平成30年度 卒業論文

論文題目
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における
あおり顔画像の正面補正

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
平成27年度入学　27115071番

名前　柴田　大地

目 次

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いる．そのため，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．このような支援システムがある中で，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が行われている．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]．また，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実験で示されている．

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．上述の関連する研究においては，固定型のカメラ，もしくは距離センサ付きカメラを使用しており，話者はカメラに映る範囲内しか動けない．質問者は広い会場内におり，固定カメラでの対応は困難であり，またセンサによる質問者位置検出も難しい．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，発表者が手に持ったまま移動するため，発表者を追従する装置を別に用意する必要はなく，常に発表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．そこで本研究では，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ことで違和感を軽減することを目指す．これにより，違和感のない自然な表示になり，より実用性のあるシステムとなることが期待できる．本論文では，第 2 章では，話者の表示方法について説明し，第 3 章では第 2 章の記述をもとに構築したシステムを用いて行った評価実験とその実験結果について述べる．第 4 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　見上げた構図から正面顔画像への加工

この研究は，上述のカメラ内蔵ハンドマイクを用いた支援システムの研究をもとに表示部分の改良を行っている．したがって，全天球カメラを用いて撮影した画像から話者を抽出するまでの説明は省略し，カメラから話者の顔が映っている画像が取得できることを前提として説明する．

2.1　概要

カメラ内蔵ハンドマイクで話者の顔を撮影する際，マイクを顔より下方に持つため，マイクと顔は図 2.1 のような関係になる．それにより，このとき撮影される画像は図 2.2 のように下から見上げる構図となる．このような画像に対して，正面顔画像となるよう補正を行う．撮影した 2 次元の画像を補正する方法として，画像のピクセルデータをそのまま 2 次元平面上で扱う方法と，3 次元空間中で扱う方法が考えられる．今回は画像を加工するに当たり，後者の方法を採用した．具体的に説明すると，コンピュータ内部で話者の頭部の 3 次元形状モデルを作成し，それに対して撮影した話者の顔画像を，モデルに対して撮影したときと同じ相対位置から投影すれば，3 次元の立体的な話者の頭部が再現できる．モデルに対して適切な位置に画像を貼り付け，顔モデル正面からレンダリングすれば，話者の顔を正面から撮影したような画像が得られるだろう．また，カメラには角度計測装置等がついておらず，撮影角度はわからない．画像をモデルに貼り付ける際，同じ位置から貼り付けるために，撮影時のカメラに対する頭部の角度が必要となる．よって，撮影される画像より，撮影角度の推定を行う．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．

