

## マーカーとアピアランスを併用した手の姿勢認識

阿部 巖<sup>†1</sup> 岡谷 貴之<sup>†1</sup> 出口 光一郎<sup>†1</sup>

画像を用いた手の姿勢認識には幅広い応用があり、様々な手法がこれまでに提案されている。最も安定かつ高精度な方法は、マーカーベースのモーションキャプチャシステムであるが、手にたくさんのマーカーを取り付ける必要があり、ユーザの負担が大きい。アピアランスベースの方法は、このような問題はないものの、精度や頑健性の面で大きく劣るか、認識可能な手の姿勢の自由度に制約がある。以上を踏まえて本研究では、少数のマーカーを使って得られる幾何学的な情報と手のアピアランスの情報を組み合わせることで、ユーザの負担と精度をバランスできる方法を述べる。

## Hand Pose Recognition Using Markers and Appearance

TAKASHI ABE,<sup>†1</sup> TAKAYUKI OKATANI<sup>†1</sup>  
and KOICHIRO DEGUCHI<sup>†1</sup>

Image-based recognition of human hand poses has a large number of applications and many methods have been proposed so far. The most stable and accurate approach is marker-based motion capture. However, it requires that many markers are attached to the hands, which deteriorates the usability of the system. Appearance-based methods, which are free from such problems, are much inferior in terms of accuracy and robustness, and often have limitations in the degrees of freedom of the hand poses that can be recognized. Based on these considerations, this study presents a method that can balance the system usability and the recognition accuracy, by combining the geometric information obtained by using a small number of markers with the appearance of the hands.

### 1. はじめに

手の姿勢認識は仮想現実、マニピュレータの操作、テレビゲームなどの様々な目的での応用が考えられ、広く研究が行われている。手は日常生活の中で、物を掴むなどある対象を物理的に操作するために、あるいはジェスチャーのようにコミュニケーションの道具としてよく使われる器官である。つまり、手を使って人やものとインタラクションを行うことは人間にとって日常的で直感的であり、このことは手をインタフェースとして用いる大きな動機である。

本研究で考える手の姿勢認識とは対象とする手の3次元空間上での位置と姿勢および各関節の角度を求めることである。手の姿勢認識は次のような理由から難しい問題である。手の姿勢変化は一般的に用いられるモデルで27自由度と、とても高次元であるため推定すべきパラメータが多い。また、手を2次元の画像上に投影すると非常に多くのセルフオクルージョンが生じる。これにより画像から認識に必要な特徴を得るのが困難になる。画像特徴の乏しさもしばしば問題になる。たとえば、テクスチャから手の部位を判別するというようなことは不可能である。手の姿勢認識と類似した問題として人体の姿勢認識があるが、人体姿勢推定と手の姿勢推定は、推定を行う環境という意味で違いがある。たとえば、人体は服の影響により背景からの切り出しが困難である一方で、手に比べると色やテクスチャの情報が豊かであることから各部位ごとのトラッキングなどが容易である。

すでに実用的に用いられている方法としてデータグローブを用いるものがある。データグローブとは関節の角度などを求めるためのセンサーを取り付けたグローブで、利用者にこのグローブを着けさせることで姿勢の測定を行う。データグローブは高精度、高速に測定ができる一方で、手の自然な動きを阻害してしまうこと、利用者個人にあわせて装置のキャリブレーションが必要があるということ、装置が高価であるという問題がある。本研究で扱うカメラを用いて認識を行うビジョンベースの手法はデータグローブを用いるのに比べて利用者への負担が少なくシステムが単純である。ビジョンベースの手法のなかでもっとも高精度な認識が行えるのはマーカーを用いたモーションキャプチャーである。この種の手法は手の各部に付けたマーカーを付ける必要がありユーザへの負担が大きいこと、マーカーの追跡を行うためのたくさんのカメラが必要という問題がある。マーカーを用いない手法としてカメラからの入力とあらかじめ用意したアピアランスを表現するモデルをフィッティングすることで姿勢を求める手法が研究されている<sup>1)2)3)4)</sup>。また、あらかじめアピアランスと姿勢パラメータの組を大量にデータベースに登録しておき、カメラからの入力をクエリとしてデータ

<sup>†1</sup> 東北大学  
Tohoku University

ベースを検索することにより姿勢を求める手法もある<sup>5)6)</sup>。これらのアピアランススペースの手法は精度が不十分であったり、認識できる姿勢の自由度に制限があるという問題がある。関連研究については2節で詳しく述べる。

