

平成 28 年度 卒業論文

全天球カメラ内蔵マイクによる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の提案

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 25 年度入学 25115062 番

小林 勇輝

目次

| | | |
|-------|------------------------|----|
| 第1章 | はじめに | 1 |
| 第2章 | 全天球カメラによる質問者の抽出 | 3 |
| 2.1 | 魚眼カメラ画像の歪み補正 | 3 |
| 2.2 | 射影方式 | 5 |
| 2.3 | 他球面画像の生成 | 9 |
| 2.4 | プレゼンテーションスクリーンへの質問者の表示 | 11 |
| 第3章 | 実験 | 13 |
| 3.1 | 実験概要 | 13 |
| 3.2 | 結果と考察 | 15 |
| 3.2.1 | 聴衆による評価結果 | 15 |
| 3.2.2 | 質問者の位置による効果の検証 | 16 |
| 3.2.3 | 発表者による評価結果 | 17 |
| 第4章 | むすび | 18 |
| | 謝辞 | 19 |
| | 参考文献 | 20 |
| | 発表論文リスト | 21 |

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，プロジェクターを用いたスライドによるプレゼンテーションの機会が増えている．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，司会者のサポートを行うシステムの研究 [1] や発表中のプレゼンテーション資料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．さて，プレゼンテーション発表の際には形式は一人，または少数の人が多数の聴衆に向けて語ることがほとんどである．一般に，発表者は多数の聴衆全てに自身の声が行き渡るようにマイクを使用する．聴衆は発表者の言動を逃がさぬようにその方向へと目を向ける．しかしながら現実には，こういった場を使用される会場は大人数を収容できる部屋であることが多く，発表者の表情や身振りといったものを聴衆が全ての確に把握するのは難しい．この問題への対処法としてビデオカメラで発表者を撮影し，もう1つのスクリーンに表示するという方法がある．しかし，この方法を実現するには会場に1台，または複数台のビデオカメラを用意する必要がある．加えて，時として発表者はその発表中に自身の立ち位置を変更することがあるため，ビデオカメラを持って発表者を追跡するスタッフなども必要となり，この解決法は現実的にはコストの面から実用的ではないと言える．

そこで，当研究室では，この問題解決へのアプローチとして発表者を発表に使用しているスクリーン上に重畳表示するプレゼンテーション支援ソフト [3] の開発を行っている．このソフトウェアは発表者を光学カメラと距離カメラにより抽出してプレゼンテーション資料に重畳表示し，自身が強調したい資料中の箇所をインタラクティブに聴衆に提示するための機能を持つ．また，発表者は小さな動きでスクリーン上の全体を動き回ることができるように設計されているため，動きが制限される壇上でも有効的に使用できる．このソフトウェアを用いて，発表者が身振り手振りやプレゼンテーション資料への書き込みで内容を補完して説明することで，聴衆の深い理解が期待できる．コスト面の問題についても使用が想定されるカメラにはマイクロソフト社の Kinect など安価で誰でも入手しやすいデバイスが存在しており，また，追加のスタッフも必要ないことから実用的であると言える．

また，プレゼンテーション発表や学校の講義の際には発表とは別に不特定の人物から質問がなされることがある．しかし，先ほど述べたように会場の多くは，大人数を収容できる部屋であり，他の聴衆全てが質問者の姿を捉えることは難しい．加えて，質問者の位置

によっては発表者自身もはっきりとその姿を見ることは難しい。当研究室の従来研究では発表者にのみ着目しており、質問者の姿の捕捉については一切考慮されていない、この問題についても前述と同様に、コスト面の問題さえ解決できるのであれば、質問者を撮影するカメラとそれを操作する人を配置し、質問者を別のスクリーンに表示することで解決できる。また、魚眼レンズ付きカメラと多数のマイクロフォンを天井に配置し、発言者の位置を推定、撮影する手法 [4] やマイクに赤外線照射機能を付加し、発言者近辺の映像をデジタルズームする手法 [5] などの先行研究もある。しかし、それらの手法でも多くのコストがかかる。全ての会場で実現するためには機材の極端な低コスト化が必須だといえる。ところで、質問者がハンドマイクを利用することは、小さな部屋での状況を除けば、一般的である。そこで、本研究では、同様の効果を低コストで実現する為にマイクにカメラを内蔵させて、質問者をスクリーンに表示することを考えた。内蔵したカメラにより、質問者の顔を抽出し、プレゼンテーションに使用しているスクリーン上に重畳表示するシステムを構築する。「質問者の声」という音声情報だけでなく、「質問者の姿」という視覚情報が加わることで発表内容への聴衆の親近感が深まり、ひいては深い理解が期待できる。また、発表者が質問者の姿を目に捉えながら質問に回答することで、発表者の熱意が高まり、回答内容の向上も期待できる。本論文では、第2章において、質問者の抽出処理について説明し、第3章では第2章で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2章 全天球カメラによる質問者の抽出

