

インタラクティブな情報支援システム実現のための 個人差に頑健なジェスチャ認識

Robust Gesture Recognition Considering Individual Variation for Interactive Information Support System

冠城 久芳[†]
Hisayoshi Kaburagi

岡谷(清水) 郁子[†]
Ikuko Shimizu Okatani

西村 拓一[‡]
Takuichi Nishimura

1. はじめに

近年、ネットワーク、計算機、携帯端末などの進歩により、「いつでも、どこでも、誰でも」情報にアクセスできるユビキタス情報環境が実現されつつあるのに伴い、莫大な情報から「いま、ここで、私が」欲しい情報を、迅速に提供することが重要となってきている。特に、博物館、美術館などのイベント空間における情報支援への応用を考えると、ユーザの意図や嗜好によって個別の情報を簡単な操作で取り出すことができる望ましい。例えば、展示品などの前に立ったユーザとインタラクションしながらユーザの欲しい情報を提供することができるといったような情報支援が求められている。このような、ユーザそれぞれの、その時点での状態や意図を反映したインタラクティブな情報支援を実現するために、筆者らは、無電源小型通信端末 CoBIT[1] を用いたユーザの位置に基づく情報支援システムを提案している [2][3]。本システムでは、ユーザはジェスチャによりシステムとのインタラクションを行う。本論文では、このシステムのため、多くのジェスチャの教示を必要としない頑健なジェスチャ認識について述べる。

動画像を用いたジェスチャ認識は、動画像から特徴を抽出し、あらかじめ教示しておいた時系列モデルとマッチングすることで行われる。複雑な背景から体格などが異なるユーザの動作を安定に抽出することは難しく、今までに数多くの手法が提案されている。例えば、肌色情報を利用して人の顔や手の位置を推定する手法 [4][5]、背景差分画像の幾何学的な特徴を用いる方法 [6]、時間差分画像の領域を分割して得られる低解像特徴を用いる方法 [7] などである。また、オプティカルフローを用いると環境の変化に依存しにくいとされている [8] が、計算量が多い。これらに対し、本システムでは、CoBIT[1] を手を持ってジェスチャを行い、この CoBIT の反射シートを、赤外線投光赤外線パスフィルタ付のカメラにより撮像することで、CoBIT を画像中の輝点として捉える。そのため、背景や照明などの環境変化に影響されることなく、容易にユーザの手の動きの軌跡を抽出することができる。

また、本システムでは、連続 DP (Dynamic Programming) [9][10] により事前に登録しておいた標準パターン (ジェスチャモデル) とマッチングを行う。ジェスチャ認

識に用いられるマッチングの手法の代表的なものとして、他に HMM(Hidden Markov Model) や HMM を改良したものを用いた手法 [11][12] が挙げられる。しかし、連続 DP には、以下が挙げられる利点がある。まず、連続 DP では、時々刻々と流れるシーケンスの中からジェスチャの始点と終点を検出して認識することができる点である。これにより、ユーザに動作の開始や終了を意識させないインターフェースを実現することができる。また、モデル間のマッチングを時間的伸縮を考慮して行うことができるため、ジェスチャの速さや大きさの変化にもある程度対応することができる。さらに、連続 DP は計算量も少ない上に、次々と入力されるフレームに同期して処理が可能であるため、実時間処理に向いているといえる。

しかし、連続 DP は一般にジェスチャ動作の個人差に弱いため、高い認識率を得るために多くのジェスチャの教示を行う必要がある。これは、情報支援システムへの応用を考えた場合には、大きな欠点である。これを解決するために、連続 DP よりもジェスチャの速度や大きさの変化に対してより頑健になるように改良した手法がある [13] が、ユーザごとのジェスチャ動作の軌道の違いやユーザの向きの変化に対して頑健に対応することはできない。

そこで、本システムでは、ジェスチャ認識のための特徴量として、CoBIT の 3 次元位置を展示品に平行な平面に射影したものを利用する。本システムでは、音声による情報支援を行うために、CoBIT を複数のカメラで撮影することによって CoBIT の 3 次元位置を推定している。3 次元位置を利用することで、位置によってスケールが変化しないために、ユーザの位置に依存しないジェスチャ認識が実現できるという利点があるが、3 次元位置そのものやそのフレーム間差分を特徴量として用いると、個人差やユーザの向きの変化に対して大きく認識率が下がるという欠点があった。そこで、本システムでは、ユーザは展示品に注目しており、展示品の方向を向いてジェスチャを行うことを利用し、動きの小さな方向の成分を無視して動作の個人差による認識率の低下を軽減する。さらに、オクルージョンなどで生じる欠損フレームを補間することにより、より安定なジェスチャの認識を可能とする。実験により、ジェスチャ動作の変化の大きな成分のみを認識に用いることで、ユーザの向きの変化にも頑健なジェスチャ認識が可能になり、欠損データの補間にオクルージョンが発生した場合に対応でき

