

前後移動時の運動視差を模した擬似視差映像の生成

平岡 勇作 宮田 慎也 坂本 竜基

本稿では、固定されたカメラとデプスカメラの映像から前後方向に移動した際の運動視差を再現した映像をリアルタイムで生成するシステムを提案する。運動視差とは観察者が運動することによって生じる見えの違いであり、メディアスペースでは前後方向の運動視差の有無が観察物の存在感に影響を及ぼすとされている。本研究では奥行情報を基にユーザと背景に領域分割し、それぞれ個別に拡大して合成することで運動視差を模した映像を生成手法の提案と、その評価を報告する。

Virtual Motion Parallax Videos as if Users Moving Back and Forth

YUSAKU HIRAOKA, SHINYA MIYATA
and RYUUKI SAKAMOTO

In this paper, we propose a system converting videos of video chat into videos with motion parallax virtually as if the users move back and forth. The motion parallax is the difference of views between when human moves from here to somewhere, and effects the telepresence of objects in mediaspace. For making videos with the motion parallax, we use the depth camera for the video segmentation divided into background and foreground. The background and foreground are magnified by each scale factor and composed into single video.

1. はじめに

計算機ネットワークによって接続された遠隔地の人間同士の協調作業やコミュニケーションを彼らがあたかも同室にいるかのように演出するアプリケーションは古くから研究がなされている。メディアスペース¹⁾は、これらを総称する概念であるが、いまだ十分に同室感を演出するにはいたっていない²⁾⁻⁴⁾。

この要因のひとつと考えられるのがカメラを通じて得られる映像は、人間の両眼から得られる映像とかけ離れており、遠隔地にいる相手の存在感を得ることができない点にある⁵⁾。

両者の大きな違いとして（固定）カメラの場合は運動視差と両眼視差⁶⁾がない点が挙げられ、これらを提供する研究がなされてきた^{7),8)}。

このうち、1対1の遠隔地環境において、片方のユーザからみて相手側に設置されたカメラを物理的に前後移動させることにより得られる運動視差はテレプレゼンスを高める効果があることが報告されている^{9),10)}。

本研究では、この先行研究のような1対1のテレビ会議環境において前後方向の運動視差を模した映像を、固定された通常のカメラとデプスカメラをもちいて擬似的に生成する手法について述べる。これにより、設置の手間やスペースの制限が軽減され、一般的なテレビ電話環境においてテレプレゼンスが増強可能になると考えられる。

2. 先行研究

近年、運動視差によるプレゼンス強化を試みた研究がなされてきている。Nakanishiらは、メディアスペースにおいて、相手側に設置してあるカメラを、そのカメラの光軸方向に物理的に動かすことが可能な環境を構築し、そのカメラの映像が従来の固定されたカメラよりもテレプレゼンスが増さっていることを示した¹⁰⁾。この環境では、ユーザは自分自身の能動的な操作によりカメラを移動させていたが、加藤らは、測距センサーを用いることにより一方のユーザの前後方向の動きに合わせて、相手方にあるカメラが前後方向に動く環境を作り、このような能動的な操作がない状態で与えられる運動視差がテレプレゼンスに及ぼす影響を調べた⁹⁾。その結果、ユーザの動きと相手側のカメラの動きが同期していればこのような運動視差はテレプレゼンスを増強させるということを示した。ただし、物理的にカメラを動かす装置やスペースが必要となるため、たとえばテレビ電話のような環境に簡単には導入しにくい。

一方、石井らは、運動視差映像コミュニケーションシステム、MoPaCo(Motion Parallax Communication video system)とよばれるシステムを提案している¹¹⁾。これは、通常のテレビ電話環境におけるユーザ同士が、あたかも同室内で対面しているかのような臨場感の高い映像コミュニケーションシステムの実現を目指すものであり、画像処理によって特別なデバイスが必要とせずソフトウェアだけで運動視差を模した映像をリアルタイムで生成するアプローチをとっている。ただし、加藤らの環境では、カメラは前後方向の運動視差を扱っているのに対して、MoPaCoでは特に明記はされていないが、横方向の運動視差を提供することを想定しているようである。実験では、通常のテレビ電話とMoPaCoを用いた条件で比較がおこなわれ、映像処理による画像の破綻や画質の低下は特に問題にならず、対面感



