

EmoCast: 顔への表情変化情報投影による話し相手の印象向上システム

浅野 恭志*大西 鮎美†寺田 努†塚本 昌彦†

概要. 会話時の表情は話し相手の印象に大きな影響を与えており、表情が豊かでないと話し相手に悪い印象を与える。本研究では、静止対象物にグレースケールの動き情報を投影して対象物が動いているように錯覚させる技術に応用し、顔のパーツが動く会話中などでも運動パターンを投影できるウェアラブル表情拡張システム EmoCast を提案する。提案システムでは、カメラで取得した顔映像をスマートフォンに送信して顔の 3D モデルを作成し、この 3D モデルにあらかじめ作成した運動パターンをマッピングして、プロジェクタで本人の顔に投影する。これにより、無表情からの表情の生成、表情の増強、表情の打消しといった機能が実現できる。評価実験により、動きのある顔への運動情報の投影で表情がより変化したと錯覚させることができ、会話時に提案システムを使用することで印象が良くなることが示唆された。

1 はじめに

会話時の表情は、話し相手の印象に大きな影響を与える。Mehrabian は、会話時に印象を与える要素として会話内容が 7%、音声トーンや話し方が 38%、そして表情や身振りが 55% であると示した [4]。このように、表情や身振りは話し相手の印象に大きな影響を与えており、表情が豊かでないと話し相手に悪い印象を与えてしまう。

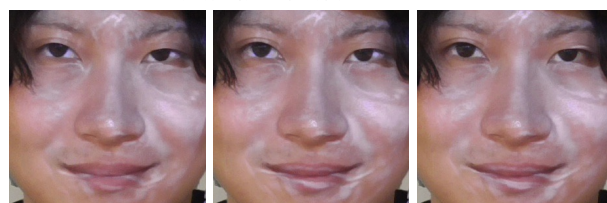
会話時の表情を変化させる先行研究では、顔の一部をディスプレイで覆ったり [8][10]、顔の部位にプロジェクションマッピングを行って塗りつぶしたりすることで [1]、理想的な表情を提示するものが多い。しかし、顔の一部を覆ったり、部分的にでも口や鼻などの部位をプロジェクションによって塗り替えると、細かい表情の変化を塗りつぶしてしまうため、投影対象者の顔自体が話し相手から見えづらくなる。

静止対象物にグレースケールの動き情報を投影して対象物が動いているように錯覚させる変幻灯がある [9]。この手法は、ユーザの顔の部位を塗りつぶさずに表情を変えるのにも有効であると考えられるが、静止物体への投影のみを対象としており、会話中などの動きのある顔への投影を対象としていない。

そこで本研究では、動きのある顔のパーツに追従して運動パターンを投影し、会話時の表情を豊かに見せることで話し相手への印象を良くするウェアラブル表情拡張システム EmoCast を提案する。提案システムでは、カメラで取得した顔映像から現在の顔の 3D モデルを作成し、この 3D モデルにあらかじめ生成した運動パターンをマッピングして、プロ



無表情を笑顔にみせる



笑顔を増強する

図 1: 提案システムの機能の一例

ジェクタで投影する。これにより、図 1 の例のような無表情からの表情の生成、表情の増強、表情の打消しといった機能が実現できる。運動パターンを投影している性質上、図のような静止画ではシステムの機能が把握しづらいため、動画¹を確認いただきたい。

本稿では提案システムのプロトタイプを実装し、評価実験として、顔に運動情報を付与して動きのある顔のパーツがより動いて見えたか、その結果会話時の印象が良くなったかを確認する質問紙調査を実施した。結果より、動きのある顔に運動情報を投影することでより表情が変化したと錯覚させることができ、会話時に提案システムを使用することで印象が良くなることが示唆された。

Copyright is held by the author(s).

