

手話認識のための動画像を用いた複雑背景における手指形状推定

浜田康志 島田伸敬 白井良明 三浦純
大阪大学大学院 工学研究科 電子制御機械工学専攻

Hand Shape Estimation using Image Sequence under Complex Backgrounds for Sign Language Recognition

Y. HAMADA, N. SHIMADA, Y. SHIRAI, and J. MIURA

Dept. of Computer-Controlled Mechanical Systems, Graduate School of Engineering, Osaka University

1. まえがき

近年、ジェスチャ認識の研究がさかんに行われている。特に手話認識は障害者の意思伝達手段の幅を広げることが期待できる。手話認識において手話話者にわずらわしさを感じさせない手指形状推定が重要であり、画像を用いた手法が適している。

文献 [1] では手指形状抽出が容易な条件下での形状推定、ゆっくり変化するジェスチャについて複雑背景下での形状の抽出と推定を行った。そこで本論文では、自然な手話の画像から手指形状の抽出と推定を行う。あらかじめサンプル画像から代表的な形状をモデルとして登録し、入力画像に最適なモデルを選択する。最適モデル選択では、入力画像から明度変化の大きいエッジ点を抽出し、それらと最も良く適合するモデルを選ぶ。しかし複雑背景では強いエッジが手指輪郭を表すとは限らない。そこで背景中のエッジ点存在確率を考慮した評価基準を用いる。

2. モデルの学習

本論文では右手のみについて形状推定を行う。サンプル画像列の各画像から手指の重心位置、移動速度、輪郭形状を求め、形状モデルとして登録する。

3. 手指形状推定

3.1. マッチング候補の限定

遷移ネットワーク [2] を用いて入力画像と照合する形状モデルを限定する。あらかじめサンプル画像列から形状モデル間の可能な遷移を学習し、遷移ネットワークを生成する。時系列画像の入力に対して、遷移ネットワークを用いて直前の画像で推定された形状から遷移可能な形状モデルに候補を限定する。

さらに時系列入力における動きの連続性の仮定により手指位置を予測し、その近傍にない形状モデルを候補から除く。また入力画像から肌色領域を抽出し、この領域内にない形状モデルもマッチング候補から除く。最終的に残ったマッチング候補モデルから入力画像に最適な形状モデルを選択する。

3.2. 最適モデル選択のための評価基準

各マッチング候補モデルについて入力画像中のエッジ点と最も良く重なる位置に合わせたのち、候補モデルの中から最適なモデルを選択し、推定結果とする。そこで候補モデルを比較するための評価基準について述べる。

従来手法では強いエッジが手指の輪郭を表すことを仮定している。しかし背景や手指領域の内部にも強いエッジが存在し、また真の輪郭であっても顔などとの境界ではエッジが弱くなる。そのため図 1(a) の入力画像に対し、図 1(b) のように形状推定を誤る。そこでモデルの輪郭と重なるエッジ点の数 n_e には背景のエッジ点が含まれる事を考慮し、モデル輪郭の長さ l と背景中のエッジ点存在確率 P_F を用いて評価基準 c_m を定義する。

$$c_m = n_e - lP_F \quad (1)$$

各候補モデルについて c_m を求め、 c_m が大きいほど入力画像により適合するモデルとする。

図 1(c) にこの評価基準を用いて推定した形状を示す。また「赤い」を意味する日本語手話単語の画像列に対する推定結果を図 2 に示す。

4. むすび

複雑背景を持つ入力の各画像について最適な形状の推定を行った。今後の課題は推定された形状の時系列を単語や文章として認識することである。

参考文献

[1] 島田 伸敬, 白井 良明, "手話認識のための画像に基づく非接触手指形状計測," 第 17 回日本エム・イー学会秋季大会, 2003.

[2] 浜田 康志, 島田 伸敬, 白井 良明, "遷移ネットワークに基づく多視点画像時系列からの手指形状推定," 信学論 (D-II), vol.J85-D-II, no.8, pp.1291-1299, Aug. 2002.



図 1: 背景中のエッジ点存在確率に基づく評価基準による推定結果

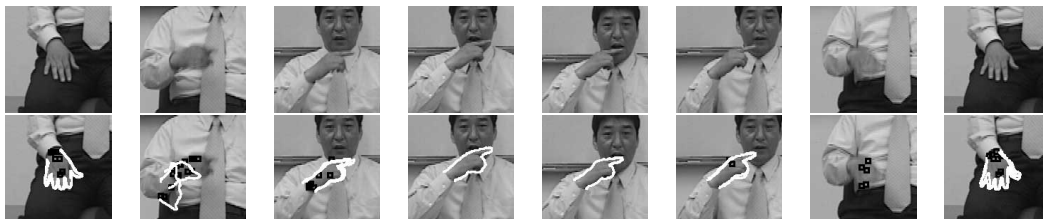


図 2: 推定結果例 (上段: 入力画像列, 下段: 推定結果画像列)。白い線は推定された形状の上位四つを表す。手指の移動速度が速いときは輪郭は重要な意味を持たないので手指位置のみを用いて追跡する。黒い四角はそのときの位置を表す)