図 1 Kinect のカメラ映像と深度センサの出力

や実在感が増したという結果が報告されている．本研究は物理的な移動物体がなくリアルタイムな画像処理を施すという点で MoPaCo と類似しているが，MoPaCo では背景があらかじめ撮影された画像を用いることが想定されており，オンタイムな背景が表示できない欠点がなく，また距離センサーを用いるため映像セグメンテーションの頑健性でも勝ると考えられる．

3. 提案システム

本研究で提案する視差を模した映像は，通常の RGB ビデオカメラ（以下，カメラとする）とカメラから被写体までの距離を測定する深度センサを入力として生成される．深度センサはセンサ側から照射した赤外光が物体に反射してセンサに戻るまでの時間計測による所謂 TOF（Time-of-Flight）方式や，赤外光のプロジェクタからパターンを照射して，赤外カメラによりそのパターンの大きさを測定する Light Coding 方式等がある．このうち，Light Coding 方式のセンサは非常に安価に入手することが可能になったため普及しており，たとえば Microsoft 社の Kinect はカメラ付きの Light Coding 方式の入手が容易なセンサである．本研究ではこれを用いて開発・実験をおこなった．図 1 は，Kinect で撮影したカメラの映像と深度センサの出力である．深度センサの出力は 2 次元のグレースケールの 1 チャンネル画像として獲得でき，その画素値の大きさによって各画素位置での深度を把握することができる．

提案手法は，本質的には MoPaCo と同様に画像を被写体領域と背景領域に領域分割をおこない，それぞれを別のレイヤとして適切に加工して合成した画像を生成する．MoPaCo

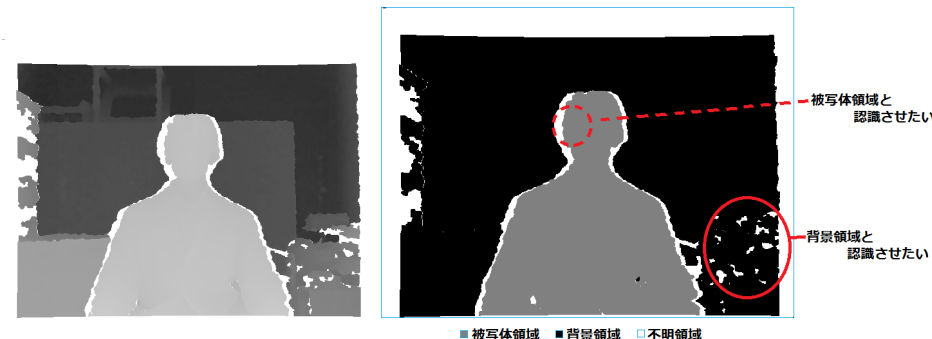


図 2 人が映っている深度情報 (左) と三値化した深度情報 (右)

と異なるのは，本研究が生成対象としているのは前後方向への運動視差を模した映像であり，先の適切な加工とは，即ち各レイヤの拡大率を前後運動をした場合のそれに設定することになる．

- (1) カメラ画像 f を被写体領域と背景領域に分割する．
- (2) カメラ画像を 2 種類の拡大率 α, β で拡大し，2 種類の画像 f_a, f_b とする．
- (3) f_a, f_b の拡大前の画素位置が被写体領域であれば f_a ，そうでないならば f_b とした画像を出力する．

拡大率 α, β は，それぞれカメラから被写体とカメラから背景までの距離，および閲覧者の初期位置からの前後移動した距離に依存して決定される．ここで利用される閲覧者の初期位置と現在位置は，閲覧者側に設置した深度センサの出力から推定することができる．

