

人行動理解のためのジェスチャ認識アルゴリズムと 新型「ASIMO」への適用

Gesture Recognition Algorithm for Understanding of Man-Machine Interaction
and Its Adaptation for New ASIMO

檜 垣 信 男*
Nobuo HIGAKI

横 山 太 郎*
Taro YOKOYAMA

藤 村 希久雄**
Kikuo FUJIMURA

要 旨

二足歩行ヒューマノイドロボット「ASIMO」の自律性を高めるため、人応答機能の一部として画像処理によるジェスチャ認識機能を開発した。人の顔と手の実空間位置の時間的空間的变化を特徴量として統計的手法を用いてジェスチャを認識する。左右それぞれ10種類のジェスチャおよびポスチャの認識が可能となった。また、手付近で観測された3次元点群の主成分分析によって腕の向きを求め、人の頭と腕の位置関係および腕の向きから、人が指示しているポイントを抽出することが可能となった。これらの処理は、階層化によって処理の複雑さを回避し、各処理間において情報を有効に利用することで計算コストを削減し、その結果、オンボードで動作する総合的な画像処理システムを実現した。

ABSTRACT

An image-based gesture recognition module has been developed as a part of biped walking humanoid robot ASIMO's man-machine interaction function in order to enhance its autonomy. ASIMO recognizes human gestures based on statistical analysis, using spatio-temporal variance of three-dimensional(3D) positions of the face and hand. It is able to distinguish 10 kinds of gestures and postures on left and right side each. It is also able to recognize a position pointed by an operator from the relative position of operator's head and hand and from the direction of the arm. This is calculated by PCA(Principal Component Analysis) on a 3D-point set observed in the proximity of his arm. These components are implemented in a layered structure to reduce complexity, and in optimally designed communication structure to minimize processing resources. This effort has resulted in a comprehensive image processing system in ASIMO's onboard computer.

1. はじめに

Honda は 2000 年 10 月末に人の身近なところで実際に人の役に立つ存在を目指し、人間の生活環境に適合できるミニマムなサイズのヒューマノイドロボット「ASIMO」を誕生させた⁽¹⁾。そして 2002 年 12 月に、その ASIMO を改良し、人応答機能とネットワーク接続機能を搭載した新型「ASIMO」を発表した(Fig. 1)。

従来の ASIMO は、これまでにさまざまな場所で受付業務や各種アトラクション等を行ってきており、次第に人社会に受け入れられつつある。しかし、従来の ASIMO は、歩行や握手等の動作をオペレータがリモート端末や携帯コントローラーを操作することで行っており、人の指示を判断して自律的に行動することはできなかった。今後、ASIMOが人社会において人の役に立つロボットとなるためには、人の行動を理解し、人と協調した作業が行えることが必要となってくる。

そこで今回我々は、従来の ASIMO のシステムを拡張し、画像処理による人応答機能を搭載することで、自律性を持

*和光基礎技術研究センター

** Honda R&D Americas, Inc.

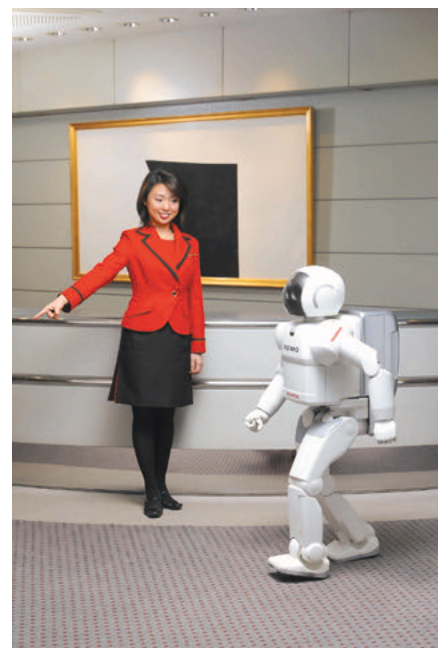


Fig. 1 New ASIMO

たせたヒューマノイドロボットの実現を目指した⁽²⁾。

一般的にヒューマノイド(人型)という形態は、ロボットが何をしているのか人にとって理解しやすく、また、一般環境において人と共存するためにも適している。しかし、人がヒューマノイドロボットに動作の指示を与える場合、相手が人型という理由から人はロボットに対して人と同じように接する傾向にあり、人とのコミュニケーション能力が他の形態のロボット以上に重要視されることになる。

人間同士のコミュニケーション手段の最たるものは音声によるものであるが、それと並んで重要な役割を果たすのがジェスチャによるコミュニケーションである。言葉では表現が冗長になりがちな大まかな位置の指示や、音声が届かないようなシーンにおいては、ジェスチャによるコミュニケーションの重要性はさらに増すことになる。また、音声では表現が難しいニュアンスを伝える場合にもジェスチャを使ったコミュニケーションは有効である。

ジェスチャ認識に関する研究は、これまでに数多く提案されているが、ロボットとの自然なインターフェースを実現するためには、身体的拘束のあるデータグローブや磁気センサなどの装着型センサを必要としないという点で、画像情報等を利用した非接触型センサによる認識手段が望ましい⁽³⁾。しかしながら、これまでに提案されている画像処理によるジェスチャ認識手法は、カメラが固定されている⁽⁴⁾、⁽⁵⁾、背景や人物に対する制限が多い⁽⁶⁾、計算コストが高く専用処理システムを必要⁽⁷⁾とするなどの課題があった。

