

2025 年度 卒業論文

論文題目  
質問者を重畳表示する  
発表支援システムにおける  
あおり顔画像の AI による正面補正

指導教員  
舟橋 健司 准教授

名古屋工業大学 工学部 情報工学科  
2022 年度入学 34714070 番

白木 陽

# 目次

第1章	はじめに	1
第2章	従来手法の概要	4
2.1	全天球カメラによる質問者の抽出	4
2.2	見上げた構図から正面顔画像への補正	7
第3章	AIによる正面顔画像への加工	10
3.1	概要	10
3.2	FacePokeによる顔画像変換	10
3.3	撮影角度の推定	14
3.4	プレゼンテーションスクリーンへの話者の表示	14
第4章	評価	16
4.1	従来手法との比較評価	16
4.1.1	評価1の概要	16
4.1.2	評価1の結果と考察	16
4.2	質問者重畳表示の有効性評価	18
4.2.1	評価2の概要	18
4.2.2	評価2の聴衆による評価結果	22
4.2.3	質問者重畳表示の評価の発表者による評価結果	24
第5章	むすび	26
	謝辞	27
	参考文献	28

## 第1章 はじめに

近年、学会や企業での発表、学校での講義など、スライドによるプレゼンテーションを行う機会が増えており、発表をより円滑、効果的に行うための支援システムへの関心が高まっている。例えば、ウェアラブルコンピュータを使用して司会者のサポートを行うシステムの研究 [1] や、発表中のスライド上でのフィードバック共有により発表者と聴衆間のリアルタイムなインタラクションを可能にするシステムの研究 [2] などがなされている。また、発表者のプレゼンテーションにおける動作や発話を評価し、発表の改善を支援する研究 [3] もある。発表者をプレゼンテーション用のスクリーン上に重畳表示するシステムの研究も行われている。梅村らは、発表者のシルエットをスライドの背景に薄い影絵として表示することで、スライドの視認を妨げることなく、発表者のジェスチャーの表示を実現している [4]。当研究室でも、発表者をスクリーン上に重畳表示するプレゼンテーション支援システムの研究 [5] を行っている。プレゼンテーションにおいて、発表者の声や発表に使用されるスライドは聴衆が発表内容を理解するのに重要であるが、発表者の態度や表情、身振り手振りなども同じく重要である。広い会場など、発表者の直接的な視認が困難な会場でも、発表者を重畳表示することで、スクリーンを通して視認できるようになり、聴衆のより深い理解が期待される。

ところで、当研究室ではプレゼンテーションを支援するために、質問者の顔を発表用スクリーンに重畳表示することで、質疑の理解を促し、また質疑を活性化させるシステム [6] の提案も行っている。広い会場で聴衆が質問をする際、質問者の位置により、発表者や他の聴衆が質問者の姿を捉えることが難しい場合がある。質問者の顔を重畳表示することで、他の聴衆が、システムを使用しない場合と比べて質問者をより身近に感じ、質問に対する興味が大きくなることが示されている。また、質問を受けた発表者も、質問内容の理解度が高まり、回答により思いが込められるこ

とが示されている。広い会場でのプレゼンテーションでは、質問者は広い会場のどこにいるのか決まっていないため固定カメラでの対応は困難である。位置、向きの移動が可能なカメラで発表者を捉えるためには、そのための装置や人が必要となる。カメラをハンドマイクに内蔵すれば、質問者が手に持つため質問者を追従する装置を別に用意する必要はなく、常に質問者に近い場所で撮影が可能である。ハンドマイク内蔵のカメラとして360度撮影が可能な全天球カメラを利用することで、カメラに対して質問者の顔がどの方向に位置していても撮影することができる。広い会場でのハンドマイクの使用は一般的であり、それにカメラが付くこと以外は一般的なプレゼンテーションの環境と違いはない。このように、ハンドマイクを用いた重畳表示システムは、手法としては広い会場でのプレゼンテーション支援に有用である一方、システムとして課題が残されている。ハンドマイクにカメラが付いているという制約上、話者の口元、もしくはそれより下にカメラが位置することになり、撮影される話者の顔が下から見上げる構図になってしまう。そのため、撮影した映像をそのままスクリーンに表示すると、正面からの顔画像と異なり違和感が生じてしまう。

当研究室では、その違和感を軽減するために撮影された下から見上げた構図の顔画像を正面画像へと補正する研究も行っている [7]。コンピュータで話者の頭部の3次元形状モデルを作成する。撮影した顔画像をモデルに貼り付け、正面からレンダリングすることで、正面から撮影した顔画像を得られるようにした。しかし、顔形状モデルを楕円体に近い単純な形状で近似しており各話者の顔と全く同じ形ではないため、補正画像の質が低いという課題が残されている。

