自己遮蔽にロバストなヒト上肢3次元運動推定

Estimation of Three-Dimensional Trajectory of Arm Movement Robust to Self-Occulusion

正会員 星野 聖^{†,††},小 渡 悟[†],神里 志穂子[†],新 垣 武 士[†]Kiyoshi Hoshino^{†,††}, Satoru Odo[†], Shihoko Kamisato[†] and Takeshi Arakaki[†]

あらまし 手話認識のための上肢3次元運動推定を目的として、3次元要素多面体から構成される人体モデルを用意し、また、過去のオブティカルフローを元に現時刻の運動推定を行うことにより、単眼 CCD カメラでも自己遮蔽に強い3次元ジェスチャ認識システムを構築した。

キーワード：上肢3次元運動推定、自己遮蔽、単眼 CCD カメラ、手話認識

1. は じ め に

手話動作の認識を非接触・無拘束で行おうとする試みが行われているが、手話認識には「手指の形状」「手指の動き」「運動の領域・位置」といった推定が必要である。これまでに、手指形状を構成する各種の特徴量から指文字認識を行うもの¹⁾、形状や色情報を使って動作追跡を行うもの²⁾、対数極座標変換により参照データを減らして追跡を行うもの³⁾などが報告されている。

しかし、実用的な手話認識システムを構築する場合、手指の形状と変化、手の軌道、手の運動領域などを同じ重要度で行うことは効率的とは言えない。たとえば、ヒトの運動知覚の観点から考えると、静止している手話動作では手の形や掌の向きの詳細な認識が重点的に行われていると予想されるし、運動を伴う手話では軌道形状の詳細な認識が重点的であると予想される。しかも、日本語手話の単語を動きの有無で分類すると、全体の84.0%が何らかの移動を伴っていることが報告されている⁴⁾。1,200語程度のフランス語手話でも、静止、直線、円、円弧、それらの組み合わせ、といった動作要素により単語が構成されていることが報告されており⁵⁾、手話認識では手指の動き、すなわち「軌道」に関するパラメータの抽出が重要であると考えられる。

一方、ビデオ動画像を用いた従来の手話認識では、2

次元映像の空間的および時間的情報を照合の手掛かりとして用いることがほとんどであった。しかし最近になって、2次元映像の画像的特徴を用いるよりは、1次元情報の集合として、すなわちヒトと同じ筋・骨格構造を持つ人体モデルの「各関節動作の時間変化パターンの集合」もしくは「各筋肉の収縮パターンの集合」として照合を行った方がロバストな手話認識が実現できる見込みが報告されている⁶⁾。例えば、「行く」や「来る」のような日常的な手話動作でも、話者により、手を直線的に移動させる場合と、弧を描いて移動させる場合があるし、軌道の微妙な時間的变化や空間的变化が動詞を修飾して、さまざまな副詞の意味変化を起こさせる⁷⁾ことが知られている。とくに空間的变化とは、同じ手話単語でありながら、軌道の幾何学的形状変化により伝えられる情報が変化することを意味する。もし、CCD映像を基に、見まねによりヒト型ロボットアーム⁸⁾を動かすことができる程度に軌道形状の幾何学的変化を推定できれば、ロバストな手話認識ができるだけでなく、筋や関節の1次元情報の集合に手話名称を与えることによって照合用テンプレート設計の手間なしに手話認識が実現できる可能性がある。

そこで本研究では、手話動作における上肢3次元運動推定を目的として、単眼 CCD カメラを用いた奥行きジェスチャ認識アルゴリズムを提案する。予備実験として、ビデオレートよりも低いサンプリング周波数でも軌道形状情報が十分に保持できるかどうかを検討するため、典型的な2種類の手話動作を高速ビデオカメラで撮影し、サンプリング周波数を落とした場合の補間処理の結果を示す。構成するシステムでは、手話者の顔や他の身体部位による自己遮蔽にロバストな処理を実現するため、あらかじめ用意した3次元人体モデルへ動画像を当てはめ

2002年6月1日受付、2002年8月15日再受付、2002年9月3日採録

[†]琉球大学 工学部 情報工学科

(〒903-0213 沖縄県中頭郡西原町字千原 1, TEL 098-895-8719)

^{††} 科学技術振興事業団さがけ研究 21・基礎的研究発展推進事業

(〒903-0213 沖縄県中頭郡西原町字千原 1, TEL 098-895-8719)

[†] Department of Information Engineering, Faculty of Engineering, University of the Ryukyus

(1, Senbaru, Nishihara-cho, Nakagami-gun, Okinawa 903-0213, Japan)

^{††} PRESTO & SORST, Japan Science and Technology Corporation

(1, Senbaru, Nishihara-cho, Nakagami-gun, Okinawa 903-0213, Japan)

図 1 手話動作「森」と「空」の 2 次元軌道
Trajectories of “wood” and “sky” in Japanese sign language.