特に ASIMO に搭載している画像処理システムでは、今回報告するジェスチャ認識の他に顔認識や障害物認識等の画像処理を行っているため、それら処理と同時に動作する高速な手法が求められる。一方、他の処理結果を利用できるという利点もあるが、一般的に複数の画像処理を組み合わせることは、機能に多様性を持たせたり信頼性を向上させることは可能ではあっても、処理が複雑となり前述のように計算コストがかかって実時間処理が難しくなるという課題が生じる。また、単機能では実現できる処理ではあっても、それらを統合的に取り扱うことは非常に難しい。

本稿では、顔認識や障害物認識等の他の画像処理と共に実時間かつオンボードで動作するジェスチャ認識処理について報告する。

2. 画像処理システム概要

Fig. 2 は新型 ASIMO に搭載されている画像処理システムの概要である。

画像処理システムは、ASIMOの頭部内に設置された二つの CCD カメラと ASIMO 本体に収められたフレームグラバ、画像処理を実行する PC ボードから構成される。また、画像処理時に用いるカメラの位置や角度等のカメラパラメータを計算するために必要な ASIMO の各関節角等の姿

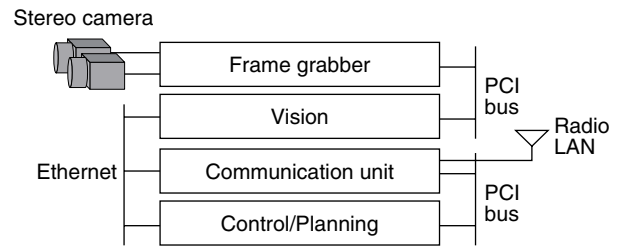


Fig. 2 Image processing system

勢情報は、制御用 ECU ボード上に記憶される。

なお、画像処理用 PC は市販されている CompactPCI バスの PC ボードであり、搭載されている CPU は Intel Pentium III-M 1.2GHz である。

画像を撮影するカメラは、ステレオ視によって距離情報を得るために 2 つの CCD カメラが用いられている。

また、カメラから出力される画像信号を取り込むためのフレームグラバは、ステレオ視によって物体の実空間上の位置を抽出する際に必要となる各画像に対応したカメラパラメータを取得するために、同期信号によって各画像とカメラパラメータとの同期を取ることが可能な仕様とした。これにより、移動カメラによって撮影された動画の画像処理を、複雑な処理を必要とせずに精度良く行うことが可能となった。

3. 画像処理アルゴリズム

3.1. 概要

画像処理部のソフトウェア構成を Fig. 3 に示す。

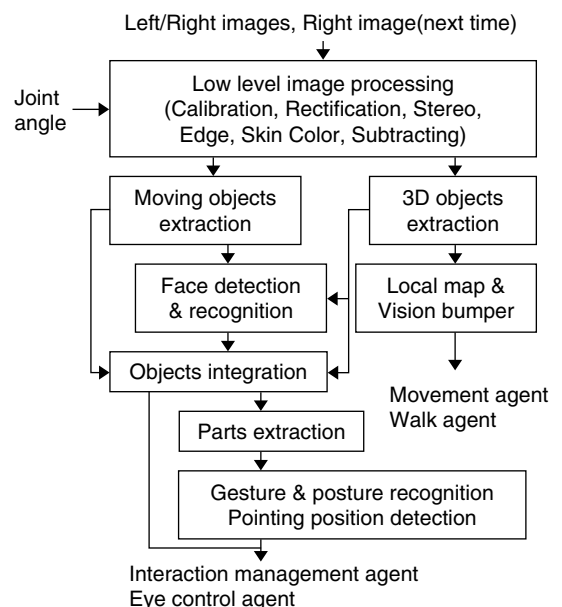


Fig. 3 Vision system software configuration

カメラから入力された画像に対して低レベル画像処理部で画像の歪を補正する処理を行ない(Fig. 4 左上段), 続いて, 距離情報(Fig. 4 右中段), エッジ情報(Fig. 4 左下段), 肌色領域情報(Fig. 4 右下段), 動き領域情報(Fig. 4 左中段)を計算する。