2.1 魚眼カメラ画像の歪み補正

一般に，ハンドマイクは顔に近い所で使用し，上下は別にして使用する向きが決まっていない．従って，使用者に不自由なく使用してもらうためには使用者がどのようにマイクを持って顔を撮影できることが求められる．そこで魚眼カメラを2つ背中合わせに設置したハンドマイクを使用し，360°どの方向に使用者がいても対応することを考える．本研究ではビデオカメラとしても利用できるマイク内蔵型全天球カメラである RICOH THETA S[6]を用いる．RICOH THETA Sは，Ricoh Company, Ltdより販売されているカメラであり，高画質な全天球映像をライブストリーミングで撮影できる．実際に RICOH THETA Sにより取得した画像の例を示す(図 2.1)．左側に一方のレンズから取得した半球画像が，右側にもう一方のレンズから取得した半球画像が表示されている．魚眼レンズはその性質上，得られる画像にゆがみが生じるため，人物を抽出する前処理として画像の歪みを補正する必要がある．変換前の座標を x, y ，対応するゆがみ補正後の画像上の座標を x', y' ，画像の幅，高さを w, h ，レンズと投影面との距離を D ，球の半径を $r (= w/2 = h/2)$ とする．歪み補正後の座標に対応する変換前の座標 x, y は以下の式 (2.1), (2.2) により算出できる．図 2.1 の一方のカメラ画像から歪みを取り除いた例を下に示す(図 2.2)．

$$x = \frac{r \left(\frac{x' - w}{2} \right)}{\sqrt{D^2 + \left(\frac{x' - w}{2} \right)^2 + \left(\frac{y' - h}{2} \right)^2}} + \frac{w}{2}, \quad (2.1)$$

$$y = \frac{r \left(\frac{y' - h}{2} \right)}{\sqrt{D^2 + \left(\frac{x' - w}{2} \right)^2 + \left(\frac{y' - h}{2} \right)^2}} + \frac{h}{2}. \quad (2.2)$$

