

K-039

指先追跡による人物ポインティングを用いた実時間情報検索 ～マルチモーダル対話型TVに向けて～

Real-Time Information Retrieval by Human Face Pointing Based on Finger Tracking –Toward Multi-modal Interactive TV–

佐古 淳†
Atsushi Sako

熊野 雅仁†
Masahito Kumano

有木 康雄†
Yasuo Ariki

藤本 雅清†
Masakiyo Fujimoto

1. はじめに

近年、インターネットの普及により、様々な情報源から情報を得ることが可能となった。しかし、依然として主な情報源はテレビニュースに依るところが大きい。現在のテレビ放送では、視聴者の知らない情報があった場合に、直ちに調べることが出来ない。そこで、視聴者が、知らない情報を直接テレビに問い合わせることができる対話型テレビが望まれる。

対話型テレビの例として、画面上に現れた人物が誰であるかを問い合わせるシステムが考えられる。しかし、画面上に複数の人物が現れた場合、音声のみを用いて「この人は誰ですか」と問い合わせても、画面上のどこに位置する人物を意図しているかわからない。

我々は今まで、音声認識とモーショントラッカーを用いた指先の指示位置というマルチモーダル情報を用いた人物情報検索システムの開発に取り組んできた [1]。しかし、モーショントラッカーを用いた方法では、ユーザがLEDを体に装着しなければならないという問題があった。そこで、本稿では、2台のカメラを用いることで、LEDなどの装置を装着することなく指先の指示位置を実時間で推定する手法について提案する。

2. 人物情報検索システムの概要

このシステムは、ニュース映像中にユーザの知らない人物が現れた場合「この人は誰ですか」と問い合わせることにより、問い合わせた人物についての情報を検索するものである。システムは以下のようなモジュールから構成されている。

- ニュース映像制御部
- 音声入力・認識部
- 指先の指示位置推定部
- 顔切り出し・認識部
- 検索結果表示部

まず、音声入力・認識部では、ユーザの音声を認識し、マイクロフォン・アレイを用いて、離れた位置にいるユーザの発話を認識する。音声認識の結果、人物検索の要求であると判定された場合、ニュース映像制御部では、再生中の映像を一時停止する。このとき、同時に指先の指示位置推定部では、ユーザがスクリーン上のどの位置を



図 1: 撮影画像と指示方向検出結果

指さしているかを認識する。顔切り出し・認識部では、指先の指示位置に存在する人物の顔を切りだして認識を行う。検索結果表示部では、認識された人物についての検索結果を表示する。

3. 実時間指先指示位置の推定手法

本節では、2台のカメラを用いて指先の指示位置を推定する手法を提案する。ユーザが指示動作を行う場合、指先の方向が必ずしも意図した指示方向と一致するとは限らない。ユーザの指示する方向は、顔と指先を通る直線の方向とほぼ一致することがわかっている。そこで、提案手法では、実時間処理を可能とする肌色抽出手法を用いて顔と手を検出し、顔と手を結ぶ直線の延長線と、スクリーンとの交点を指先の指示位置とする。

まず、前処理として、あらかじめカメラ・キャリブレーションを行い、カメラの回転行列 R および並進ベクトル t を求めておく。そして、以下の各節で説明する手法に基づいて指先の指示位置を推定する。

3.1 肌色抽出・ノイズ除去

まず、肌色抽出に有効な [2] の手法に基づいて、RGB色空間の入力画像を、 $luv \cdot HSV \cdot RGBY \cdot Wr$ 色空間に変換する。次に、それぞれの色空間内でしきい値処理を施し、肌色部分とそれ以外の部分に分類し、2値化を行う。このとき、小さな湧き出し領域が多量に発生するため、一定の大きさ以下の領域はノイズとみなして取り除く。

3.2 画像上の指先と顔の座標検出

ノイズ除去を行った場合、画像上には、手と顔の領域が残される。ここで、実時間処理を行うために、顔と手

†龍谷大学理工学部

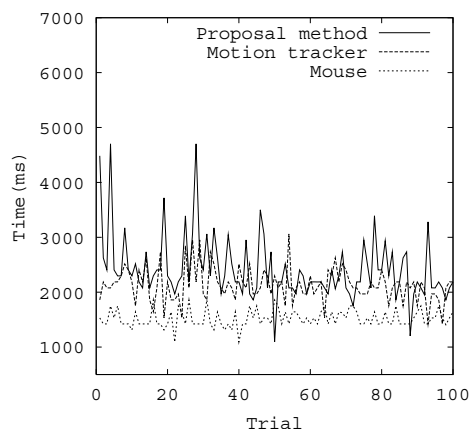


図 2: 試行実験結果

の判別処理は行わず、画像上の肌色領域全体の線形回帰直線を求め、直線上、スクリーンに最も近い肌色領域の点を指先の座標と定義する。また、スクリーンから最も遠い肌色領域の点を顔の座標と定義する。図 1 は、スクリーンの横から人を撮影した *Camera1* と、上から人を撮影した *Camera2* の画像であり、それぞれの画像の下に求められた肌色抽出領域と線形回帰直線を示している。