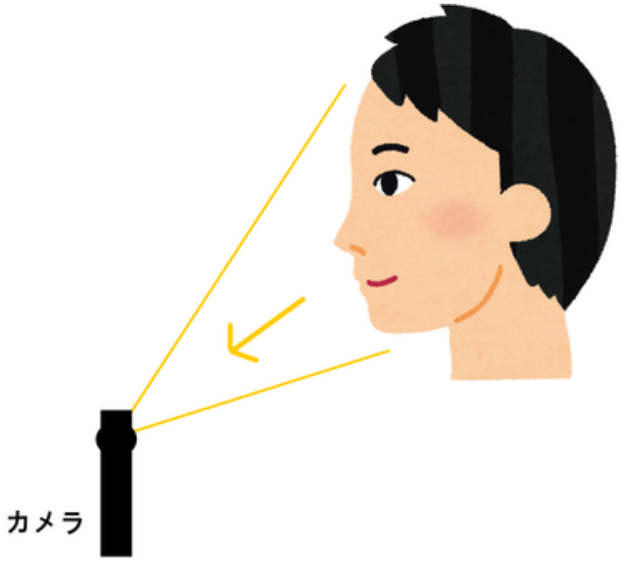

図 2.1: 見上げた構図の画像撮影時のイメージ図

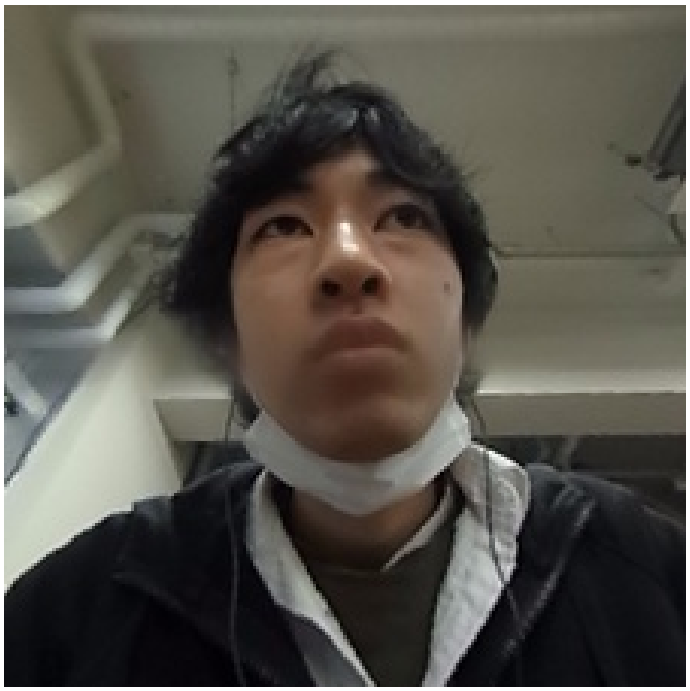

図 2.2: 見上げた構図で撮影される画像の例

2.2　3 次元顔形状モデルの作成

まず，話者の画像を貼り付けるための，話者の頭部の 3 次元形状モデルをコンピュータ内部の 3 次元空間上に作成する必要がある．そのための手法として，2 次元の顔画像から CNN を用いて顔形状を推定し，3 次元形状モデルを作成する研究 [10] がある．しかし，リアルタイムで 3 次元モデルを作成するにはコンピュータの処理能力が必要となり使用環境が限られる．また，話者の顔形状に忠実な 3 次元モデルを作成してしまうと，撮影した話者の画像を貼り付ける際にずれが生じた場合，大きく目立つ可能性が考えられる．しかし，逆に楕円体のような，簡単すぎる形を 3 次元モデルとして採用してしまうと，大まかな補正しかできず，正面から見るような画像にならない．そこでここでは，一般的な日本人成人男性の顔形状を目指し，簡単で汎用的な 3 次元顔形状モデルを作成する．モデルは，楕円体をもとにして，鼻，顎，頬骨部分を人の顔に近づくよう変形させる．この手法であれば，リアルタイムで顔モデルを作成するという計算時間の必要な処理がないため，使用環境が性能の高いコンピュータに限られることはない．また，凹凸の多い複雑な形状を避けることで，多少のずれを許容して画像を投影できるだろう．以上の考えをもとに 3 次元の顔形状モデルを作成した．ワイヤーフレームモデルを図 2.3，図 2.4 に，サーフェスモデルを図 2.5，図 2.6 に示す．

