distance from the face to the hand

6.1 ジェスチャの切り出しアルゴリズム

H1 によると、何か意図を持ったジェスチャは何回か繰り返される。そこで繰り返しパターンを持つ手の動きを検出する。Fig. 5 は、左手で「来い」を意味するジェスチャを繰り返し行ったときの顔から手までの距離の時間変化 $d(t)$ である。Fig. 5 の縦軸は顔から手までの画像上での距離（単位画素）を表し、横軸は時間（単位は 1/30 秒（ビデオレートの 1 フレーム時間））を表す。グラフは綺麗な周期関数になっている。そこで、まずその関数に対し次式により自己相関を計算し、極大点と極小点を求める (Fig. 6)。

$$\kappa = \frac{\sum_t (d(t) - \bar{d})(d(t + \tau) - \bar{d})}{\sqrt{\sum_t (d(t) - \bar{d})^2 (d(t + \tau) - \bar{d})^2}} \quad (1)$$

そして、極大点・極小点・極大点の順になる区間の区間の平均と分散を計算する。ジェスチャが繰り返されたときは、谷の数が多く区間区間の分散の値が低くなるので、谷の回数が多いとき、分散値が低いとき、ジェスチャが繰り返し行われたと判定し、区間区間の平均値 $\bar{d}$ をジェスチャの周期とする。今回、谷の数のしきい値を 4 回、分散値のしきい値を 2.0 とした。この谷の回数は、繰り返し回数に相当するので、今回のシステムでは意志がすぐに通じないときは、人間は 4 回以上連続的にジェスチャを繰り返すものと仮定していることになる。意志が伝わらないときにこの程度の繰り返しは期待できると考えられる。ただし、この間はジェスチャを中断することはできない。

次に、距離の時間変化グラフを 1 周期分ずらしたグラフと元のグラフとの誤差 $diff_i$ を計算し、誤差が初めてあるしきい値を下まわった時刻をジェスチャの開始時刻とする。

$$diff_i = \frac{\sum_{i=-\bar{d}/4}^{\bar{d}/4} |d(t+i) - d(t+\bar{d}+i)|}{\bar{d}/2} \quad \bar{d}/4 \leq i \quad (2)$$

以上の処理によって得られた周期と開始時刻を用いて Fig. 7 のように分割し、それぞれの区間内の位置データをまとめて一つの区間ジェスチャとする。そして、区間ジェスチャの平均 $\bar{g}_j$ を未知ジェスチャのパターンとする。

$$G_i = \{g_1^i, g_2^i, \dots, g_{\bar{d}}^i\} : i = 1, \dots, M \quad (3)$$

Fig. 6 Autocorrelation result

Fig. 7 Segmentation

$$\bar{G} = \{\bar{g}_1, \bar{g}_2, \dots, \bar{g}_\delta\} \quad (4)$$

$$\bar{g}_j = \frac{1}{M} \sum_{i=1}^M g_j^i \quad (5)$$

Fig. 8 にその結果を示す。太線が切り出した未知ジェスチャを、その他の線が区間ジェスチャを表している。座標系は顔の位置を原点としている。

このジェスチャの切り出しアルゴリズムは、ある一定時間区間に得られるデータに対して適用する。仮にその区間に未知ジェスチャが切り出せなければ、この区間には繰り返しジェスチャが存在しないと考える、次の時刻からデータを一定区間集めてデータを更新し、同様にその区間のデータに対してアルゴリズムを適用する。

この方式では、一つのジェスチャがあるパタンの繰り返しの場合には、その要素パタンをジェスチャとしてしまう可能性がある。したがって、繰り返し回数によって違う意味を与えるようなジェスチャには対応できないことになる。一般には大きな問題ではないと考えられるが、提案方式の期限事項である。

