

# キネクトを用いた映像視聴時の弱表情の検出について On the Detection of Small Expressions in Video Viewers with a Kinect Sensor

サイモン クリッピングデル<sup>†</sup> 望月貴裕<sup>†</sup> 佐野 雅規<sup>†</sup> 菅沼 睦<sup>‡</sup> 亀山 渉<sup>‡</sup>

Simon CLIPPINGDALE<sup>†</sup> Takahiro MOCHIZUKI<sup>†</sup> Masanori SANNO<sup>†</sup> Mutsumi SUGANUMA<sup>‡</sup> Wataru KAMEYAMA<sup>‡</sup>

## 1. まえがき

視聴者の嗜好を考慮したコンテンツ提供を目指し、映像視聴時の反応を取得する方法について研究を進めてきた。視聴実験を行い、反応の一つとして顔の小さな変形「弱表情」を分析した[1]。

映像視聴中の被験者 40 人の顔をキネクトセンサで撮影し、キネクトが追跡する特徴点 (FP) の配置から、フレーム毎の頭部姿勢の影響を除去した「FP 残量」(つまり顔の変形分) を推定した。FP 残量の主成分分析 (PCA) を行い、PCA 係数の時系列から(弱)表情を検出した[2]。

検出の精度を把握するために、アノテータ 3 人がキネクト映像を見て(弱)表情発生時を指定してもらうことで「正解データ」を作成し、検出器の出力と比較した。

今回は、実時間で顔変形データを解析するシステムの開発に向けて、FP 残量データを射影する PCA 空間の構築方法が、検出精度に及ぼす影響を検討したので、その結果を報告する。

## 2. (弱)表情検出システムの構成

システム全体の構図を図 1、(弱)表情検出器の内部構成を図 2 に示す。キネクトが出力する「High Detail Face Points」の中で、弱表情に大きく関与すると考えられる目・鼻・口周辺に位置する 18 点の FP を追跡し、Iterative Bundle Adjustment により頭部全体の動きを除去し、表情に伴う変形と思われる「FP 残量」をフレーム毎に推定する。

FP 残量は 36 次元ベクトルであり、約 10Hz の頻度で取得する。オフラインで解析するとき、FP 残量データの主成分分析 (PCA) を行い、エネルギーの大きい順に 10 個の PCA 係数を使用し、その 10 次元ベクトルの 1 秒平均をとった時系列の変化量から(弱)表情を検出する (詳細は [2] を参照)。

### 2.1. 実時間での(弱)表情検出についての課題

視聴実験にて、キネクトからの FP データを記録し、実験のセッション終了後にオフラインで残量推定、PCA、解析した。そのときに、図 1 のように、PCA 空間を個人毎に構築することにより、(弱)表情に関して個人的な癖が反映されると考えられる。

しかし、将来にリアルタイム且つオンラインで(弱)表情を検出・解析する場合に、初めてのユーザーが現れたら、個人に合った PCA 空間を事前に構築することができない。

<sup>†</sup> NHK 放送技術研究所

<sup>‡</sup> 早稲田大学

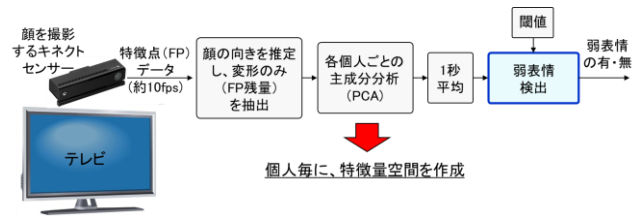


図 1. (弱)表情検出システムの構成概要

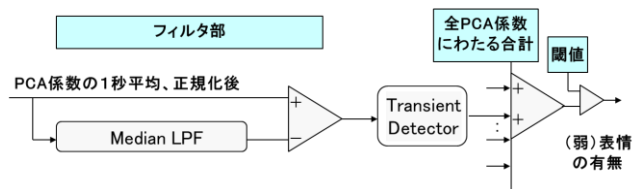


図 2. (弱)表情検出器の内部構成

もし個人に合っていない PCA 空間を使用したら、どの程度に(弱)表情検出の精度が低下するかが課題になる。そこで、その状況をシミュレーションするためにデータを解析した。

## 3. PCA 空間による検出精度の実験

2.1 節の課題を受け、被験者 40 人の FP 残量データを 2 つの条件下に得られた (弱)表情検出の精度を比較した。

被験者毎に、視聴実験で取得したその被験者の FP 残量データを PCA 空間に射影して、求めた PCA 係数の 1 秒平均をとって、(弱)表情検出器に入力して、検出器の閾値に連れて、検出器の出力を「正解データ」と、1 秒間隔毎に比較した。ここで「正解データ」は、アノテータ 3 人のなかに 2 人以上が 1 秒以内に何らかの(弱)表情が発生したとの指定があったとした。

今回の実験で比較した条件は、PCA 空間の構築方法の下記の 2 つの条件であった：

### 1. 「個人の PCA 空間」

その被験者のすべての FP 残量データ (10 次元ベクトル、約 10Hz で取得された) をデータ行列に並べて、PCA を行うことにより、PCA 空間を構築した (ここで、「すべてのデータ」は視聴してもらったテスト映像の 10 個×約 5 分ずつであった [1])；

### 2. 「個人以外の PCA 空間」

その被験者以外の 39 人のすべての FP 残量データをデータ行列に並べて、PCA を行うことにより、PCA 空間を構築した。

「個人の PCA 空間」は各個人の表情が張る，エネルギーが最大な 10 次元空間で，個人の癖を反映することが期待される．一方，「個人以外の PCA 空間」は他の被験者の表情が張る，エネルギーが最大な 10 次元空間なので，個人に限った癖を反映しないので検出精度が低下する可能性があると考えられる．