我々はマーカーに加えてアピアランス情報を補助的に用いる、一般的なマーカーベースのモーションキャプチャーのように多くのマーカーやカメラを必要としない、ユーザへの負担が小さいシステムでの認識手法を検討する。マーカーの数を減らしたことによる性能の劣化や、多くのカメラを用いない場合にセルフオクルージョン等が原因で生じるマーカーの欠損を補い精度良く認識を行うためにアピアランス情報を用いる。

## 2. 関連研究

手の姿勢認識はコンピュータビジョンの分野で主要なテーマのひとつであり、様々な研究が行われてきた。

Park と Yoon は LED を取り付けられたグローブと複数のカメラを用いて手の姿勢認識を行う遠隔地間会議のためのシステムを提案した<sup>7)</sup>。

マーカーを用いない手法として、アピアランスを表現する様々な手モデルを入力とフィッティングすることにより認識を行う手法が研究されている。Stenger らは二次曲面の組み合わせからなるモデルを用いて高速な射影とオクルージョンの計算が可能な手法提案した<sup>2)</sup>。Wu らは単純な二次元のモデルであるカードボードモデルを用いた<sup>3)</sup>。Bray らはデフォーマブルスキンモデルという精密なモデルと3次元センサーを用いた手法を提案した<sup>4)</sup>。Wang らは特殊なパターンで着色した手袋を用い、類似画像検索に似たアプローチで姿勢を求める手法を提案した<sup>5)</sup>。この手法では着色した手袋を用いることで紛らわしい姿勢を区別することが可能になり、単一のカメラで高い精度での推定を可能としている。

高自由度の手の運動学的なモデルに対して、実際の手に生じる拘束を導入することで自由度を下げたり認識のコストを下げる研究が行われている<sup>8)</sup>。このような拘束を学習する手法として、Lin らはデータグローブを用いてサンプリングしたデータに主成分分析 (PCA) を適用して、低次元の部分空間を構築し、データを構築された部分空間上で線形な多様体の集合として表現した<sup>9)</sup>。

## 3. マーカーとアピアランスを併用した手の姿勢認識

我々の手法の基本的な考えは、マーカーベースのモーションキャプチャーにアピアランス情報を補助的に用いることで、できるだけ少ない数のマーカーとカメラで認識を行おうとい

うものである。マーカーの数が少なくなればユーザへの負担が小さく使いやすいシステムになる。また、必要なカメラの数が少なくなれば様々な環境での利用が可能になる。

マーカーは手の各リンクの中央と手の甲に取り付ける。それぞれのカメラで色情報を用いてマーカーの検出を行い、カメラ間エピソード拘束を用いて対応を求めてマーカー座標の3次元復元を行う。復元されたマーカーの座標と手の運動学的なモデルを最適化によりフィッティングすることで姿勢を求める。この際、3次元復元に失敗したマーカーをエッジ等のアピアランス情報をもちいて補完して姿勢認識にもちいる。

## 参 考 文 献

- 1) E.Ueda, Y.Matsumoto, M.Imai, and T.Ogasawara. Hand pose estimation for vision-based human interface. In *Robot and Human Interactive Communication, 2001. Proceedings. 10th IEEE International Workshop on*, pages 473–478, 2001.
- 2) B.Stenger, P.R.S. Mendonça, and R.Cipolla. Model-based 3D tracking of an articulated hand. 2001.
- 3) Y.Wu, J.Y. Lin, and T.S. Huang. Capturing natural hand articulation. 2001.
- 4) M.Bray, E.Koller-Meier, and L.VanGool. Smart particle filtering for 3D hand tracking. In *Automatic Face and Gesture Recognition, 2004. Proceedings. Sixth IEEE International Conference on*, pages 675–680. IEEE, 2004.
- 5) R.Y. Wang and J.Popović. Real-time hand-tracking with a color glove. In *ACM SIGGRAPH 2009 papers*, pages 1–8. ACM, 2009.
- 6) V.Athitsos and S.Sclaroff. Estimating 3D hand pose from a cluttered image. In *Computer Vision and Pattern Recognition, 2003. Proceedings. 2003 IEEE Computer Society Conference on*, volume2. IEEE, 2003.
- 7) Jun Park and Yeo-Lip Yoon. Led-glove based interactions in multi-modal displays for teleconferencing. *International Conference on Artificial Reality and Telexistence*, 0:395–399, 2006.
- 8) 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, IEEE Transactions on*, 19(7):677–695, 2002.
- 9) J.Lin, Y.Wu, and T.S. Huang. Modeling the constraints of human hand motion. In *Human Motion, 2000. Proceedings. Workshop on*, pages 121–126. IEEE, 2002.