Fig. 8 Segmented gestures and their mean

Fig. 9 Route path

6.2 連続 DP 用のしきい値の設定

今後切り出したジェスチャを連続 DP を用いたマッチング法で認識するためには、ジェスチャが行われたかどうかを判定するためのしきい値が必要である。そのしきい値を式 (6) のように平均ジェスチャと区間ジェスチャの差の最大値に重みを掛けた値に設定する。この計算方法は、DP 計算の際 Fig. 9 のように遷移してきたときの計算方法と一致する。今回は w を 1.4 の値に設定した。

$$th_{dp} = w \max_{i=1, \dots, M} \{|\bar{G} - G_i|\} \quad (6)$$

$$|\bar{G} - G_i| = \sum_j^{\delta} |\bar{g}_j - g_j^i| \quad (7)$$

6.3 片手・両手の未知ジェスチャの切り出し

手を使ったジェスチャには片手を用いたジェスチャと両手を用いたジェスチャが存在するため、未知ジェスチャを切り出すためには右手・左手・両手のジェスチャの 3 種類に対して 6.1 節の切り出しアルゴリズムを適用しなければいけない。そこで、Fig. 10 のような流れで適用する。

はじめから判定処理を行っているのは右手と左手だけで、両手は距離値データを集めるのみである。ここで、両手の距離値データは両方の手の距離の和とする。そして、ある片手の未知ジェスチャが得られたとき、もう一方の手の状況を確認しに行き、得られたジェスチャに近い周期である程度繰り返されている

Fig. 10 Flowchart of unknown gesture obtaining algorithm

たら、このジェスチャは両手によるものと判断して、このときはじめて両手の切り出しアルゴリズムを行う。両手によるジェスチャではないと判断されると片手のジェスチャとする。

7. 呈示する行動の選択と人間の反応の認識

7.1 行動の選択

未知ジェスチャが切り出されたとき、どのような行動をユーザが指示しているかわからないが、何かしらの行動を要求していることはわかる。そこで、何らかの行動を選択して人間に呈示し、反応を見る。初めに選択した行動が、人間が意図したものであれば、インタラクションの回数が減り、効率的である。ここでは、そのために次のような選択処理を行う。入力パターンは、認識に用いる特徴量に関して既知パタンのどれとも類似していないため、未知ジェスチャとなっている。そこで、この特徴量の限定を弱めて、そのなかで類似しているものはないか探索する。最初の認識の場合には、顔を中心とした座標系での位置を特徴量としている。したがって、顔に対して異なった場所で、同じような手の動きをしても、違うジェスチャと判定される。そこで、このジェスチャ全体の並進位置ずれを除いた移動量のパターンを調べ、一番一致しているものを行動として選択する。具体的には、これを以下のようにして行う。

はじめに記憶しているジェスチャの相対座標値標準パターンから移動量標準パターンをそれぞれ作成する。次に、それぞれの移動量標準パターンに対して、切り出されたジェスチャの移動量を順番に2サイクル入力して連続DPを計算する。そして、それぞれのジェスチャに対して、連続DPの出力のなかで最小値を求める。以後、この最小値を最小DP値と呼ぶ。この最小DP値が低いほど切り出されたジェスチャと類似していると考えて、呈示順序を決める。ただし、片手によるジェスチャが切り出されたときは、右手・左手の記憶ジェスチャと比較し、両手によるジェスチャが切り出されたときは、両手の記憶ジェスチャと比較する。

以上の方法で類似した既知のジェスチャがない場合は、ロボットに可能な行動のなかから順番に選ぶようにする。後述の車椅子の場合は、前進・後退・右に回る・左に回るの四つの行動が基本である。そのため、命令はCome・Back・Right・Leftの4種類に限られるので、この順に行動が選ばれるようにした。

7.2 人間の反応の認識

ユーザに呈示した行動が正しかったかどうか判断するために、ヒューリスティック H2 に基づき、前回の切り出されたジェスチャと今回の切り出されたジェスチャの二つを比較して、パタンの類似と大小関係を見ることにより、ユーザの反応を認識する。以下、前回のジェスチャをジェスチャA、今回のジェスチャをBとする。

まずそれぞれのジェスチャのデータから重心を計算し、重心から最も遠いところに位置するデータとの距離(d^A, d^B)を求める。次に、式(8)のように距離の比率 ϵ を用いてジェスチャBを正規化する。

$$\widetilde{g_i^B} = \epsilon(g_i^B - g_B) + g_B \quad (8)$$

$$\epsilon = \frac{d^A}{d^B} \quad (9)$$

そして、ジェスチャAを標準パターンとして、正規化したジェスチャBのデータを順番に2サイクル入力し、最小DP値 dp_{min} を求める。ただし、DP計算に用いる特徴量は相対座標値である。比較判定は、最小DP値と比率 ϵ 、ジェスチャAの連続DP用のしきい値 th_A 式(6)を用いて、 $dp_{min} \leq th_A$ かつ $\epsilon \leq 1 - th$ ならスケールアップ、 $dp_{min} \leq th_A$ かつ $1 - th \leq \epsilon \leq 1 + th$ なら一致、 $dp_{min} \leq th_A$ かつ $\epsilon \geq 1 + th$ ならスケールダウン、その他は異なるパタンのジェスチャとする。 th は正の定数である。今回の実験では0.15とした。