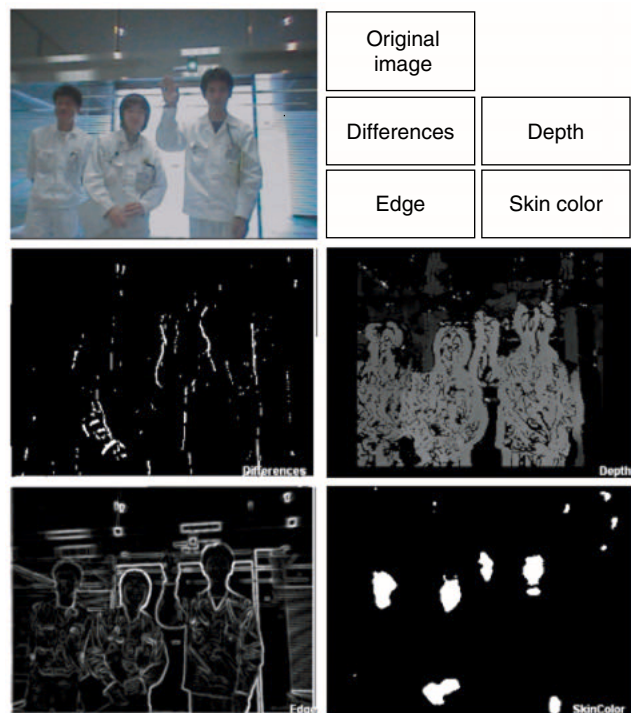


Fig. 4 Preprocessing images

距離情報は, ブロック相関を用いたステレオ処理によって求める。エッジ抽出は, Sobel オペレータを用いて行う。肌色領域は, HLS(色相, 明度, 彩度)空間上でしきい値操作により抽出する。動き領域情報は, 画像とカメラパラメータ変化量から背景の動きを推定して背景位置合わせを行った後, 差分計算によって抽出する。

これらの情報を元に, 移動体抽出, 3Dオブジェクト抽出, 顔検出のオブジェクト抽出処理を行い, 一旦, オブジェクトの統合処理により各フレーム内でのオブジェクトの統合及びフレーム間でのオブジェクトの対応処理を行う。

次に, 各オブジェクトもしくは追跡対象となっているオブジェクトの部位抽出処理を行い, 最後にジェスチャとポスチャの認識, 指示位置認識の認識処理を行った後, 処理結果が人応答処理部へと出力されロボットの行動に反映される。

このように本システムでは, オブジェクト抽出部, 部位抽出部, 認識部と処理を階層化し, 処理結果を各階層毎に取りまとめることで処理の複雑さを回避した。また一方では, 他処理からの情報を有効に利用することで, 各処理での探索範囲を限定するなどして計算コストの削減を図り, オンボードで 100 ~ 200msec/frame の動作を実現した。

以下において, オブジェクト抽出からジェスチャとポスチャの認識, 指示位置認識までの各処理について述べる。

3.2. 移動体抽出

ここでは, ジェスチャ認識の対象となる人(移動体)の抽出を行う。まず, 低レベル画像処理で得られた距離, エッジ, 動きの各情報を元に移動体を覆う大まかな直方体領域を設定し, その後, その領域内の画像情報からより詳細な移動体の形状や位置を抽出していく。

距離方向へは, 距離情報と動き情報から動きのあるエッジ点を最も多く含むカメラからの距離 d を抽出し, その値を基準に処理対象距離範囲を設定する。

続いて, この対象距離範囲内に存在する画素(もしくはエッジが抽出された画素(Fig. 5 の青色で表示された画素))を抽出し, 画像の縦軸方向にここで抽出された画素数を数え, その数が最大となる線を処理領域の中央線とする(Fig. 5)。この中央線から距離 d に応じて左右に人の肩幅程度の広さを取り, 左右の処理領域境界を設定する。

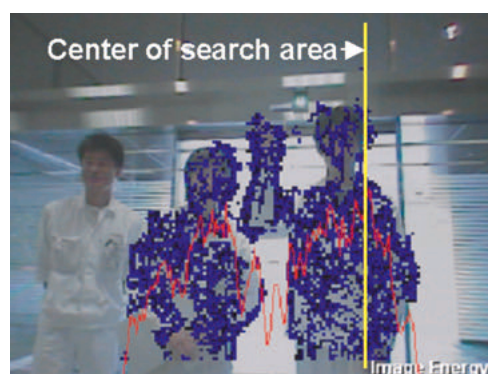


Fig. 5 Center of search area (left-right)

最後に, 縦方向の処理範囲は, 処理対象距離 d とカメラパラメータの値から, 床と床からの高さが人の身体を十分覆うことのできる範囲を設定する。

以上の処理によって設定された探索領域内にある画像情報から移動体のより詳細な形状や位置を抽出する。

画像情報から対象物を抽出する手法としては, 画素の特徴量のクラスタリングに基いた領域分割法や, 検出されたエッジを連結させる輪郭抽出法⁽⁸⁾, 予め定義したエネルギーを最小化するように閉曲線を変形させる動的輪郭モデル(Snakes⁽⁹⁾)等がある。

本システムでは, 複雑なシーンにおいて顔や手等の対象物の詳細な部位を抽出する必要があることから, ノイズに強く対象物の輪郭形状が抽出できる Snakes を用いて移動体の抽出を行った。

Snakesとは, エネルギー最小化原理に基いた輪郭モデルであり, 式(1)で計算されるエネルギーが最小になるように $v(s)=(x(s), y(s))$ で表される輪郭の形を変えていく。

$$E_{snake} = \int_0^1 E_{int}(v(s)) + E_{image}(v(s)) + E_{con}(v(s)) ds \quad (1)$$

ここで、画像エネルギー $E_{image}(v(s))$ の計算において、前述の探索領域内にあるエッジの強度を用いる。

抽出された移動体 i の輪郭上もしくは輪郭内に存在する点の位置から移動体の重心位置 (x_i, y_i) を計算し、移動体の角度 θ_i を求める。また、輪郭内点の画素が持つ距離情報から移動体までの距離 d_i を計算し、実空間上での移動体の位置を求める。既に抽出された移動体の輪郭内部の画素が持つ距離情報を更新(視差をゼロにする)し、処理範囲設定からの処理を繰り返すことで複数人の抽出を行う。