ところで、近年、ディープラーニングを用いた画像生成技術が急速に発展している。特にディープラーニングを用いた生成AIは、顔画像の生成や変換において顕著な進歩を遂げており、単なる画像フィルタリングから、複雑な顔の向きや表情の変換まで実現可能になっている。生成AIの中でも、拡散モデルやGANなどの手法は、高い品質で現実的な画像を生成することができ、特に顔画像の補正や変換タスクにおいて優れた性能を示している。このような技術の進展により、どんな形状をした顔の画像でも、高品質な正面顔画像へと変換できる可能性が高まっている。ディープラーニングを用いた生成AIの技術を用いることで、従来の3次元モデルを用い

た手法よりも高品質な正面補正が可能になるだろう。本研究では、ハンドマイクに取り付けられた全天球カメラで撮影された下から見上げた構図の顔画像を、AIを用いて正面画像へと補正する手法を提案する。これにより、違和感のない自然な表示にして、より実用性のあるシステムとすることを目指す。本論文では、第2章では、従来手法について説明する。第3章では、話者の表示方法について説明し、第4章では第3章の記述をもとに構築したシステムを用いて行った体験評価とその結果について述べる。第5章では本研究のまとめや今後の課題について述べる。

## 第2章 従来手法の概要

本章では従来手法について説明する。従来手法では全天球カメラ内蔵ハンドマイクを使用して質問者の顔を抽出し、その顔画像を3次元形状モデルを用いて正面顔画像へと補正してスクリーンに表示している。

### 2.1 全天球カメラによる質問者の抽出

使用者が持ち方を気にせずマイクを使用したい。将来的には全天球カメラを内蔵した従来型形状のハンドマイクの開発も想定したい。ここでは360度全ての方向を撮影できるマイク内蔵型カメラ RICOH THETA S [9] を使用する。図 2.1 に RICOH THETA S の画像を、図 2.2 に RICOH THETA S を持った話者の例を示す。全天球カメラから取得される画像、及びそれを等距離射影方式として歪み補正した画像を図 2.3 に示す。

全天球カメラは一般に半球が写るカメラを2つ備えており、撮影された画像は2つに分かれている。そこでカメラより取得した画像の境目に質問者がいる場合も想定する必要がある。取得画像を各軸において30度ステップで回転させて、別方向から見た場合の画像を生成し、生成された画像の歪みを補正する。取得画像から話者が撮影された画像を抽出するために顔認識を行っている。認識には OpenCV の標準ライブラリ関数である `CascadeClassifier::detectMultiScale` を使用している [8]。複数の顔が認識された場合、質問者は最もカメラに近い位置にいると考え、顔と認識した範囲が最も大きい画像を質問者が映っている画像とする。例として、取得画像を水平に90度回転後に歪み補正を行った画像（図 2.4）と、垂直に60度回転後歪み補正を行った画像（図 2.5）を示す。



図 2.1: RICOH THETA S

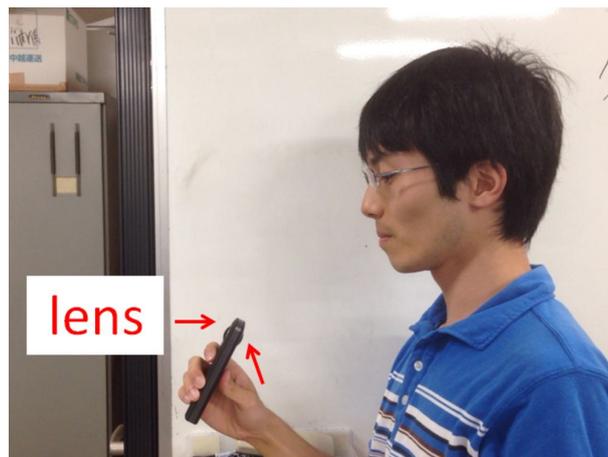


図 2.2: RICOH THETA S を持った話者の例



図 2.3: 取得画像と歪み補正後の画像



図 2.4: 水平に 90 度回転後, 歪み補正



図 2.5: 垂直に 60 度回転後, 歪み補正

## 2.2 見上げた構図から正面顔画像への補正

カメラ内蔵ハンドマイクで話者の顔を撮影する際は一般的に、マイクを顔より下方に持つため、マイクと顔は図 2.6 のような関係になる。このため、カメラにより撮影される映像中の話者の顔が、図 2.7 のように下から見上げる「あおり」の構図になってしまう。この映像をスクリーン上に表示すると、正面からの顔画像と異なり、表示に違和感が生じてしまう。より自然な表示にするため、「あおり」をなくした表示を目指して映像の補正を行う。