比較結果がスケールダウン、または一致なら、ユーザに呈示した行動で正しかったと判断し、スケールアップなら行動は誤りであったと判断する。異なるジェスチャの場合は、まったく情報が得られないのでこれ以上同じ行動を続けても何も進展しないと判断し、別の行動をとる。

ただし、ロボットはジェスチャの切り出し、および、ジェスチャの認識を常に同時に行っている。したがって、ユーザがロボットの行動に対する応答を、顔によるジェスチャ、「うなずき」「顔を振る」などで返してきたときでも、そのジェスチャが既知のジェスチャであれば認識でき、ユーザの反応を理解することができる。ロボットの行動に対する回答として考えられないジェスチャが認識された場合は、情報が得られないので、次の選択行動をとる。

8. 未知動作の学習実験

8.1 実験システム

本研究で試作した知的車椅子の概観をFig. 11に示す。カメラ(Evi-G30 SONY)から撮影された画像を画像処理装置(IMAP-VISION)[16]に入力する。画像処理装置上で、顔を検出するときは顔の候補領域・探索領域の検出、ジェスチャの認識のときは顔と手の追跡を実時間でやり、ホストコンピュータ(PC-AT 互換機 CPU: MMX Pentium 233MHz OS: FreeBSD2.2.5)は画像処理装置から必要なデータを受け取り顔の検出・ジェスチャ認識を行う。そして、必要に応じて画像処理装置へのデータの送信やカメラの制御、ロボット(MC-16 SUZUKI)の移動などを行う。顔の候補領域・探索領域の検出と顔と手の追跡は、それぞれ1/15、1/6秒で処理を行って

Fig. 11 System overview

(a) Original image (b) User detection result

Fig. 12 Face recognition to detect the user (person A)

いる。ただし、顔と手の追跡処理時間は、顔のテンプレートのサイズが 50×50 のときに得られた結果である。

8.2 実験結果

顔認識・手検出の実験

Fig. 12 に示すように 3 章に述べる方法で、人間の顔を検出し、それが主人であるか認識できた。そして手の部分を検出・追跡できた。以後は、本論文の中心であるそれ以降のインタラクションによるジェスチャ認識の実験について述べる。

実験 1

はじめにユーザがロボットがあらかじめ記憶していない、「来い」という意味を持つ両手によるジェスチャ (Fig. 13) を行ったときの実験結果について示す。はじめの観測でロボットは Fig. 14 の実線で表す未知のジェスチャのパターンを切り出した。このとき両手によるジェスチャは Circle のみしか記憶していなかったため、まったくの未知動作ということで、あらかじめ定められた順番により Come を行動として選択し、ロボットは 50 [cm] 前進しユーザの応答を観察した (Fig. 15 (b))。ユーザが Fig. 14 の破線で表すジェスチャで応答したため、ロボットは先ほどの行動は正しかったと理解し、ユーザの要望通りにユーザの元に向かった (Fig. 15 (c))。そして、このジェスチャを前進の意味のものと登録した。

Fig. 13 Repeated gesture

Fig. 14 Obtained unknown gestures

実験 2

実験 1 の後、先ほどと同じ Fig. 13 のジェスチャを行ったときの実験結果を以下に示す。ロボットは、先ほどは記憶していなかったために認識することはできなかったが、実験 1 で Fig. 13 のジェスチャを記憶したので、Fig. 16 において時刻 $t = 25$ 辺りで認識することができた。Fig. 16 は、Circle ジェスチャと Fig. 13 の両手を用いて行うジェスチャ (Coming gesture) の認識結果を示している。縦軸が DP 値を、横軸が時間を表している。時刻 $t = 25$ 辺りで、Fig. 13 のジェスチャの DP 値が、そのジェスチャ固有のしきい値を下回ったので認識された。その結果、ユーザの元に直行した。

