

Behind-the-Mask: 顔映像が取得可能な頭部搭載型ディスプレイ

暦本 純一^{*†} 浦垣 啓志郎[‡] 山田 献二郎[§]

概要. ヴァーチャルリアリティ(VR)で標準的に使われている頭部搭載型ディスプレイ(HMD)は、装着者の顔を隠してしまう。このため、VR空間を介しての他人とのコミュニケーションや、サロゲートロボットへの接続などの場合に、装着者本人は映像を見ることができ、装着者自身の顔を見ることができない。テレビ会議では容易に行えるような対面コミュニケーションがHMDを使うと達成できない。この問題を解決するため、本研究では赤外線反射ミラーを利用した光学系によりHMD装着中の利用者の顔を再現する機構を提案する。表情形成に重要な目周辺と口周辺の映像を取得し、それを利用者の頭部三次元モデルと結合する。HMDレンズ手前に配された赤外線反射ミラーは可視光を透過するので、装着者の視界の変化はないが、赤外光に対しては鏡として機能するのでHMD左右に配されたカメラで左右の目の周辺映像を取得できる。この機構は従来のHMDに装着することも可能であり、視線や表情の認識にも使用でき、多くのVRアプリケーションに適用可能である。

1 はじめに

バーチャルリアリティ(以下VR)は、コンピュータにより生成された人工的な世界を、あたかも現実であるかのように提示し操作させる技術である。VRの応用分野として、コミュニケーションを拡張するシステムが多く提案されている。たとえば臨場感通信システムでは、複数の参加者が仮想空間を共有し、仮想空間内での共同作業を実現する[11]。また、VRによるテレプレゼンス応用では、サロゲートロボットが遠隔地におもむき、周辺の環境情報を取得する。取得した情報をVR空間中に再現すれば、参加者が仮想的に遠隔地に没入し、その場で起きていることを体験するシステムが構築できる。さらに遠隔地の他の参加者とコミュニケーションをはかることができる。

これらのシステムを構築する上での課題となるのが、VR参加者の顔情報取得である。VRシステムで一般的な頭部搭載型ディスプレイ(HMD)を装着すると、利用者の顔が隠されてしまう。テレビ会議の基本機能である遠隔地との対面型コミュニケーションが、HMDの装着により不可能となる。CAVE[3]のようなスクリーン型VRシステムでも、利用者が液晶シャッターグラスを装着するので顔映像を取得することができない。サロゲートロボットを介して遠隔地に没入する場合も、没入する利用者の顔映像が取得できないと、「誰が」そのサロゲートロボットに入っているのかを伝えることが難しい。すなわち、**VR(HMD)と対面型コミュニケーションは両立し**

ないことが現状のVRにおける大きな課題である。

コミュニケーションを直接の目的としないVR応用でも、参加者の顔が隠されてしまうことは問題となる。VR参加者がどんな世界を体験することが外部からはわかりにくい。最近ではVR参加者を取り囲む環境そのものを外部スクリーンに表示し、そこに参加者の姿も重畳表示する試みが増えてきている。しかし、参加者の顔がHMDによって隠されているので、参加者がシステムに対してどんな反応をしているのかを表情から知ることができない。

この課題を回避する一つの方法として、利用者の顔映像そのものではなく、アバターのようCGキャラクターを用いることが考えられる。後述するように、アバターの表情を、利用者の顔からセンサーにより取得した情報を利用して変形すると、参加者の感情などある程度までは表現することができる。しかし、通常のテレビ会議でアバターを使うことが一般的ではないように、対面型のコミュニケーションの代わりとなるわけではない。

以上の問題は、HMDを装着しても顔画像を取得できる方法があれば解決できる。

Liらは、HMD内部に配した複数の伸縮センサー(ストレインゲージ)から顔表面の変形を推定し、顔形状の復元を試みている[9]。この方法では顔の3次元モデルは復元できるものの、眼球の方向(視線情報)など、皮膚の変形とは直接関係ない情報を復元することはできなかった。

Kimuraらは、顔周辺に配した複数のカメラにより顔映像を復元するメガネ型ディスプレイを提案している[8]。この方式では顔の周囲に7個のカメラを配し、それらから得られる映像を合成して顔映像としている。メガネ型という構造からカメラの設置位置に大きな制約があり、どのカメラからも顔の正

Copyright is held by the author(s).

