

ジェスチャーユーザインタフェースのための P型フーリエ記述子を用いた手の姿勢推定

Hand-Pose Estimation Using P-type Fourier Descriptors for Gesture-Based User Interface

○立見駿介*, 張山昌論*, 亀山充隆*

○Shunsuke Tatsumi*, Masanori Hariyama*, Michitaka Kameyama*

*東北大学大学院情報科学研究科

*Graduate School of Information Sciences, Tohoku University

キーワード: ジェスチャーユーザインタフェース (Gesture-Based User Interface),
手の姿勢推定 (Hand-Pose Estimation), P型フーリエ記述子 (P-type Fourier Descriptors),
3次元手形状モデル (3-D Hand Model), 複数カメラ (Multi-camera)

連絡先: 〒 980-8579 仙台市青葉区荒巻字青葉 6-6-05
東北大学 大学院情報科学研究科 亀山・張山研究室 立見駿介

Tel.: (022)795-7155, Fax.: (022)263-9167, E-mail: s_tatsumi@ecei.tohoku.ac.jp

1. はじめに

近年, ヒューマンインターフェースへの応用を目的の1つとして, 任意の姿勢における手の姿勢を計測し計算機への入力へ利用しようとする試みが行われている。これは手の姿勢を計測することで, ジェスチャーを用いて入力機器に触れることなく機器の操作を行ったり 3D コンテンツの操作をより直感的に行うことが出来ると考えられているからである。手の姿勢を取得するための手法としては, 手にセンサを取り付けたり¹⁾, マーカや特殊な手袋を用いるなどの手法があり, このような方法はシンプルで実時間での処理に適しているが, 使用状況が限られたりユーザに不自由を与える場合がある。そのためカメラから取得した画像のみを用いて手の姿勢を計測する手法が研究されており, 大きく次の2つの手法

に分けられる。1つ目は画像から姿勢の特徴を抽出し, 3次元手形状モデルの姿勢パラメータを変化させて画像と特徴が一致させることで姿勢を推定する手法²⁾である。この手法は精度の高い姿勢推定を行うことが出来るが, オクルージョンに弱いという特徴がある。2つ目はあらかじめ様々な手の姿勢画像をデータベースに登録しておき, 入力と類似度の高い画像を求めることで姿勢を推定する手法³⁾である。この手法は事前に様々な姿勢の手の画像を登録しておくことでオクルージョン強い推定を行えるが, 手は姿勢変化の自由度が大きいいため探索対象の姿勢の数が多くなり推定精度と推定速度にトレードオフが生じるという特徴がある。そこで探索姿勢を直前に推定された姿勢の近傍姿勢のみに限定する手法⁴⁾や, 姿勢のクラスタリングを行い階層構造を用いて探索木を構成することにより探

索範囲を絞る手法⁵⁾等が提案されている。

本研究では3次元手形状モデルを用いて事前に認識するジェスチャ姿勢の画像を作成しておく2つ目の手法をとる。入力画像と事前に作成した画像とでマッチングを行うことにより手の姿勢を推定し、ジェスチャを認識して機器への入力に用いることで、マーカレスかつ直感的な片手のみによる3Dコンテンツの操作を行うことを目的とする。本稿ではP型フーリエ記述子による特徴表現を利用した手の姿勢推定手法を提案し、ジェスチャを認識して3Dコンテンツの操作を行う。ジェスチャは平行移動・拡大・縮小と回転に対応するものを定義し、それぞれ手の3次元空間の位置の変化から移動量、回転動作の推定姿勢から回転量を取得して操作を実現する。

2. P型フーリエ記述子による輪郭特徴表現

2.1 ジェスチャ認識処理の概要

ジェスチャ認識処理の流れをFig.1に示す。ジェスチャ認識処理は事前に行う処理とジェスチャ操作時に行う処理の2つに分かれている。事前に行う処理では、まずユーザの手と3次元手形状モデルとでキャリブレーションを行った後、モデルを用いてジェスチャ認識に必要な様々な姿勢の手の画像を作成し、データベースを作成する。ジェスチャ操作時に行う処理では、まずユーザのジェスチャをKinect⁶⁾を用いて撮影し、RGB画像とDepth画像を取得する。次に取得したRGB画像とデータベース内の画像とでマッチングを行い、手の姿勢推定を行う。その後姿勢推定結果の時系列を基に入力されたジェスチャを認識し、3Dコンテンツにジェスチャに対応する操作を行うことでジェスチャ操作を実現する。