Fig. 15 Robot's action in the experiment 1

Fig. 16 Result of gesture recognition

実験 3

実験 1 と同じように未知の左手による Fig. 17 のジェスチャを行ったときの実験結果を示す。そのジェスチャはユーザからみて「右の方向に行け」「右の方向を向け」を意味する。ロボットは、Fig. 18 の実線で表す未知のジェスチャのパターンを切り出した。実験 1 とは異なり片手による記憶ジェスチャのなかには 4 種類の行動を意味するものが含まれているので、7.1 節で述べたように移動量パタンの類似度の順に未知動作に対する行動選択が Fig. 19 に示す順番に決められた。そこで、まず左に 7.5 度回転しユーザの応答を観察した (Fig. 20 (b))。ユーザの応答が手による否定のジェスチャであったために、次の行動として右に 7.5 度回転を行った (Fig. 20 (c))。このときのユーザの応答が Fig. 18 の破線で表すジェスチャであったため、ロ

Fig. 17 Repeated gesture

Fig. 18 Obtained unknown gesture

ボットは先ほどの行動が正しいと理解し、その方向への回転を続けユーザの要望通りにユーザから見て右に 30 度回転した (Fig. 20 (d))。

9. ま と め

ジェスチャの意味を推定し、その推定した意味を行動で呈示し、それに対する反応を見るという人間とのインタラクションを通じ、未知のジェスチャでも認識できる方法を提案した。そして、知的車椅子システムに実装し、有効性を確認した。

このインタラクションの過程で最も問題になるのは、ジェス

Fig. 19 Behavior ordering result

Fig. 20 Robot's action in the experiment 3

チャの意味の推定部分である。人間の場合は、多くの知識を用いて、相手のジェスチャの意味をかなり推測できる。ロボットの場合はこれが難しい。現在は手の運動パターンが見かけ上一番近いものを推定候補としているだけである。車椅子の場合は行動が四つしかないで、これで推定が間違っている、少しインタラクションが増えるだけで大きな問題ではない。しかし、より多機能のロボットの場合は問題になる。作業目的や現在の状況などから、人間がロボットに伝えたいだろうことを限定するような処理をジェスチャ認識に合わせて検討する必要がある。また、ロボットが人間の意図を理解するとき、今回は人間のジェスチャの大小関係から判断した。しかし、人間はジェ

スチャの大小関係のみでなく、手を早く／遅く動かすといったジェスチャのスピードからも相手の意図を理解することができる。今後は、さらにいろいろな情報をもとに人間の意図を理解するシステムの検討も必要である。その他、今回のシステムでは単純な行動しか実現できていないが、行動スキルを組合せたり、副詞、形容詞を付加してより複雑な行動を行えるシステムも検討していきたい。