る処理と、現在の位置推定に過去の運動情報を参照する処理とを併用した。計算機内部にあらかじめ構築された 3 次元人体モデルに 2 次元画像を当てはめる手法については、従来、各種のものが提案されてきた⁹⁾が、本研究では、とくにトップダウン処理とボトムアップ処理の両方で同モデルを併用することにより、自己遮蔽対策を試みた。

2. 予 備 実 験

CCD カメラを用いた動画画像処理では、通常で 1 秒あたり 30 フレーム、最高でも 60 フレームの処理しか期待できない。手話認識を実時間で行うためには、そこで、ビデオレート以下の画像情報でも必要な軌道情報が獲得できることが前提となる。そこで、本章では、動作速度や軌道面積が一般的な手話動作の例として、直線軌道から構成される「森」と曲線軌道から構成される「空」の 2 種類を高速ビデオカメラで撮影し、サンプリング周波数を落として補間処理をした場合の精度について検討する。使用したモノクロ高速ビデオカメラは、日本ローパー製 MotionMeter 500 (エンハンストメモリー付) で、サンプリング周波数 125Hz で、熟練手話者の顔から腰が画面いっぱいになるように撮影した。手首位置の軌道形状は、

WINalyze 3D (Mikromak 社製) を用いてオフラインで推定した。

図 1 に、125Hz の 2 次元軌道座標と、それぞれ 60Hz、30Hz、15Hz に間引いた際の 2 次元軌道座標を示す。図中、上段は「森」、下段は「空」の手話動作である。直感的な印象では、サンプリング周波数を 1/8 に間引いても、軌道形状はそれほど変わっていない。

そこで、30Hz と 15Hz の 2 次元軌道座標について、3 次 B スプライン補間を用いて 125Hz に内挿した結果を図 2 に示す。図中、上段は「森」、下段は「空」の手話動作であり、それぞれ丸印が 125Hz の元座標、実線が低い周波数から 125Hz に補間した後の座標である。極めて良好な当てはめ精度が得られているのが直感的にわかる。また、同図に示したデータのなかで、元座標と補間座標の差の標準偏差は、最大の図 2 (a) 右側でも 0.4 以下であった。これらの結果から、3 次の B スプライン程度の補間処理により、十分な精度の手話軌道が獲得できることがわかる。本システムで構築する単眼 CCD カメラによる上肢 3 次元運動推定システムでも、15Hz 以上の処理時間が実現できれば充分であると見なすことができる。

図2 手話動作「森」と「空」のオリジナル軌道と補間結果
Original and interpolated trajectories of "wood" and "sky" in the sign language.

3. 上肢運動推定システム

3.1 3次元人体モデル

単眼カメラにより奥行き運動を伴う手話動作要素推定を行うために、3次元人体モデル (Human Based Model. 以下、HBMと略す) を生成する。HBMは、頭部、胸部、左右の上腕部、前腕部に相当する6個の剛体の連結構造で成り立っており、各部は3次元要素の多面体で近似される。HBMの体節点は、頭部、胸部、右肩、右肘、右掌、左肩、左肘、左掌の9箇所である。各体節点は、体節位置 $\hat{\mathbf{w}}$ 、探索すべき範囲を示す体節点探索領域 $\mathbf{w}$ 、さらに、マウスマーキングによる照合処理で得られた2次元上の上腕最大長 L_{upper} 、前腕最大長 $L_{forearm}$ の情報を持つ。ジェスチャ部位に相当する肩と肘には、それぞれ3自由度を設けている。

初期フレーム画像内のヒトの各関節点の初期位置をマウスマーキングにより照合処理を施す。なお、ヒトのジェスチャは自由度が高く、完全な動き情報を得ることは困難であるが、時間軸方向に滑らかに変化する、すなわち近い時刻間では、極端に大きな変化がないと仮定すると、比較的単純な関節物体の比較的滑らかな動きと見なすことができる。そこで、マウスマーキングによる正