* 神戸大学工学部

Faculty of Engineering, Kobe University

† 神戸大学大学院工学研究科

Graduate School of Engineering, Kobe University

¹ <https://youtu.be/h2qkCgmjvMA>

2 関連研究

2.1 顔に表情を提示する研究

表情を豊かにするために、顔の上に表情を提示する研究は多数行われている。石井らの HappyMouth は、ユーザの選択した表情や会話内容を提示するマスク型ディスプレイで、会話の促進を目的としている [10]。また Kumasaki らは、ユーザの表情の印象を変化させるマスク型ディスプレイ TransEmotion を提案している [8]。このデバイスはユーザの口の形や動きに同期してマスク型のディスプレイに CG で生成した口唇を表示し、対話相手がおもつユーザの印象を変化させることを目的としている。しかし、これらのデバイスは顔の一部をディスプレイで覆うため、ユーザの表情を直接見せず、あらかじめ用意した映像を表示するため、投影対象者の顔自体が話し相手から見えづらくなる。

そこで本研究ではプロジェクタを用いて表情変化の運動情報のみを顔に投影することで、ユーザの顔の一部を覆うことなくユーザの表情を変化させ、相手に与える印象を良くするデバイスを提案する。

2.2 顔へのプロジェクションマッピング

プロジェクタを用い、表情変化を支援する研究はいくつか存在する。Bermano らが提案する Make-UpLamp は、舞台上で演じる俳優や女優の外見を変化させるシステムであり、顔の特徴やしわを強調するようなエフェクトを使用できる [1]。また渡辺らは、高速カラープロジェクタと高速カメラを組み合わせ高速に顔にプロジェクションマッピングを行い、バーチャルメイクなどに応用可能なシステムを提案している [11]。

これらの研究は顔のパーツの動きに追従してプロジェクションマッピングを行う点で本研究と共通するが、しわや顔の部位といった、新たな顔のパーツの映像をユーザの顔に重ねて投影し、ユーザの見た目を別のものに塗り替えるため、細かな表情の変化が塗りつぶされてしまう。また、鼻や口といった顔の部位を新たに塗り替えるために、適切な色をもつ解像度の高い映像を投影できる高性能な大型プロジェクタが必要となってしまう、システムを利用できる場所が限られる。よって本研究で提案するデバイスは、運動情報のみ投影することで、ユーザの細かな表情の変化を塗りつぶすことなく表情を豊かにすることを目指し、さらに会話で使用することを想定しているため、構成デバイスを持ち運び可能な大きさにする。

2.3 運動情報を投影して表情を変化させる手法

運動情報のみをユーザに投影し、ユーザの顔の部位を塗り替えずに投影する技術として、吹上らの変幻灯が有効であると考えられる [9]。変幻灯は、静止

対象物にグレースケールの運動パターンを投影し、静止対象物が動いているかのように錯覚させる技術である。この技術の仕組みは、色や形、動きの情報を独立に分析した後に脳内で統合する人間の視覚特性に基づいており、運動情報のみをプロジェクタから投影し、静止対象物の色や形を塗り替えることなく動いているように知覚させる。この技術を用いることで、顔の部位の色や形を塗りつぶすことなく、表情が変化しているように見せることができる。また、運動情報は空間解像度が低くても認知でき、グレースケールでも違和感が生じないため、高性能でない小型プロジェクタでも投影する映像に違和感が生じにくい。

変幻灯は静止物体への投影のみを対象としているが、本研究では会話中など動きのある顔への投影を対象とする。そのため、顔のパーツの動きに追従して運動パターンを投影できるよう、リアルタイムに顔映像を処理し、顔の変化に応じて映像をマッピングして投影する必要がある。本研究では、動きのある会話時の顔にリアルタイムで表情の運動パターンを投影して表情が変化しているように錯覚させ、会話時の印象を良くすることを旨とする。

3 提案システム

提案システムでは、表情が豊かに見える映像を会話時の顔に投影して、話し相手への印象を良くすることを旨とする。提案システムは、以下のような機能に対応できるように設計する。

表情の生成

無表情の顔に特定の表情を生成する

表情の増強

表情が少ししか変化しない場合に表情を増強する

表情の打消し

出すべき表情が間違っている場合に表情をキャンセルする

提案システムの使用例を以下に示す。

- 顔の表情がでない人が、周りの人に面白く思っていると思わせたい状況で、特定の表情を生成する機能を活用することで、その人が笑顔で話を聞いているように周りの人に感じさせる
- 顔の表情がでにくい人が、話し相手に感じている印象をより深く伝えたい状況で、表情を増強する映像を活用することで、その人が話をよく聞いているように周りの人に感じさせる
- 適切な顔の表情を作ることができない人が、笑ってはいけない状況で笑ってしまった時に周りに笑っていることを隠したい状況で、表情を打消す機能を活用することで、その人が