* 東京大学

† ソニーコンピュータサイエンス研究所

‡ 東京工業大学

§ 名古屋大学

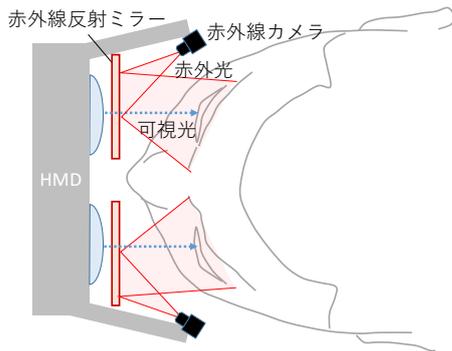


図 1. Behind-the-Mask 構成

面映像が取得できない。その結果、得た画像を大きく歪ませる必要があり、再現される顔映像の品質は満足できるものではなかった。

Vesso は、HMD 内に設置したカメラにより口と眼球周辺の画像を撮像し、それを表情パラメータに変換する HMD の (クラウドファンディングによる) 提案である¹。HMD 内に設置されたカメラと顔との距離をとることができず、眼球周辺の一部映像しか撮像することができない。この制限により、得られた画像はアバター表情の制御に限定されており、顔映像そのものの再現は達成していない。その他、ストレインゲージにより簡易的に表情を取得する提案 [4] などがあるが、いずれも本研究で目指しているように顔映像そのものを再現するものではない。

これらの既存研究で未解決だったのは、HMD で隠されている顔画像をいかに取得するかという方法であった。本研究では、カメラと顔表面までの距離を保つために、赤外線反射ミラーを利用する機構を提案する。この構成により、利用者が可視光である VR 映像を通常どおり観測しながら、HMD 側面に配されたカメラにより顔映像を取得することができる。

2 Behind-the-Mask システム構成

提案システムの構成を図 1 に示す。本構成の特徴は、HMD に組み込まれたカメラである。両眼側面に配した赤外線カメラと、HMD のレンズ直前に配した赤外線反射ミラーにより、両眼付近の顔映像を撮像する。この映像と、別に撮像した顔下半分 (HMD に隠されていない部分) とを、あらかじめ計測された利用者の顔三次元形状とを合成して、利用者の顔映像を再現する。

2.1 顔形状と顔テクスチャ取得

装着者の顔の三次元形状と顔画像を計測しておく。表情は目、眉付近と口元の変形によって大きく変化

¹ www.kickstarter.com/projects/veeso/veeso-the-first-face-tracking-virtual-reality-head 論文執筆時点ではクラウドファンディングはキャンセルされている。



図 2. 撮像された三次元顔形状 (左) およびテクスチャ画像 (中). (右) はテクスチャが連続するように再配置したもの。



図 3. 顔画像取得のためのカメラと赤外線反射ミラー

するが、それ以外の部分はあまり変形しないという考えに基づき、この形状と、HMD に搭載されたカメラからの映像を合成して表情を生成するという方針をとる。

三次元計測には ARTEC 社のハンドヘルド型スキャナおよび計測ツール ARTEC Studio を用いた。この計測システムは、対象となる三次元形状の周辺を撮像し、複数の計測データを自動的にステッチングして全体形状とする。また、同時に対象物の二次元画像も撮像し、テクスチャ画像として保存する (図 2)。撮像された三次元形状とテクスチャ画像は .obj 形式で保存される。計測された生の三次元形状からポリゴン数を削減する。スキャナが生成するテクスチャ画像は図 2(中) のように連続していないので、頭部を取り囲む円筒形に投影したものをテクスチャ画像として保存する (図 2(右))。この処理は利用者あたり一回行えばよい。

