This article is a technical report without peer review, and its polished and/or extended version may be published elsewhere.

第 25 回日本バーチャルリアリティ学会大会論文集（2020 年 9 月）

# 頭部装着型ディスプレイに組み込んだ反射型光センサによる表情認識のための機械学習の検討

A Study on Machine Learning Approaches for Facial Expression Recognition with Embedded Photo Reflective Sensors in Head-Mounted Display

中村文彦 1)
Fumihiko NAKAMURA

1) 慶應義塾大学大学院 理工学系研究科（〒 223-8522 神奈川県横浜市港北区日吉 3-14-1, f.nakamura@imlab.ics.keio.ac.jp）

**概要**: 頭部装着型ディスプレイ（HMD）の普及とともに，VR 環境に没入するアプリケーションが多数提供されるようになった．このようなアプリケーションにおいてユーザの表情の認識は重要である．本研究では，HMD に組み込んだ反射型光センサによる顔面形状の計測と機械学習を組み合わせて HMD 装着者の表情を認識する手法を検討する．また，訓練データの収集を自動化する手法についても検討する．

**キーワード**： 表情認識，機械学習，反射型光センサ

## 1. 序論

2010 年代前半に消費者向けの HMD (Head-Mounted Display) が登場して以降，Virtual Reality (VR) を活用したアプリケーションが多数提供されるようになった．このようなアプリケーションでは，一般的にユーザが自己を投射する身体としてアバターが用いられる．アバターを介したコミュニケーションでは，音声やテキストといった言語情報だけでなく，ジェスチャや表情といった非言語情報も不可欠な要素である．ユーザの腕や足のジェスチャは認識されアバターに反映されている一方で，表情認識はあまり利用されていない．表情認識では顔画像を用いることが多いが，HMD を着用している時はディスプレイによって顔の大部分が隠されているため，顔画像の取得が難しい．

そこで，著者らは HMD を着用したユーザの表情を反射型光センサを用いて認識する技術を提案してきた (図 1) [7][8]．HMD に組み込んだ反射型光センサによって表情を計測し，機械学習を用いて表情を認識した (図 2)．反射型光センサは，光を照射し，計測対象から反射された光を受光することで，センサから計測対象までの距離を計測できる素子である．反射型光センサは，小型，軽量，低電力といった特徴があり，HMD への組み込みが容易である．表情によって光センサ値は異なるので，各表情のセンサ値を機械学習することでユーザの表情を認識した．

光センサによる表情認識の研究の中で，機械学習モデルの精度や学習データの収集法を検討してきた．著者らはこれまでに，目の周辺と頬を計測し表情認識する手法 [7] と口周辺と頬を計測し口元の形状を認識する手法 [8] を提案した．目周辺と頬を計測する手法では，機械学習モデルとして NN

図 1: 表情のアバターへのマッピング [7]

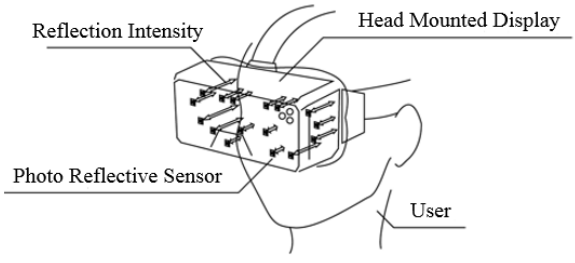


図 2: 反射型光センサによる表情の計測 [7]

ⓒ 2020 日本バーチャルリアリティ学会

図 3: (左) 目周辺を計測する表情認識デバイス [7]，(右) 反射型光センサによる表情認識システム構成 [7]

(Neural Network) を用いた．さらに，学習プロセスを自動化するために，ユーザに提示したアバターの表情をまねさせて訓練データを自動的に収集する手法を提案した．口周辺と頬を計測する手法では，SVM (Support Vector Machine) を用い，学習プロセスの自動化のためにユーザの音声を用いて訓練データを自動的に収集する手法を提案した．このように機械学習のモデルや訓練データを自動収集する手法についての検討を進めてきた．

## 2. 関連研究

表情認識はコンピュータビジョンの分野において長い間研究されている．一般的には，顔画像から目や口等の顔のパーツを認識する．RGB カメラを用いて細やかな表情まで認識する手法 [1] や，RGBD カメラを用いて顔面形状を再構成する手法が提案されている [2]．これらの手法は明瞭な顔画像が必要であるので，顔の大部分がディスプレイによって覆われている HMD 装着時には利用が難しい．HMD 装着者の表情認識のために，HMD に取り付けた RGB-D カメラと接触センサの情報から HMD 下の顔面形状を推定する手法も提案されている [3]．画像処理による表情認識では高精度に顔の動きが認識できるが，計算量が高コストであることや精度がカメラの性能に依存すること等の問題がある．