まず話者の3次元顔形状モデルを作成し、そのモデルに対して撮影された話者の顔画像を図 2.8 のように適切な位置に貼り付けることで、見上げた構図の画像を補正する。これを正面からレンダリングすることで、話者の顔を正面から撮影したような画像を得る。顔形状モデルは、話者の顔形状に忠実であるほど正確な補正が可能だと考えられるが、画像を貼り付ける際にずれが生じた場合に大きく目立つ可能性も考えられる。一方で楕円体のような単純すぎるモデルだと大まかな補正しかできず正面から撮影したような画像にならない。そこで楕円体を元にして、鼻、顎、頬骨部分を人の顔に近づけるよう変形させて3次元顔形状モデルを作成する。図 2.9 に作成した3次元顔形状モデルを示す。また、作成した顔形状モデルに対して適切な位置に話者の顔画像を貼り付けるために、撮影された画像より、顔に対するカメラの方向を推定する必要がある。カメラ方向の変化に対応して、画像中の両目間の距離と、目から鼻の下端までの距離の比率が変化する。これを用いて角度の推定を行う。この手法により得られた正面補正画像を、もとの取得画像の話者の顔の位置に適切な大きさに貼り付けたものをシステムの最終的な出力とする。話者の頭部に対してカメラを水平より下方30度の方向に構えたときの、正面補正前のカメラの取得画像を図 2.10 左に、正面補正後の画像を図 2.10 右に示す。

