

QoL 評価に向けた軽度認知症の表情変化計測

迫 匠一郎†

近藤 一晃‡

小泉 敬寛‡

中村 裕一‡

† 京都大学大学院工学研究科電気工学専攻

‡ 京都大学学術情報メディアセンター

1 はじめに

高齢化社会が進んだ昨今，認知症者の数は年々増加している．アルツハイマー型の認知症などは基本的に治せることはないため，治療するのではなく進行を遅らせる，または生活の質 (Quality-of-Life:QoL) を高く保つことが重要とされている．通常 QoL の計測は認知症のケアマッピングのように対象者の振る舞いを短時間 (数分) 毎につぶさに観測して記録する方法が用いられるが，これには高い技能と多大な労力を要する．

このような問題背景から，我々は認知症者の QoL を自動的に測る方法について取り組んでいる．これにより人手で行ってきた作業を部分的に計算機に置き換える，人手の結果をバックアップするエビデンスとして用いる，計測者に依らない客観的な計測 [1] が可能となる，といったメリットが得られる．QoL が最もよく現れる振る舞いとしてはまず表情が挙げられ，得に笑顔の頻度と QoL には強い相関がある [2]．しかし日常生活下の認知症の表情を対象とした計測・認識に関してはこれまでほとんど取り扱われていない．そこで本稿では既存手法を用いたときの結果を分析することで，QoL 計測のための表情認識における課題点を明かにする．

2 日常生活における表情認識の難しさ

QoL 計測のための表情認識には，(1) 認知症者を対象とすることと，(2) 日常生活の記録を用いることという大きく 2 つの課題があると考えている．ただし健康者と比べても表情の表出傾向は変わらずむしろ感情がそのまま表情に出やすい，という認知症に携わる医師や介護スタッフの体験から，(1) の認知症者を対象とすることはあまり支障にはならないと考えている．

一方で (2) は考慮すべき問題である．従来から非正面の顔に対する表情認識が難しいことは知られているが，それに加えて我々は発話中の表情認識が重要かつ課題ではないかと考えている．なぜなら表情の変化はコミュニケーション時に多く発生する一方で，発話による顔形状の変化が表情認識に悪影響を及ぼす可能



図 1: 認知症者の対話シーン記録の様子 (認知症者役 (左)，スタッフ役 (右)．認知症者の正面にある机上のカメラで表情を記録)

性があるからである．例えば，口周辺の形状が激しく変化することでその領域における顔特徴の検出に失敗したり，単に発話しているだけでも関わらずある特定の時刻だけに着目するとまるで表情が変化したかのように見えたりすることが考えられる．

3 表情認識実験

3.1 表情計測と笑顔識別

上記で述べたような問題が実際にどの程度深刻なのかを，発話時・非発話時の映像データに既存の表情認識手法を適用する．すなわち発話時・非発話時では表情認識結果にどのような差が出るかを検証する．

実際の認知症者を撮影対象とするには倫理上の手続きが必要となる．そこで本実験では認知症者のケアを日々行っている病院の介護スタッフ 2 名が，各々認知症者役およびスタッフ役を模擬した対話シーンを撮影対象とした (図 1)．具体的には配布物を読んでいる認知症者に大して何について書いてあるか問いかけるシーンである．認知症者役の人物の表情はカラー画像 640×480pixel，深度画像 320×240 pixel の小型の RGB-D カメラを正面に設置して撮影した．

表情認識は顔検出・顔向き正規化を行った上で目尻・口角といった顔器官の特徴点を検出し，それらの相対的な一のパターンを事前登録したパターンと照合することで行う．本稿では発話時・非発話時の表情それぞれに対して笑顔・非笑顔の 2 クラス識別を行った．なお発話の有無および笑顔かどうかの真値は目

†Shoichiro Sako ‡Kazuaki Kondo ‡Takahiro Koizumi ‡Yuichi Nakamura

†Graduate School of Engineering,Kyoto University

‡Academic Center for Computing and Media Studies,Kyoto University

表 1: 発話時・非発話時における笑顔の識別結果の違い

非発話時の識別結果 (単位は frame 数)			非発話時の識別結果 (単位は frame 数)		
認識結果	真値		認識結果	真値	
	笑顔	非笑顔		笑顔	非笑顔
笑顔	837	51	笑顔	218	82
非笑顔	121	451	非笑顔	267	273
適合率=94.3%			適合率=72.7%		
再現率=87.4%			再現率=45.0%		
F 値=0.453			F 値=0.278		