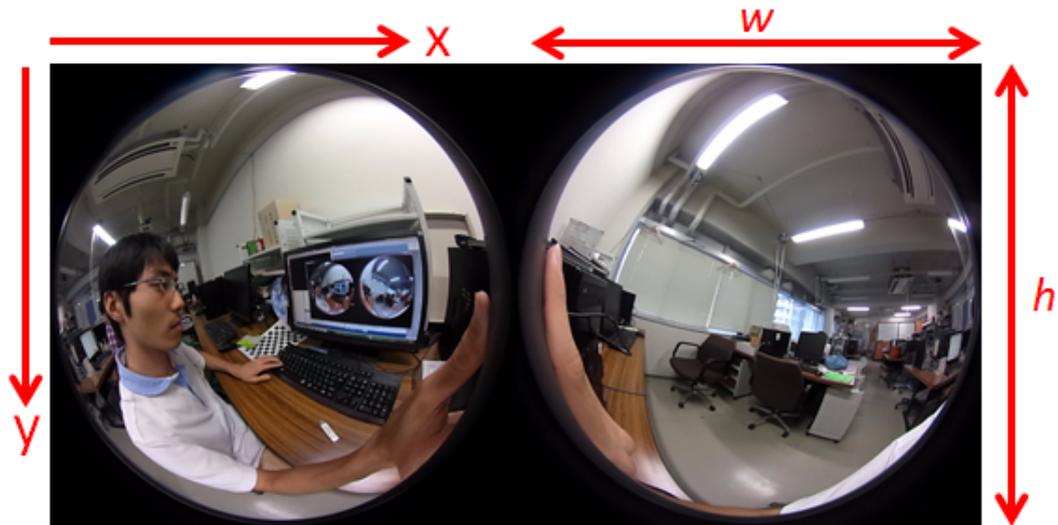


図 2.1: カメラからの取得画像

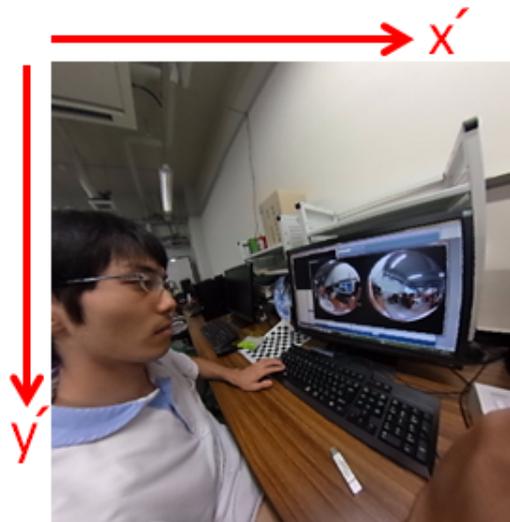


図 2.2: 歪み補正後の画像

ところで本研究では、質問者が持ち方を気にせずにマイクを使用できることを目標にしているため、図 2.1 のように片側のカメラに姿が完全に入るとは限らない。そこで、2つの画像から取得した画像を用いて別方向から見た場合の画像を生成する。まず次節で、魚眼カメラでよく用いられる射影方式について説明する。2.3 節では別方向から見た場合の画像生成について述べる。最後に、2.4 節で実際のプレゼンテーションスクリーンへの質問者の表示について述べる。

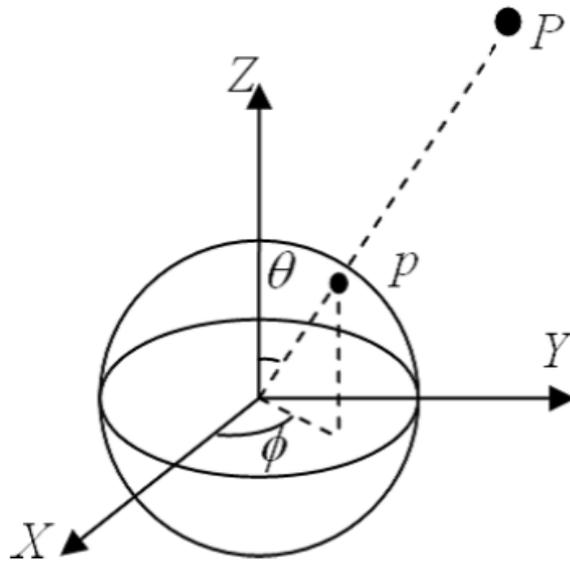


図 2.3: 球面モデルの球面投影

2.2 射影方式

魚眼レンズの変換において主に採用される射影方式は等距離射影方式である．RICOH THETA Sでは，製品の射影法式を公開していない．そこでこの節では，等距離射影方式を前提としてよいかどうかを検証する．以下では球面モデルに基づく魚眼カメラの校正方法を説明する [7]．まず，球面モデルの球面投影を考える (図 2.3)．空間に半径 f の球と点 P があるとき， P と球の中心を結ぶ直線が球面と交わる点 p を，点 P の球面への投影点とする．球面画像は，焦点を球の中心として，全天周視野を表現できる．原点を球の中心においたカメラ座標系において点 P の座標が

$$P = [X_c \ Y_c \ Z_c], \quad (2.3)$$

であるとき，この点の球面画像への投影点 p の座標は図 2.3 に示す天頂角 θ と方位角 ϕ を用いて，

$$p = [f \sin \theta \cos \phi \ f \sin \theta \sin \phi \ f \cos \theta]^T, \quad (2.4)$$

$$= \frac{f}{\sqrt{X_c^2 + Y_c^2 + Z_c^2}} [X_c \ Y_c \ Z_c], \quad (2.5)$$

と表すことができる．従ってスケールファクタを除いて P は p に等しい．実際に画像処理を行う際には， $f = 1$ とし，単位球への投影である正規化球面座標を