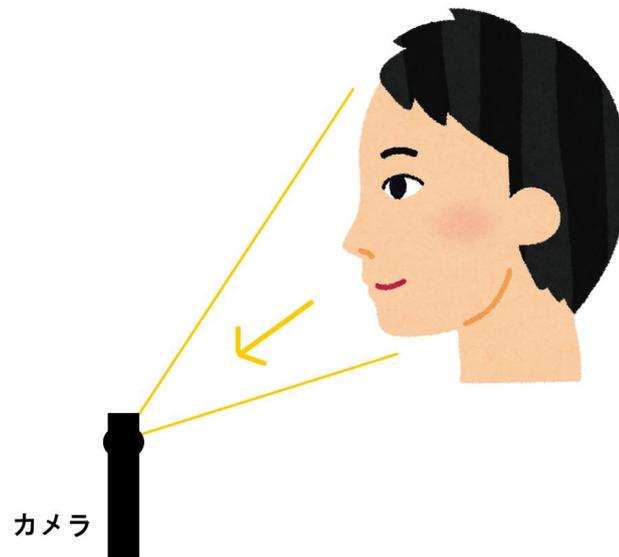


図 2.6: カメラ内蔵ハンドマイクと話者の関係



図 2.7: 見上げた構図の顔画像例

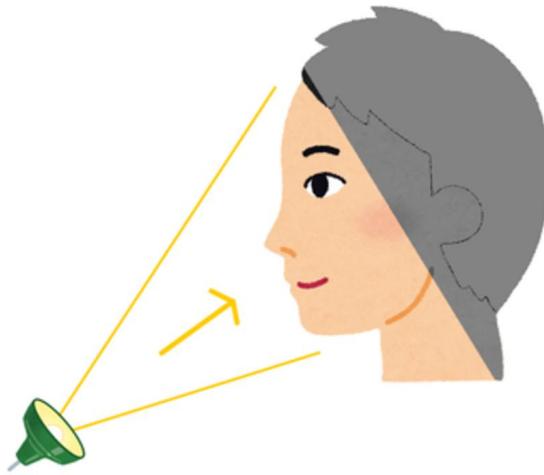


図 2.8: 画像の貼り付けのイメージ図

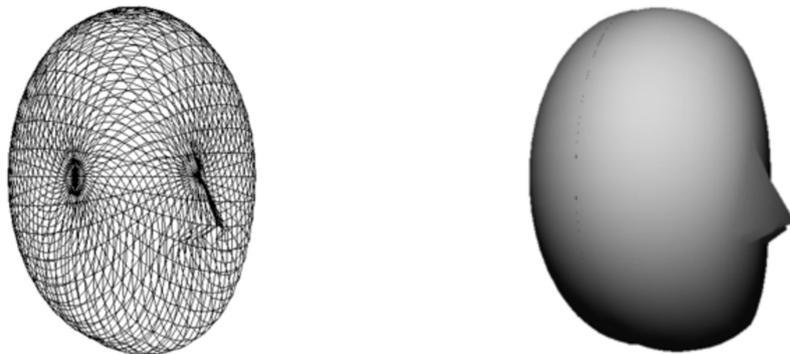


図 2.9: 作成した3次元顔形状モデル

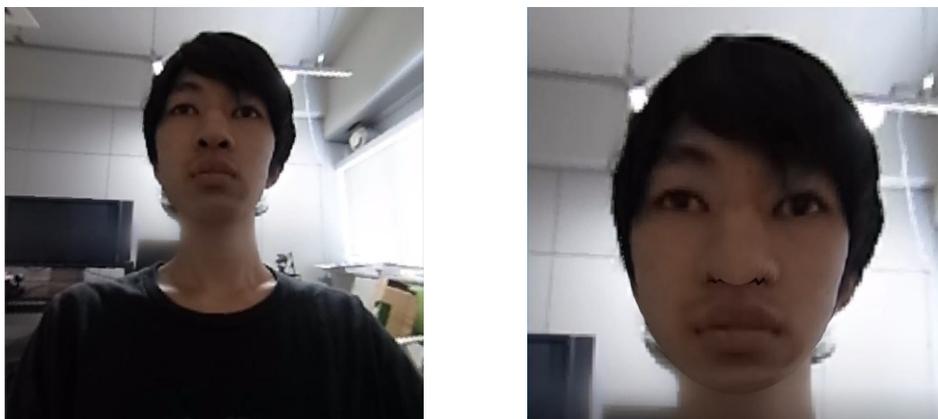


図 2.10: 撮影角度 30 度ときの取得画像と補正画像

## 第3章 AIによる正面顔画像への加工

本章では、AIを用いてあおり顔画像を正面顔画像へと補正する手法について説明する。

### 3.1 概要

従来手法 [7] では、第2章に述べた通り3次元形状モデルを用いて補正を行っていたが、顔形状モデルが話者の顔と全く同じ形ではないため、補正画像の質が低いという課題が残されている。そこで、本研究では、より高品質な補正を実現するため、深層学習を活用した画像変換技術の導入を検討し、AIを用いて補正を行う手法を提案する。既に顔画像を変換する様々なAIベースのツールがある。本研究では、GitHub上で公開されている顔画像変換ツールFacePoke [10]を利用する。FacePokeでは、顔のパーツの位置と、顔画像の向きを変えることができる。このツールを用いれば、見上げた構図の顔画像を正面顔画像に変換できるだろう。FacePokeで顔画像を変換する際、顔を回転させるために、撮影時のカメラに対する頭部の角度の情報が必要となる。しかし、カメラ自体には撮影角度を計測する機能はなく、またFacePokeにも顔画像の向きを推定する機能はない。そこで撮影される画像より、撮影角度の推定を行った後に、見上げた構図の顔画像を正面顔画像に変換する。以上の考えをもとに、補正を自動的に実現するための手法を以下に示す。

### 3.2 FacePokeによる顔画像変換

FacePokeは2024年にJulian Bilcke氏によってGitHub上で公開されたオープンソースソフトウェアである [10]。フロントエンドはReactで実装されており、バックエンドはPythonで構築されている。FacePokeは、AIを用いて顔画像を変換するツールであり、顔画像の表情や向きを変えることができる。具体的には、顔画像を

入力し、マウス操作で顔のパーツの位置や顔の向きを調整することで、希望する顔画像を生成することができる。調整できるパラメータは、目、口、眉毛、顔の輪郭などの位置や大きさ、および顔の回転角度（ピッチ、ヨー、ロール）である。FacePokeはWebブラウザ上で動作するため、インストールの必要がなく、手軽に使用できる。FacePokeの操作画面を図3.1に示す。図3.1は図2.7の顔画像を入力し、顔画像変換を行っている様子である。

FacePokeの主な特徴と機能としては以下が挙げられる。まず、ユーザーフレンドリーなWebインターフェースを提供しており、専門的な知識がなくても直感的に顔画像の編集が可能である。次に、顔のメッシュベースの変形により、自然な顔の変形を実現することができる。さらに、リアルタイムでプレビューが可能であるため、ユーザーは調整結果をすぐに確認しながら作業を進められる。FacePokeは複数の顔認識モデルをサポートしており、様々な顔画像に対する汎用性が高い。加えて、パラメータのエクスポート機能により、編集結果をプログラマ的に再利用することが可能である。

FacePokeは画像生成エンジンとしてLivePortrait [11]を利用している。LivePortraitは、単一のソース画像とドライビングビデオからポートレート動画を生成する効率的なフレームワークである。その核心は、3D implicit keypointsを用いた動きの表現にあり、以下の主要なコンポーネントで構成される：

- Base Model: face vid2vid [12]を改良し、ConvNeXt-V2バックボーン [13]やSPADEデコーダ [14]を採用することで、表現力と汎用性を向上させている。
- Stitching Module: 生成された顔領域を元の画像空間へシームレスに統合し、肩の境界などの不自然なズレを解消する。
- Retargeting Modules: 小規模なMLPを用いて、目の動き（Eyes retargeting）や口の開閉（Lip retargeting）を独立して制御することを可能にし、異なる人物間でのモーション転送（Cross-id reenactment）における精度を向上させている。