低次元データの処理による表情認識も研究されている．筋電センサを用いて笑顔と怒りの表情を認識する手法 [6] が開発されている．また，反射型光センサを用いて皮膚表面の変形を計測し表情を認識する手法も提案されている．Shoji らはマスク型インタフェースに光センサを取り付けて目・眉・口等の動きを計測し，着ぐるみの顔に反映した [4]．また，Masai らは光センサを取り付けたメガネ型デバイスと機械学習を用いて 6 表情を認識した [5]．反射型光センサによる表情認識は計算量が少なくセンサも小型・軽量なので HMD に統合しやすいと考えられる．そこで，著者らはこれまでに反射型光センサを用いた HMD 装着者の表情を認識する手法を提案してきた．

## 3. 反射型光センサによる表情認識

HMD に組み込まれた反射型光センサと機械学習を用いて表情を認識した，また，中間表情を学習・推定し，アバターに階調値を持った表情変化を反映した．表情の学習時に，アバターの表情をまねすることで自動的に訓練データを収集した．システム構成を図 3 右に示す．

表情認識デバイスを市販の HMD を改造して作成した (図 3 左)．Oculus Rift DK2 に反射型光センサ (Kodenshi, SG-105) を目の周辺 14 か所と口の周辺 2 か所に取り付けた．

表情の計測は Masai ら [5] と同様の原理で行った．表情を変化させた時に起こる顔面の皮膚形状の形状の変化を光センサで計測した．表情の認識では，表情を識別しその強度を推定するために NN による表情識別と NN による各表情の線形回帰を組み合わせた．表情識別の NN は入力層 (16 ノード) と隠れ層 (100 ノード) と出力層 (5 ノード) で構成された．各表情の線形回帰での NN は入力層 (16 ノード) と隠れ層 (100 ノード) と出力層 (1 ノード) で構成された．認識した表情をアバターに反映する際には，表情識別結果と各表情の強度を用いて表情を合成し中間表情を再現した．

また，訓練データの収集を自動化するためにユーザにアバターの表情をまねさせる手法を試みた．各表情を段階的に変化させるアバターをユーザに提示し，その表情を都度真似するように教示した．このプロセスを通して，各表情のラベルとその強度をラベル付けした訓練データを自動的に取得した．

反射型光センサによる表情識別精度について評価実験を行った．本実験では，データセットを収集する条件を変えた時の識別精度を評価した．条件は，再装着 (デバイスを再装着したか) と自動収集 (データセットを提案した自動収集法によって取得したか) であった．識別精度を，再装着なし＋自動収集なし，再装着あり＋自動収集なし，再装着あり＋自動収集なし，の 3 条件下で比較した．被験者は各条件で 20 代の成人男性 10 名であり，個人毎に学習・識別を行った．識別した表情は無表情，喜び，怒り，驚き，悲しみの 5 つであった．再装着なしの場合は，デバイスを一度も取り外さずに合計 5000 サンプル (1 表情につき 1000 サンプル) 取得した．再装着ありの場合は，デバイスを装着し 1 表情につき 100 サンプル取得した後に，デバイスを取り外し再装着するサイクルを 10 セット繰り返し，合計 500 サンプル (1 表情につき 1000 サンプル) 取得した．訓練データとテストデータは前半と後半のそれぞれ 2500 サンプルとした．

ⓒ 2020 日本バーチャルリアリティ学会

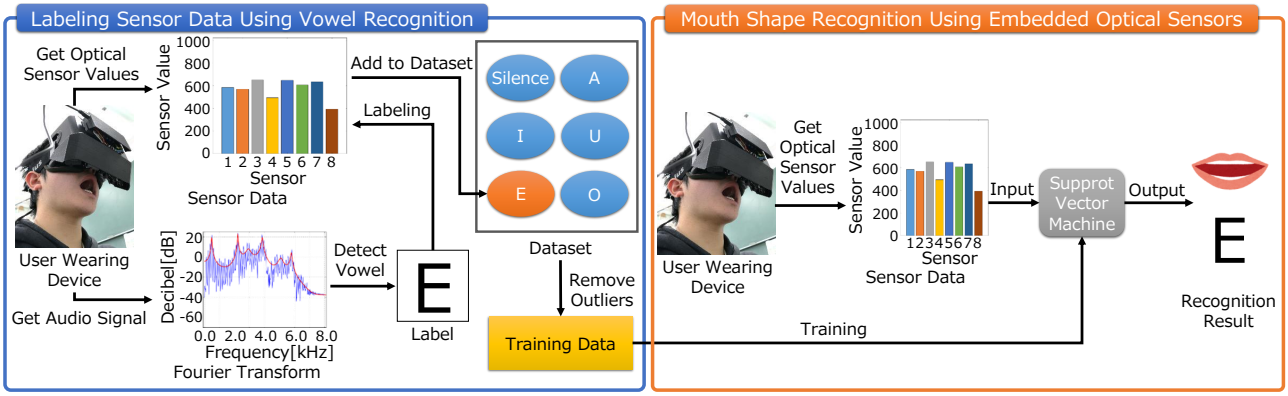


図 4: 口元形状認識システム [8]