2.2 顔画像撮像

赤外線反射ミラーとして、本研究ではセラテック社の赤外線カットフィルター IRC4-76A を、HMD 内に設置できるように切断加工したものを利用している。このミラーは 750nm より長い波長 (近赤外領域) は鏡面反射し、それよりも短波長 (可視光) 領域は透過させる。ミラーを HMD レンズ前に添付することで、装着者は VR 映像を通常どおり見ることができると同時に、側面のカメラにより顔映像を取得することができる (図 3)。

側面のカメラは可視光カットフィルターを装備したもので、HMD 内に赤外線 LED を配し、顔表面が均等に照明されるようにしている。HMD 内部にカ



図 4. 取得された画像



図 5. 撮像映像の顔テクスチャ画像への貼りこみ (左: グレースケール画像 右: カラー化処理をほどこしたものの)

メラを組み込む場合と比較して、ミラーで折り返す分、カメラから顔表面への距離が長く、表情取得に重要な眼球と眉毛を含む顔領域を撮像することができる。また、側面方式では既存のHMDにカメラを組み込むことが可能であることも特徴である。本研究ではHTC社のVIVEに装着して実験している。

図4に撮像された画像例を示す。赤外線ミラーと両眼脇のカメラによる構成は、筆者らが以前に開発した眼鏡型の視線認識装置[7]を参考にしている。本構成でも、顔表情の取得に加えて、撮像した眼球映像を解析して視線情報および瞬目などの眼球情報を認識することが可能であり、視線応用VRの撮像装置として使用できる。

2.3 動画の変形処理

取得した顔画像は顔の斜め方向から撮像されたものとなり、歪がある。また、顔画像と合成する顔テクスチャは、図2で示したように、顔の三次元形状に貼り込めるように展開したもので、目や眉の位置があうように撮像画像を変形する必要がある。このずれは単純なアフィン変換やホモグラフィ変換では補正できない。

そこで、撮像画像を非線形に変形させて顔テクス

チャに合わせる処理を行う。具体的には、Beierらがイメージモーフィング用に開発したfeature based image metamorphosis手法[2]を利用する。この手法では、複数の特徴線を顔テクスチャ画像とカメラ画像上に定義し、特徴線が対応するように画像を非線形に変形させる。変形した動画画像を顔テクスチャ画像と合成し、最終的な顔テクスチャとする。本研究の場合は、撮像用のカメラと顔の位置関係は一定なので、モーフィング用の補助線は一度設定すればよい。

モーフィングの実装にはOpenGLのフラグメントシェーダー(GLSL)を用いた。動画画像の各画素ごとに上記の変形処理を行う必要があるが、シェーダーで実現したことにより、処理はGraphics Processing Unit(GPU)上で高並列に実行される。左右の眼付近のビデオ映像、口付近のビデオ映像の3画像について変形処理を行い、テクスチャと合成する処理性能は50フレーム/秒以上であり、入力ビデオのフレームレート(30fps)に対して十分な速度である。

今回の試作では、上記のような特徴線の対応を手作業で行っているため、HMDの装着中のずれが顔画像の画質に影響を与えてしまう。Huberらの顔画像処理[5]で用いられているような局所特徴量による自動対応を検討中である。

2.4 着色

近赤外線カメラによって取得された画像は単色なので、顔のテクスチャに合わせるためにはカラー化する必要がある。3次元スキャナで顔形状を計測する際に同時に取得された顔テクスチャはカラー画像なので、この画像と色合いと違和感がないように取得映像を着色する必要がある。

これを解決するため、白黒画像自動カラー化フィルターを適用する。カラーである顔テクスチャもいったんグレースケール化し、(もともとグレースケールである)赤外線画像と輝度がなめらかに接続されるように調整し合成する。このようにして得たグレースケールの顔テクスチャ動画画像に対して、カラー化フィルターを利用して色情報を復元する。