LivePortraitの詳細な技術仕様および処理フローについて以下に記述する。

- LivePortrait は、顔動画生成において高い品質と効率性を実現するため、複数の先進的な技術を統合している。まず、3D implicit keypoints と呼ばれる暗黙的な3次元特徴点を用いることで、顔の複雑な動きを効率的に表現している。この手法により、明示的な3Dモデルを必要としないため、計算コストを削減しながらも自然な顔の変形を実現できる。
- Base Model の構造では、ConvNeXt-V2をバックボーンとし、最新の畳み込みニューラルネットワーク技術を活用している。ConvNeXt-V2は、高い表現力を有しながらも計算効率に優れた特性を備えており、様々な解像度と品質の顔画像に対応可能にしている。デコーダには SPADE (Spatially-Adaptive Normalization) を採用することで、生成画像の詳細度を向上させ、より自然な肌のテクスチャと顔の特徴を保持している。
- Stitching Module は、生成された顔領域と元の画像背景をシームレスに統合する重要な役割を果たす。このモジュールは、顔と背景の境界領域における色合いと明るさの不一致を補正し、肩や首周辺の不自然なズレを解消することで、最終的な動画の視覚的品質を大幅に改善している。
- Retargeting Modules は、微細な顔の動きを高精度で制御することを可能にする。目の動き (Eyes retargeting) と口の開閉 (Lip retargeting) をそれぞれ独立して制御する小規模な MLP により、異なる人物間でのモーション転送精度が向上し、より表現力豊かな動画生成が実現できる。これにより、クロスアイデンティティ (Cross-id reenactment) における動作の正確性が大幅に向上している。

FacePoke のバックエンドにおいて、LivePortrait は顔画像変換の中核を担っており、高品質な顔動画生成を実現するための強力な基盤を提供している。FacePoke は、これらの技術を活用することで、ユーザーフレンドリーなインターフェースを通じて、高度な顔画像編集機能を提供している。

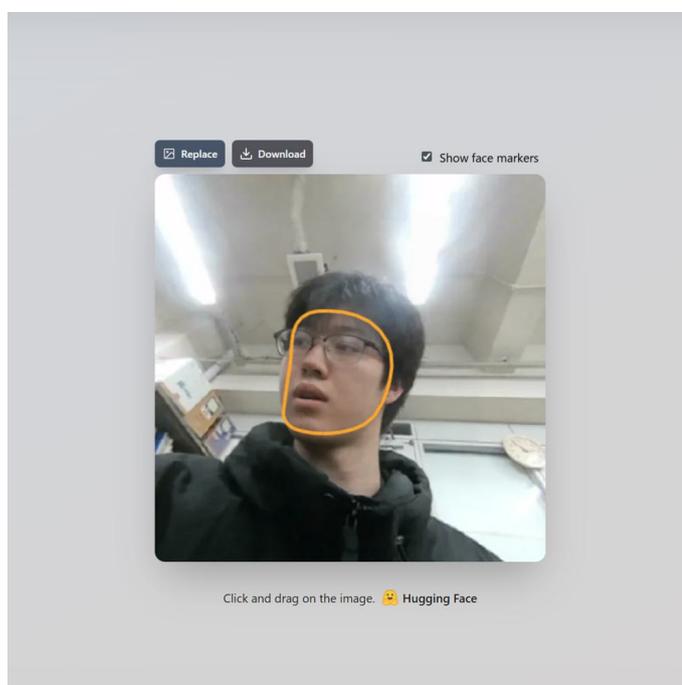


図 3.1: FacePoke の操作画面

### 3.3 撮影角度の推定

撮影角度の推定には、Dlib ライブラリを使用する [15]. Dlib は、顔認識や顔特徴点検出などの機能を提供する C++ ライブラリであり、Python から利用可能である. Dlib を用いて、顔画像から顔特徴点を検出し、それらの特徴点をもとに撮影角度を推定する. 具体的には、Dlib の顔特徴点検出器を使用して、顔画像から 68 個の顔特徴点を検出する. 次に、これらの特徴点を用いて、頭部の姿勢推定を行う. 頭部の姿勢推定には、PnP (Perspective-n-Point) アルゴリズムを使用する. PnP アルゴリズムは、3D 空間上の点と 2D 画像上の対応点を用いて、カメラの位置と姿勢を推定する手法である [16]. 本研究では、顔の 3D モデルを用意し、Dlib で検出した顔特徴点と対応させることで、PnP アルゴリズムを適用する. これにより、撮影角度 (ピッチ, ヨー, ロール) を推定することができる.

### 3.4 プレゼンテーションスクリーンへの話者の表示

2.2 節の FacePoke のバックエンドを利用し、2.3 節の撮影角度推定を組み込むことで、あおり顔補正を自動的に実現する手法を構築した. 具体的な手順は以下の通りである. まず、カメラ内蔵ハンドマイクで話者の顔画像を撮影する. 次に、撮影された顔画像から Dlib を用いて顔特徴点を検出し、PnP アルゴリズムを適用して撮影角度を推定する. 図 2.7 に示す見上げた構図の顔画像に対して、撮影角度推定を適用した結果を図 3.2 に示す. その後、推定された撮影角度をもとに、FacePoke で顔画像を正面顔画像に変換する際の回転角度を設定する. 具体的には、ピッチ角度を用いて、顔画像を上下方向に回転させる. これにより、見上げた構図の顔画像を正面顔画像に補正することができる. 最後に、補正された正面顔画像をプレゼンテーション画面に重畳表示する. 図 2.7 に示す見上げた構図の顔画像に対して、本手法を適用した結果を図 3.3 に示す.