Snakes を使って複数移動体を抽出した結果を Fig. 6 に示す。各矩形は画像上での移動体の探索領域を、輪郭内部の点は各移動体の画像上での重心位置を表している。その横の数値は各移動体までの距離と角度を示している。



Fig. 6 Extraction of humans

本手法では、このように移動体を覆う大まかな直方体領域を設定した後、詳細な形状や位置の抽出を行っているため、複数人が並んで立っているようなシーンにおいても、個々の移動体を高速に抽出することが可能である。

3.3. 部位抽出

次に、ジェスチャ認識に使用する人の部位(顔と手の位置)を抽出する。

3.3.1. 顔位置抽出

顔位置抽出は、抽出された移動体の輪郭情報から頭頂部を抽出し、その頭頂部を基準点として探索範囲を設定して行う。基準となる頭頂部位置は、移動体重心を中心として左右方向におおよそ人の頭の幅だけ広げた範囲内にある輪郭のうち最も上部に位置する点とする。この点を基準にして顔の探索範囲を設定する。なお、画像上での探索範囲のサイズは、移動体までの距離 d_i に対応して変動する。距離方向の探索範囲についても、移動体距離 d_i を中心にして、おおよそ人の顔を覆う範囲を設定する。

よそ人の顔を覆う範囲を設定する。

画像上での顔位置 (x_p, y_p) は、上記探索範囲内にある肌色領域の重心位置とする。実空間上での顔位置 (X_p, Y_p, Z_p) は、 (x_p, y_p) と移動体距離 d_i 及びカメラパラメータを用いて計算できる。

また、この後の手位置抽出のために、顔位置抽出後に顔位置を中心に顔が移っていると推測される範囲を再び設定し、その範囲内にある肌色情報を削除しておく。

3.3.2. 手中心位置抽出

手中心位置 (x_h, y_h) も顔と同様、探索領域内にある肌色領域の重心位置とする。

手の探索範囲は、顔位置 (x_p, y_p) を基準にして手が届く範囲として設定する。なお、抽出対象としている移動体の近くに別の移動体が存在する場合には、隣り合う移動体重心の中心を探索範囲の境界とすることで、対象としない移動体の手や顔の誤検出を防いでいる。

実空間上での手中心位置 (X_h, Y_h, Z_h) は、顔位置と同様に計算される。

3.3.3. 手先位置抽出

手中心位置 (x_h, y_h) は、その抽出手法からも大まかに手の位置を求めただけであるため、次にこの位置を元にしてジェスチャ認識に必要な手先位置 (x_{hp}, y_{hp}) を推定する。

まず、3.3.2. で求めた手中心位置 (x_h, y_h) を中心にして、手先の探索範囲を設定する。探索範囲は、人の手を十分覆うことのできる大きさとする。画像上での領域サイズは、これまでの処理と同様に移動距離に応じて変化する。

次にこの領域内に含まれる点のうち、肌色情報を持つ画像上における上下左右の端点を抽出し、上下端点間もしくは左右端点間の距離のどちらか長い方を手が伸びている方向として、手先が上下のどちらか(もしくは左右のどちらか)の端点であると仮定する。

最後に、手中心位置 (x_p, y_p, z_p) と顔位置 (x_h, y_h, z_h) の位置関係から、上下のどちらか(もしくは左右のどちらか)が手先かを判定する。具体的には、手中心位置が顔から遠い場合には、手を伸ばしているものとして顔から遠い方の端点を手先位置とし、手の中心位置が身体に近い場合は、肘を曲げていると判断して顔に近い方の端点を手先位置としている。実空間上での手先位置 (X_{hp}, Y_{hp}, Z_{hp}) は、手先位置として抽出した点の画像上での座標 (x_{hp}, y_{hp}) 及びその点の距離(視差)から計算できる。

Fig. 7 は手と顔の探索範囲と抽出結果である。赤色の枠は、それぞれの探索範囲を示している。なお、左右で手の探索範囲の大きさが異なるのは、あるフレームで手が抽出された場合、次のフレームでは前フレームの手位置を中心に探索範囲を設定するので、探索範囲を狭くするためである。