また，MoPaCo では背景差分を用いて領域分割をおこなっていたが，提案手法では深度センサをもちいて領域分割をおこなう．つまり，被写体はカメラに近い領域に存在するという前提から，各画像における画素位置に相当する深度センサの出力が比較的近接している画素を被写体領域とラベル付けし，そうでない領域を背景領域として設定する．

しかし，実際には各画素位置における深度は必ずしも安定して獲得することはできない．これは，以下のようなセンサの制約に起因する．

- (1) カメラと深度センサ間の視差
- (2) パターンを照射するプロジェクタと深度センサ間の視差
- (3) パターンをうまく反射できない箇所の存在

よって，実際には深度センサによる領域分割は，被写体領域，背景領域，不明領域の 3 種



図 3 カメラからの映像 (左) と背景画像 (中) と明度を合わせた背景画像 (右)

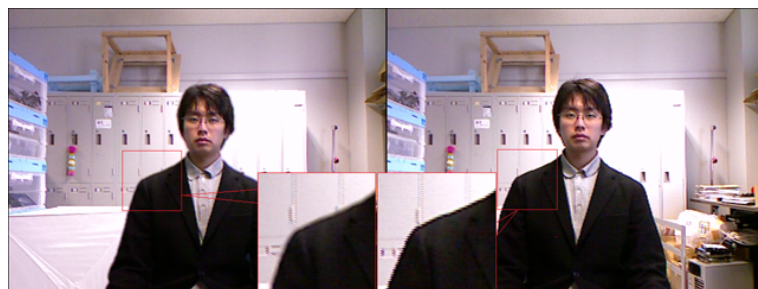


図 4 フィルタを掛けた後 (左) と掛ける前 (右)

類が存在することになり、この不明領域が被写体なのか背景なのかを推定する必要がある。この推定は、あらかじめ被写体がない状態で背景だけの画像を取得しておき、これと不明領域との画素毎の差分によって推定する。すなわち、不明領域において背景差分した結果が一定値以内であれば背景領域、それ以外は被写体領域として推定する。

しかし、被写体がない場合の映像と被写体がいる映像では、同じ箇所を映した背景部分でも AGC の働きにより輝度が大きくことなる。そこで、現在のフレームと背景として撮影した画像の輝度をあわせる処理をした上で背景差分をおこなう。

このように被写体領域と背景領域に分けて拡大することで目的の映像を生成することが可能であるが、領域同士の境目は不自然さが目立つ。そこで、境目の部分にだけ平均化フィルターをかけると見た目が自然になる (図 4)。

以上の処理はすべて画素毎に独立して処理することができるため、並列化可能である。実験でもちいたプロトタイプシステムでは CUDA を用いて実装し、Core2Duo2.4GHz の計算機で 640*480 画素の解像度の入力映像に対して 30fps 以上で処理することができた。

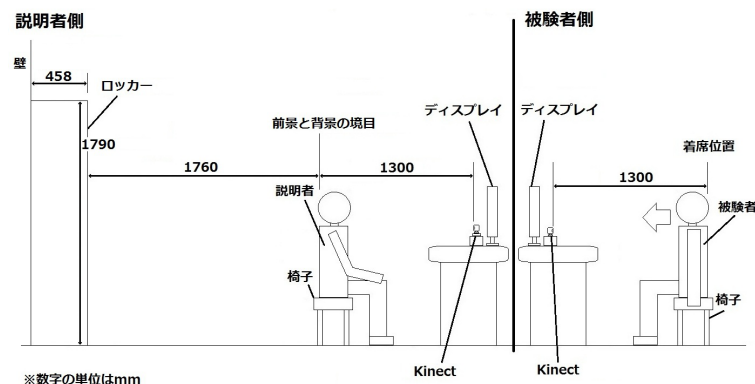


図 5 実験環境

4. 実 験

4.1 実験概要

提案した手法を用いて生成された映像がテレプレゼンスを強化するかどうかを検証する実験をおこなった。実験は 2 台の PC と Kinect、および音声チャット用のデバイスを図 5 のように設置した環境でおこなった。