図 3.2: 撮影角度の推定例



図 3.3: 補正後の正面顔画像例

## 第4章 評価

第3章の提案をもとに，Windows PC上でシステムを作成した．本手法の有効性を評価するために，2通りの検証を行った．

### 4.1 従来手法との比較評価

#### 4.1.1 評価1の概要

従来手法と本手法の補正画像を比較し，本手法の有効性を評価した．ピッチ角が30度，45度，60度それぞれの画像を用意し，従来手法と本手法それぞれで補正画像を生成する．補正前の画像，従来手法で補正した画像，本手法で補正した画像の3種類に対して比較評価した．

#### 4.1.2 評価1の結果と考察

評価1で使用した補正前の画像，従来手法で補正した画像，本手法で補正した画像を図4.1，図4.2，図4.3に示す．各図の左から順に，補正前の画像，従来手法で補正した画像，本手法で補正した画像である．従来手法で補正した画像を見ると，顔の輪郭とその外側の背景との境界が不自然であるように見える．特にピッチ角が大きくなるにつれて，その不自然さが顕著になる．一方，本手法で補正した画像を見ると，顔の輪郭と背景との境界が自然であり，従来手法と比較して違和感が少ないことが分かる．また，ピッチ角が大きくなっても，顔の輪郭と背景との境界が比較的自然に保たれている．以上のことから，本手法は従来手法と比較して，あおり顔画像の正面補正において有効であると考えられる．



図 4.1: ピッチ角 30 度の画像 (左から補正前, 従来手法, 本手法)



図 4.2: ピッチ角 45 度の画像 (左から補正前, 従来手法, 本手法)



図 4.3: ピッチ角 60 度の画像 (左から補正前, 従来手法, 本手法)

## 4.2 質問者重畳表示の有効性評価

### 4.2.1 評価2の概要

プレゼンテーションにおいて、質問者を重畳表示する発表支援システムにおけるあおり顔画像のAIによる正面補正手法の有効性を評価するために、正面補正前の画像による発表支援と、正面補正後の画像による発表支援を比較し、どちらが質問者をより身近に感じ、質問に対する興味が大きくなるか評価した。RICOH THETA Sと同様に360度撮影が可能な全天球カメラ、Insta360 Air [17]を使用した。なお、体験評価ではPCにUSBケーブルで有線接続した。図4.4にInsta360 Airの画像を、図4.5にInsta360 Airを取り付けたハンドマイクを持った話者の例を示す。従来手法で使用したCascadeClassifier::detectMultiScaleと同様に顔認識を行うDNNモジュールResNet-SSD [18]を使用し、Insta360 Airのカメラより取得した画像から質問者の抽出を行い、正面補正したのちその姿をスクリーンに表示する。Insta360 Airのカメラから取得した画像、及びそれを歪み補正した画像を図4.6に示す。参加者は9名である。まず、発表者役に簡単なプレゼンテーションをしてもらい、その後質疑応答を2回行う。これを1セットとして計6セット行う。このうち3セットは、2回行う質疑応答のうち、1回目は正面補正前の画像を使用し、2回目は正面補正後の画像を使用する(パターンA)。残りの3セットは順序を逆にし、1回目に正面補正後の画像を使用し、2回目に正面補正前の画像を使用する(パターンB)。各セット終了後、発表者役及び聴衆役に対して、正面補正前後の画像を比較し、後述する観点について評価を行ってもらう。発表者は1セット毎に違う人を充てるものとし、また各セットで質問者は1名のみであり、セット毎に交替する。部屋の中における聴衆役の配置についても1セット毎にランダムに再配置する。聴衆役は、

- 意識的に質疑応答の内容を聞きたいと思ったか
- 質疑応答内容の理解度はどうだったか
- 質問者が身近に感じられたか

の3点について正面補正後の場合に対して評価を行う。評価は正面補正前の場合を基準(3)とし、5段階評価(1:悪い~5:良い)で行う。発表者は、

- 質問者をより身近に感じられたか
- 質問内容の理解度はどうだったか
- 質問に対する回答に思いがより込められたか

の3点について聴衆と同様の5段階評価で回答する。図4.7, 図4.8に評価時の様子を示す。



図 4.4: Insta360 Air



図 4.5: Insta360 Air を取り付けたハンドマイクを持った話者の例



図 4.6: 取得画像と歪み補正後の画像



図 4.7: 質問者重畳表示の評価の様子 (部屋全体)



図 4.8: 質問者重畳表示の評価の様子 (スクリーン部分)