ユーザのジェスチャを取得するためのセンサとしてKinectを用いているが、これはRGB画像とDepth画像の取得が容易であること、ライ

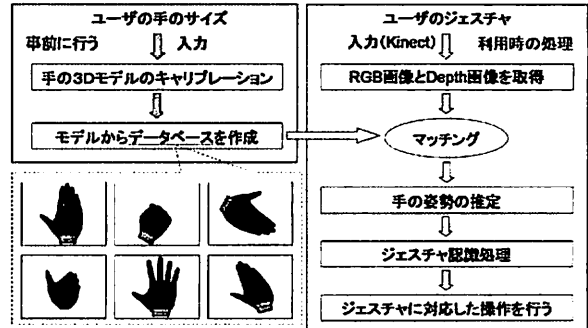


Fig. 1 ジェスチャ認識処理の流れ

ブラリを利用することでハンドトラッキングが容易に可能であることが理由であり、センサがKinectでなくても問題はない。なおKinectとは2010年にMicrosoft社から発売された同社製機器用のゲームコントローラであり、RGBセンサ、Depthセンサ、マイク等を備えている。

2.2 P型フーリエ記述子を利用したマッチング手法

手の姿勢を推定するためにマッチングを行う。画像に基づくマッチングでは輪郭線、シルエット、輝度、奥行き、特徴点、エッジ等様々な画像特徴が用いられるが、本稿では輪郭線に着目し、輪郭線をP型フーリエ記述子で表現し特徴量とするマッチング手法を提案する。まずP型フーリエ記述子について説明する。曲線をFig.2のように折れ線近似すると、座標 Z_i と座標 Z_{i+1} の関係は偏角 θ_i を用いて次式で表される。

$$w_i = \exp(j\theta_i) \quad (1)$$

w_i を周期関数として離散フーリエ変換を求めると次式のようなになる。

$$C_k = \frac{1}{N} \sum_{i=0}^{N-1} w_i e^{-j\frac{2\pi}{N}ki} \quad (2)$$

C_k をP型フーリエ記述子という。P型フーリエ記述子は平行移動、拡大・縮小に対して不変であるという特徴を持つ。

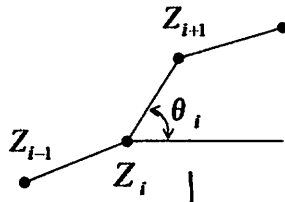


Fig. 2 曲線の折れ線近似

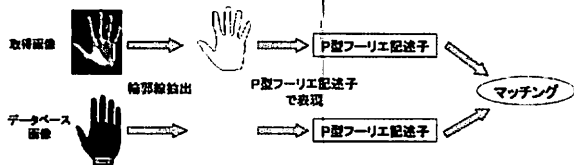


Fig. 3 P型フーリエ記述子を利用したマッチング手法

P型フーリエ記述子を利用したマッチング手法の概要を Fig.3 に示す. 取得した画像とデータベース内の画像それぞれで輪郭線を抽出してP型フーリエ記述子で表し, 2つの特徴の低次成分のユークリッド距離を相違度としてマッチングを行う. マッチングによって相違度が最小となるデータベース画像を求め, その画像の姿勢を推定姿勢とする. P型フーリエ記述子を用いることで, 平行移動, 拡大・縮小の影響を受けずに姿勢推定ができる, 低次成分のみを用いることで大まかな形状の認識ができる, 低次成分のみを用いるので計算量が少なく済むといった利点がある.