この実験は、先行研究⁹⁾¹¹⁾の実験内容と結果に基づいて設計した。まず、被験者と説明者を IP 電話によって接続し、24 インチのディスプレイにフルサイズで説明者の映像を表示した。説明者は、被験者にお城の説明をするむねを伝え、手に持ったお城の写真の説明をする。被験者は、初期状態では椅子に深く腰掛けた状態であるが、説明者に説明される前に画面を見続けたまま少し身を乗り出し、すぐにまた元の姿勢に戻ってもらった。初期状態においてディスプレイから被験者までの距離は 1100mm であったが身を乗り出した場合は 300mm 程度近づく恰好になった。1 つの実験につきお城の説明を 3 種類おこない、一回の説明につき一回身を乗り出す作業をしてもらった。説明は、質疑応答も含むため一方的なコミュニケーションではなく TV 電話的な双方向のコミュニケーションになるように心がけた。この実験を以下の 5 種類の映像に対しておこない、それぞれの説明の後にアンケートに答えてもらった。映像 2 と映像 4 は、図 5 にある距離から算出した倍率で拡大した映像であるが、映像 3 と映像 5 はそれよりも高倍率の映像とした。



図 6 実験映像の比較

- 映像 1 : 生の映像
- 映像 2 : デジタルズームの映像 (通常倍率)
- 映像 3 : デジタルズームの映像 (高倍率)
- 映像 4 : 提案手法の映像 (通常倍率)
- 映像 5 : 提案手法の映像 (高倍率)

映像 2, 3 のデジタルズームとは、被写体の頭部位置にあわせて単にバイリニア補間法で拡大をした映像であり、運動視差はないが拡大はする映像である。ただし、倍率が結果に強く影響するかもしれないため、通常倍率とそれよりもズームの倍率が高めた二種類を設定した。提案手法も倍率の影響をしらべるために倍率をかえた二種類を見せることにした。それ

ぞれの映像は被験者によってランダムな順で選び、カウンターバランスをとった。アンケートは以下項目を設定し、各項目に対して 9 段階のリッカート尺度を用いた。

- 映像は十分きれいだと感じた
- 音声は会話を行う上で問題が無かった
- お城の説明は良く分かった
- 実際に机越しに対面して会話している感じがした
- 説明者を机越しに眺めている感じがした
- 説明者から机越しに眺められている感じがした
- 机越しに対面している距離感を感じた

4.2 結 果

各映像のアンケート結果の平均値を図 7 に示す。これによると、映像の綺麗さでは映像 1 やデジタルズームが有利であり、提案手法の映像はやや問題のある映像として捉えられたようだ。しかし、説明に関してはどの手法も高得点をあげていることから、資料を参照しながらコミュニケーションをとる程度には提案手法も綺麗であったようである。一方、プレゼンスに関する質問では高倍率な提案手法の映像が最も高い得点を得た。運動視差つきの映像が高得点を得ることは先行研究で報告されている結果と矛盾しないため、本提案手法が運動視差を模した映像として機能していたことを示唆している。しかし、通常の倍率の運動視差付き映像よりも高倍率なデジタルズームのほうが高い得点を獲得する結果となっているため、同じ倍率であれば運動視差付き映像のほうが良いが、デジタルズームでもそれなりに高倍率にすればテレプレゼンスを高める結果となるのかもしれない。ただし、今回は被験者が少ないため、上記結論は評価実験を十分に重ねたうえで結論付ける必要がある。