図 2: 作成したデバイス

あまり笑っていないように周りの人に感じさせる

3.1 システム構成

本研究では、会話における話し相手への印象を良くすることを目的としている。会話で用いるためには、デバイスは持ち運びできるような小型化する必要がある。また、会話時は特に口元が大きく動くため、顔のパーツの動きに追従して、表情を強めたいタイミングでリアルタイムに表情を増強する必要がある。図 2 に提案システムのプロトタイプ構成を示す。プロトタイプがヘルメット型であるのは、顔全体に投影を行うことで様々な表情を拡張するため、部分的な顔のパーツの拡張に限定すればシステムは小型化できる。システムはヘルメット型デバイスとスマートフォンで構成されており、ヘルメットに小型プロジェクタと Web カメラを固定している。ユーザは小型プロジェクタや Web カメラをスマートフォンに接続し、システムを利用する。

提案システムの処理の流れを図 3 に示す。スマートフォンのストレージに、あらかじめ作成した表情の運動パターンと、映像内の顔の特徴点を表す UV 座標を格納する。システムが起動すると、デバイスに搭載したカメラが顔の映像を撮影し始める。スマートフォン内では、撮影した顔の映像を機械学習モデルに渡して顔の特徴点を取得し、3D モデルを作成する。3D モデルの作成と並列して、カメラやマイクなどから音声や表情などの情報を受け取る。これらの情報を分析し、トリガとなる情報を捉えた際に、あらかじめ作成した運動パターンを、顔の特徴点の UV 座標を用いて顔の 3D モデルにマッピングする。

表情の運動パターンの作成方法

提案システムで用いる運動パターンは、無表情からある表情に変化する様子をカメラで撮影し、無表情画像との差分を取って作成する。この方法は、グレースケールの静止画像と静止画像を動的に変形した画像列との差分を計算して作成する変幻灯の運動誘導パターンを参考にしている。

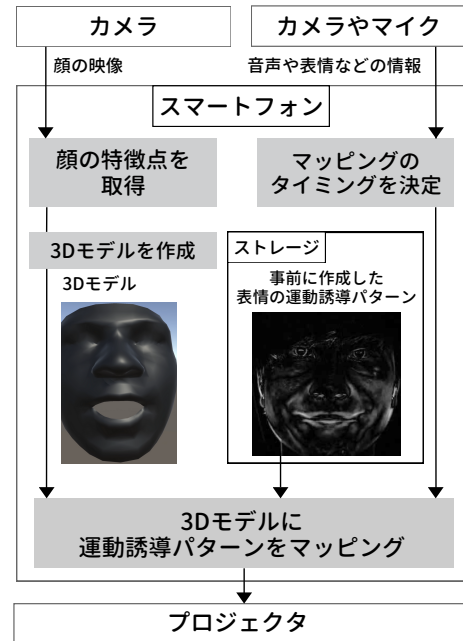


図 3: 提案システムの処理の流れ

各感情に対応する表情の作成には、Sato らの日本人の感情と表情の関係に関する研究にて提示されている、基本 6 感情(怒り・嫌悪・恐怖・喜び・悲しみ・驚き)に対応するシナリオをもとにした表情生成方法を採用した [5]。図 5 のように、基本 6 感情に対応する表情に変化する様子をカメラで撮影し、無表情の画像との差分を取り、図 4 に示すような基本 6 感情に対応する表情の運動パターンとした。

さらにより効果的な映像の作成を目指し、表情の変化によって大きく動き、運動情報の適切な差分が取れない輪郭や顔の部位といった投影時に白くなりすぎる箇所の映像を削除した。また、3D モデルの輪郭部分をユーザの顔のサイズよりも大きくしてから投影することで顔全体にプロジェクタの光が当たるようにし、プロジェクタの投影光による違和感を軽減した。

投影モデルの作成方法

会話時には様々な顔のパーツが動くため、顔のパーツの動きに追従して映像を投影する必要がある。装着者の顔と投影する映像の間に大きなズレがあると、顔の動きと投影する映像が別々で知覚されてしまう [9]。そのため、投影モデルを作成する処理は高速に行う必要がある。

提案システムでは、機械学習モデルで顔の特徴点を取得し、Unity 内で顔の 3D モデルを作成する。スマートフォンを用いた投影モデル作成を高速化するため、機械学習モデルに MediaPipe の FaceMesh