3.3 指先・頭部の 3 次元座標推定

上記の処理によって求められた撮影画像上の座標から、指先と顔の 3 次元座標を推定する。まず、指先の 3 次元座標を求める。指先の 3 次元座標は、以下の式 (1) と式 (2) で表される直線の交点として計算される。

$$L_1(s_1) = R_1^{-1}(s_1 m_1 - t_1) \quad (1)$$

$$L_2(s_2) = R_2^{-1}(s_2 m_2 - t_2) \quad (2)$$

ここで、 m_1, m_2 は *Camera1, 2* における指先の座標である。実際には、 L_1, L_2 は交差しながることが多いため、2 直線の最近接点を指先の 3 次元座標とする。2 直線の最近接点は、式 (3)(4) によって求められる [3]。

$$s_1 = \det\{(P_2 - P_1), V_2, V_1 \times V_2\} / |V_1 \times V_2|^2 \quad (3)$$

$$s_2 = \det\{(P_2 - P_1), V_1, V_1 \times V_2\} / |V_1 \times V_2|^2 \quad (4)$$

ただし、 $V_i = R_i^{-1} m_i, P_i = -R_i^{-1} t_i$ である。同様にして顔の 3 次元座標も求めることができる。

3.4 指先の指示位置推定

指先の 3 次元座標と顔の 3 次元座標を通る直線とスクリーンとが交わる点を推定値とする。ただし、カメラのノイズや手ぶれによる指示位置のゆらぎを防ぐため、過去 5 フレーム分の推定値を積み付け平均し、その値を指示位置とした。

4. 実験

実験手法としては、スクリーン上に四角形の領域をランダムな位置に表示して、その領域に指示位置が到達し、

1 秒間領域内にとどまるまでの時間を計測した。指示位置が 1 秒間領域内を出なければ、領域は次のランダムな位置に移動する。ただし、領域内に指示位置が 1 秒間連続して存在しなければ、指示位置が 1 秒間領域内にとどまるまでの時間を計測した。なお、指先による指示位置は、スクリーン上で矢印のカーソルとして表示されている。実験に使用したスクリーンは、縦 212cm、横 284cm の大型スクリーンであり、被験者とスクリーンの距離は 4m である。また、計算機は Xeon2.4GHz を使用した。

試行実験としては、被験者を一人として、これを 100 回繰り返した。図 2 にその結果を示す。図 2 には、同じ実験をモーショントラッカーを用いた場合とマウスを用いた場合についても示している。図 2 より、モーショントラッカーと比べ、指先による指示の到達時間は、全体的には若干遅くなっている。しかし、体感速度としてはほぼ同じ速度に感じられる程度であった。また、表 1 に、100 回の試行実験に対する計測時間の平均を示す。

表 1: 領域指示にかかった時間

手法	平均時間 (ms)
提案手法	2420
モーショントラッカー	2132
マウス	1520

表 1 より、提案手法とモーショントラッカーの時間差は、0.3 秒程度であり、体感速度がほぼ同じであったことが伺える。原因としては、計算機の処理速度が考えられる。また、マウスの指示が最も早かった。ただし、マウスは、マウスパッド上の微小な動きでカーソルが大きく移動するため、手の動作よりも有利である。しかし、その差は 1 秒程度であり、人物情報の検索指示については大きな支障はないと思われる。

5. まとめ

本稿では、モーショントラッカーなどの装置を装着することなく、2 台のカメラを用いて指先の指示位置を推定する手法を提案した。しかし、現状では、両手が入ると正確に指示位置が推定できないという問題や、指先と視線により意図される指示位置と推定された位置が感覚としてずれるという問題がある。今後は、これらの問題に取り組んでいく予定である。

参考文献

- [1] 井上 徹, 藤本 雅清, 山本 夏夫, 有木 康雄, 熊野 雅仁, 堂下 修司: “ニュース映像に対する発話内容と人物問い合わせシステム” 第 1 回情報科学技術フォーラム, FIT2002, K-60, pp.487-488, 2002-09.
- [2] G. Gomez. On selecting colour components for skin detection. Proc. of the ICPR, vol. 2, pp. 961-964, Aug. 2002.
- [3] 齋藤 真希子, 小池 英樹, 佐藤 洋一: “複数視点画像に基づく手の 3 次元位置ならびにジェスチャの実時間計測” ヒューマンインターフェース学会 ヒューマンインターフェースシンポジウム HIS99, pp. 423-428, October 1999.