5. ま と め

本稿では、カメラと深度カメラの情報を統合することによって前後方向の運動視差を模した映像を生成する手法について述べた。本手法は、画像を被写体と背景に領域分割し、それぞれの倍率を適切に設定して拡大し、合成することで頑健に運動視差付きの映像生成が可能である。この提案手法をもちいて被験者実験を行った結果、やや映像生成の精度の点で問題があるものの高倍率な拡大率ではテレプレゼンスを高める結果になることが示唆された。今後は被験者を増やした実験と、映像の高精度化に取り組みたい。

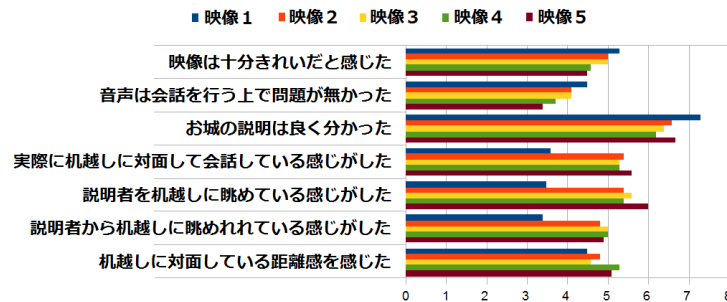


図7 アンケート結果

謝 辞

研究初期の段階で有益な示唆を頂戴した大阪大学工学研究科中西 英之准教授に感謝する。

参 考 文 献

- 1) Bly, S.A. and Harrison, S.R. and Irwin, S.: Media spaces: bringing people together in a video, audio, and computing environment, *Communications of the ACM*, pp. 28-46 (1993).
- 2) Jancke, G. and Venolia, G.D. and Grudin, J. and Cadiz, JJ and Gupta, A.: Linking public spaces: technical and social issues, *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp.530-537 (2001).
- 3) Fish, R.S. and Kraut, R.E. and Chalfonte, B.L.: The VideoWindow system in informal communication, *Proceedings of the 1990 ACM conference on Computer-supported cooperative work*, pp.1-11 (1990).
- 4) Heath, C. and Luff, P.: Media space and communicative asymmetries: Preliminary observations of video-mediated interaction, *Human-Computer Interaction*, Vol.7, No.3, pp.315-346 (1992).
- 5) Heath, C. and Luff, P.: Disembodied conduct: communication through video in a multi-media office environment, *Proceedings of the SIGCHI conference on Human factors in computing systems: Reaching through technology*, pp.99-103 (1991).
- 6) Cutting, J.E. and Vishton, P.M.: Perceiving layout and knowing distances: The integration, relative potency, and contextual use of different information about depth, *Perception of space and motion*, Vol.5, pp.69-117 (1995).

- 7) Towles, H. and Chen, W.C. and Yang, R. and Kum, S.U. and Kelshikar, H.F.N. and Mulligan, J. and Daniilidis, K. and Fuchs, H. and Hill, C.C. and Mulligan, N.K.J. and others: 3d tele-collaboration over internet2, *International Workshop on Immersive Telepresence, Juan Les Pins*, (2002).
- 8) 末永剛 and 松本吉央 and 小笠原司: 非拘束な運動視差提示 3次元ディスプレイの提案と評価, *ヒューマンインタフェース学会論文誌*, Vol.9, No.2, pp.49-56 (2007).
- 9) 加藤慶, 村上友樹, 中西英之: 可動式カメラによる社会的テレプレゼンスの強化, *情報処理学会論文誌*, Vol.52, No.4, pp.1635-1643 (2011).
- 10) Nakanishi, H. and Murakami, Y. and Nogami, D. and Ishiguro, H.: Minimum movement matters: impact of robot-mounted cameras on social telepresence, *Proceedings of the 2008 ACM conference on Computer supported cooperative work*, pp. 303-312 (2008).
- 11) 石井亮, 小澤史郎, 向内隆文, 松浦宣彦: MoPaCo: 単眼カメラを用いた運動視差映像コミュニケーションシステム (複合現実感, 仮想都市), *電子情報通信学会技術研究報告. MVE, マルチメディア・仮想環境基礎*, Vol.110, No.238, pp.73-78 (2010).