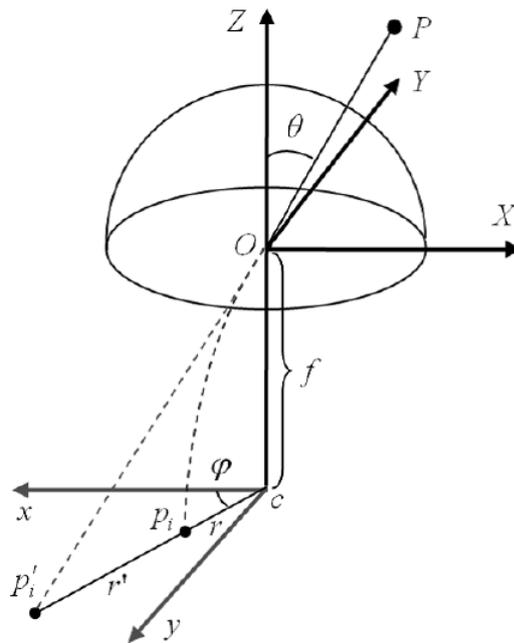


図 2.4: カメラレンズと平面画像

$$p = [\sin\theta\cos\phi \quad \sin\theta\sin\phi \quad \cos\theta]^T, \quad (2.6)$$

と表す.

カメラレンズと平面画像の関係を図 2.4 に示す. 点 P は平面画像上の点 p_i' に投影される. 等距離射影方式を用いた場合,

$$r' = f\theta, \quad (2.7)$$

$$\sin\phi = \frac{y}{r'}, \quad (2.8)$$

$$\cos\phi = \frac{x}{r'}, \quad (2.9)$$

であるから, 点 p の座標 (x', y', z') は

$$x' = \sin\left(\frac{r'}{f}\right) \times \frac{x}{\sqrt{x^2 + y^2}}, \quad (2.10)$$

$$y' = \sin\left(\frac{r'}{f}\right) \times \frac{y}{\sqrt{x^2 + y^2}}, \quad (2.11)$$

$$z' = \cos\left(\frac{r'}{f}\right), \quad (2.12)$$

と表される。

カメラより取得した画像と平行に xy 軸をとり，奥行きを z (図 2.5) とする．また，球の中心を $x = y = z = 0$ とする．カメラからの取得画像(図 2.6)とカメラの取得画像に上記変換式を適用して求めた球を x 軸回りに 90 度回転させた画像(図 2.7)，及び y 軸回りに 60 度回転させた画像(図 2.8)を示す．回転に使用する変換式は 2.3 節にて述べる．等距離射影方式を前提とした変換を行っても，接合部分の違和感は少ないことがこれらの画像よりわかる．よって本研究では等距離射影方式を前提として変換を行う．