#### 4. 実験結果と考察

2 種類の PCA 条件下での(弱)表情検出の精度を F 値として，図 3 に示す．横軸は検出器の閾値，縦軸は F 値としての精度である．個人の PCA 空間使用条件のグラフは“PCA\_self”のラベル，個人以外の PCA 空間使用条件のグラフは“leaveoneout”のラベルが付いている．

両グラフは，閾値ごとに，全被験者にわたって合計的な F 値を示す．個人の PCA 空間使用条件下の最大 F 値 (0.620) は個人以外の PCA 空間使用条件下の最大 F 値 (0.614) より大きい，グラフ同士の差分が小さい．

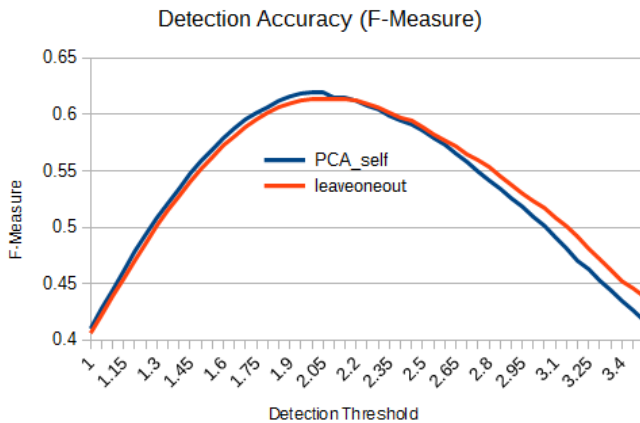


図 3. 検出精度 (F 値) 対検出器の閾値.  
条件：個人の PCA 空間使用 (PCA\_self)，  
個人以外の PCA 空間使用 (leaveoneout)

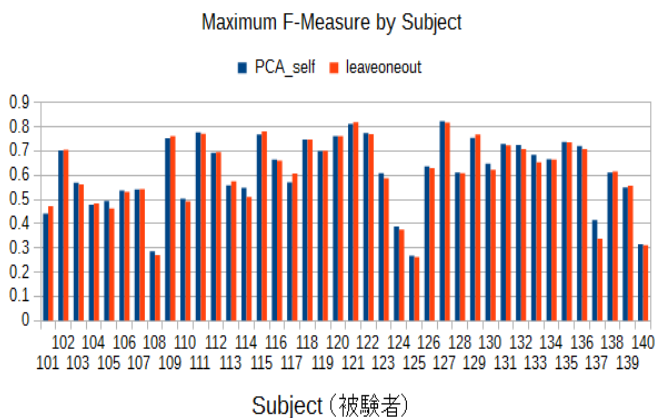


図 4. 最大検出精度 (F 値) 对被験者.  
条件：個人の PCA 空間使用 (PCA\_self)，  
個人以外の PCA 空間使用 (leaveoneout)

また，両条件下で得られた F 値が被験者によって大きく違う．被験者ごとに，閾値にわたっての最大精度 (F 値) を図 4 に示す．最大 F 値は被験者に大きくよるが，条件 (PCA 空間) にはほぼよらない．

これらの結果から，事前用意不可である個人 PCA 空間の代わりに，事前用意可能な汎用 PCA 空間を使用しても，期待できる(弱)表情検出の精度がほぼ変わらない．今回の解析では，「個人以外の PCA 空間」はこの汎用 PCA 空間をシミュレートした．汎用 PCA 空間を既存の特徴点データから構築すれば，初めてのユーザでも直ぐに使える(弱)表情検出システムが可能になり，その個人に合わせた PCA 空間を使用した場合に比べてほぼ精度が変わらないことが期待できる．

#### 5. まとめと今後の課題

本報告では，映像視聴時に視聴者の顔特徴点 (FP) をキネクトセンサで取得し，(弱)表情を検出するシステムについて述べた．本システムは頭部姿勢を推定，除去して，特徴点集合の変形 (FP 残量) を残し，主成分分析 (PCA) を行う．エネルギーの大きい順で幾つか (現在 10) の PCA 係数の 1 秒平均をとり，変化を抽出することにより(弱)表情を検出する．

そこで，システムをオンラインで使用して，初めてのユーザが現れた場合に，そのユーザの FP 残量データがまだ無いので，入力される FP 残量データをどの PCA 空間に射影すれば良いかという課題が出る．本報告では，“Leave One Out” クロスヴァリデーションに基づいた解析により，他のユーザのデータから構築された PCA 空間を用意して使用すれば，個人に合わせた PCA 空間使用時に比べて，(弱)表情検出の精度がほぼ変わらないことを示した．理由は，(弱)表情に伴う FP 残量の殆どが他の人物の他の表情の主成分の線形的組み合わせにより近似ができるからであると考えられる．

今後の課題のひとつは，今回の実験で使った 2 条件下で，PCA 空間への射影エラーなどの解析から，各条件下で表現できなかった成分を処理し，各被験者の個人的な癖を指定できるかを検討していく．

もう一つの課題は視聴者が複数人で一緒に映像を視聴するときの表情処理にある．どの表情が視聴中の映像への反応，どの表情が映像と関係の無いものを識別することが課題である．

最後に，特徴点集合の変形から情動へのマッピング (個人差を含む) の推定が大きな課題である．

#### 参考文献

- [1] クリピングデル 他，“映像視聴時における弱表情の変化点とその個人差検出についての一検討”，映像情報メディア学会年次大会，24-D-1 (2016 年)．
- [2] クリピングデル 他，“キネクトセンサによる映像視聴時の弱表情検出についての一検討”，映像情報メディア学会冬季大会，11-C-1 (2017 年)．