[†] 東京農工大学大学院工学府, Tokyo University of Agriculture and Technology

[‡] 産業技術総合研究所, National Institute of Advanced Industrial Science and Technology(AIST)

ることを示す。

本稿の構成は、2節で本システムの概要を示し3節で新たに頑健な認識法を提案する。4節の実験で本手法の有効性を示し、5節でまとめる。

2. CoBITを用いたインタラクティブな情報支援システムの概要

本システムの概要是、図1のように、無電源小型通信端末CoBIT[1]と環境ユニットから構成する。環境ユニットとは、イベント空間においてCoBITを装着したユーザに展示品に関する情報を提供するために、展示品のそれぞれに対して設置するサブシステムである。

CoBITは、環境やユーザが提供するエネルギーのみで駆動し、ユーザと環境側の装置との間で情報の送受信を実現することができる無電源で動作する通信端末である。非常に小型で、イヤホン、太陽電池、反射シートから構成される。一方、環境ユニットは、複数の赤外光投光・赤外線パスフィルタ付カメラと、方向制御可能な赤外光LED光源から構成される。

CoBITに装着されている反射シートは、赤外光投光・赤外線パスフィルタ付のカメラで撮像することにより、画像上で輝点として容易に抽出できる。そこで、複数のカメラをあらかじめキャリブレーションしておけば、複数のカメラから得られる画像から輝点を抽出し、これらをエピポーラ拘束に基づき対応付けることにより、ひとつの展示品の前にいる複数のユーザに対し、複数カメラで得られたCoBITの像から各ユーザの3次元位置を推定することが可能である[3]。

そして、各ユーザに対して光源をひとつずつ割り当て、各ユーザの方向に向けて個別の情報を送信する。具体的には、LED光源の方向をCoBITの位置に向け、音声情報を変調した光線を照射する。そのため、複数カメラと光源との相対的な位置・姿勢はあらかじめキャリブレーションしておく。そして、この光線をCoBITが受信し、ユーザはイヤホンから音声情報を聞くことができる。

このとき、[Yes][No]のような意思を表す簡単なジェスチャによってユーザとシステムがインタラクションを行なうことにより、ユーザそれぞれがそのときに欲しい情報を取り出す。このとき、CoBITの3次元位置から本稿で提案する特徴抽出法により個人差やユーザの向きに頑健な特徴を抽出し、連続DPによりジェスチャ認識を行う。

3. 個人差とユーザの位置や向きに頑健な特徴抽出法の提案

3.1 3次元位置の2次元平面への射影

本システムでは、インタラクションに用いるジェスチャを縦振りと横振りという単純なものに限定しており、また、ユーザは、展示品の方向を向いているということを前提としている。

縦振りのジェスチャは、ユーザはどこを向いてもユーザの持つCoBITは上下方向の動きしかしないのでユー

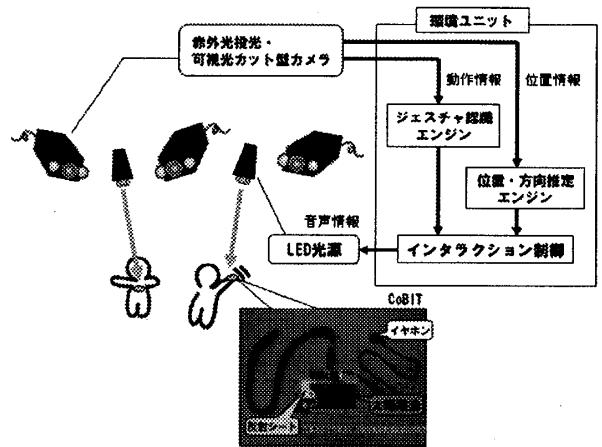


図1: システムの構成。

ザの向きには依存せずにジェスチャを認識することができる。一方で、横振りのジェスチャでは、正面を向いてジェスチャを行った場合と他の方向を向いてジェスチャを行った場合のCoBITの3次元空間中での軌跡が異なるため、正面を向いて教示したジェスチャとは異なったジェスチャとして見なされてしまい、認識率の低下をまねく。しかしながら、向きを変えて横振りのジェスチャを行った場合のCoBITの軌道でも、展示品の置いてある正面方向から見れば、ユーザの向きが多少異なっていても、横振り方向の成分は比較的大きな動きとして捕らえることができると考えられる。