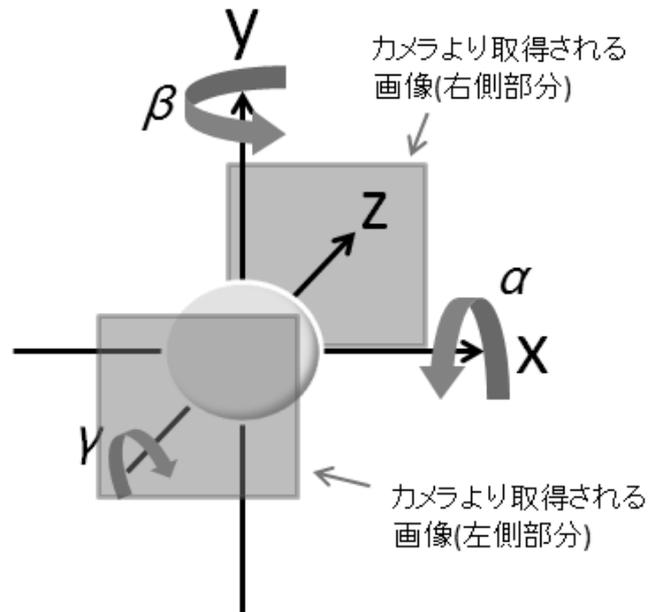


図 2.5: 軸の定義

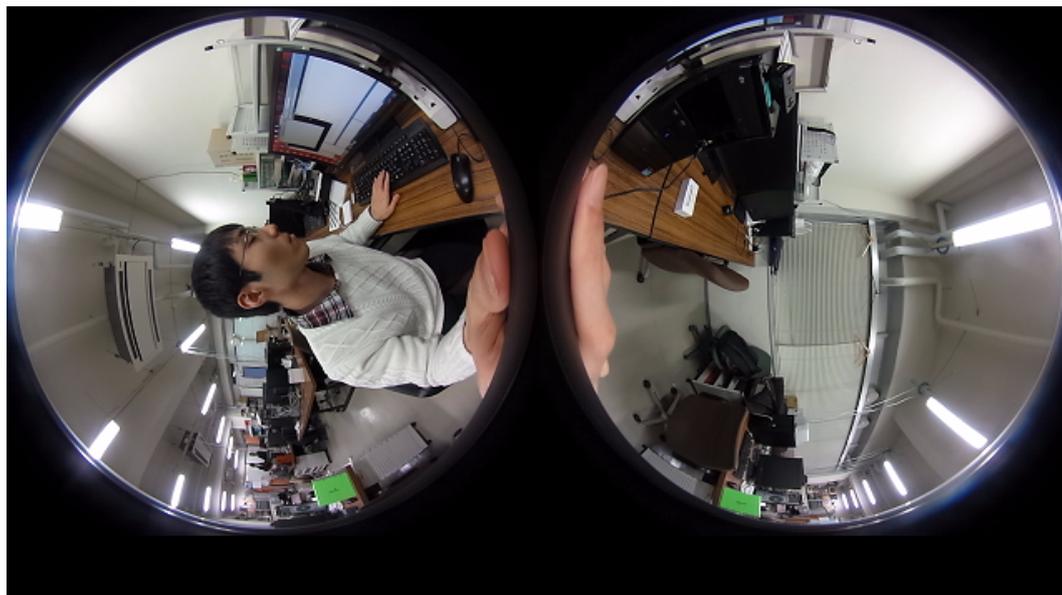


図 2.6: カメラより取得した画像

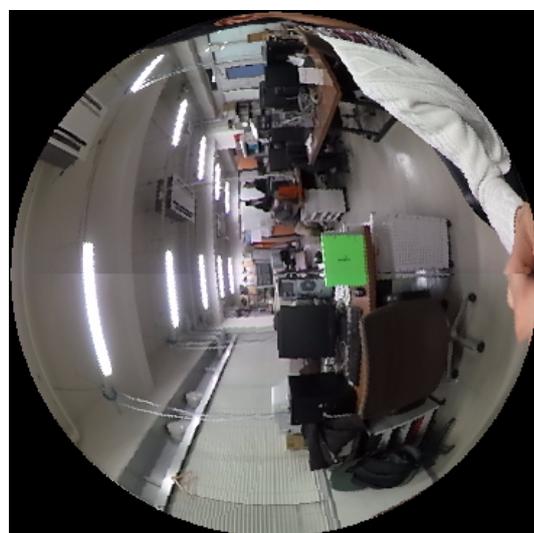


図 2.7: x 軸回りに 90 度回転した画像

図 2.8: y 軸回りに 60 度回転した画像

2.3 他球面画像の生成

2.2 節にて述べた等距離射影の変換式及び極座標による球の回転公式を用いて別方向から見た場合の球面画像の生成を行う． x 軸周りの回転を θ 回転， y 軸周りの回転を ϕ 回転とする．回転後の座標を x'', y'' とし， θ 回転， ϕ 回転に使用する式を下に示す．

- θ 回転

$$y'' = y' \cos \theta - z' \sin \theta, \quad (2.13)$$

$$x'' = x', \quad (2.14)$$

- ϕ 回転

$$y'' = y', \quad (2.15)$$

$$x'' = x' \cos \phi - z' \sin \phi. \quad (2.16)$$

回転後，2.1 節にて述べた歪み補正を行い，スクリーン上へ表示する透視投影画像を生成する．例として $\theta = -\pi/2$ ，または $\phi = \pi/3$ 回転することで対象物が正面に位置する画像となるカメラ取得画像 (図 2.9, 図 2.11) と，実際に回転した後に 2.1 節の歪み補正を行った画像 (図 2.10, 図 2.12) を示す．

