

動的色閾値を利用した唇検出に関する一検討 A Study on Lip Detection Method Using Dynamic Color Threshold

二階 雅弘[†] 西村 広光[†]
Masahiro Nikai Hiromitsu Nishimura

1. はじめに

手話は口話と同様に、イントネーションや感情を込めた表現が可能である。具体的には、「非常においしい」という手話に感情をこめて伝えることもあれば、感情をこめなで手話を行うことも可能である。

我々は、これまでに画像による手話認識手法^[1]を構築してきた。本報告では、手話におけるイントネーションや感情を表現する手法の中で、最も基本的かつ大切な要素である唇の動きに着目し、唇検出を高精度に行う技術について検討した。

従来の唇検出は、対話型ロボットによる読唇の基盤となる発話検出^[2]や、セキュリティ分野では個人識別に向けた検出^[3]が行われてきた。本研究はこれまで検討されてきた分野とは異なり、手話における非言語情報を対象とした。

本研究は、手話認識への応用を目指し、唇動作の認識の前処理となる唇検出として、動的な色閾値を利用した唇領域の抽出法を検討した。

2. 撮影環境

本研究の最終目標は HMD などのウェアラブルデバイスでの手話認識を想定している。そこで、撮影機材には Logicool 社の C930e の Web カメラを使用した。撮影環境を図 1 に示す。Web カメラと被写体との距離は、2 者の一般的な対話距離に基づき 1m とした。Web カメラの高さは唇が画像中央付近になるように床から 1.4m とした。解像度は 1920×1080 とした。

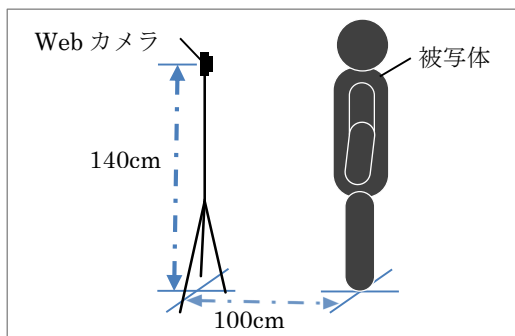


図 1 撮影環境

3. 色閾値の動的探索

唇領域抽出の精度向上には、動的輪郭法の一つである Snakes^[4]を参考にした。一般的な Snakes が輪郭点をエネルギー収束によりコントロールするのに対し、本手法は抽出した唇領域の形状をもとに、色閾値をコントロールする点で性質が異なる。考案アルゴリズムを(1)から(5)に示す。

- (1) 平均唇の色値を用いて唇候補領域を抽出する。
- (2) 抽出した唇候補領域の面積に応じて、閾値変動フラグに「狭窄」または「弛緩」を設定する。

- (3) 閾値変動フラグを元に色相の閾値を変動させ、新たに唇候補領域を抽出する。抽出の対象となる範囲は(1)で抽出した唇候補領域の接線矩形サイズとした。
- (4) (3)で得られた唇候補領域ごとに理想唇との形状特徴の差分を算出する。形状特徴の差分結果が前回の差分よりも小さい場合には、閾値変動を続行する。差分結果が前回よりも大きい場合には、色閾値の変動対象とする上限値または下限値の弛緩完了フラグ、または狭窄の完了フラグを立てる。また、変動させた色閾値を前回抽出時の色閾値に戻す。未完了のフラグが存在する場合には、フラグに応じて閾値変動フラグに狭窄または弛緩を設定する。
- (5) 全唇候補領域の下限値と上限値が保持する狭窄完了フラグと弛緩完了フラグが完了でない場合には、(3)に戻り色閾値の変動と唇候補領域の再抽出を行う。

(1)から(2)の処理は、平均唇から得られた色閾値を用いて大まかな唇領域候補領域の抽出を行う。平均唇から得られた色閾値は、1人につき 20 枚の撮影を 10 人に実施して得られた 200 枚の画像群 A に含まれる唇の色相、彩度、明度とした。

