

素手による手話動画像からの手領域追跡

5 J-5

今川 和幸†‡ 呂 山† 猪木 誠二†

† 郵政省 通信総合研究所 ‡ 松下電器産業（株）

1. はじめに

現在、新たなヒューマンインターフェースや聴覚障害者と健常者との間のコミュニケーション支援を目指して、カメラにより撮影した人物動画像から手領域を抽出し、手形状、手振り、手話を認識する研究が盛んに行われている。しかしながら、ほとんどの場合、一様背景、手に手袋等のマーカーを装着するなど、手を抽出しやすい環境を仮定していた。特に、手話の場合、顔の前で手を動かす場合も多く手にマーカーを装着する必要がある。しかし、実際には素手による手話動作から手領域を抽出し追跡することが望ましい。そこで、我々は、色情報を利用して手顔領域を抽出し、さらに動き情報を利用して手領域を追跡する手法を、実際の手話動画像に適用して検討した。本稿では、我々の用いた手法と実際の手話動画像に適用した実験結果について述べる。

2. 手顔領域の抽出

まず、手話動画像の第1フレームの画像から色情報を用いて手顔領域を抽出する。なお、抽出後の画像は、顔の前に存在する手の動きがわかる画像である必要がある。また、様々な照明環境のもとでも、あまり影響を受けない画像であることが望ましい。そこで、Swain らによるヒストグラム逆投影法 (Histogram Backprojection) [1] を改良した以下の手法を提案する。

図1に本手法の概念図を示す。

1. 本手法を適用する原画像を用意する(①)。
2. マニュアルで、抽出したい手顔領域を囲む(②)。
3. $d \times d \times d$ の空間(bin)に分割した3次元色空間を用意する(③)。なお、3次元色空間として、今回はRGB色空間を用いる。
4. 原画像をスキャンし、画素の色に従い3次元色空間のbinへ投票する(④)。投票する画素が抽出したい領域であれば t 票、背景領域であれば b 票 ($b < 0$) 投票する。なお、今回の実験では、

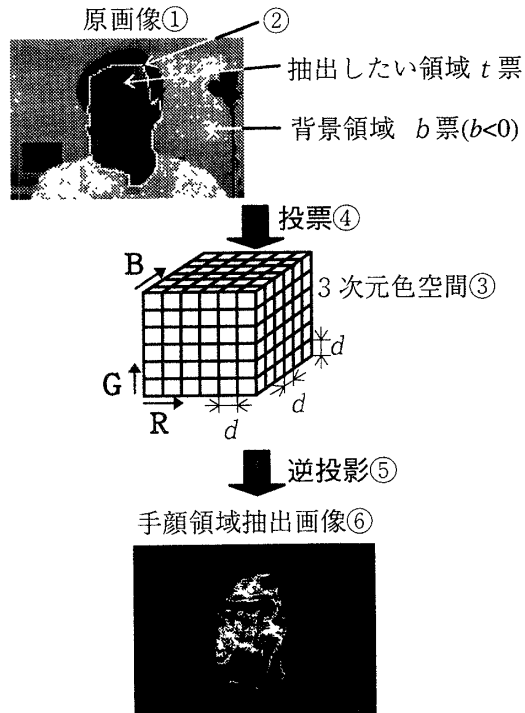


図1 手顔領域の抽出法

5. それぞれの bin の投票結果を V 、出力画像の明度の範囲を D とした場合、それぞれの bin の重心位置の色 C の出力値 R が

$$R(C) = \text{Max} \left(\frac{D}{V}, 0 \right)$$

となるように、原画像に逆投影する(⑤)。なお、各 bin の重心位置以外の色をもつ画素の場合、PRISM方式[2]による補間演算を行う。以上の処理の結果、顔の前の手の存在がわかる画像(⑥)を得ることができる。

なお、実際の手話動画像へは次のように適用する。一般に、手話動作中は照明環境などがそれほど変化しないと考えられる。そこで、各 bin の重心位置の色をテーブル値とする3次元ルックアップテーブル(LUT)を用意する。そして、手話動画像の第1フレームの画像を用いて、上記手法により3次元LUTのテーブル値を設定する。第2フレーム以降の画像は設定したLUTを用いて出力画像へ変換する。なお、今回は、 $d=32$ のRGB 3次元LUTによりビデオレートで出力画像へ変換する独自の色変換ボードを用いて手話動画像から手顔抽出動画像へ変換した。

Tracking Hands by Skin Color from Sign Language Image Sequences.