図 4: 笑顔の運動パターン

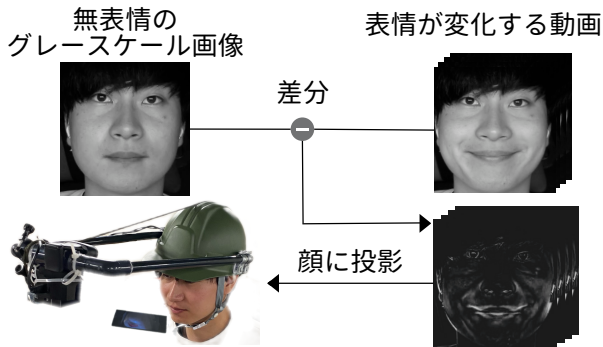


図 5: 表情の運動パターンの作成方法

を用い [3], TensorFlowLite の GPUDelegate という高速化手法を取り入れた。FaceMesh はオープンソースの機械学習ライブラリである MediaPipe のモデルの 1 つであり、顔の 478 個の特徴点を高速に取得できる。このモデルはエッジデバイスに最適化されているため、今回使用した Android 端末上でも高速に顔の特徴点を取得できる。また、GPUDelegate はエッジデバイス内の GPU を活用して推論時間を短縮する高速化手法であり、XNNPACK や GPUDelegate などの高速化手法を比較し、最も高速だった GPUDelegate を取り入れることで、顔の特徴点の取得にかかる時間を短縮した。

提案システムでは、3D モデルの作成後、運動パターンのフレーム毎に 3D モデルの顔の特徴点と運動パターンの顔の特徴点をリアルタイムでマッピングして投影モデルを作成する。フレーム毎に高速に処理する必要があるため、運動パターンの顔の特徴点はあらかじめ保存した。FaceMesh を用いて取り出した運動パターンの顔の特徴点を運動パターンの映像とともにバイナリファイルとして Android のストレージに保存し、リアルタイムなマッピング処理を軽量化した。

表情を増強する機能

提案システムを使用することで様々な機能を実現できるが、本研究では以下の機能を実装した。

- ユーザーが希望するタイミングでシステムを動かし、特定の表情を生成する
- ユーザーの顔映像を分析し、表情を増強する

この機能のうち、表情を増強する機能について説明する。ユーザーの表情を増強する際に投影する表情とユーザーの表情が一致しない場合、ユーザーの意図しない印象を話し相手に与えてしまい、誤解を生じさせてしまう。そこで、ユーザーの表情が変化したタイミングで素早くユーザーの表情を判別して適切な運動パターンを投影する必要がある。

表情の判別方法に FACS というものがある [2]。FACS (Facial Action Coding System) は Action Unit という表情筋動作の単位の組合わせを用いて感情を推定する方法であり、高精度で表情に基づく感情推定ができる。一方で、日本人の顔面運動が西洋人に比べて全体的に減少する文化的差異が報告されているように [6], 文化圏の違いや顔の筋肉の個人差によって、表情と感情の対応関係が変化してしまう。

そこで提案システムでは、FACS に Sato らが提示した日本人の感情と Action Unit との関係に関するデータ [5] を組み合わせることで、日本人の表情と感情の対応関係に即した手法を実装した。特定の感情に対応する Action Unit が動作した際に Sato らによって示された表情の強度を合算する。表情の強度を各強度の総和で除算し、基本 6 感情のうちのどれに該当するかを判別する。

4 評価実験

提案システムで顔に運動情報を付与し、以下の 2 項目を評価した。

項目 A 動きのある顔がより動いて見えたか

項目 B 会話時の印象がよくなったか

4.1 実験方法

装着者は表情を作るのが苦手な 20 代男性で、170 lx の照明環境下で提案デバイスを装着させた。装着者には実験中、常に感情を出さないように伝えた。今回の評価実験では、装着者とは異なる人物 (本論文の第一著者) の顔から作成した笑顔の運動パターンを投影し、笑顔の表情を増強する効果を被験者に見せた。各被験者が見る装着者の状態は、会話あり・なし、システムによる投影あり・なしを組み合わせた計 4 状態で、見せる状態の順番は被験者ごとにランダム化した。実験で提示した 4 状態については、



(a) 会話のない状態 (b) 会話のある状態

図 6: 実験時の顔の様子



図 7: 実験の様子

動画²を確認いただきたい。装着者の会話なしと会話ありの顔の例を図6に示す。顔のパーツの動きのない会話なしの状態と、動きのある会話ありの状態を比較し、錯視の効果を定量的に調べる。実験の様子を図7に示す。被験者は装着者と机を挟んだ170cmの距離で装着者の表情を見せ、表情や印象の変化について表1に示す質問紙調査を行った。被験者は20代の男女10名である。