3. 3次元手形状モデルに基づく姿勢推定

3.1 3次元手形状モデル

本稿では3次元手形状モデル (Fig.4) に基づいた手の姿勢推定を行う. 姿勢推定に3次元手形状モデルを用いる利点としては, 取得情報の欠損やノイズおよびオクルージョンに強い推定が行えることが挙げられる. これらは推定したい手部分の情報取得できない場合であると考



Fig. 4 3次元手形状モデル

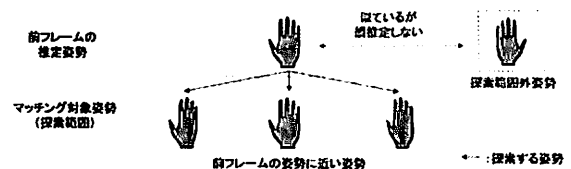


Fig. 5 変化姿勢の予測

えることができる. このときモデルを使用しない場合は情報が取得できなかった手部分の姿勢は推定することが出来ないが, モデルを使用することにより取得できた情報を用いて姿勢推定を行い推定姿勢のモデル情報から情報が取得できなかった手部分の姿勢を推定することが出来る.

3.2 変化姿勢の予測

マッチングを行う際にデータベース内の全ての姿勢とマッチングを行うと, 取得情報の欠損やユーザの手とモデルとの形状の差異などの原因から, 特徴量は似ているが大きく異なる姿勢を推定してしまうことがあり, 問題となる. そこで前フレームで推定された姿勢から近い姿勢のみをマッチング対象として探索することでありえない姿勢を推定することを防ぐ (Fig.5). これはフレーム間隔が小さいならば隣接したフレーム間における手の姿勢は近いものになるという考えに基づいた手法である. 姿勢推定開始の際はあらかじめ定めた開始姿勢を認識し, そこから姿勢探索を開始する. 変化姿勢の予測の利点としてはありえない姿勢を探索することが無くなり精度が向上すること, 計算量が削減できることが挙げられる.

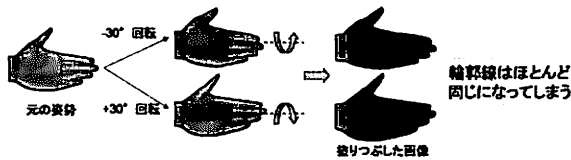


Fig. 6 見分けにくい回転動作

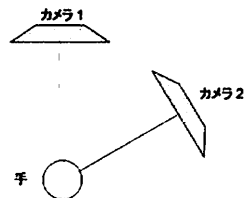


Fig. 7 カメラの配置図

3.3 複数カメラの利用

本稿ではジェスチャによる3Dコンテンツの直感的な回転操作を実現するために手の回転動作の認識を行っている。しかし回転動作ではオクルージョンが多く発生し、単一のカメラでは見分けることの出来ない動作が発生することが問題となる (Fig.6)。そこで本稿ではオクルージョンを軽減する為に2台のカメラを利用する (Fig.7)。これにより単一のカメラでは見分けることが出来ない動作を見分けることが出来る。

4. ジェスチャ認識

4.1 ジェスチャ認識処理

ジェスチャ認識処理の流れを Fig.8 に示す。ジェスチャ認識処理では手の姿勢推定結果の時系列を記録し、その時系列から平行移動・拡大・縮小、回転に対応するジェスチャを認識する。その後認識したジェスチャに応じて移動量を RGB 画像と Depth 画像から取得、もしくは回転量を推定姿勢から取得し、取得した情報に基づいて3Dコンテンツの操作を行う。

各操作に対応したジェスチャは全てのジェスチャで共通の基本となる姿勢と各ジェスチャに固有のジェスチャ開始姿勢の2つから成り、ジェ

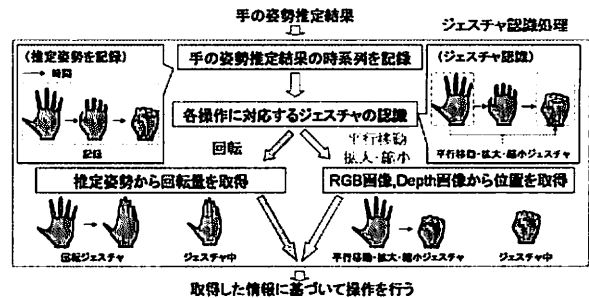


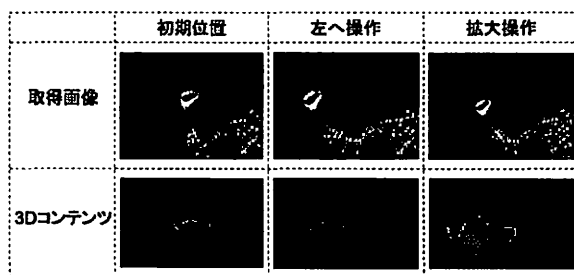
Fig. 8 ジェスチャ認識処理の流れ

スチャの認識は基本となる姿勢が認識された後、一定フレーム間内にジェスチャ開始姿勢が認識されることで成立する。またジェスチャ開始後はジェスチャ開始姿勢の手形状のまま手の移動、回転動作を行うことで3Dコンテンツの操作を行い、ジェスチャ開始姿勢の手形状とは異なる手形状になったときジェスチャが終了したと判断して操作をやめる。