Kazuyuki Imagawa, Shang Lu, Seiji Igi.
 † Communications Research Laboratory, Ministry of Posts and Telecommunications.
 ‡ Matsushita Electric Industrial Co., Ltd.
 4-2-1, Nukui-Kitamachi, Koganei, Tokyo 184, Japan.

3. 動き領域の抽出

次に、顔より手の方が動きが大きいことを利用して、手顔抽出画像から手領域を分離し、手を追跡する。詳細を以下に示す。

1. 時刻 t の画像と時刻 $t+1$ の画像との間の差分をとる。
2. ある閾値を用いて二値化してラベリングを行い、領域を抽出する。そして、領域を構成する全画素数が、ある閾値より大きくなる領域を取得する。この領域が手の候補となる領域となる。
3. 照明などの影響により、実際には手の領域候補が分裂している場合がある。そのため、抽出した動き領域の重心位置を特徴パラメータにし、K平均法によるクラスタリングにより領域を統合する。なお、クラスタリングの終了条件は、領域の重心間の最小距離がある閾値を越えた場合とする。

4. 手の位置の追跡

動き領域抽出の結果、動画像の各フレームで手領域が存在する候補となる位置をいくつか得ることができる。しかしながら、手が静止している場合、前述の方法だけで、静止位置を得ることは不可能である。また、両手が動作している場合、右手、左手それぞれに対し追跡する必要がある。そこで、前フレームの手の位置を予測値として、右手、左手それぞれに対し、予測値から最も近い候補を次の手の位置とする。なお、予測値からある距離以内に候補となる位置が得られない場合、予測値を次の手の位置とする。

5. 追跡実験

本手法の有効性を検証するために、実際の手話動画像を用いて手の位置の追跡実験を行った。

5.1. 実験に用いた手話動画像

手話動画像として、以下の動画像を用意した。

1. CCDカメラで撮影し、書き込み可能レーザーディスクに保存した3名のろう者動画像。
2. ビデオジャーナル「手話ジャーナル[3]」の2名のろう者動画像。

上記いずれの場合も、背景とカメラの位置が固定で、肌色以外の長袖の服を着用している人物の正面上半身動画像を用意した。そして、各人物が手話を行っている30秒間の動画像を追跡実験に用いた。

5.2. 実験システムの構成

以下に実験システムの構成を示す。まず、用意した手話動画像から前述の色変換ボードを用いて手顔領域を抽出する。次に、手顔抽出画像を毎秒30frames/sで市販の汎用画像処理ボード(IMAGING

TECHNOLOGY Series 150/40)に入力する。入力された画像は、640×480画素の8bit画像に変換される。汎用画像処理ボードでは、時系列差分、二値化、ラベリング処理を行い、ラベリングの結果をパソコンに転送する。パソコンでは、クラスタリング、手の位置の予測を行い、手領域を追跡する。なお、追跡結果として得られる手の位置を元の手話動画像上にオーバーレイ表示した。

なお、以上の処理は、1フレームあたり1/30秒以内であるため、ビデオレートで手領域を追跡することが可能となる。

5.3. 実験結果

追跡実験の解析では、オーバーレイ表示されている手の位置が、実際の手領域に含まれているか否かを1フレーム毎に目視で判定した。なお、両手が画面上で接触している場合、本システムでは両手まとめて一つの領域と処理される。そこで、この場合は両手のどちらかにオーバーレイ表示されている手の位置が含まれていれば追跡成功とした。その結果、追跡率が61%であった。なお、手が静止している場合と顔の動きが激しい場合、追跡に失敗することが多かった。

次に、追跡実験の結果から、右手、左手のいづれかが顔の前に存在しているフレームのみを抜き出し、追跡率を調べた。なお、手の一部が画面上で顔と接触している場合、手が顔の前に存在していた。その結果、追跡率が62%であった。以上の結果、手が顔の前にある場合も同等の追跡がなされていることがわかった。

6. おわりに

本報告では、手話動画像から手を追跡することを目的として、色情報と動き情報を用いて追跡する手法を実際の手話動画像に適用し、その結果を報告した。今後は、本結果を生かし、より安定した追跡手法を開発していく予定である。

謝辞

本研究を進めるにあたり、実験に協力して頂いた埼玉県所沢市聴覚障害者協会の方々に感謝致します。

参考文献

- [1] M. J. Swain, D. H. Ballard: "Color Indexing", Int. J. Comput. Vision, Vol 7, No. 1, pp. 11-32 (1991)
- [2] Kanamori: "Fast color processor with programmable interpolation by small memory (PRISM)", J. of Electronic Imaging, Vol. 2, No. 3, July (1993)
- [3] サインファクトリー編: "手話ジャーナル 1996 春号", 手話ジャーナル, No. 1, (1996)