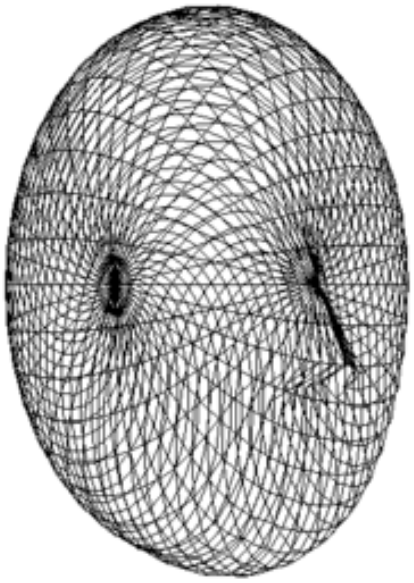

図 2.3: ワイヤーフレームモデル

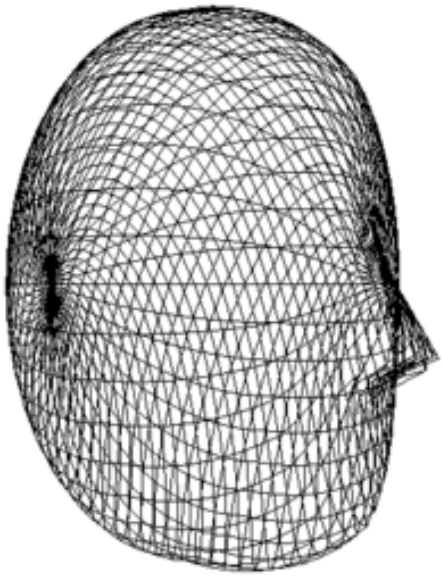

図 2.4: 別角度から見たワイヤーフレームモデル

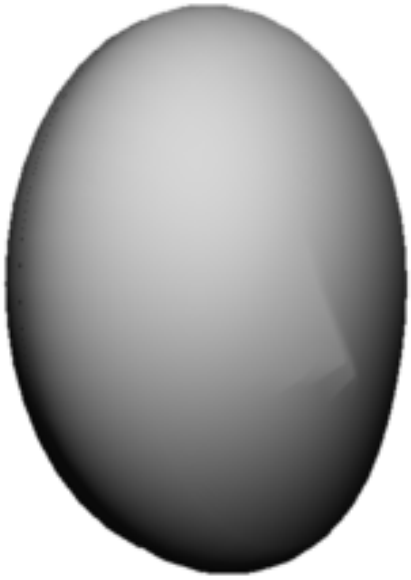

図 2.5: サーフェスモデル

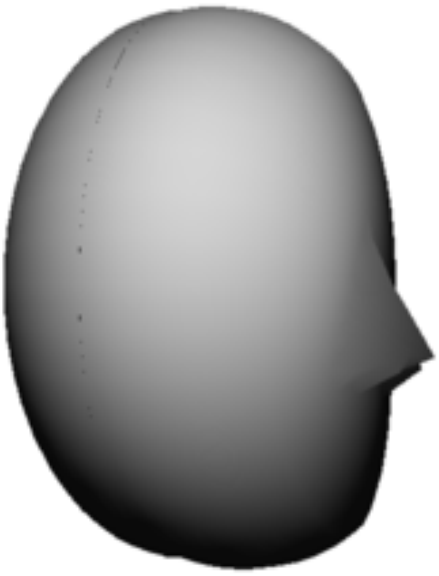

図 2.6: 別角度から見たサーフェスモデル

2.3　撮影角度の推定

3 次元空間中に話者の顔を作るため，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必要がある．そのため，撮影される画像より，顔に対してカメラがどこに位置しているのか推定を行う．具体的に説明するために，話者の顔形状モデルの中心を原点とした 3 次元座標を図 2.7 のように定義する．この 3 次元座標において，XZ 平面に対する，カメラと座標原点を結んだ直線との成すピッチ角 θ を求める．また話者は，顔に対してマイクを正面に構えるものとし，カメラは YZ 平面にあると仮定する．