項目Aの顔に運動情報を付与したことで動きのある顔がより動いて見えたかは、表1のQ1の笑顔度を用いて評価する。今回の評価実験では、笑顔の運動パターンを投影し、笑顔の表情を増強する効果を評価するため、装着者の笑顔の表情がシステムによってどれだけ増強したかを比較する。

項目Bの印象がどのように変化したかは、表1のQ2~Q6を用いて評価する。Q2~Q5を用いて印象が良くなったかを調査し、Q6でどのような印象を抱いたかを調べる。Q2とQ4は会話に関する内容のため、会話ありの状態でのみ質問した。

システムのリアルタイム性や問題点は、表1のQ7を用いて評価する。システムのリアルタイム性が十分であるかを評価するために、違和感を覚えた箇所を調査した。

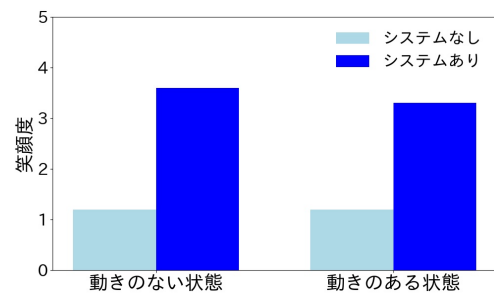


図 8: 笑顔度の評価結果

4.2 実験結果

項目Aの動きのある会話時の顔が笑顔に見えたかについて、Q1の回答で得られた笑顔度の平均値を図8に示す。会話なしにおいて、システムなしは1.2、システムありは3.6であった。また、会話ありにおいて、システムなしは1.2、システムありは3.3であった。よって、動きのある顔でも運動情報を投影することでより動いて見え、動きのある顔でも、動きのないときと同様な錯視の効果が得られることが示唆された。

項目Bの印象の変化については、Q2~Q5の回答で得られた回答の平均値を2に示す。全ての質問で動きの有無に関わらず、システムなしよりもありの方が点数が高かった。したがって、システムを利用した装着者の印象が良くなったことが示唆された。

装着者に投影されている映像の違和感について、違和感を覚えたと答えた被験者は、会話なしでは10名中6名、会話ありでは4名だった。そのように回答した理由について、適切なタイミングで投影することが違和感を減らしたと答えた被験者が複数いた。このことから、会話時の動きのある顔に投影した方がより自然であることが示唆された。また、作成した映像が装着者の顔と異なっていると指摘した人はおらず、第一著者の顔で作成した映像でも装着者の顔に適切にマッピングされていると考えられる。

5 議論

今後、より違和感のない映像を投影するためにどのようなシステムを作成すべきかについて、被験者の回答から違和感を覚えた内容とその原因について議論する。

システムの速度について、口元の追従や瞬きのある目元などの映像のずれに違和感を覚えたと答えた人は10名中1名であり、適切な速度で装着者の動きに追従できていたと考えられる。一方で、プロジェクタの光量について、3名が映像が白すぎることに違和感を指摘した。これは運動情報以外の不必要な情報もプロジェクタから投影されていたことが原因であると考えられる。より効果的な錯視を生む

² https://youtu.be/qG_dGsIvrNY

表 1: 質問紙調査の内容

項目	内容
A	Q1. 笑顔度はどうか (1. 全く笑顔に見えない, 5. 強い笑顔に見える)
	Q2. 話を聞いているように感じたか (1. 全く聞いていない, 5. とても聞いている)
	Q3. 愛嬌よく感じたか (1. 全くそう思わない, 5. とてもそう思う)
B	Q4. リアクションはどうだったか (1. 全くない, 5. とても大きい)
	Q5. よく笑う人と感じたか (1. 全く感じない, 5. 強く感じた)
	Q6. どのような印象を感じたか (自由記述)
	Q7. 違和感を覚えたか (自由記述)

表 2: 印象に関する評価結果

会話システム	なし		あり	
	なし	あり	なし	あり
話を聞いているか	-	-	1.6	2.8
リアクション	-	-	1.5	2.7
話を聞いているか	1.3	3.4	1.1	3.1
よく笑う人か	1.1	3.2	1.4	3.0