#### 4.2.2 評価2の聴衆による評価結果

評価2の聴衆による評価結果を表4.1に示す。各項目は発表者と質問者を除く聴衆は7人により、6セット評価されており、すなわち、のべ42回評価されている。表4.1は各項目においてこの42回のうち、1から5の各評価をつけた回数を示している。「意識的に質疑応答の内容を聞きたいと思ったか」「質疑応答内容の理解度はどうだったか」の2項目においては、正面補正後の画像があることで、聴衆が質疑応答に対してより関心を持ち、内容の理解度も向上したことが示唆される。一方、「質問者が身近に感じられたか」については、正面補正後の画像があることで、質問者を身近に感じられたという評価は少なかった。これは補正時の処理で生じたフレームレートの低下により映像が滑らかでなく、表情の変化が分かりにくくなったこともあり、違和感を覚えた聴衆が多かったためと考えられる。本手法は質問者の提示において聴衆の関心および理解度を高める点で有効であるだろうと結論づけられる。

聴衆による評価結果を、パターンAとパターンBの順序で分けて集計した結果を表4.2と表4.3に示す。各項目はそれぞれのパターンでのべ21回評価されている。パターンAとパターンBでU検定（判定基準： $p < 0.05$ ）を行った結果、3項目全てにおいて有意差は認められなかった（「意識的に質疑応答の内容を聞きたいと思ったか」： $p = 0.807$ 、「質疑応答内容の理解度はどうだったか」： $p = 0.868$ 、「質問者が身近に感じられたか」： $p = 0.979$ ）。したがって、正面補正前後の提示順序による影響は認められなかったと考えられる。

表 4.1: 質問者重畳表示の評価の聴衆による評価結果

評価項目	評点				
	1	2	3	4	5
意識的に質疑応答の内容を聞きたいと思ったか	0	5	31	4	2
質疑応答内容の理解度はどうだったか	0	1	28	10	2
質問者が身近に感じられたか	0	22	7	10	3

表 4.2: 質問者重畳表示の評価の聴衆による評価結果 (パターン A)

評価項目	評点				
	1	2	3	4	5
意識的に質疑応答の内容を聞きたいと思ったか	0	2	17	2	0
質疑応答内容の理解度はどうだったか	0	0	15	6	0
質問者が身近に感じられたか	0	10	5	6	0

表 4.3: 質問者重畳表示の評価の聴衆による評価結果 (パターン B)

評価項目	評点				
	1	2	3	4	5
意識的に質疑応答の内容を聞きたいと思ったか	0	3	14	2	2
質疑応答内容の理解度はどうだったか	0	1	13	4	2
質問者が身近に感じられたか	0	12	2	4	3

#### 4.2.3 質問者重畳表示の評価の発表者による評価結果

評価2の発表者による評価結果を表4.4に示す。発表者による評価結果を見ると、聴衆と同様に、「質問者をより身近に感じられたか」に対しては、低い評価が多かった。また、質問内容の理解度や、「質問に対する回答に思いがより込められたか」についても、正面補正後の画像があることで向上したとは言い難い結果となった。これは、評価環境の特性に起因すると考えられる。発表者側からは会場全体が見渡せる配置であり、体験評価が中程度の広さだったため、質問者を直接視認しやすかった。したがって、正面補正による画像提示の効果が相対的に小さく、それほど高評価につながらなかったのだろうと考えられる。

発表者による評価結果を、パターンAとパターンBの順序で分けて集計した結果を表4.5と表4.6に示す。各項目はそれぞれのパターンでのべ3回評価されている。パターンAとパターンBでU検定（判定基準： $p < 0.05$ ）を行った結果、3項目全てにおいて有意差は認められなかった（「質問者をより身近に感じられたか」： $p = 0.763$ 、「質問内容の理解度はどうだったか」： $p = 1.000$ 、「質問に対する回答に思いがより込められたか」： $p = 0.371$ ）。したがって、正面補正前後の提示順序による影響は認められなかったと考えられる。

表 4.4: 質問者重畳表示の評価の発表者による評価結果

評価項目	評点				
	1	2	3	4	5
質問者をより身近に感じられたか	0	4	1	1	0
質問内容の理解度はどうだったか	0	0	6	0	0
質問に対する回答に思いがより込められたか	0	1	5	0	0

表 4.5: 質問者重畳表示の評価の発表者による評価結果 (パターン A)

評価項目	評点				
	1	2	3	4	5
質問者をより身近に感じられたか	0	2	1	0	0
質問内容の理解度はどうだったか	0	0	3	0	0
質問に対する回答に思いがより込められたか	0	1	2	0	0

表 4.6: 質問者重畳表示の評価の発表者による評価結果 (パターン B)

評価項目	評点				
	1	2	3	4	5
質問者をより身近に感じられたか	0	2	0	1	0
質問内容の理解度はどうだったか	0	0	3	0	0
質問に対する回答に思いがより込められたか	0	0	3	0	0

