

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

柴田大地¹⁾ 小林勇輝¹⁾²⁾ 舟橋健司¹⁾

1) 名古屋工業大学 2) 現在：株式会社システムサーバー

Superimposing Questioner on Presentation Screen Using Whole-Sky Camera on Microphone

Daichi Shibata¹⁾ Yuki Kobayashi¹⁾²⁾ Kenji Funahashi¹⁾

1) Nagoya Institute of Technology 2) Present: SYSTEM SERVER CO., LTD.

d.shibata.946@stn.nitech.ac.jp

y.kobayashi.447@nitech.jp

kenji@nitech.ac.jp

概要

プレゼンテーションの場では不特定の人物から質問されることがある。このとき会場が広いと、発表者や聴衆が質問者を視認するのは難しい。スタッフが質問者を撮影してスクリーンに映すことは可能だろうが、そのための人員やカメラ機材などが必要になる。ところで、質問者がハンドマイクを使用することは、広い会場では珍しくない。そこで、我々はマイクにカメラを内蔵させて、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提案する。質問に対する理解や親近感の高まりが期待できる。

1 はじめに

学会や企業での発表、学校での講義などのプレゼンテーションの場では、一人の発表者が、多数の聴衆に向けて語るのが一般的である。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、発表者の表情、身振り手振りなどの情報も同じく重要である。しかし、聴衆が多数の場合、当然プレゼンテーションを行う会場も広くなり、これらの情報を聴衆が全体的に把握するのは難しい。この問題を解決するために我々は、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 [1] を提案している。このシステムは、発表者を光学カメラと距離カメラにより抽出し、プレゼンテーション資料に重畳表示する。また、プレゼンテーション

資料中の任意の箇所をインタラクティブに強調表示する機能を持つ。このシステムを用いた発表者が、身振り手振りやプレゼンテーション資料への書き込みを行い内容を補完して説明することで、発表内容に対する聴衆のより深い理解が期待できる。ところで、プレゼンテーションや学校での講義の際には、聴衆から質問がなされることがある。しかし、上述の通り多人数に対する発表を行う場合、会場は広くなるため、発表者や他の聴衆全てが質問者の姿を視認できないことが想定される。上述の従来研究では発表者のみに着目しており、質問者の存在については一切考慮されていない。上述のシステムにおける発表者と同様に、質問者についても表情や身振りなどをスクリーンに表示する機能を提供することができれば、質疑応答の内容のより深い理解が促され、さらに効果的なプレゼンテーションに繋がることが期待される。

著者らは先行研究として、全天球カメラ内蔵マイクを使用し、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提案している [2]。しかしながら、撮影した映像を歪み補正のみでスクリーンに表示すると、通常マイクは顔よりも下に構えるため、正面からの顔画像と異なり違和感が生じてしまう。そこで本論文では話者の表示に焦点を当て、下から見上げた構図の顔画像を正面画像へと補正する方法について述べる。広い会場でのマイクの使用は一般的であり、質問者の撮影や抽出を低コストで実現できると考えられる。「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、発表や質疑の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、回答内容が向上することも期待できる。

2 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できることを目標とするため、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 [3] を使用する。RICOH THETA S の画像を図 1 に、RICOH THETA S を持った話者の例を図 2 に示す。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 3) に示す。また、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。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 [4]。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 4) と、垂直に 60 度回転後歪み補正を行った画像 (図 5) を示す。

3 見上げた構図から正面顔画像への補正

ハンドマイクにカメラが付いているという制約上、一般的には話者の口元より下にカメラが位置することにな



図 1 RICOH THETA S

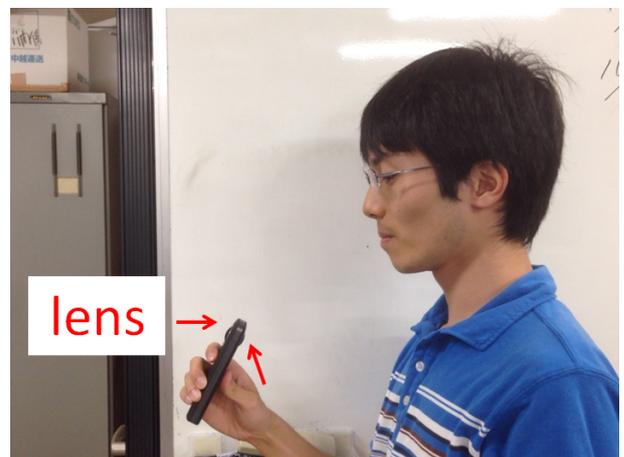


図 2 話者が全天球カメラを持った例

る。これにより、カメラにより撮影される映像中の話者の顔が、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。

そこで、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 6 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と、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が考えられる。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