確な関節位置情報を利用する方法を採用した。後に詳細を述べるが、この初期情報をもとにして、時間差分法と一般化勾配法¹⁰⁾によりオプティカルフローを求め、現在と過去のオプティカルフローの平均と分散から体節点を決定する。

3.2 確率的手法を用いた領域推定

本研究では、ベイズ推定により各特徴点の探索領域と特徴点位置の推定を行う。推定すべきパラメータ $\hat{\mathbf{w}}$ は、片腕あたり肩座標 (x_s, y_s, z_s) 、肘座標 (x_e, y_e, z_e) 、掌座標 (x_h, y_h, z_h) の9パラメータである。観測値は、座標とオプティカルフローの大きさの (x, y, u, v) とする4次元ベクトルを $\mathbf{x}$ とする。

時刻 t における観測された特徴を $\mathbf{x}_t$ 、状態を表すパラメータを $(\hat{\mathbf{w}}_t)$ で表すと、探索領域と特徴点位置推定は式(1)に示す $\mathbf{x}_t$ から推測できる $\mathbf{w}_t$ の情報を表す事後確率密度 $P(\mathbf{w}_t|\mathbf{x}_t)$ に基づいて行うことになる。

$$P(\mathbf{w}_t|\mathbf{x}_t) = \frac{P(\mathbf{x}_t|\mathbf{w}_t)P(\mathbf{x}_t)}{P(\mathbf{w}_t)} \quad (1)$$

ここで、 $(\mathbf{w}_t)$ は状態を表すパラメータであり、総探索範囲の個数を n とした $\mathbf{w}_t = w^1, \dots, w^n$ の生起確率を示す。また、 $P(\mathbf{x}_t|\mathbf{w}_t)$ は推定オプティカルフローの生起分布が正規分布と仮定した場合の条件付確率を示す。このとき、 $P(\mathbf{x}_t|\mathbf{w}_t)$ は式(2)とする。

$$P(\mathbf{x}_t|\mathbf{w}_t) = \sum_{i=Min}^{Max} \sum_{j=Min}^{Max} \frac{1}{2\pi\hat{\sigma}_x(t)\hat{\sigma}_y(t)} \times \exp \left[-\frac{1}{2} \left\{ \left(\frac{i - \hat{\mu}_x(t)}{\hat{\sigma}_x(t)} \right)^2 + \left(\frac{j - \hat{\mu}_y(t)}{\hat{\sigma}_y(t)} \right)^2 \right\} \right] \times |\mathbf{F}_{(i,j)}(t)| \quad (2)$$

ここで、 Min, Max は、探索領域 $\mathbf{w}_t$ 内の範囲 n を示し、 $\mathbf{F}_{(i,j)}(t)$ は時刻 t のときの座標 (i, j) のオプティカルフローを示す。また、 $\hat{\mu}_x(t)$ 、 $\hat{\mu}_y(t)$ 、 $\hat{\sigma}_x(t)$ 、 $\hat{\sigma}_y(t)$ は、現時刻 t から過去3フレーム¹¹⁾のオプティカルフローの平均値と分散値を示す。事後確率分布を閾値により受容領域と棄却領域に分割し、この受容領域を候補領域として追跡を行う。

候補領域から、以下に示す手順により掌、肘、肩の決定を行う。第1に、候補領域内の重心点を中心として4つの領域に分割する。ここで領域を4分割したのは、体節の候補点が1部に偏ってしまわないためであり、反対に、多すぎる分割は非効率的だからである。第2に、各領域の重心と候補領域の外周までのEuclid距離を求め、各領域内で最長となる4点を選択する。ここで、前節の手順にしたがってベイズ推定を行うと、オプティカルフローの平均値と比べ、重心点が過去の体節点側へ、すなわち少ない動きとして推定される傾向がみられた。とくに、オプティカルフローの平均値が小さく、分散が大きくなる回転運動ではその傾向が顕著となる。予備的検討の結果、重心点のずれは、運動方向への最長外周までの距離とおおむね一致していた。第3に、これら4点と、

体躯に向かって1つ手前の体節点との間のそれぞれ距離を求める。最後に、1時刻前の対応する体節間距離と最も近いものを選び、体節点と決定する。

3.3 自己遮蔽対策

ボトムアップ的処理だけでは、自己遮蔽が生じると、体節点決定法の信頼性は低下してしまう。しかし、トップダウン的処理を適応すれば、大まかな体節点の推定は可能である。また、抽出できなかった体節点に関してもその推定位置近傍にオプティカルフローが存在する場合、推定位置から最近点となるオプティカルフローの検出点を体節点と決定することができる。