## 第5章 むすび

本研究では、発表支援システムにおける質問者のあおり顔画像を AI を用いて正面補正する手法を提案した。従来手法では、3次元形状モデルを用いて補正を行っていたが、顔形状モデルが話者の顔と全く同じ形ではないため、補正画像の質が低いという課題が残されていた。本手法では、深層学習を活用した画像変換技術を導入し、GitHub 上で公開されている顔画像変換ツール FacePoke を利用して補正を行った。本手法は従来手法と比較して、より質の高い正面補正ができることが示された。さらに、本手法を用いて発表支援システムを構築し、体験評価を行った。体験評価では、発表者側には大きな有効性が認められなかったものの、聴衆側には質問者への関心や理解度の向上が確認でき、有効性が示唆された。今後の課題としては、実装に伴うフレームレート低下等の問題を改善し、提示の自然さや親近感の向上を図ることで、さらに有効性を高めることが挙げられる。本研究が円滑で効果的なプレゼンテーション支援に寄与することを期待したい。

## 謝辞

本研究を進めるにあたって、日頃から多大な御尽力を頂き、ご指導を受け賜りました名古屋工業大学、舟橋健司 准教授、伊藤宏隆 助教に心から感謝致します。最後に、本研究に多大な御協力を頂きました舟橋研究室諸氏に心から感謝致します。

## 参考文献

- [1] 岡田智成, 山本哲也, 寺田努, 塚本昌彦, “ウェアラブル MC システム: 司会進行を支援するウェアラブルシステムの設計と実装”, コンピュータ ソフトウェア, Vol.28, No.2, pp.162-171, May 2011.
- [2] 井上良太, 白松俊, 大園忠親, 新谷虎松, “発表中の資料へのフィードバックに基づくインタラクティブプレゼンテーションシステムの実現”, 情報学論, Vol.56, No.10, pp.2011-2021, October 2015.
- [3] 趙新博, 由井藺隆也, “ノンバーバル表現に注目したプレゼンテーション支援システムの提案”, 研究報告ヒューマンコンピュータインタラクション (HCI) , Vol.2014-HCI-157, No.42, pp.1-6, March 2014.
- [4] 梅村恭司, 梅村真由, “Kuroko: 話者シルエットを活用するプレゼンツール”, 情報処理学会シンポジウム論文集, Vol.2012, No.3, ROMBUNNO.1EXB-25, March, 2012.
- [5] Kenji Funahashi, Yusuke Nakae, “Getting Yourself Superimposed on a Presentation Screen”, Proceedings of the 2nd ACM symposium on Spatial user interaction, pp.138-138, 2014.
- [6] Yuki Kobayashi, Kenji Funahashi, “Superimposing Questioner on Presentation Screen Using Microphone with Whole-Sky Camera”, ICAT-EGVE 2016, pp.3-4, 2016.
- [7] 柴田大地, 小林勇輝, 舟橋健司, “全天球カメラ内蔵マイクを利用して質問者を重畳表示するプレゼンテーション支援システム”, NICOGRAPH2019 講演論文集, 2019.

- [8] OpenCV, CascadeClassifier::detectMultiScale  
[https://docs.opencv.org/3.0-beta/modules/objdetect/doc/cascade\\_classification.html?highlight=cv2.grouprectangles](https://docs.opencv.org/3.0-beta/modules/objdetect/doc/cascade_classification.html?highlight=cv2.grouprectangles)
- [9] Ricoh Company, Ltd RICOH THETA S  
<https://theta360.com/ja/about/theta/s.html>
- [10] FacePoke  
<https://facepoke.ai/>
- [11] Jianzhu Guo, Dingyun Zhang, Xiaoqiang Liu, Zhizhou Zhong, Yuan Zhang, Pengfei Wan, Di Zhang, “LivePortrait: Efficient Portrait Animation with Stitching and Retargeting Control”  
<https://arxiv.org/abs/2407.03168>
- [12] Ting-Chun Wang, Arun Mallya, Ming-Yu Liu, “One-Shot Free-View Neural Talking-Head Synthesis for Video Conferencing”  
<https://arxiv.org/abs/2011.15126>
- [13] Zhuang Liu, Hanzi Mao, Chao-Yuan Wu, Christoph Feichtenhofer, Trevor Darrell, Saining Xie, “ConvNeXt V2: Co-designing and Scaling ConvNets with Masked Autoencoders”  
<https://arxiv.org/abs/2301.00808>
- [14] Taesung Park, Ming-Yu Liu, Ting-Chun Wang, Jun-Yan Zhu, “Semantic Image Synthesis with Spatially-Adaptive Normalization”  
<https://arxiv.org/abs/1903.07291>
- [15] Dlib C++ Library  
<http://dlib.net/>
- [16] OpenCV, Perspective-n-Point (PnP) pose computation  
[https://docs.opencv.org/3.4/d5/d1f/calib3d\\_solvePnP.html](https://docs.opencv.org/3.4/d5/d1f/calib3d_solvePnP.html)

[17] Insta360 Air

<https://www.insta360.com/product/insta360-air>

[18] OpenCV, Deep Neural Networks (dnn module)

[https://docs.opencv.org/4.x/d2/d58/tutorial\\_table\\_of\\_content\\_dnn.html](https://docs.opencv.org/4.x/d2/d58/tutorial_table_of_content_dnn.html)