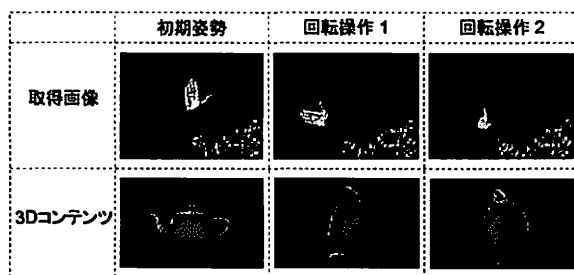
4.2 ジェスチャ操作の実験

提案手法を用いて、ジェスチャによる3Dコンテンツの操作を行った。実験には2台の Kinect を用いて画像を取得し、Intel Core2 Duo E8400 3.00GHz CPU を搭載した PC で処理を行った結果、平均 2.4fps 程度の処理速度となった。操作結果を Fig.9 に示す。

平行移動・拡大・縮小操作においては概ね意図したとおりの操作が実現できた。しかし手を動かしていないつもりでも多少3Dコンテンツが動いてしまったり、終了時にジェスチャ中の姿勢以外の姿勢に変化している際の動きが操作に反映されることが問題として挙げられた。回転操作においては3Dコンテンツに対する回転動作は結果として実現できるが、ユーザが意図するであろう滑らかな回転操作を実現することは出来なかった。どちらも取得した情報をそのまま3Dコンテンツの操作に用いたことが原因であり、取得情報からジェスチャでない動作を認識する等の処理が必要であると考えられた。ま



(a) 平行移動・拡大・縮小操作



(b) 回転操作

Fig. 9 実験結果

た回転操作時に誤推定を起し、実際の動作と異なる姿勢変化経路が取得されることが問題として挙げられた。そこで各カメラの信頼度を考慮して重み付けを行うなどによる姿勢推定精度の向上、姿勢変化経路までを考慮した姿勢推定等が必要であると考えられた。

5. まとめ

本稿では手のジェスチャによって 3D コンテンツの操作を行うために手の姿勢を推定する手法として、3次元手形状モデルに基づいた手の姿勢推定手法および P 型フーリエ記述子を利用したマッチング手法を提案した。今後の課題として姿勢推定精度の向上、処理速度の向上、非ジェスチャの認識等が挙げられる。

参考文献

- 1) 澤田秀之, 橋本周司, 松島俊明: 運動特徴と形状特徴に基づいたジェスチャー認識と手話認識への

応用, 情報処理学会論文誌, Vol.39, No5, pp.1325-1333, 1998

- 2) 陳維英, 藤木隆司, 有田大作, 谷口倫一郎: 複数カメラを用いた実時間三次元手形状推定, 画像の認識・理解シンポジウム (MIRU2006), 2006
- 3) 今井章博, 島田伸敬, 白井良明: 実時間手指推定のためのロバストな輪郭画像照合, コンピュータビジョンとイメージメディア, Vol.2002, No.34(2002-CVIM-133), pp.63-70, 2002
- 4) 島田伸敬, 白井良明, 久野義徳: 確率に基づく探索と照合を用いた画像からの手指の三次元姿勢推定, 電子情報通信学会論文誌 D-II J79-D-II/7, pp.1210-1217, 1996
- 5) 藤本光一, 松尾直志, 島田伸敬, 白井良明: 輪郭部分特徴の階層構造学習による三次元手指姿勢推定的高速化, 画像の認識・理解シンポジウム (MIRU2010), pp.63-70, 2010
- 6) Kinect - Xbox.com, <http://www.xbox.com/ja-JP/kinect/>, 2013/10/24