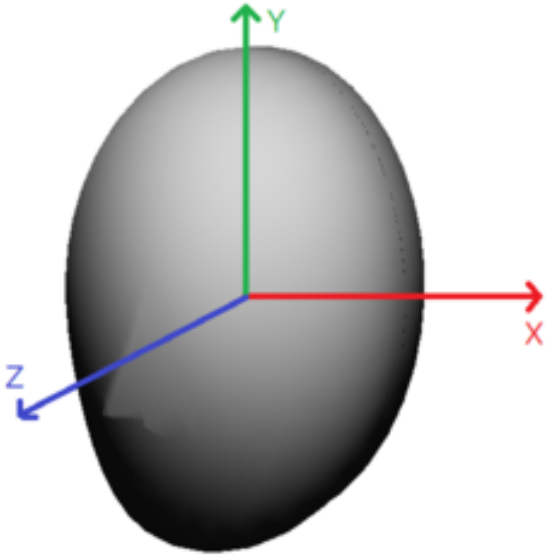

図 2.7: 3 次元顔モデルの座標

ピッチ角 θ の推定を，撮影画像中の顔器官点の座標を用いて行う．話者の両目間の距離を W，目から鼻の下端までの距離を H とする (図 2.8)．話者の顔を下方から撮影した場合，画像中の目から鼻の下端までの距離の実測値は，撮影する角度により，値 H と異なる．ここで，画像中の目から鼻の下端までの距離の実測値を H' とすると，顔の形状が平面であるとしたとき，H と H' および角 θ は図 2.9 のような関係になる．すなわち，撮影された画像中において，H は H' と θ を用いて次の式で表される．

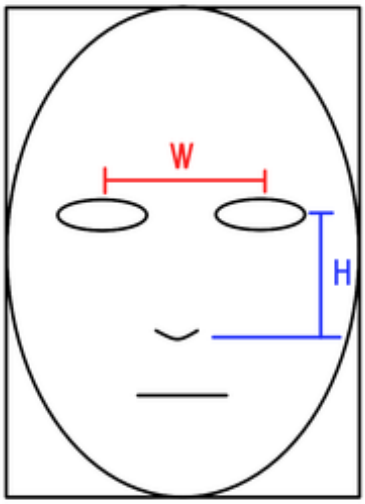

図 2.8: 両目間の距離 W と目から鼻の下端までの距離 H の定義

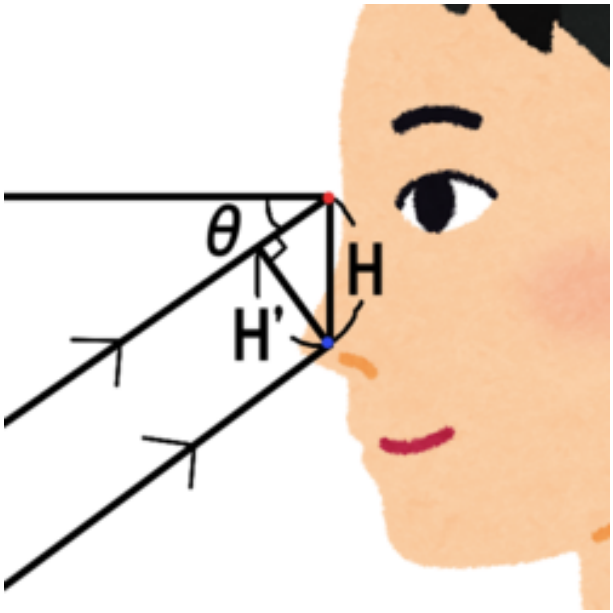

図 2.9: 距離 H および距離 H’ とピッチ角 θ の関係

$$H = H' \times cos(\theta) \tag{2.1}$$

したがって，両目間の距離と目から鼻の下端までの距離の比 R は，次のようになる．

$$R = \frac{H' \times cos(\theta)}{W} \tag{2.2}$$

また，これらの式よりピッチ角 θ は次式の通りである．

$$\theta = cos^{-1}(R \times \frac{W}{H'}) \tag{2.3}$$

これにより，実際の話者の顔における両目間の距離と目から鼻の下端までの距離の比 R がわかれば，ピッチ角 θ を推定することができる．

上述した撮影角度推定手法の精度の確認を行った．ピッチ角が 10 度から 50 度までの 10 度刻みで撮影し，そのとき推定された角度の平均値を記録した．また，顔器官の検出は，C++の機械学習ライブラリである dlib[9]，およびその学習済みモデルを使用した．表 2.1 に推定結果を示す．結果より，おおよそ推定角度が合っていると言える．50 度より大きい角度に関しては，顔器官点の検出精度が低く，角度の推定ができていない．