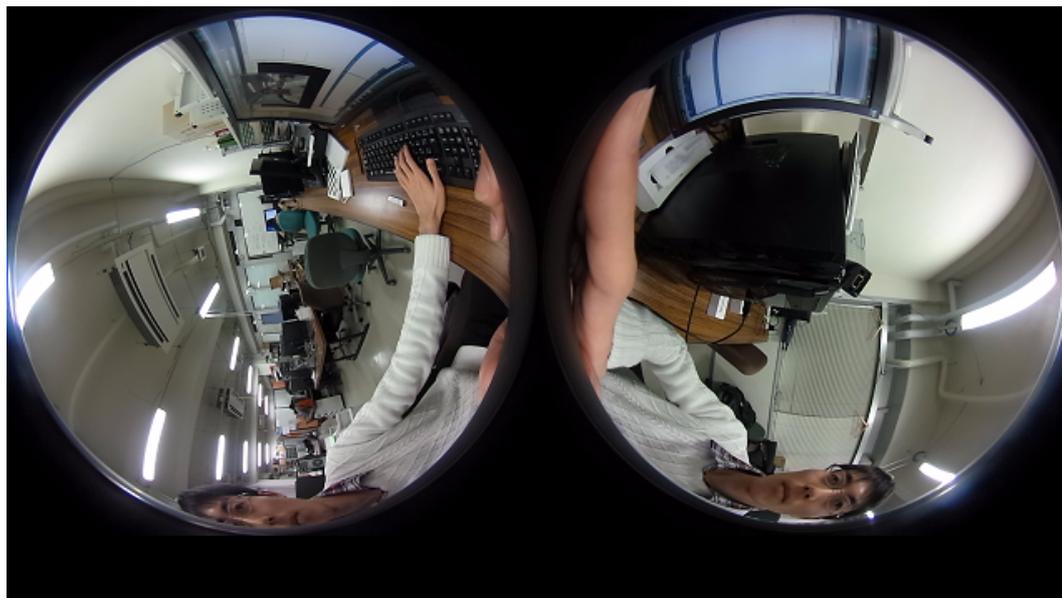
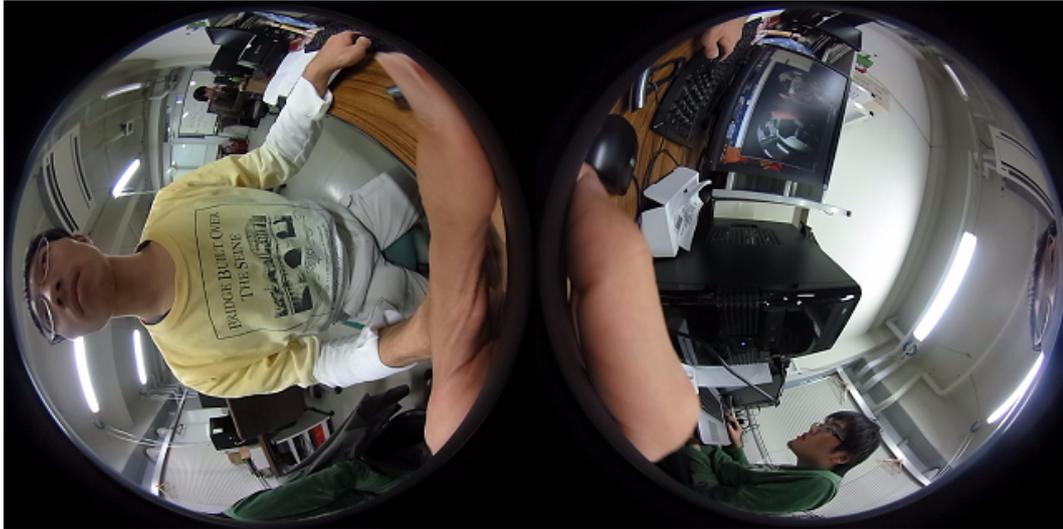


図 2.9: $= -\frac{\pi}{2}$, 取得画像



図 2.10: $= -\frac{\pi}{2}$, 回転後に歪み補正をした画像

図 2.11: $= \frac{\pi}{3}$, 取得画像図 2.12: $= \frac{\pi}{3}$, 回転後に歪み補正をした画像

2.4 プレゼンテーションスクリーンへの質問者の表示

前述の 2.2, 2.3 節の手法を用いて 30 度ステップでカメラからの取得画像を回転させていく。各ステップにおいて生成した画像に 2.1 節の歪み補正を行い, OpenCV を使用して顔認識を行う。画像には質問者以外の聴衆も映っており, 質問者以外の顔も認識される可能性が高い。しかし, 質問者は最もカメラレンズに近い位置にいるはずであるため, 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。採用した画像を図 2.13