視で行った。

3.2 表情認識の結果と考察

被験者映像から発話と非発話時のフレームをそれぞれ抽出し、それらに対して笑顔か非笑顔を示すタグを手で付与した。これらの各フレームに対して表情認識処理により笑顔か非笑顔の認識を行なった。

認識結果を表 1 に示す。認識結果から、発話時における表情認識の結果は非発話時に比べて適合率、再現率ともに低い値となっていることがわかる。このことから、本稿では表情認識に失敗する主な原因となる以下の 2 点について原因を調べて考察を行なった。

1. 顔認識や顔部位検出の失敗
2. 発話による顔形状変化の影響による失敗

1. のような失敗は、想定外の表情が現れた場面や、顔を正面から捉えられていない場面で発生していた。例えば、図 2 のように、口元が突き出た表情では、下唇と顎の合間にわずかな凹凸が生じる。この凹の部分で“口”として検出され、下唇部分において誤った識別結果となった。また、図 3 のように顔が大きく傾げられた場合は、顔自体の検出ができなため顔器官の特徴点が得られず、表情認識が行なえなかった。

次に、2. が原因となる誤認識の場面を調査した。図 4 に示すように発話において母音が“u” および“o” の際に、口元が突き出た形状となっていた。そのため、発話中かつ笑顔であったにも関わらず笑顔と認識されなかった。これは、非発話時の笑顔に比べて、これらの発話時には口角が上がらないためであると考えられる。逆に、母音が“i” および“e” の発話の際には、図 5 のように口角が上がることで、発話中かつ笑顔でないにも関わらず笑顔と認識されていた。

以上のように、発話による口角の形状の変化が、表情識別に大きく影響していることを確認できた。その

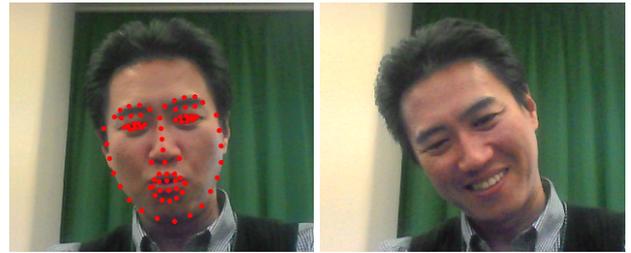


図 2: 下唇部分が間違った 図 3: 顔の検出に失敗して部分に誤対応している例 特徴点が現れない例

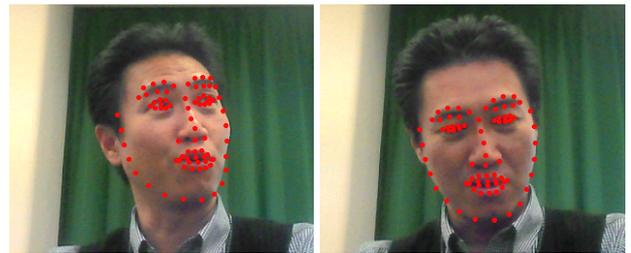


図 4: “u” 音発話時の表情 図 5: “i” 音発話時の表情の例

ため、発話時の表情認識を行うためには、顔全体の複合的な要素 [3] を見る手法を検討することや、前後の時間帯の表情認識結果や音声などの情報を追加することで表情認識の精度を改善する手法を考えていく必要があると考えられる。

今後はこのような表情認識精度を改善する手法を模索するとともに、QoL 計測のために必要な計測環境についても検討を進めたい。

参考文献

- [1] Kezheng Lin, “Facial Expression Recognition Based on Geometric Features and Geodesic Distance”, International Journal of Signal Processing Image Processing and Pattern Recognition vol7, No1, pp323-330, (2014)
- [2] 花沢明俊ら, “認知症老人の笑顔画像計測による情動機能回復評価”, 一般社団法人電子情報通信学会信学技報, pp39-42, (2012)
- [3] 檜山洋一, “FACS を用いた複数感情の表情合成手法”, 電子情報通信学会総合大会講演論文集 (2008 年), S-74-75, (2008)