2.4　プレゼンテーションスクリーンへの話者の表示

2.2 節の手法を用いて作成した 3 次元の顔モデルに，2.3 節の手法にて推定した撮影角度に合わせて映像を投影するように話者の画像を貼り付ける（図 2.10）．この 3 次元モデルを顔の正面からレンダリングすることで，話者を正面から撮影したような画像を得る．ピッチ角 θ が 30 度，45 度，60 度のそれぞれの場合において，カメラを用いて撮影した画像と，それをシステムを用いて正面から見えるよう補正した画像，およびそのときの話者とカメラの関係を示す画像を図 2.12 から図 2.20 に示す．

表 2.1: 推定した撮影角度の平均 [度]

撮影角度	0 度	10 度	20 度	30 度	40 度	50 度
推定した角度の平均値	0.0	9.9	19.5	30.5	42.5	54.2

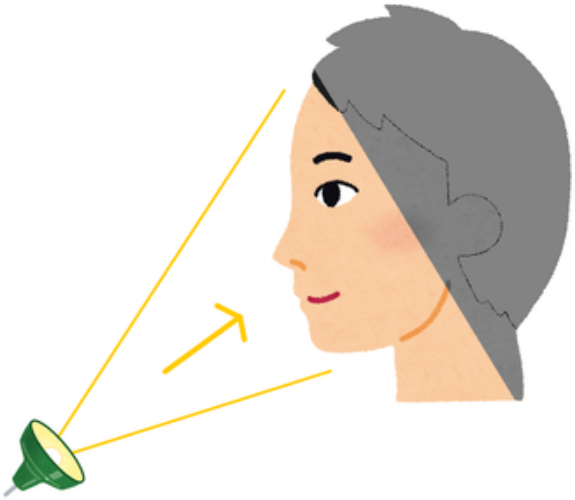

図 2.10: 画像の貼り付けのイメージ図

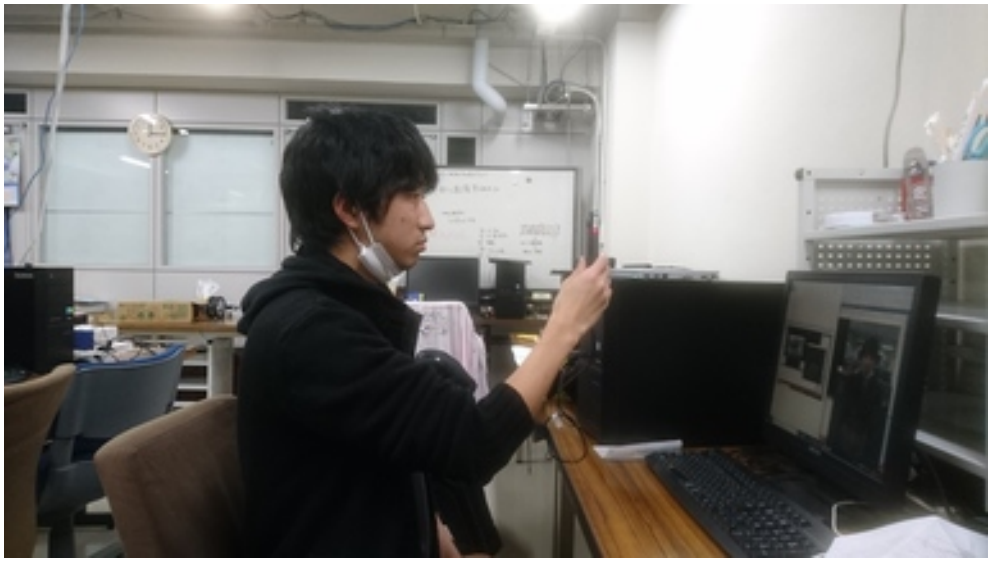

図 2.11: 顔とカメラの関係（ピッチ角 0 度）

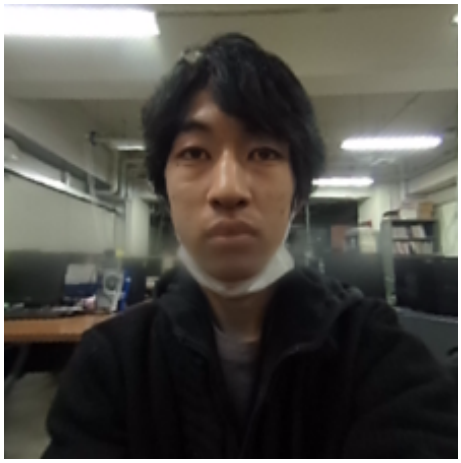