このように、探索範囲を前フレームの抽出結果や移動体の抽出結果を利用することで、顔と手の抽出を高速に精度良く行うことが可能となった。

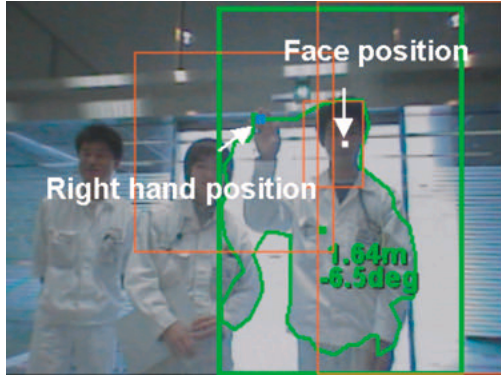


Fig. 7 Extraction of face & hands position

3.4. ジェスチャとポスチャの認識

3.3.で抽出した実空間上の顔位置および手先位置から、それらの位置関係の時間的変化を観測することでジェスチャおよびポスチャを認識する。

ジェスチャの認識手法としてはHMM(隠れマルコフモデル)を用いた手法があげられるが⁽⁴⁾, 認識率を高めるためには学習データが大量に必要であり⁽³⁾, また3次元へ拡張した場合には実時間処理が難しいといった課題がある⁽¹⁰⁾。そこで、本システムでは、処理が高速であり、統計的に最適な判定が行えるために間違っただ判断をしないという理由から、統計的手法の一つであるベイズの手法⁽¹¹⁾を用いてジェスチャの推定を行う。ジェスチャを特徴付けるための特徴ベクトル $\mathbf{x}$ は、実空間上での顔を原点としたときの手先位置 $(X_{ht}, Y_{ht}, Z_{ht}) = (X_{ht} - X_f, Y_{ht} - Y_f, Z_{ht} - Z_f)$ の数フレーム間の平均 $(\bar{X}, \bar{Y}, \bar{Z})$ と標準偏差 $(\sigma_x, \sigma_y, \sigma_z)$ の6つのパラメータにより構成される。

ベイズの定理より、特徴ベクトル $\mathbf{x}$ が与えられたときにそれがジェスチャ i である確率密度 $p(\omega_i|\mathbf{x})$ は、式(2)で求められる。

$$p(\omega_i|\mathbf{x}) = \frac{p(\mathbf{x}|\omega_i)p(\omega_i)}{p(\mathbf{x})} \quad (2)$$

ここで、 $p(\mathbf{x}|\omega_i)$ はジェスチャ i が与えられたときに画像が特徴ベクトル $\mathbf{x}$ を含む条件付き確率密度であり、式(3)で表される。

$$p(\mathbf{x}|\omega_i) = \frac{1}{2\pi\sqrt{|\Sigma|}} e^{-(\mathbf{x}-\bar{\mathbf{x}}\Sigma^{-1}(\mathbf{x}-\bar{\mathbf{x}}))/2} \quad (3)$$

なお、特徴ベクトル $\mathbf{x}$ は共分散行列 Σ をもち、期待値 $\bar{\mathbf{x}}$ の正規分布に従うものとしている。

また、 $p(\omega_i)$ は ω_i の事前分布の確率密度を表す。 $p(\omega_i)$ が期待値 ω_{i0} 分散 $V[\omega_{i0}]$ の正規分布であるとすれば、 $p(\omega_i)$ は式(4)のように書ける。

$$p(\omega_i) = \frac{1}{\sqrt{2\pi}V[\omega_{i0}]} e^{-(\omega_i-\omega_{i0})^2/2V[\omega_{i0}]} \quad (4)$$

式(2)の分母は ω_i に依存しないので、式(3)、(4)より、 ω_i の事後分布の確率密度は式(5)のように表される。

$$p(\omega_i|\mathbf{x}) \propto p(\mathbf{x}|\omega_i)p(\omega_i) = \frac{1}{\sqrt{2\pi}^3 \sqrt{|\Sigma|V[\omega_{i0}]}} e^{-(\omega_i-\omega_{i0})^2/2V[\omega_{i0}] - (\mathbf{x}-\bar{\mathbf{x}}\Sigma^{-1}(\mathbf{x}-\bar{\mathbf{x}}))/2} \quad (5)$$

式(5)から、抽出された特徴ベクトル $\mathbf{x}$ を与えたときにそれがジェスチャ i である確率を全ジェスチャに対して求め、その中で確率が最大となるものをそのフレームでのジェスチャ認識結果とし、さらに数フレーム間の結果から最終的な出力結果を得る。

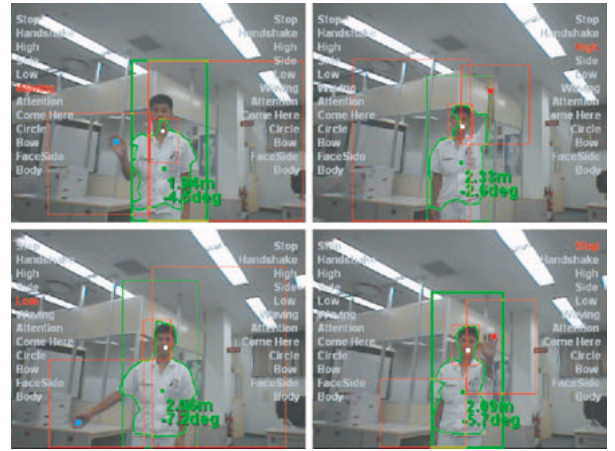


Fig. 8 Gesture & posture recognition results

現在、この手法を用いてHandshake(握手), Hand circling(手を身体の前で回す), Bye-Bye(手を身体の前で左右に振る Fig. 8 左上), Hand swing(手を頭の上で左右に振る), High-hand(手を上げる Fig. 8 右上), Side-hand(手を身体の前で左右に振る Fig. 8 左下), Low-hand(手を身体の前で左右に振る Fig. 8 右下), Come Here(おいでおいで), Stop(手の平を前に突き出す Fig. 8 右下), Face-side(手を顔の前で左右に振る)の計10種類のジェスチャおよびポスチャが左右それぞれの手で個別に認識可能である。Fig. 8の画像の左右端に赤で表示されているジェスチャまたはポスチャが認識結果である。