のようにプレゼンテーションスクリーンに重畳表示する。

重回帰分析による推定

主観的に知覚したベクシオン強度とその方向の違いから不快感強度を推定するモデル

$$Z = aX + bY + c$$

Z: 不快感強度
X: ヨー回転に対応するベクシオン強度
Y: ピッチ回転と不明確な回転に対応するベクシオン強度の合計値
a,b,c: 定数



図 2.13: スクリーンへの質問者の表示

第3章 実験

3.1 実験概要

以上の提案をもとに Windows PC 上で実験システムを作成した。本研究ではプレゼンテーションにおいて、質問者の姿を発表スクリーン上に表示することで、質疑応答における聴衆の発表内容への理解を深めることを目的としている。本実験システムでは入力デバイスに Ricoh Company, Ltd より販売されているマイク内蔵型カメラ RICOH THETA S を用い、PC に USB ケーブルで有線接続する。RICOH THETA S のカメラより取得した画像を元に質問者の抽出を行い、その姿を発表スクリーン上に表示する。実験はプレゼンテーション会場において、聴衆の大部分、及び発表者が質問者の顔を、背後からのために、あるいは遠いために直接視認できないという想定のもと、椅子に着席した人が 160 人程度入る部屋を使用して行う。被験者は大学教員 2 名、学会発表経験のある学生 2 名を含む 11 人である。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、その後質疑応答を行う。これを質問者のスクリーンへの表示なしとありの 2 回を 1 セットとして行い、発表者役及び聴衆役に後述する観点について比較、採点してもらう。この手順を計 5 セット行う。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るか

の 3 点について質問者表示ありの場合に対して評価を行う。評価は質問者表示なしの場合を基準 (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るか
- 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 3.1, 図 3.2, 図 3.3 に実験の様子を示す。



図 3.1: 部屋の後ろから見た様子



図 3.2: 部屋の前から見た様子



図 3.3: 発表者とスクリーン

3.2 結果と考察

3.2.1 聴衆による評価結果

聴衆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 の各評価をつけた回数を示している．聴衆は発表者と質問者を除く 9 人で構成され，実験が 5 セット行われていることから，各項目はのべ 45 回評価されている．全体的に 1, 2 の評価がほぼないことから，本実験システムの導入により一定以上の効果が得られたことが分かる．特に，質問者が身近に感じられるか，という質問に対しては全ての回答で 4 以上の評価が得られたことから，スクリーンに質問者が映ることで聴衆は質問者とのその後のコミュニケーションが図れるのではないかと考えられる．しかしながら，質疑応答内容の理解度が高まったか，については 5 の評価が 1 つもなく，あまり大きな効果はあげられていないといえる．原因として，内容の理解度は質問の内容自体に左右される上，音声情報による影響が大きいことから，本実験システムの導入の影響が少なかったのだと考えられる．

表 3.1: 聴衆による評価結果 (各評点を回答したのべ人数)

| 評価項目 | 評点 | | | | |
|------------------|----|---|----|----|----|
| | 1 | 2 | 3 | 4 | 5 |
| 質問内容への興味が湧き易いか | 0 | 0 | 23 | 17 | 5 |
| 質疑応答内容の理解度が高まったか | 0 | 1 | 24 | 20 | 0 |
| 質問者が身近に感じられるか | 0 | 0 | 0 | 20 | 25 |

3.2.2 質問者の位置による効果の検証

3.2.1 節の聴衆の評価を質問者より前にいた聴衆の評価と質問者より後ろにいた聴衆の評価の2つに分け、質問者の位置によって効果に違いがあるかどうかを検証する。表3.2に質問者より前にいた聴衆に対するのべ26回の評価の分布を、また表3.3に質問者より後ろにいた聴衆に対するのべ19回の評価の分布を示す。質疑応答内容の理解度が高まったか、及び質問者が身近に感じられるか、の項目については質問者より前に座っているか後ろに座っているかの違いによる影響はみられなかった。質疑応答の内容に興味が高まったか、については質問者より前に座っている人の方が評価3が少なく、代わりに評価5が多いという結果であった。これは、スクリーンへの質問者の表示がない場合に、質問者の後ろ姿が見える聴衆よりも質問者を直接見ることができない聴衆の方が実験システム導入による差を実感したからだと考えられる。すなわち、実験システムの導入によって、新たに質問者が実際に話している姿が見えるようになった聴衆の方が、話している内容に興味を持つようになる可能性が高いのだと推察できる。この推察を裏付けるものの一つとして、実験参加者の意見の中に「質問者の背中が見える場合には、顔がスクリーンに表示されても、表示していない場合とあまり差は感じられない」という意見があった。総評として、質問者が比較的会場の後ろにいる場合の方が本実験システムの有効性は高いといえる。