(3)から(5)の処理は、色閾値の変動による唇候補領域の再抽出と理想唇との形状比較を繰り返し行い、唇候補領域ごとに最適な色閾値を動的に探索する。(3)から(5)の処理では色相閾値のみを色閾値として探索する。これは色相、彩度、明度において、色相のみが唇と肌の境界線付近において色値の変化が激しく、唇の輪郭抽出に最適であると考えたためである。色閾値である色相の初期値は平均唇から得られた色閾値を用いた。

(4)処理で用いる理想唇は、平均唇を求めた画像群 A から得られた唇の中でも、より唇らしい形状をしていた領域を複数選び、平均化したものである。唇の形状特徴は、唇候補領域の矩形占有率(以下、占有率と略記する)、左右対称率、上下対称率とした。理想唇の形状特徴は、各形状特徴の最小値と最大値の平均値とした。理想唇と形状特徴の差分算出では、占有率、左右対称率、上下対称率で絶対値による差分を合計した。

4. 評価実験

開発した色閾値の動的探索アルゴリズムの唇抽出精度の調査と非唇領域の変化の調査を目的とした評価実験を行った。評価実験では、平均唇の色閾値(以下、固定色閾値と略記する)を用いた唇検出と動的探索で得られた色閾値(以下、動的色閾値と略記する)による唇検出の二種類の唇抽出法を用いた。被験者には、唇検出前の入力画像、固定色閾値による唇検出結果画像、動的色閾値による唇検出結果画像を目視させた。3 枚の画像は横一列に表示し、左端は唇検出前の入力画像が常に表示され、固定色閾値による唇検出結果画像と動的色閾値を用いた唇検出結果画像は、ランダムに中央または右端に表示されるようにした。被験者に

[†] 神奈川工科大学大学院

は、どちらが固定色閾値による唇検結果画像なのか動的探索を用いた唇検出結果画像か知らせないこととした。

本評価実験で唇検出する画像は、あらかじめ撮影した 20 代の男女 10 名の動画画像から閉口状態のフレームをランダムに 20 枚選択した 200 枚の画像データ群 B を用いた。200 枚の画像を見せる順番はランダムとした。本評価実験は 20 代の男性 9 名を対象に行った。

4.1 唇抽出精度の調査実験

画像中に含まれる唇領域を対象に、唇領域の抽出精度について評価を行った。しかし、被験者ごとに唇領域を認識する感覚の度合いが異なるため、唇領域の認識基準が必要となった。以下に、唇領域の認識基準と条件を示す。

条件1: 唇は上唇下唇と別々の領域として捉えず、唇全体で捉えること。

条件2: 唇全体として捉えた時の唇候補領域が、唇面積の半分以上且つ、唇面積の倍以内の領域は唇領域として扱うこと。

唇領域の認識後、被験者には二種類の唇検出手法による唇領域形状の変化の有無で異なる回答をさせた。2 つの唇検出結果画像を比較し、唇検出結果画像中の唇領域形状に変化がある場合には、どちらがよりの確に唇領域を抽出しているかを回答させた。唇領域が存在しない、または唇領域形状に変化が無い場合には、変化なしとして回答させた。

唇抽出精度の結果を表 1 示す。回答結果の枚数は、同一画像の重複を含めた 1800 枚である。唇領域に変化ありとなった回答数は 895 枚、変化なしとなった回答数は 905 枚であった。唇領域に変化が起きた 895 枚の内、よりの確に唇領域を検出した数は、固定色閾値が 36.6%にあたる 328 枚であり、動的色閾値は 63.4%にあたる 567 枚であった。

開発した色閾値の動的探索アルゴリズムにより、全体の約 30%において唇領域の抽出精度が向上した。全被験者の回答結果が一致した唇検出結果において、動的色閾値により精度が向上した成功画像を図 2、失敗画像を図 3 に示す。

表 1 唇検出精度の結果

		数(割合)	
変化あり	動的色閾値	567	(31.5%)
	固定色閾値	328	(18.2%)
変化なし		905	(50.3%)
合計		1800	(100.0%)