ために、投影する映像の明るさや投影する部位を調整する必要があると考えられる。

装着者の装着感について、装着者にシステムを使用した感想を答えさせた。装着感については、デバイスの重量が大きく、1時間を超える長時間の装着が難しいとの指摘があった。提案手法は、様々な表情を表現するために顔全体に映像を投影することを選択しており、プロジェクタの位置が顔から離れてしまうために、装着者に重さを感じさせていた。より長時間の装着が可能であるデバイスを作成するために、今後は特定の表情に特化し、部分的な顔のパーツにのみ投影を行うことで小型にする。

装着者が感じる眩しさについて、提案デバイスでは目の部分を黒い映像にして、装着者が眩しくないように投影を行っていた。しかし、長時間の装着によって、プロジェクタと顔の位置関係が変わってしまい、目に白い映像が入るようになってしまうと眩しいという指摘があった。今後、プロジェクタから投影する映像の位置や角度を自動で機能する機能を実装し、プロジェクタと顔の位置関係が変わってしまっても、装着者にとって眩しくないデバイスの作成をする。

6 まとめ

本論文では会話時の印象を良くするために、表情が豊かに見える映像を顔に投影するウェアラブル表情拡張システム EmoCast を提案した。提案システムでは、カメラで取得した顔映像から現在の顔の3Dモデルを作成し、この3Dモデルにあらかじめ作成した運動パターンをマッピングして、プロジェクタ

で投影する。これにより、無表情からの表情の生成、増強、打消しといった機能が実現できる。提案システムの有効性を評価した結果、動きのある顔に対しても運動情報を投影することでより動いているように錯覚させることができ、会話時にシステムを使用することで印象が良くなったことが示唆された。しかし、装着者を観察した被験者が感じる違和感や装着者の長時間装着には課題があった。今後は、より効果的な錯視を生む投影方法や装着者の装着感の向上についてより詳しく調査し、日常利用できるようにシステムを改良していく。

謝辞

本研究の一部は、JST CREST(JPMJCR18A3)の支援によるものである。ここに記して謝意を表す。

参考文献

- [1] A. H. Bermanno, M. Billeter, D. Iwai, and A. Grundhöfer: Makeup Lamps: Live Augmentation of Human Faces via Projection, *ComputerGraphics Forum*, Vol. 36, No. 2, pp. 311–323 (May 2017).
- [2] P. Ekman and W. V. Friesen: Facial Action Coding System, *Environmental Psychology & Nonverbal Behavior* (1978).
- [3] Google: MediaPipe | Google for Developers. <https://developers.google.com/mediapipe>. (Accessed 2023-08-04).
- [4] A. Mehrabian: Communication without Words, *Communication theory*, pp.193 – 200 (2017).
- [5] W. Sato, S. Hyniewska, K. Minemoto, and S. Yoshikawa: Facial Expressions of Basic Emotions in Japanese Laypeople, *Frontiers in psychology*, Vol. 10, No. 259, pp. 1–11 (Feb. 2019).
- [6] C. H. J. Tzou, P. Giovanoli, M. Ploner, and M. Frey: Are There Ethnic Differences of Facial Movements between Europeans and Asians?, *British journal of plastic surgery*, Vol. 58, pp. 183–195 (Mar. 2005).
- [7] T. Kawabe, T. Fukiage, M. Sawayama, and S. Nishida: Deformation Lamps: A Projection Technique to Make Static Objects Perceptually

- Dynamic, *ACM Trans. Appl. Percept.*, Vol. 13, No. 2, pp. 1–17 (Mar. 2016).
- [8] R. Kumazaki and A. Inoue: Development and Evaluation of a Mask-type Display Transforming the Wearer’s Impression, *In Proceedings of the 31st Australian Conference on Human-Computer-Interaction (OzCHI ’19)*, pp. 568–571 (Dec. 2019).
- [9] 吹上大樹, 河邊隆寛, 西田眞也: 変幻灯—錯覚を利用した光投影による実物体のインタラクティブな動き編集—, 情報処理学会研究報告, Vol. 2017-CVIM-206, No. 8, pp. 1–6 (Mar. 2017).
- [10] 石井綾郁, 小松孝徳, 橋本 直: HappyMouth: マスク型デバイスによる対面コミュニケーション能力の拡張, 情報処理学会研究報告, Vol. 2018-HCI-177, No. 7, pp. 1–7 (Mar. 2018).
- [11] 渡辺義浩: ダイナミックプロジェクションマッピングによる虚実融合, 日本画像学会誌, Vol. 61, No. 1, pp. 14–21 (Feb. 2022).