表 3.2: 質問者より前にいた聴衆の評価分布 (各評点を回答したのべ人数の割合)

| 評価項目 | 評点 | | | | |
|------------------|----|----|-----|-----|-----|
| | 1 | 2 | 3 | 4 | 5 |
| 質問内容への興味が高まったか | 0 | 0 | 46% | 39% | 15% |
| 質疑応答内容の理解度が高まったか | 0 | 4% | 54% | 42% | 0 |
| 質問者が身近に感じられるか | 0 | 0 | 0 | 47% | 53% |

表 3.3: 質問者より後ろにいた聴衆の評価分布 (各評点を回答したのべ人数の割合)

| 評価項目 | 評点 | | | | |
|------------------|----|---|-----|-----|-----|
| | 1 | 2 | 3 | 4 | 5 |
| 質問内容への興味が高まったか | 0 | 0 | 58% | 37% | 5% |
| 質疑応答内容の理解度が高まったか | 0 | 0 | 53% | 47% | 0 |
| 質問者が身近に感じられるか | 0 | 0 | 0 | 47% | 53% |

3.2.3 発表者による評価結果

発表者による評価結果を表 3.4 に示す。発表者による評価でも聴衆と同様に、質問者が身近に感じられるという点に関しては十分な効果が得られた。また、若干ではあるが質問に対する回答に思いが込められやすいという結果が得られており、質問者の理解を深めるにあたって効果が期待される。

表 3.4: 発表者による評価結果 (各評点を回答したのべ人数)

| 評価項目 | 評点 | | | | |
|----------------------|----|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 |
| 質問内容の理解度が高まったか | 0 | 0 | 2 | 3 | 0 |
| 質問者が身近に感じられるか | 0 | 0 | 0 | 3 | 2 |
| 質問に対する回答に思いがより込められたか | 0 | 0 | 1 | 4 | 0 |

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、聴衆や発表者が質問者の姿を直接視認することが難しいという問題を解決する為のシステムを提案した。実験では、本システムがプレゼンテーションの質疑応答において聴衆が理解を深めるための一助となるかどうかを検証した。被験者に質問者の姿がスクリーン上に表示されている場合とされていない場合について比較してもらい、その差を評価してもらった。その結果、本システムの導入により一定以上の効果が得られるということが実証された。一方、ユーザに実際に体験してもらい、フィードバックを得たことで、今後の開発のために必要な項目を洗い出すことができた。本システムはスクリーン上に表示する質問者の画像サイズが固定されており、拡大/縮小する機能を持ち合わせていなかった。実験の中でその機能の必要性が浮き彫りとなったことから、導入を検討したい。また、見上げた構図の顔を目にする状況は一般には少ないため、多少の違和感があった。使用者のカメラの持ち方によっては顔の認識精度が低いことがあり、本研究の目標である質問者が持ち方を気にせずにマイクを使用できることが完全には実現できていない。よって、今後の課題として、斜め下から見上げた画像を正面から見た画像に補正することも必要である。

プレゼンテーションという場は聴衆に何かを伝えるためにある。しかしながら、発表の時間だけで全てを伝えきるのは難しい。そのため、質疑応答によって聴衆の疑問に答え、可能な限り理解を深めることは重要である。本研究がそ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,” 情処学論 Vol.56, No.10, pp.2011-2021, October 2015.
- [3] Kenji.F, Yusuke.N, “Getting Yourself Superimposed on a Presentation Screen,” *Proceedings of the 2nd ACM symposium on Spatial user interaction*, Honolulu, Hawaii, USA, pp.138-138, October 2014.
- [4] 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.
- [5] 市村哲, 福田正城, 田胡和哉, “赤外線照射マイクを追尾する質問者自動撮影システム,” 情処学論 Vol.49, No.1, pp.141-149, January 2008.
- [6] Ricoh Company, Ltd RICOH THETA S
<https://theta360.com/ja/about/theta/s.html>
- [7] 中野誠士, 李仕剛, 千葉則茂, “球面モデルに基づくしま模様パターンを用いた魚眼カメラの校正,” *IEICE Trans. Inf. & Syst.(Japanese edition)* Vol.90, No.1, pp.73-82, January 2007.

発表論文リスト

投稿論文

1. Yuki Kobayashi, Kenji Funahashi, “Superimposing Questioner on Presentation Screen Using Microphone with Whole-Sky Camera,” *ICAT-EGVE 2016*, Arkansas, USA, December 2016.