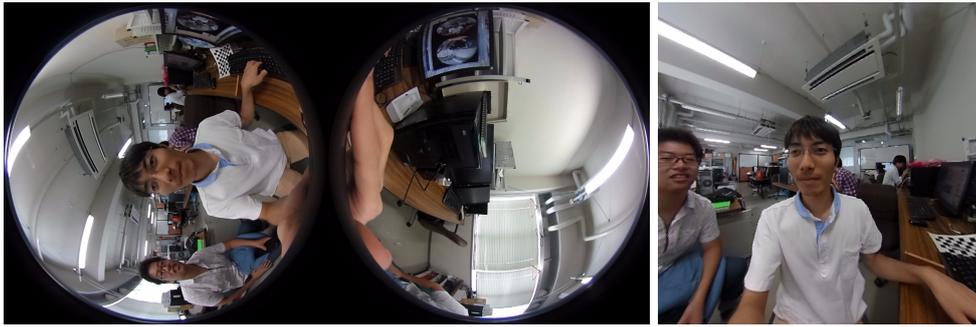


図3 取得画像と歪み補正後の画像



図4 水平に90度回転後, 歪み補正



図5 垂直に60度回転後, 歪み補正

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7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で

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8左に、補正後のシステムの出力画像を図8右に示す。

4 実験と考察

本システムの有効性を検証するために評価実験を行った。実験1では、スクリーンに質問者を重畳表示することで、質疑応答において発表者本人や聴衆の意識がどのように変化するか調査して、その有効性を評価する。

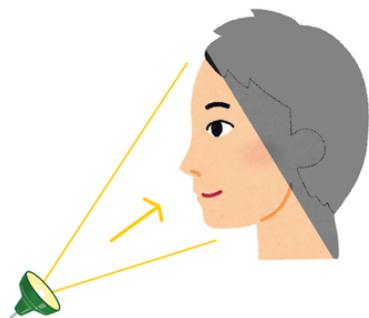


図6 画像の貼り付けのイメージ図

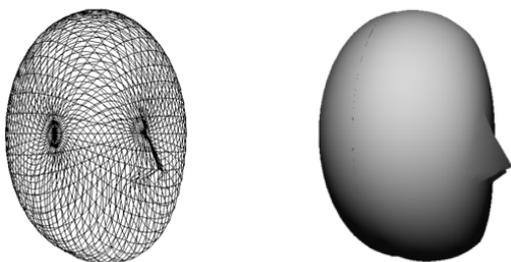


図7 作成した3次元顔形状モデル



図8 撮影角度30度のときの取得画像と補正画像

なお、この時点では実験システム設置の事情により3節の正面補正を行っていない。実験2では正面に補正した映像の評価を行う。

4.1 実験1

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で発表者役と質問者役が質疑応答を行う。実験の様子を図9と図10に示す。質疑応答はスクリーンへの質問者表示なしの場合とありの場合の2回行う。質問者表示なしの場合を基準3として、ありの場合の評価を

3つの項目について5段階(1:悪い~5:良い)で被験者に回答してもらった。聴衆役が各評点を回答した回数を表1に、発表者役が各評点を回答した回数を表2に示す。どの項目であってもほとんどが3以上の評点となり、また、評点全体の65%以上が4または5であったことから本システムの有効性が実証された。

4.2 実験2

まずマイク内蔵型カメラを顔より下方に構えた状態で、3通りの撮影角度の見上げた構図の映像を撮影する。角度推定や正面補正はリアルタイムに処理しており、動画として保存する。それらの補正を行った各映像を被験者8名に評価してもらう。撮影角度30度の図8に加えて、撮影角度を45度としたときの例を図11に、60度としたときの例を図12に示す。評価時には、別のカメラによる顔正面映像を用意し、補正の評価の基準として使用する。話者を正面から撮影した映像を評点1相当(完



図9 実験の様子



図10 実験中のスクリーンの様子

表1 聴衆による評価（各評点を回答したのべ人数）

評価項目	評点				
	1	2	3	4	5
質問者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

表2 発表者による評価（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



図11 撮影角度45度の際の取得画像と補正画像



図12 撮影角度60度の際の取得画像と補正画像

全に補正されている，正面映像と遜色ない），見上げた構図の映像を評点5相当（全く補正されていない，補正の効果がない）の5段階評価とし，補正した映像を評価してもらう．結果は表3の通りで平均評価値が2.3であり，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がわかった．しかし，角度が大きくなるほど評価が悪くなっている．原因として，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考えられる．

表3 被験者による評価結果（各評点を回答したのべ人数）

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

5 まとめ

本論文ではプレゼンテーションにおける質問者に着目し，質疑応答を支援するシステムを提案した．具体的には，マイクに内蔵された360度撮影が可能な全天球カメラを用いて質問者をスクリーンに重畳表示することで支援する．またその際，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して表示することで，より自然な表示を目指した．実験では，従来のプレゼンテーションと比較して，聴衆や発表者の理解度が高まっていることや，質問者をより身近に感じられることを確認できた．また，正面から撮影した映像と下から撮影した映像と比較を行い，違和感が軽減されていることを確認した．

参考文献

- [1] 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.
- [2] Yuki Kobayashi, Kenji Funahashi, “Superimposing Questioner on Presentation Screen Using Microphone with Whole-Sky Camera”, Proceedings of ICAT-EGVE 2016, pp.3-4, 2016.
- [3] Ricoh Company, Ltd., RICOH THETA S, <https://theta360.com/ja/about/theta/s.html>. 参照: 2019-9-27.
- [4] OpenCV API Reference, Haar Feature-based Cascade Classifier for Object Detection, https://docs.opencv.org/3.0-beta/modules/objdetect/doc/cascade_classification.html?highlight=cv2.grouprectangles. 参照: 2019-9-27.