図 2.12: カメラの取得画像（ピッチ角 0 度）

図 2.13: システムによる補正画像（ピッチ角 0 度）

図 2.14: 顔とカメラの関係（ピッチ角 30 度）

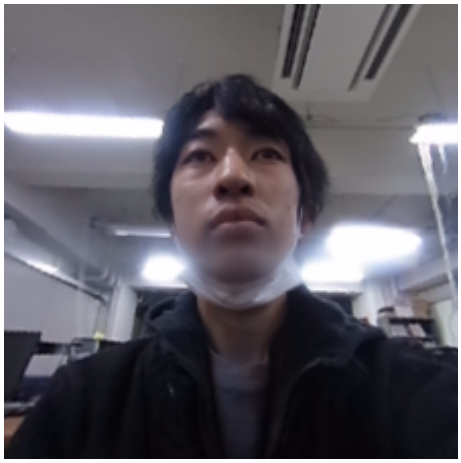

図 2.15: カメラの取得画像（ピッチ角 30 度）

図 2.16: システムによる補正画像（ピッチ角 30 度）

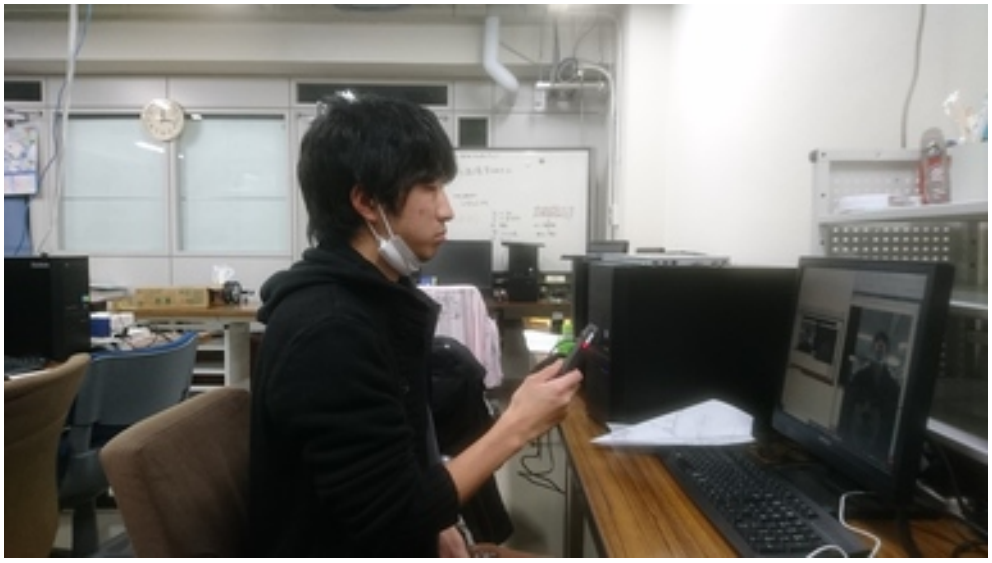

図 2.17: 顔とカメラの関係（ピッチ角 45 度）

図 2.18: カメラの取得画像（ピッチ角 45 度）

図 2.19: システムによる補正画像（ピッチ角 45 度）

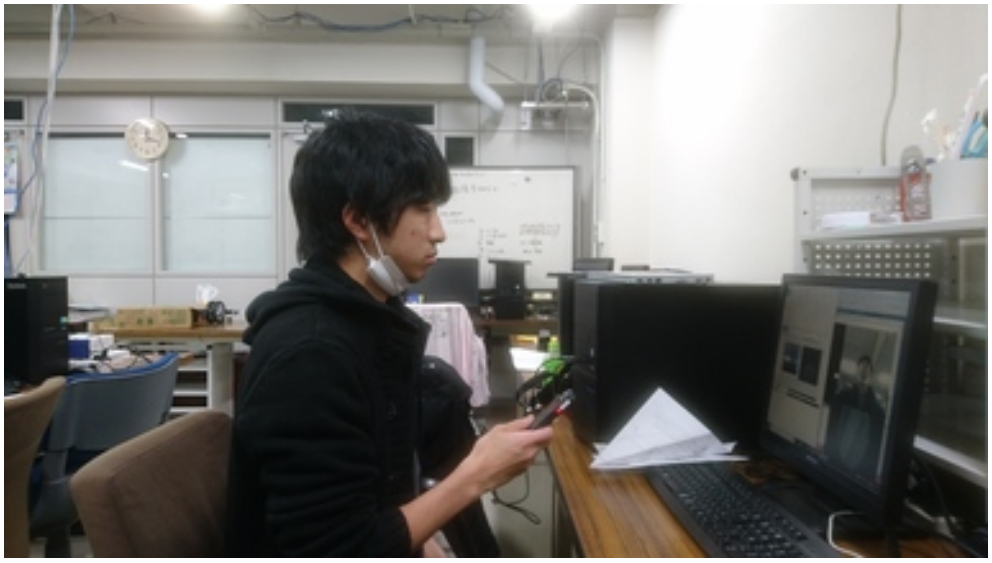

図 2.20: 顔とカメラの関係（ピッチ角 60 度）

図 2.21: カメラの取得画像（ピッチ角 60 度）

図 2.22: システムによる補正画像（ピッチ角 60 度）

第3章　実験

3.1　実験概要