また、縦振り、横振りというような一方向にのみ動く単純なジェスチャでも、個人差がある。ユーザによって、直線的であったり、湾曲していたり、手がぶれてしまったりすることがある。このような、ユーザが行うジェスチャの個人差も、動作の変化の大きな成分以外の方向成分を無視することで、軽減できるはずである。

そこで、本手法では、推定したCoBITの3次元位置を展示品に平行な平面に射影する(図2)。このように、ジェスチャ動作の変化の大きな成分のみを認識し、ジェスチャ動作に關係のない方向の動きの変化(ユーザの前後方向の変化)を無視することにより、ユーザの向きの変化やユーザの持つCoBITの動きのブレなどから生じるジェスチャ動作の個人差を吸収してジェスチャを認識することができる。

本システムでは、展示品に平行な平面に射影したCoBITの座標の時間差分(現在のフレームと1つ前のフレームの輝点位置との差)をジェスチャ認識のための特徴量とする。これは、平面上に射影したCoBITの位置のフレーム間差分を特徴量に用いることで、平面上に射影されたCoBITの動きをそのまま表現することができるためである。

3.2 輝点座標のフレーム間補間

ユーザがCoBITを持ってジェスチャを行なう際、CoBIT上の反射シートの向きや位置によっては、カメラの

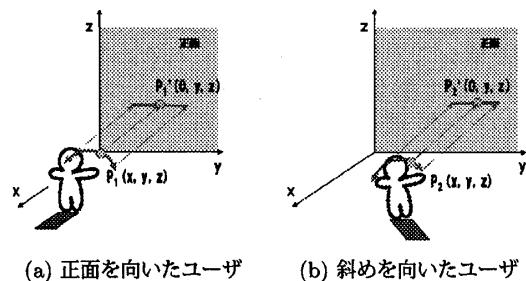


図 2: 展示品に平行な平面への射影.

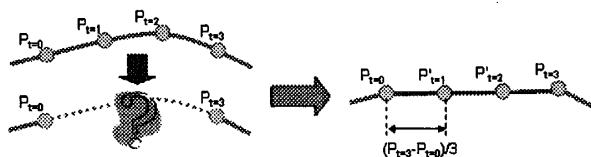


図 3: 輝点座標のフレーム間補間の例.

画像に輝点が映らないことがある。また、一つの展示品の前に多くのユーザがいる状況では、オクルージョンも発生することが考えられる。このような場合にはCoBITの3次元位置を推定することができず、データが欠損するため、認識結果が非常に悪化する。そこで、特徴量を求める前に、オクルージョンなどで観察できなかったCoBITの位置を、そのフレームの前後のフレームを用いて補間する。具体的には、欠損する前後の時刻で推定した3次元位置を線形補間したものを、欠損した時刻でのCoBITの3次元位置とする。これにより、ジェスチャ認識率の低下を防ぐことができる。欠損データの線形補間の例を図3に示す。

4. 実験

本手法の有効性を検証するために、本手法を情報支援システムに実装し、実際に被験者がジェスチャを行ったときの認識率を調べた。実験環境は、図4に示したようになっている。ジェスチャの教示は事前に各ジェスチャにつき1回のみ行なった。このとき、教示者は、展示品の正面から1mの場所でジェスチャの教示を行った。実験では、被験者はジェスチャの教示位置と、教示位置から横に50cmずれた位置でジェスチャを行い、それぞれの位置で展示品の方向を向いてジェスチャを行う。また、試行するジェスチャは[Yes]という意味の縦振りジェスチャと[No]という横振りジェスチャの2種類とした。ジェスチャを行なう被験者は教示者1人を含めた7人で、[Yes]と[No]のジェスチャをそれぞれの位置で20回ずつ行なった。

比較のため、CoBITの3次元位置のフレーム間差分をそのまま特徴量としたもの、3次元位置を平面に射影した後にフレーム間差分をとったものを特徴量としたもの、そして、平面に射影したものにさらに欠損フレーム補間を加えたのもの(提案手法)の3つの方法でジェス