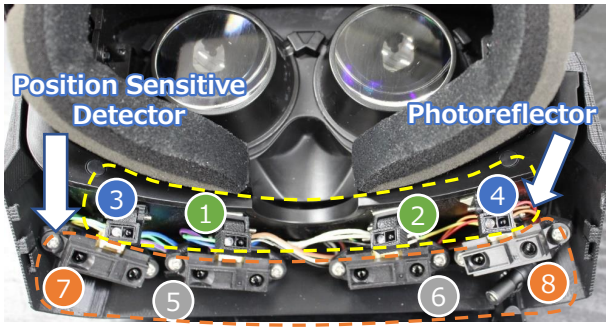


図 5: 口元形状認識デバイス [8]

各条件下での表情毎の認識精度は表 1 に示す．再装着なし+自動収集なしの識別精度は全体で 90.4%，再装着あり+自動収集なしの識別精度は全体で 88.6%，再装着あり+自動収集なしの識別精度は 83.9%となった．自動的に訓練データセットを取得した場合，再装着ありで手動でデータセットを取得した場合と比較して 4.7%の識別率の低下が見られた．笑顔は全被験者で誤識別が少なかった．また，怒りと悲しみの識別率が低かったが，これらは互いを誤識別することが多かった．

表 1: 各条件での表情の認識精度

| Condition | | Neutral | Happy | Angry | Surprised | Sad |
|---|---|---|---|---|---|---|
| 再装着 | 自動収集 | | | | | |
| なし | なし | 96.6% | 92.6% | 92.1% | 93.8% | 76.7% |
| あり | なし | 95.7% | 98.9% | 80.0% | 92.1% | 76.2% |
| あり | あり | 96.4% | 93.0% | 78.5% | 85.5% | 66.2% |

## 4. 反射型光センサによる口元形状認識

HMD に組み込まれた反射型光センサと機械学習を用いて口元形状を認識した，また，中間の口元形状を推定しアバターに反映した．表情の学習時に，音声から認識した母音をラベルとしてセンサ値に付与することで自動的に訓練データを収集した．システム構成を図 4 に示す．

口元形状認識デバイスを市販の HMD を改造して作成した (図 5)．Oculus Rift DK2 の外装底部に 4 個の反射型光センサ (LBR-127HLD) と 4 個の PSD (Position Sensitive Detector) 距離センサ (SHARP GP2Y0A21YK) とマイクロホン (audio-technica AT9904) を取り付けた．反射型光センサは上唇と上頬を計測し，PSD 距離センサでは下唇と頬を計測した．

口元周辺の動きの計測は Masai ら [5] と同様の原理で行った．発声時に起こる顔面の皮膚形状の変化を光センサで計測した．機械学習モデルとして線形カーネルの確率推定付き多クラス分類 SVM (Support Vector Machine) を用いて 8 個の光センサの値から口元形状を認識した．また，無表情と母音の口元形状を基底として口元形状を合成することで中間の口元形状を再現した (図 6)．口元形状の合成には SVM から出力された各クラスへの所属確率を重みとして用いた．

音声を用いてセンサ値にラベルを自動的に付与した．人間が声を出す時には声に応じて口唇や頬といった口元周辺の形状は変化する．また，母音は唇の円形度合いや顎の開閉度合い等の口元形状によって生じる音である．そこで，音声から認識した母音をセンサ値と自動的に対応付けし，データセットを自動的に収集した．しかし，母音認識を利用して収集されたデータセットは外れ値を含み得る．そこで，マハラノビス距離を用いてデータセットから外れ値を除去した．外れ値除去されたデータセットを訓練データとして用いて機械学習した．

組み込み式光センサによる口元形状の認識精度について評価実験を行った．実験では，手動学習 (手動で訓練データをラベル付けすること) と自動学習 (音声を用いて自動で訓練データをラベル付けすること) の 2 条件下で比較した．実験が終わるまでデバイスを取り外して再装着することはなかった．認識対象は無音時と日本語 5 母音の発音時の 6 個の口元形状であった．実験の被験者は各条件で 20 代の日本人男性 5 名であった．手動学習では各被験者で訓練データは 900 サンプル (150 サンプル * 6 口元形状) を収集し，テストデータは各被験者で 1200 サンプル (200 サンプル * 6 口元形状) を収集した．自動学習では各被験者でデータセットは 900 サンプル (150 サンプル * 6 口元形状) を収集し，テストデータは各被験者で 1200 サンプル (200 サンプル * 6 口元形状) を収集した．このとき，テストデータは手動で訓練データをラベル付けした．