第 2 章の提案をもとに，WindowsPC 上で実験システムを作成した．入力デバイスに Ricoh Company, Ltd より販売されているマイク内蔵型カメラ RICOH THETA S[8] を使用し，PC に USB ケーブルで有線接続する．被験者に RICOH THETA を顔より下方に構えてもらい，カメラで撮影した映像をもとに話者の抽出し，正面から撮影されたような映像に補正する．そして，補正した映像を被験者に評価してもらう．

システムを用いて補正した映像を被験者に評価してもらうため，まず評価用映像を撮影する．想定されるシステムの使用時のように口を動かしている様子を撮影するため，話者には用意した文章を読んでもらう．映像撮影時の様子を図 3.1，図 3.2 に示す．様々なピッチ角に対して評価してもらうため，30 度，45 度，60 度それぞれのマイクの持ち方において映像を撮影する．なお，評価用映像は各ピッチ角に対して男性 1 名の計 3 名である．正面補正はリアルタイムで処理が可能であるが，同一の補正映像に対して複数の被験者に評価してもらうため，補正した動画を保存しておく．また，それぞれの評価用映像男性に対して，RICOH THETA S とは別のカメラにより話者の顔正面映像を用意する．以下，説明を容易とするためにこれら 3 つの映像を，映像 A，映像 B，映像 C として次のように定義する．