カラー化フィルターには多く研究事例があるが、本研究ではCNN(convolutional neural networks)に基づくIizukaらの手法[6]を利用している。この手法では、まずカラー画像サンプルをグレースケール化して大量のカラーとグレースケール画像の学習セットを準備する。この学習セットにより、グレースケールを入力し、それと対応するカラー画像を出力するようにニューラルネットを学習させる。学習済みのニューラルネットを、上述の方法で得た遠赤外線によるグレースケール画像に適用し、カラーの顔テクスチャ画像を得る。

図5(右)はこのようにしてカラー化した顔テクスチャ画像である。肌色や唇の色が再現されているの



図 6. 顔画像生成例

が確認できる。現状の学習セットは一般のカラー画像とそのグレースケール化画像である。顔テクスチャ画像に特化したデータにより学習することで、さらに自然なカラー化が期待できる。

この方式の現状の大きな課題は速度である。[6]で公開されているプログラムをそのまま用いると、変換速度は 224×224 の画像で約 5fps であった。画素数を増やすとさらに処理時間を要し、その解像度に対応した学習セットも必要となる。本研究では、顔テクスチャを縮小しカラーに変換したのち、その色成分のみを元の顔テクスチャ画像に適用している。顔テクスチャ画像を 224×224 のグレースケール画像に変換し、上記のフィルタにより同一サイズのカラー画像を得る。この画像の各ピクセルを YUV 変換する。YUV の Y 要素は輝度、U と V が色情報なので、もとの顔テクスチャ画像の Y 成分はそのまま使用し、UV 成分と結合する。この方法で、輝度情報はオリジナルの顔テクスチャ画像と同じ解像度で、色情報のみ低解像度の画像を得ることができる。

3 結果と応用可能性

以上を統合して再構成された 3 次元顔を図 6 に示す。三次元形状取得時に顔を閉鎖してしまったため目付近の三次元形状が乱れているが、提案方式により、HMD を装着していても利用者の顔映像を取得できることが確認できた。本方式は多くの VR 対面コミュニケーションに適用可能だと考えている。また、以下の応用を検討している。

サロゲート顔への適用: 遠隔コミュニケーション支援のため、遠隔地に参加者の顔ディスプレイや顔の型をとった整形スクリーンを置くことを考える [1, 10]。参加者の顔を表示し、首の動きに追従してスクリーンを駆動することで、表情や視線などの非言語情報が伝達できる。従来は参加者側は通常のディスプレイで遠隔地を観測し、顔表情をカメラで取得していた。本研究を適用すると、参加者 HMD が装着して遠隔地の環境に没入でき、遠隔地の環境をより容易に把握することができる。

表情/視線の取得: 本方式では、顔映像にもとづいて使用者の視線や表情の認識が可能であると考えている。表情や視線情報により、VR 空間内の事物

を視線で指定したり、利用者の興味などを推定することが可能になる。

4 結論

本研究では、装着者の顔映像を取得可能な HMD の構成を提案し、試作システムについて報告した。コンピュータアプリケーションの多くは何らかの意味で他者とのコミュニケーションを支援するものである。HMD 装着者との対面コミュニケーションが可能になれば、更に多くの VR 応用システムが構築可能であると期待している。