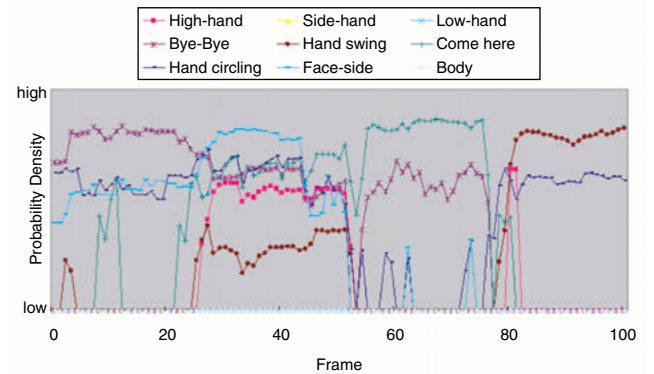


Fig. 9 Probability density

Fig. 9に、Bye-Bye, Face-side, Come here, Hand swingとジェスチャおよびポスチャを行った場合の確率密度を示す。グラフから、そのときに行っているジェスチャおよびポスチャである確率密度が、他のジェスチャおよびポスチャを示す確率密度より明らかに高い値を示し、認識が正しく行われていることがわかる。

3.5. 指示位置認識

指示位置認識処理は Fig. 10 の手順で行う。

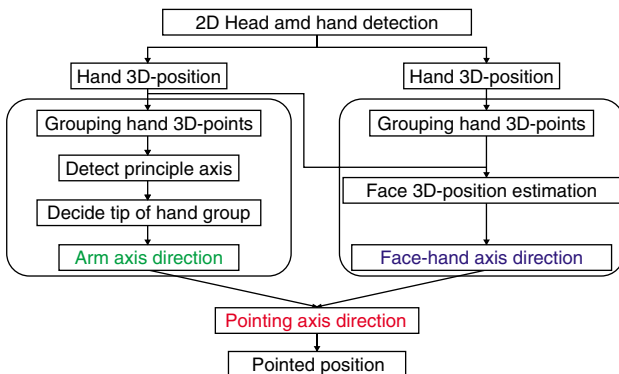


Fig. 10 Flowchart of pointing position detection

まず、ある画像取得フレームについて、部位抽出により得られた手先、頭の位置を基準としてステレオ像中での腕先、頭に対応する点を獲得する。具体的には、基準位置から3次元距離の一定範囲内の物を腕先、頭とみなし、それぞれを頭グループ、手グループと呼ぶ。

頭グループについては重心を求め頭中心とする。手グループについて3次元座標値の主成分分析を行い主軸を求める。計算は次のように行う。グループ中にN点あり、その各点の座標値を平均が0になるよう正規化した物を A_n とする。正規化されたグループを $A = [A_1, A_2, \dots, A_N]$ のように行列で表し、この行列について分散共分散行列を求め、その固有値、固有ベクトルの計算を行う。最大の固有値に対応する固有ベクトルが手グループの分散の最も大きい主軸となる。

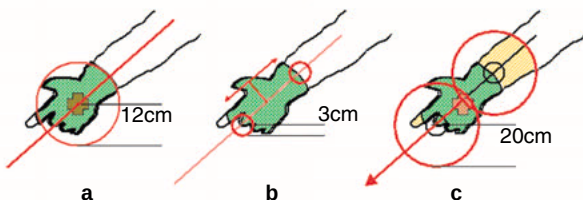


Fig. 11 Arm direction detection

得られた主軸について腕の先端方向を求める。方法としては重心位置から主軸に沿って2方向を半径3 cm以内に3次元点の無い地点を探していく (Fig. 11b)。3 cm以内に

3次元点の無い地点で20cm以内に含まれる3次元点の少ない側を先端方向とする (Fig. 11c)。

最後に、頭グループと手グループの位置関係を確認し身体構造上取り得ない方向が先端となっている場合には逆方向に修正する。逆方向でも取り得ない場合にはその画像取得フレームでの処理はできなかったものとする。

3フレーム以上について同様の処理を行い、頭中心位置、腕軸の方向、腕位置全ての分散が一定値以内になった時に、腕と頭の姿勢が特定されたとする。なお、腕の処理は左右の双方について行いより早く特定された側、かつより遠くに特定された側を指示している手とする。

腕と頭の姿勢特定後、指示者は指示をしている手を見てると仮定して顔位置の推定を行う。頭を身長から推定される半径を持つ球と仮定し、頭中心から指示手方向の頭部の表面を求め、首をZ軸に垂直に回転させたと考え、Z値を頭中心の値に戻して目位置とする (座標系は Fig. 12 参照)。

目位置と手位置を結ぶベクトルを頭手ベクトル($V_{\text{face-hand}}$)、腕方向を腕ベクトル(V_{arm})とし、頭手ベクトルと腕ベクトルの合成ベクトルを指示ベクトル(V_{pointing})とする (Fig. 12)。合成は式(6)のとおりとした。

$$V_{\text{pointing}} = \frac{V_{\text{arm}}}{|V_{\text{arm}}|} + \frac{V_{\text{face-hand}}}{|V_{\text{face-hand}}|} \quad (6)$$