各条件での口元形状の認識結果を表 2 に示す．手動学習での認識率と自動学習での認識率は全体で平均してそれぞ

ⓒ 2020 日本バーチャルリアリティ学会

図 6: アバターへの多様な口元形状の反映 [8]

れ約 99.9%と約 96.4%となった．自動学習での認識率は手動学習と比較して約 3.5%低下したが，自動学習でも高い認識率を達成した．被験者 C では手動学習と自動学習を比較すると，認識率が大きく低下した．これは，被験者 C は実験中に特定の口元形状を同じような形で作らなかったことが原因と考えられる．音声は口元周辺の形状だけでなく舌などによっても変わるので，被験者 C は実験中に母音を発声するときに口元ではなく舌を動かしたと思われる．そのため，データセットの特定のクラスの訓練データの分散が大きくなり十分に外れ値を除去できなかったため，認識精度が低下したと考えられる．

表 2: 各条件での口元形状の認識精度

| 被験者 | A | B | C | D | E |
|---|---|---|---|---|---|
| 手動学習 | 100.0% | 99.3% | 100.0% | 100.0% | 100.0% |
| 自動学習 | 100.0% | 86.1% | 95.7% | 100.0% | 100.0% |

## 5. 結論

本研究では，HMD に組み込まれた反射型光センサと機械学習を用いてユーザの表情を認識する手法を検討してきた．目周辺を計測する表情認識手法や，口周辺を計測する口元形状の認識手法を提案し，光センサによる表情認識のための機械学習モデルや訓練データ取得の自動化などを開発した．現時点でも，高い精度で表情や口元形状の認識は可能であるが，再装着した場合の認識精度の低下や訓練データ取得プロセスの改善などの課題がまだある．

**謝辞** 本研究の一部は，JSPS 科研費（16H05870）の助成を受けたものです．

## 参考文献

[1] Tian Ying-li, Kanade Takeo, Cohn, Jeffrey F.: Recognizing Action Units for Facial Expression Analysis, IEEE Trans. Pattern Anal. Mach. Intell., Vol.23, No.2, pp. 97–115, 2001.

[2] Michael Zollhöfer, Matthias Nießner, Shahram Izadi, Christoph Rehmann, Christopher Zach, Matthew Fisher, Chenglei Wu, Andrew Fitzgibbon, Charles Loop, Christian Theobalt, and Marc Stamminger: Real-time Non-rigid Reconstruction Using an RGB-D Camera, ACM Trans. Graph., 33(4), pp. 156:1-156:12, 2014.

[3] Hao Li, Laura Trutoiu, Kyle Olszewski, Lingyu Wei, Tristan Trutna, Pei-Lun Hsieh, Aaron Nicholls, and Chongyang Ma: Facial Performance Sensing Head-mounted Display, ACM Trans. Graph., Vol. 34, No. 4, pp. 47:1-47:9, 2015.

[4] Rika Shoji, Toshiki Yoshiike, Yuya Kikukawa, Tadahiro Nishikawa, Saori Taigetsu, Ayaka Suketomo, Tetsuaki Baba, and Kumiko Kushiyama: Mimicat: Face Input Interface Supporting Animatronics Costume Performer's Facial Expression, ACM SIGGRAPH 2012 Posters, SIGGRAPH '12, pp. 72:1-72:1, 2012.

[5] 正井克俊, 杉浦裕太, 尾形正泰, クンツェカイ, 稲見昌彦, 杉本麻樹: AffectiveWear: 装着者の日常的な表情を認識する眼鏡型装置，日本バーチャルリアリティ学会論文誌, 21(2), pp. 385-394, 2016.

[6] Gruebler, Anna and Suzuki, Kenji: Design of a wearable device for reading positive expressions from facial emg signals, IEEE Transactions on Affective Computing, Vol.5, No.3, pp.227–237, 2014.

[7] Katsuhiro Suzuki, Fumihiko Nakamura, Jiu Otsuka, Katsutoshi Masai, Yuta Itoh, Yuta Sugiura, and Maki Sugimoto: Recognition and Mapping of Facial Expressions to Avatar by Embedded Photo Reflective Sensors in Head Mounted Display; 2017 IEEE Virtual Reality (VR), pp. 177-185, 2017

[8] Fumihiko Nakamura, Katsuhiro Suzuki, Katsutoshi Masai, Yuta Itoh, Yuta Sugiura, and Maki Sugimoto: Automatic Labeling of Training Data by Vowel Recognition for Mouth Shape Recognition with Optical Sensors Embedded in Head-Mounted Display, Proceedings of ICAT-EGVE 2019, pp. 9-16, 2019.

ⓒ 2020 日本バーチャルリアリティ学会