参考文献

- [1] Michael Naimark, Talking Head Projection. <http://www.naimark.net/projects/head.html>.
- [2] Thaddeus Beier and Shawn Neely, Feature-based Image Metamorphosis, SIGGRAPH '92, pp. 35–42, 1992.
- [3] Carolina Cruz-Neira, Daniel J. Sandin, Thomas A. DeFanti, Surround-screen Projection-based Virtual Reality: The Design and Implementation of the CAVE, SIGGRAPH '93, pp. 135–142, 1993.
- [4] Kurara Fukumoto, Tsutomu Terada, and Masahiko Tsukamoto, A Smile/Laughter Recognition Mechanism for Smile-based Life Logging, AH '13, pp. 213–220, 2013.
- [5] P. Huber, Z. Feng, W. Christmas, J. Kittler, and M. Ratsch, Fitting 3D Morphable Models using Local Features. IEEE International Conference on Image Processing (ICIP) 2015, 2015.
- [6] Satoshi Iizuka, Edgar Simo-Serra, Hiroshi Ishikawa, Let There Be Color!: Joint End-to-end Learning of Global and Local Image Priors for Automatic Image Colorization with Simultaneous Classification. ACM Trans. Graph., 35(4):110:1–110:11, July 2016.
- [7] Yoshio Ishiguro and Jun Rekimoto, GazeCloud: A Thumbnail Extraction Method Using Gaze Log Data for Video Life-Log, ISWC '12, pp. 72–75, 2012.
- [8] Shinji Kimura, Masaaki Fukuomoto and Tsutomu Horikoshi, Eyeglass-based Hands-free Videophone, ISWC '13, pp. 117–124, 2013.
- [9] Hao Li, Laura Trutoiu, Kyle Olszewski, Lingyu Wei, Tristan Trutna, Pei-Lun Hsieh, Aaron Nicholls and Chongyang Ma, Facial Performance Sensing Head-mounted Display. ACM Trans. Graph., 34(4):47:1–47:9, 2015.
- [10] Kana Misawa, Yoshio Ishiguro and Jun Rekimoto, LiveMask: a telepresence surrogate system with a face-shaped screen for supporting nonverbal communication, AVI '12, pp. 394–397, 2012.
- [11] Haruo Takemura and Fumio Kishino, Co-operative Work Environment Using Virtual Workspace, CSCW '92, pp. 226–232, 1992.

WISS2016 採録判定時のコメント

採録区分：ショート採録

判断理由：

ヘッドマウントディスプレイ着用者の顔表情取得手法の提案です。提案手法（顔面撮影用の小型赤外線カメラと赤外線反射ミラーをHMD内に設置）の新規性・妥当性や実装手法は、ショート採録（本論文は4ページです）に十分であると評価されました。一方、リミテーションや他の技術に対する優位性については加筆を要する部分もありますので、フルペーパー化の際の参考にしてください。

レビューサマリ：

全体の構成：

本論文は、ヘッドマウントディスプレイ（HMD）を装着した際に、装着者の顔が隠れ、表情などを読み取ることができないという問題に対して、HMDに装着された赤外線カメラおよび赤外線反射ミラーを用いて顔映像を取得、予め撮影された顔モデルと合成することによってHMD装着者の顔映像を再現するものです。

論文は序章を除いて主にシステム構成や実装、実行結果等の技術的な部分にフォーカスして構成されています。アイデアはシンプルですが、実現の際に問題となる点（画像変形やグレースケールカラー変換）についても整理して述べられており、理解性も高い論文になっています。

改良のためのコメント等：

・技術的なリミテーションや他の技術に対する優位性について説明が不足しています。例えば、眼球撮影用に赤外線カメラや赤外線ミラーを用いることはEye Trackerにおいては一般的なので言及すべきでしょう。

参照すべき文献等：

・遠隔協調型複合現実感における目領域映像の重畳提示 岡本、北原、大田 (2009)

<http://jglobal.jst.go.jp/public/20090422/200902275156935779>

・協調型複合現実空間のためのアイコンタクトの復元：視線認知実験による評価 岡本、大田 (2004)

<http://ci.nii.ac.jp/naid/110003273899>

本論文に対する各査読者の詳しいコメントは以下のページを参照のこと：

<http://www.wiss.org/WISS2016Proceedings/oral/23.html>

***本ページは論文本体ではありません**

【Behind-the-Mask: 顔映像が取得可能な頭部搭載型ディスプレイ】