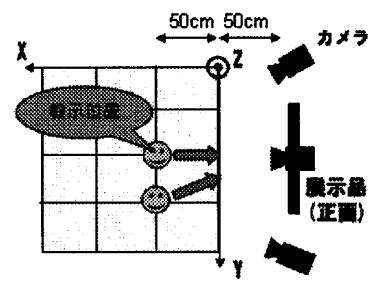
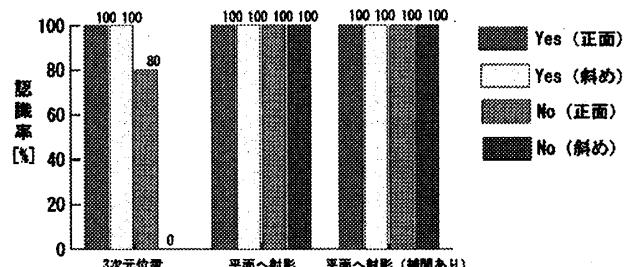
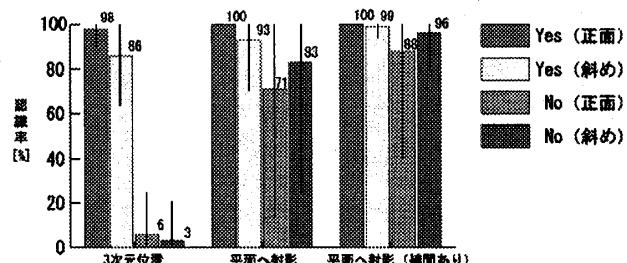


図 4: 実験環境(上方より見た図).



(a) 教示者によるジェスチャの認識率



(b) 教示者以外の被験者6人によるジェスチャの平均認識率

図 5: 実験結果(特徴量ごとのジェスチャ認識率の比較)

チャ認識を行い、ユーザの位置・向き・個人差による認識率の比較を行なった。結果を図5に示す。

図5(a)のグラフは、ジェスチャの教示者がジェスチャを行った際の認識率を示したもので、図5(b)のグラフは、教示者以外の6人がジェスチャを行った際の認識率の平均を示したものである。また、それぞれの特徴量に対する4つの棒グラフは、左から、正面でYesのジェスチャを行った際の認識率、斜め向きでYesのジェスチャを行った場合の認識率、正面でNoのジェスチャを行った際の認識率、そして、斜めでNoを行った際の認識率を表している。また、各棒グラフに、縦棒で認識率の最大と最小を示してある。

図5(a)の教示者がジェスチャを行なった場合について考察する。CoBITを縦に振る[Yes]のジェスチャについては、いずれの特徴量を用いた場合にもユーザの向きに関わらず認識率は100%である。一方で、横振りである[No]のジェスチャでは、3次元位置のフレーム間差分を特徴量として用いた場合に認識率が低下している。これは、

ユーザが斜めを向いてジェスチャを行なった場合に認識できないためである。縦方向のジェスチャでは、CoBITの3次元空間中の軌跡はユーザの向きが変わったとしても変化しないが、横方向のジェスチャでは、ユーザの向きの変化によって大きく変化してしまうためである。これにより、教示者本人がジェスチャを行っても認識できなくなっている。これに対し、提案手法では、ユーザの向きに関わらず、100%の認識率が得られている。これは、展示品に平行な平面に射影することで、ユーザの向きの変化による3次元空間中の軌跡の変化を無視できるためである。

図5(b)は、教示者以外のユーザがジェスチャを行った場合の結果である。これを見ると、特徴量に3次元位置のフレーム間差分をそのまま特徴量に用いた場合には、ユーザが正面を向いている場合の[Yes]のジェスチャ以外では、大きく認識率が低下してしまっている。これは、ジェスチャの個人差が原因であると考えられる。展示品に平行な平面に射影した場合には、[Yes]のジェスチャは向きに関わらず比較的よい認識率が得られたが、[No]では認識率の改善はみられるものの、データの欠損が発生しやすいため、認識率は80%程度である。これに対し、展示品に平行な平面への射影に加え、フレーム間補間を行なうことにより、[No]でも向きにかかわらず約90%の認識率が得られている。

これらの結果から、展示品に平行な平面へ射影した特徴抽出と欠損フレームを補間する提案手法により、高い認識率が得られることが示された。これらのことから、ジェスチャの個人差や、ジェスチャを行う際のユーザの向きの変化に対して、提案手法が有効であると考えられる。