但し、頭と指示手の距離が身長から推定される腕長さの7割以下の場合には肘が曲がっている物として腕ベクトルを指示ベクトルとしている。

手位置を基点とした指示ベクトルを延長した線と床面の交点を指示位置であるとする (Fig. 12 赤点線)。

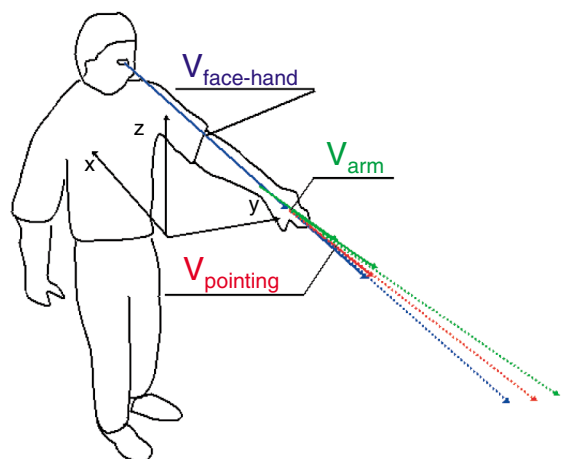


Fig. 12 Pointing gesture overview

以上のアルゴリズムを実機に組み込み、本手法の精度検証を行った。ステレオカメラは被験者の1.5m前方に設置した (ASIMOを被験者の1.5m前方に立たせた)。

被験者の前方1mの距離にある横7点の指示をした画像

について、腕ベクトルのみを用い指示床面を推定した場合、頭手ベクトルを用いた場合、指示ベクトルを用いた場合の3通りの方法で指示位置の検出結果を Fig. 13 に示す。

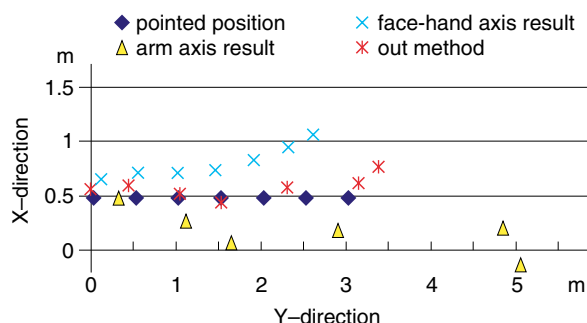


Fig. 13 Result of pointing position detected by each vector

平均の誤差は、腕ベクトルのみを用い指示床面を推定した場合 0.93cm, 頭手ベクトルを用いた場合 0.34cm, 一方、指示ベクトルを用いた場合は 0.25cm の誤差であった。頭手ベクトルと、腕ベクトルの合成である指示ベクトルを用いることにより認識結果が向上していることがわかる。

なおこの実験時には、特別に検出手法を意識することなく床面の指示動作を行っており、本手法を用いることによって自然な指示動作で腕ベクトル、頭手ベクトルから求める位置よりもより、意図した位置に近いポイントを検出することが出来るようになった。

また、6人の被験者(Fig. 14 S1～S6)について腕を伸ばして左斜め前の9点(1.4m から 4.2m 先)を3度指示してもらい検出の精度を求めた(Fig. 14)。

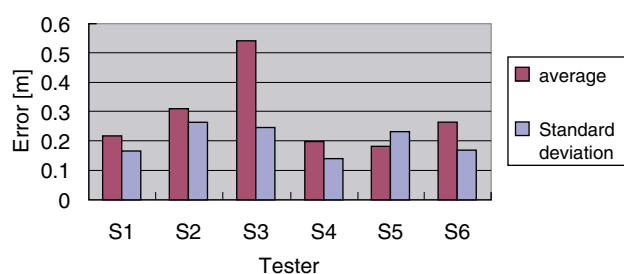


Fig. 14 The detection error for every subject

全点の平均は 0.28m(指示距離別の内訳は 1～2 m 誤差 0.15m, 2～3 m 誤差 0.25m, 3～5 m 誤差 0.45m), 最大の誤差を持つ被験者でも 0.54m, 他の被験者は誤差 0.18m から 0.31m の範囲であった。多少の個人差はあるがおおむね安定した検出能力を持っていると考えられる。