最後に、今回はジェスチャの研究という点からジェスチャに限ったが、音声言語が使える環境なら、それと組み合わせれば、より効率的なシステムが実現できると考えられる。

謝 辞 本研究の一部は、文部省科学研究費 (09555080)、倉田奨励金、栢森情報科学振興財団の補助を受けた。ここに深く感謝する。

参 考 文 献

- [1] 特集「人間共存型ロボット」, 日本ロボット学会誌, vol.16, no.3, 1998.
- [2] T. Takahashi, S. Nakanishi, Y. Kuno and Y. Shirai: "Helping Computer Vision by Verbal and Nonverbal Communication," International Conference on Pattern Recognition, vol.2, pp.1216-1218, 1998.
- [3] H.J. Boehme, A. Brakensiek, H.M. Gross, U.D. Braumann and M. Krabbes: "Neural Architecture for Gesture-Based Human-Machine-Interaction," International Gesture Workshop, pp.219-232, 1997.
- [4] 竹林洋一: "音声自由対話システム TOSBURGII-ユーザ中心のマルチモーダルインタフェースの実現に向けて", 電子情報通信学会論文誌, vol.J77-D-II, no.8, pp.1417-1428, 1994.
- [5] 高橋和子, 竹澤寿幸: "マルチモーダル対話システムにおけるインタラクション機構", インタラクション'99 論文集, pp.89-96, 1999.
- [6] V.I. Pavlovic, R. Sharma and T.S. Huang: "Visual Interpretation of Hand Gestures for Human-Computer Interaction: A Review," Pattern Analysis and Machine Intelligence, vol.19, no.7, pp.677-695, 1997.
- [7] 足立佳久, 中西知, 久野義徳, 島田伸敬, 白井良明: "顔の視覚情報処理を用いた知的車椅子", 日本ロボット学会誌, vol.17, no.3, pp.423-431, 1999.
- [8] B. Takacs and H. Wechsler: "Detection of Faces and Facial Landmarks using Iconic Filter Banks," Pattern Recognition, vol.30, no.10, pp.1623-1636, 1997.
- [9] 堀修: "統計的手法を用いた様々な背景からの顔領域抽出", 情報処理学会研究報告コンピュータビジョンとイメージメディア 106-19, pp.139-146, 1997.
- [10] 田中英治, 山口修, 福井和広, 前田賢一: "知的なインタフェースのための視覚機能実現-動画を用いた顔認識システム-", 情報処理学会研究報告ヒューマンインタフェース 75-10, pp.51-56, 1997.
- [11] T. Kurita and S. Hayamizu: "Gesture Recognition using HLAC Features of PARCOR Images and HMM based Recognizer," International Conference on Automatic Face and Gesture Recognition, pp.422-427, 1998.
- [12] T. Watanabe and M. Yachida: "Real Time Gesture Recognition Using Eigenspace from Multi Input Image Sequences," International Conference on Automatic Face and Gesture Recognition, pp.428-433, 1998.
- [13] 村瀬洋, S.K. Nayar: "二次元照合による三次元物体の認識とその学習-パラメトリック固有空間表現-", 電子情報通信学会技術研究報告, vol.93, no.431, pp.31-38, 1993.
- [14] B. Moghaddam and A. Pentland: "Maximum Likelihood Detection of Faces and Hands," International Workshop on Automatic Face and Gesture Recognition, pp.122-128, 1995.
- [15] 西村拓一, 向井理朗, 野崎俊輔, 岡隆一: "低解像度特徴を用いた複数人物によるジェスチャーの単一動画からのスポッティング認識", 電

子情報通信学会論文誌, vol.J80-D-II, no.6, pp.1563-1570, 1997.
 [16] 岡崎信一郎, 藤田善弘, 許昭倫, 山下信行: “1 ボード 超高速動画画像

処理システム IMAP-VISION”, 第 3 回画像センシングシンポジウム講演論文集, pp.201-206, 1997.

村嶋照久 (Teruhisa Murashima)

1973 年 4 月 12 日生。1997 年, 大阪大学工学部電子制御機械工学科卒業。1999 年, 同大学大学院修士課程修了。
 (日本ロボット学会正会員)

島田伸敬 (Nobutaka Shimada)

1969 年 12 月 30 日生。1992 年大阪大学工学部・電子制御機械工学科卒業。1994 年同大学大学院工学研究科電子制御機械工学専攻博士前期課程修了。1997 年同後期課程修了。博士(工学)。同年, 大阪大学工学部電子制御機械工学専攻助手。コンピュータビジョンとそのヒューマンインタフェースへの応用に関する研究に従事。情報処理学会, 電子情報通信学会各会員。

久野義徳 (Yoshinori Kuno)

1954 年 4 月 13 日生。1977 年, 東京大学工学部電気工学科卒業。1982 年, 同大学大学院博士課程修了。同年, (株) 東芝入社。1987~1988 年, カーネギーメロン大学計算機科学科客員研究員。1993 年 4 月より大阪大学工学部電子制御機械工学科助教授。コンピュータビジョンおよびそのロボットやヒューマンインタフェースへの応用に関する研究に従事。工学博士。情報処理学会, 電子情報通信学会, 日本機械学会, 人工知能学会, 計測自動制御学会, IEEE 各会員。
 (日本ロボット学会正会員)

白井良明 (Yoshiaki Shirai)

1941 年 8 月 3 日生。1964 年名古屋大学工学部機械工学科卒業。1969 年東京大学大学院工学系博士課程修了, 工学博士。同年, 電子技術総合研究所入所, コンピュータビジョン, ロボティクスの研究に従事。1971 年~1972 年, MIT AI ラボ客員研究員。1988 年大阪大学工学部電子制御機械工学科教授。人工知能学会, 電子情報通信学会, 情報処理学会, 日本機械学会各会員。
 (日本ロボット学会正会員)