ひとつの探索領域の中心点が他の探索領域の中に入った場合は、自己遮蔽と判断し、HBMを用いた幾何学的拘束と直前フレームまでのジェスチャ追跡の結果より、体節点推定を行う。手順は以下の通りである。

- a. 過去フレームまでのジェスチャ追跡法(ボトムアップ的処理)により体節点位置候補を推定する。
- b. 直前フレームからの運動情報より一定範囲を体節点候補領域とする。
- c. HBMによる拘束条件を満たす領域を体節点可動領域とする。
- d-1. 手順 b, c の条件を満たす領域を走査し、その領域にオプティカルフローが存在する場合、手順 a の体節点位置候補と最も近い点を体節点と決定する。
- d-2. オプティカルフローが存在しない場合、手順 a から c を満たす領域の重心点を体節点と決定する。
- e. 手順 a から d で体節点が決まれない場合は、手順 a の推定位置を体節点と決定する。

3.4 単眼カメラによる3次元動作推定方法

単眼カメラによる3次元動作推定手法を実現するために、平面座標系から空間座標系を形成する。単眼カメラにより取り込まれた2次元情報は、3次元情報を得るために、仮想的に奥行き軸方向成分を設ける必要がある。そこで、カメラ視点に対する直交座標系(Z軸)を平面座標系に与え、空間座標系を形成する。

ここで、一般にヒトが一連の動作を行う際、肩位置はさほど大きな変化を示さないの、そのため本手法では、肩点をローカル座標系の原点とする。また上腕の運動は、肩位置を中心としたXZ平面での回転運動 θ_{s1} と、XY平面での回転運動 θ_{s2} とし、前腕の運動は、肘位置を中心としたXZ平面での回転運動 θ_{e1} と、XY平面での回転運動 θ_{e2} とする。

ジェスチャ認識のパラメータとしては、追跡システムから得られた時刻 t での特徴点の座標値、肩座標 (x_s, y_s) 、肘座標 (x_e, y_e) 、掌座標 (x_h, y_h) を基にして、 θ_{s1} 、 θ_{s2} 、 θ_{e1} 、 θ_{e2} の回転量、ならびにZ軸方向の座標値 z_s 、 z_e 、 z_h を推定することで得る。

図3 ヒト3次元動作の実時間CG表示
Real-time CG of human three-dimensional motions.

式(3)から式(6)に、肩位置を原点とした場合の上腕の回転角度と肘座標を示す。

$$\theta_2 = \sin^{-1} \left(\frac{y_e - y_s}{L_{upper}} \right) \quad (3)$$

$$\theta_1 = \cos^{-1} \left(\frac{x_e - x_s}{L_{upper} \cos \theta_2} \right) \quad (4)$$

$$x_e = x_s + L_{upper} \cos \theta_1 \cos \theta_2 \quad (5)$$

$$y_e = y_s + L_{upper} \sin \theta_2 \quad (6)$$

なお、 z_e は $\mathbf{w}_t$ 推定時に算出される。同様に、肘位置を原点として前腕の回転角度と掌座標を算出する。

4. 評価実験

本手法の有効性を検証するための評価実験を行った。

実験は室内で行い、画像中では、被験者以外の移動物体はないものとし、カメラの視野は被験者のジェスチャが、適切に入るように設定した。また、照明は室内の天井に設置されている蛍光灯のみを用いた。使用した動画は、画像サイズ 320 × 240, RGB 各 256 階調のカラー画像を AVI 形式で保存したものを使用した。ただし、肩の位置、肩から肘までの長さ、肘から手先の長さ、最初のフレームにおける腕の状態を既知とした。

本手法にて、静止、直線、円、円弧のジェスチャについて検証した。また、これらを組合せたジェスチャについて追跡実験を行った。

その結果、図 3 に示すように、入力画像中のヒトの動作は 1 秒あたり 15 フレーム以上で、良好な精度で、かつ遅延のないジェスチャ CG として復元でき、また掌点により肩点が遮蔽した場合でも、次フレーム以降において追跡を再開することを示すことができた。