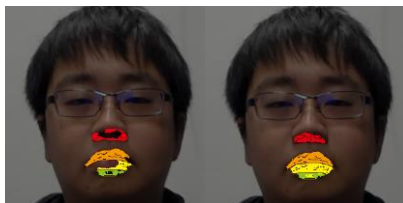


図 2 唇検出成功例(左:固定色閾値 右:動的色閾値)

4.2 非唇領域の調査実験

唇領域を除いた非唇領域の変化について調査実験を行った。被験者には固定色閾値による唇検出結果画像と動的色閾値による唇検出結果画像を比較することなく、一枚画像

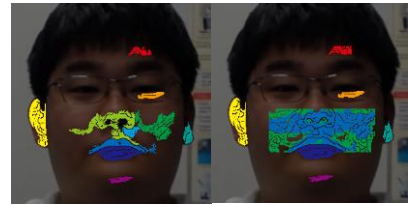


図 3 唇検出の失敗例(左:固定色閾値 右:動的色閾値)

に含まれる非唇領域について回答させた。回答は「とても多い」「多い」「少ない」「とても少ない」「(非唇領域が)無い」のなかで直感的に当てはまる項目を選択させた。

各回答項目の変化を全体として捉えるため、固定色閾値と動的色閾値における回答項目ごとに重み付けと集計を行った。その結果、固定色閾値では 5086 に対して動的探索閾値では 5343 であり、非唇領域が増加している結果となった。回答項目ごとの割合を示した結果が図 4 である。

非唇領域が増加する要因として、動的探索アルゴリズムにおける唇候補領域の再抽出処理手順が考えられる。唇候補領域の再抽出時、唇候補領域ごとに閾値の変動と抽出を行った後にラベリング処理を行っている。そのため、唇候補領域の接線矩形が他の唇候補領域と重なる、または隣接する場合において、再抽出した複数の唇候補領域が一つの大きな領域になってしまうのである。解決策として、唇候補領域の再抽出ごとにラベリング処理することを検討中である。

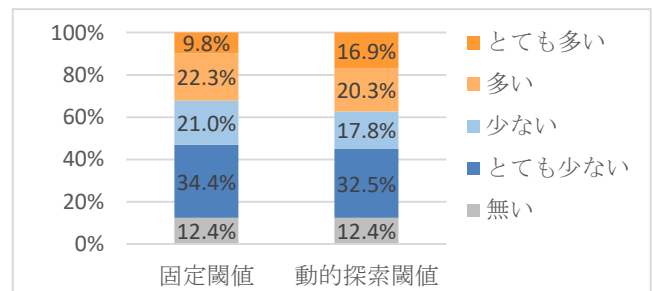


図 4 回答項目ごとの割合

5. おわりに

本研究では、手話にイントネーションや感情を込めた表現方法の一つである唇の動きに着目し、唇候補領域の抽出精度向上を目的とし、動的色閾値を利用した唇検出法を検討した。評価実験の結果、色閾値を動的探索することによって、全体の約 30%において抽出精度が向上していることがわかった。しかし、非唇領域が全体的に増加している結果となった。

今後、色閾値の動的探索処理における唇候補領域の再抽出範囲を制御し、非唇領域の面積肥大を抑制する手法を検討し、さらなる精度向上を実現する。

参考文献

- [1] 小澤 辰典 ほか, “複数の弱識別統合による手話認識性能の向上”, 信学ソ大, B-20-3(2017).
- [2] 甲斐 寛規 ほか, “口唇領域の抽出と認識による発話検出”, CVIM, 2011-CVIM-177(2011).
- [3] 白澤 洋一 ほか, “口唇の動き特徴を用いた個人識別に関する検討”, 映像情報メディア学会誌, Vol.60, No.12, pp1964~1970(2006).
- [4] M.Kass ほか, “Snakes: Active contour models”, International Journal of Computer Vision, 321-331(1988).