映像 A：話者を正面から撮影した映像

映像 B：話者を下から撮影した見上げた構図の映像

映像 C：映像 B をシステムにより補正した映像

被験者による評価は，映像 A を評点 1 相当，映像 B を評点 5 相当の 5 段階評価とし，映像 C の評点として一番近いものを選んでもらう．それぞれの角度に関して，最初

図 3.1: 撮影時の話者を後ろから見た様子

図 3.2: 撮影時の話者を横から見た様子

に，評価の基準となる映像を映像 A，映像 B の順に見せた後，映像 C を見せ，評価を行ってもらう．実験は，大学生および大学院生の計 8 名に対して行った．

3.2　結果と考察

被験者による評価結果を表 3.1 に示す．表 3.1 は各項目において，1 から 5 の評点を付けた人数を示している．またそれぞれの角度の平均値と，評点全体の平均値を表 3.2 に示す．全体的に評点 4 や 5 が少ないことや，評点全体の平均値が 2.3 であることから，下方から撮影された顔映像を正面から撮影した映像に近づけられたこと

表 3.1: 被験者による評価結果 [人]

撮影した角度＼評点	1	2	3	4	5
30 度	4	3	1	0	0
45 度	0	7	1	0	0
60 度	0	2	2	4	0
合計	4	12	4	4	0

表 3.2: 評点の平均値

撮影した角度	平均値
30 度	1.5
45 度	2.1
60 度	3.3
評点全体	2.3

がわかる．しかし，角度ごとの評点の平均点は，30 度が 1.5，45 度が 2.1，60 度が 3.3 であり，角度が大きくなるほど評価が悪くなっている．これは，顔形状モデルが話者の顔と全く同じ形ではないため，厳密な補正ができていないことが原因であると考えられる．また，60 度の場合は他と比べて評点 1 や 2 の数が少ないが，これはあおり画像における話者の顔の検出精度が低く，映像を適切な角度から投影できないためであると考えられる．

第4章　むすび

本研究では，プレゼンテーションの質疑応答において，質問者をプレゼンテーション用のスクリーンに重畳表示するシステムの改良を提案した．具体的には，下から見上げた構図の質問者の顔画像を正面画像へ補正することで，スクリーンに重畳表示する際の違和感の軽減を図った．実験では，システムを用いて補正した映像を被験者に視聴してもらい，顔を正面から撮影した映像および下から撮影した映像と比較し評価してもらった．その結果，全体的に見上げた構図の顔画像が正面画像へと補正されているという評価が得られた．しかし，見上げる角度が大きくなると話者の顔が検出できずに補正が正確にできないこともある．話者の顔を検出できる撮影角度の範囲を拡大することが今後の課題である．また，本研究で作成した3次元顔形状モデルは，一般的な日本人成人男性の顔をモデルに作成しており，また検証も数人の日本人男性しか行っていない．そのため，女性や日本人でない話者の場合，自然な表示ができないことが考えられ，それらに対応することも必要である．

本研究が円滑で効果的なプレゼンテーションの一助とな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，大囿忠親，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”，情処学論, Vol.56, No.10, pp.2011-2021, October 2015.

[3] 趙新博，由井薗隆也，“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, October 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] 西口敏司，東和秀，亀田能成，角所考，美濃導彦，“講義自動撮影における話者位置推定のための視聴覚情報の統合”，電学論．C，電子・情報・システム部門誌, No.124, Vol.3, pp.729-739, March 2004.

[8] Ricoh Company, Ltd RICOH THETA S
https://theta360.com/ja/about/theta/s.html

[9] dlib C++ Library
http://dlib.net/

[10] Aaron S. Jackson, Adrian Bulat, Vasileios Argyriou, Georgios Tzimiropoulos, "Large Pose 3D Face Reconstruction from a Single Image via Direct Volumetric CNN Regression", ICCV 2017, pp.1031-1039, Octomber 2017.