5. まとめ

本研究では、インタラクティブな情報支援システムを実現することを目的として、ユーザの個人差や向きによるジェスチャ認識について検討した。イベント空間などの情報支援への応用を考えると、ジェスチャの教示ができるだけ少なくすることが不可欠である。そこで、1回のジェスチャ教示で、ユーザの個人差を吸収でき、さらにユーザの位置や向きにも頑健にジェスチャ認識を行なうことができる特徴量を提案した。

本システムでは、個別情報支援を実現するために、複数の赤外投光・可視光カットカメラを用いてCoBITを撮影し、取得した画像からCoBITの3次元位置の推定を行っている。このCoBITの3次元位置を、展示品に平行な平面に射影したものをジェスチャ認識のための特徴量とした。これにより、ジェスチャ動作の変化の大きな成分のみを認識に利用することができ、ユーザの向きに頑健なジェスチャ認識を実現した。また、オクルージョンなどの理由で欠損してしまったフレームに対して、欠損フレームの前後のフレームを用いて線形補間を行った。これにより、ユーザの位置や向き、個人差に関わらず認識率が90%以上という高い認識結果が得られるようになり、提案した特徴量の有用性を示すことができた。

今後の課題として、安定してインタラクションができる情報支援システムの構築が挙げられる。本システムでは、ユーザの状態を移動中・停止中・ジェスチャ中の3状態に分け、状態遷移の条件をつけることによって、ユーザの状態の判定を行っている。そして、ジェスチャ認識は、ユーザがジェスチャを行っている状態であると判定した場合にのみ行う。現状では、ユーザの状態の判定に失敗することが多いため、ユーザとのインタラクションを行うためには、安定にユーザ状態を判定できるように改善することが非常に重要である。また、インタラクション中にジェスチャが認識できなかった場合の対処するために、情報支援システムにおいてジェスチャのオンライン教示が可能かどうかについても検討を行っていく予定である。

謝辞 本研究の一部は科学研究費補助金(No.17700174)によった。

参考文献

- [1] 西村他，“無電源小型通信端末を用いた位置に基づく状況支援システム”，情処研報，Vol. ICII-2, pp. 1-6, 2002.
- [2] I. S. Okatani et al, “Location-based Information Support System Using Multiple Cameras and LED Light Sources with the Compact Battery-less Information Terminal(CoBIT),” Proc. of Int. Workshop on CV in HCI, pp. 142-152, 2004.
- [3] 林他，“複数カメラを用いたインタラクティブな個別情報支援システム”，信学全大, pp.D-12-59, 2005.
- [4] 入江他，“濃淡値の時系列変化を利用した画像からの手振り検出”，画像の認識・理解シンポジウム, Vol.I, pp.285-290, 2002.
- [5] 富永他，“ジェスチャ認識のための多視点カメラによる人物位置推定および手領域抽出手法の提案”，情処研報, Vol. CVIM-87, pp.85-92, 2001.
- [6] 大和他，“カテゴリー別VQを用いたHMMによる動作認識法”，信学論, Vol.J77-D-II, no.7, pp.1311-1318, 1994.
- [7] 西村他，“低解像度を用いた複数人物によるジェスチャ認識の単一動画像からのスポットティング認識”，信学論, Vol.J80-D-II, No.6, pp.1563-1570, 1997.
- [8] 島他，“動き情報と情報圧縮を用いたロバストなジェスチャ認識手法”，信学論, Vol.J81-D-II, No.9, pp.1983-1992, 1998.
- [9] 西村他，“動作者適応のためのオンライン教示可能なジェスチャ動画像のスポットティング認識システム”，信学論, Vol.J81-D-II, No.8, pp.1822-1830, 1998.
- [10] J. Alon, et al., “Accurate and Efficient Gesture Spotting via Pruning and Subgesture Reasoning,” Proc. of ICCV HCI, pp.189-198, 2005.
- [11] 武士他，“HMMを用いた手振り認識”，信学技報, Vol.PRMU96-8, pp.53-59, 1996.
- [12] 佐藤他，“複数の観測列候補を扱う隠れマルコフモデルによる移動ロボット用ジェスチャ認識”，信学技報, Vol.PRMU2001-115, pp.15-20, 2001.
- [13] 吳他，“ジェスチャ認識のための連続DPの改良”，システム制御情報学会論文誌, Vol.14. No.6, pp.283-290, 2001.