指示位置認識結果の一例を Fig. 15 に示す。Fig. 15 左側はステレオ画像に認識結果を重ね合わせた図で、右側はその認識結果の鳥瞰図である。

このように、腕の向きに加えて頭と腕の位置関係を利用することにより、安定して精度の高い指示位置の抽出が可能となった。

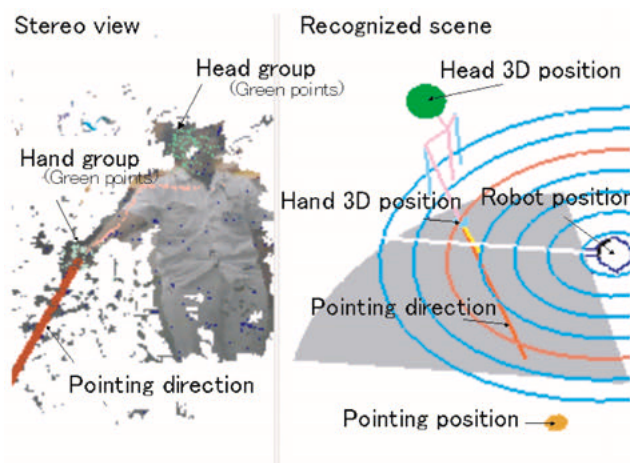


Fig. 15 A Pointing gesture recognition result

4. 実験結果

これらジェスチャ認識機能をASIMOに搭載することで、ジェスチャによる呼び寄せや指示位置への移動などの動作が可能となった。



Fig. 16 Demonstration images a: recognize "high-hand" posture, b: start to follow, c: recognize "stop" posture, d: stop walking, e: recognize pointing position, f: start to walk, g: walk to pointing position, h: arrive and stand by

具体的には、手を高く上げている、もしくはおいでおいでをしている人を見つけるとその人に追従し、追従している人から止まれの指示が出されたら、その場で止まることが可能である。(Fig. 16 左列)

また、握手を求められると握手するために右手を出し、その他のジェスチャとポストチャに対しては、人と同じジェスチャとポストチャを行う。

また、音声などにより「あっちに行って」等の移動を求められる入力があれば、人が指している位置を認識して、その場所まで移動するといった機能が可能となった。(Fig. 16 右列)

5. まとめ

二足歩行ヒューマノイドロボット ASIMO の自律性向上を目指し、画像処理によるジェスチャ認識機能をASIMOに搭載した。

- 1) 動き情報を元にジェスチャ認識の対象となる人(移動体)を抽出し、続いて色情報を利用して人の頭と手を抽出した。この頭と手の実空間位置の時間変化から、統計的手法を用いてジェスチャとポストチャの認識を行い10種類のジェスチャとポストチャの認識が可能になった。
- 2) また、手付近で観測された3次元点群の主成分分析によって腕の向きを求め、頭と腕の位置関係と腕の向きから、人が指示する方向及び指示しているポイントを抽出した。

これにより、これまではオペレータがリモート端末から逐次操作、指示を与えていた動作の一部をジェスチャによって与えることが可能となり、ASIMOの操作性及び自律性を高めることができた。

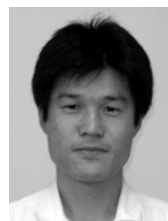
また、本システムでは、オブジェクト抽出部、部位抽出部、認識部と処理を階層化し、処理結果を各階層毎に取りまとめることで処理の複雑さを回避した。更に、他処理からの情報を有効に利用することで、各処理での探索範囲を限定するなどの工夫によって計算コストの削減を図った。その結果、オンボードで約 100 ~ 200msec/frame の速度で動作可能な総合的な画像処理システムを世界で初めて実現させることができた。

参考文献

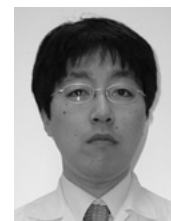
- (1) 広瀬真人, 竹中透: ヒューマノイドロボット ASIMO の開発, Honda R&D Technical Review, Vol.13, No.1, p.1-6 (2001)
- (2) Sakagami, Y., Watanabe, R., Aoyama, C., Matsunaga, S., Higaki, N., Fujimura, K.: The intelligent ASIMO: System overview and integration, IROS, p.2478-2483(2002)
- (3) 谷内田正彦, 岩井儀雄: インタラクションのための人物動作解析と認識, 信学技法, PRMU99-142(1999)

- (4) 高橋正貴, 小倉信彦, 渡辺澄夫: 隠れマルコフモデルによる混合動作の認識, 信学技法, PRMU 2000-151 (2001)
- (5) 高橋和彦, 坂口竜己, 大谷淳: 実時間非接触非装着型3次元人物姿勢推定法に関する一考察, 信学論(D-II), Vol.J83-D-II, No.5, p.1305-1314(2000)
- (6) 石井浩史, 望月研二, 岸野文郎: 人物合成のためのステレオ画像からの動作認識法, 信学論(D-II), Vol.J76-D-II, No.8, p.1805-1812(1993)
- (7) 高橋勝彦, 関進, 小島浩, 岡隆一: ジェスチャー動画像のスポッティング認識, 信学論(D-II), Vol.J77-D-II, No.8, p.1552-1561(1994)
- (8) Ballard, D., Brown, C.: セグメント化画像, コンピュータ・ビジョン, 東京, 日本コンピュータ協会, p.147-205 (1987)
- (9) Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active Contour Models, Proc.of 1st ICCV, p.259-268(1987)
- (10) Muller, S., Eickeler, S., Rigoll, G.: Crane Gesture Recognition Using Pseudo 3-D Hidden Markov Models, In 4th IEEE International Conference on Automatic Face and Gesture Recognition, p.398-402(2000)
- (11) 金谷 健一: ベイズ推定, 空間データの数理, 東京, 朝倉書店, p.149-150(1995)

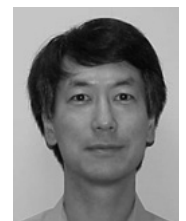
■ 著 者 ■



榎垣 信男



横山 太郎



藤村 希久雄