5. ま と め

本研究では、手話認識のための上肢 3 次元運動推定を目的として、単眼 CCD カメラでも自己遮蔽に強い 3 次元ジェスチャ認識システムを提案した。まず予備実験として、上肢運動を高速ビデオカメラで撮影し、低い周波数に間引いた軌道座標と、それを高い周波数に補間した座標とを比較し、1 秒あたり 15 フレームの処理速度でも、手話認識のための上肢運動推定が可能であることを確かめた。次に、片腕あたり 9 自由度を持つ 3 次元要素多面体から構成される 3 次元人体モデルを用意し、2 次元画像を当てはめる方式のジェスチャ認識システムを構築した。顔や他の身体部位による自己遮蔽に対してロバストな推定が可能のように、3 フレーム前までのオプティカルフローを元に現時刻の運動推定を行った。評価の結果、構築したシステムは自己遮蔽の影響を受けにくいだけでなく、1 秒あたり 15 フレーム以上の処理速度により、3 次元上肢運動推定とコンピュータグラフィックス描画とを実現できた。

本研究の一部は、通信・放送機構 ギガビットネットワーク活用研究開発制度、および文部科学省科学研究費特定領域研究 (A)「脳研究の総合的推進に関する研究」の援助を受けて行われた。

【文 献】

- 1) 長嶋祐二, 藤井昌紀, 長嶋秀世: “手の機能的な構造特徴量に基づく指文字認識,” テレビジョン学会誌, **48**, 8, pp.1053-1057 (1994)
- 2) 山本正信: “身振りの自動認識,” 映情学誌, **51**, 8, pp.1128-1131 (1997)
- 3) 小渡 悟, 星野 聖: “ヒトの視覚系に基づく時系列画像中の注視物体追跡,” インタラクティブシステムとソフトウェア, **8**, pp.147-152 (2000)
- 4) 内藤一郎, 加藤雄士: “効率的な検索方法の確立を目的とした日本手話の形態的特徴の基本的検討,” 信学論 (A), **J79-A**, 2, pp.337-345 (1996)
- 5) S.Gibet, J.Richardson, T.Lebourque, A.Braffort: “Corpus of 3D natural movements and sign language primitives of movement,” Proc. Intl. Gesture Workshop 1997, pp.111-121 (1997)
- 6) 星野 聖, 藤井智史, 猪木誠二: “バリアフリー・コミュニケーションのためのギガビットネットワーク活用技術に関する研究,” 通信・放送機構 平成 13 年度研究発表会予稿集, pp.267-272 (2001)

- 7) E.Klima and U.Bellugi: “The signs of language,” Cambridge, Harvard University Press (1979)
- 8) 添石幸憲, 小渡 悟, 星野 聖: “ロボットハンド制御のための手指形状推定,” 信学技報, NC 2001-25, pp.21-25 (2001)
- 9) 草地良規, 山澤一誠, 竹村治雄, 横矢直和: “動画画像中の歩行者追跡による映像入れ換え合成,” 映情学誌, **52**, 9, pp.1357-1365 (1998)
- 10) M.V.Stinivasan: “Generalized gradient schemes for the measurement of two-dimensional image motion,” Biol. Cybern., **63**, pp.421-431 (1990)
- 11) 星野 聖, 小渡 悟, 神里志穂子, 新垣武士: “手話認識のための 3 次元手指運動推定,” 信学技報, NLP 2000-170, pp.43-49 (2001)

星野 聖 (ほしの きよし) 1993 年, 東京大学医学系大学院博士課程修了。同年, 東京医科歯科大学医学部助手。1995 年, 琉球大学工学部助教授。1998 年より科学技術振興事業団さきがけ研究 21, 2002 年より基礎的研究発展推進事業研究者を兼任。視覚系の情報処理機構, ヒト運動解析の研究に従事。医学博士, 工学博士。正会員。

小渡 悟 (おど さとる) 1998 年, 琉球大学大学院工学研究科修士課程修了。同年, 同大学大学院理工学研究科博士後期課程入学。現在, 在学中。コンピュータビジョン, パターン認識, ニューラルネットの研究に従事。工学修士。

神里志穂子 (かみさと しほこ) 2001 年, 琉球大学大学院理工学研究科博士前期課程修了。同年, 同研究科博士後期課程入学。現在, 在学中。感性情報, ヒト運動解析の研究に従事。工学修士。

新垣 武士 (あらかき たけし) 2000 年, 琉球大学工学部情報工学科卒業。同年, 同研究科博士前期課程入学。現在, 在学中。ヒト運動解析の研究に従事。