J-051

メガネ型端末を用いたハンズフリービデオフォンの開発

Development of Eyeglass-based Hands-free Videophone

木村 真治†　　堀越 力†
Shinji Kimura　　Tsutomu Horikoshi

1. まえがき

携帯電話によるテレビ電話では，端末を自分の手で持ち，インカメラを自分に向けて撮影しながら，通話相手の顔を画面上で見るのが一般的な利用スタイルである．しかし，この利用スタイルでは常に片方の手が塞がってしまう．また，歩行時などモバイル環境における利用では，自分の姿がカメラのフレームに常に収まるよう撮影することは困難であり，長時間利用する場合には，端末を保持する腕の疲れを誘発する．更に，携帯電話は画面サイズが小さいため，その画面上では通話相手の表情変化が分かりづらい．このように，従来の携帯電話におけるテレビ電話では，お互いの顔を見ながらコミュニケーションを円滑に行う上での課題が多くあった．

また，昨今 Google Project Glass[1]に代表されるように，メガネ型端末に関する開発や商用化が進んでいる．我々も，スマートフォンに代わる将来的な携帯電話は，普段身に付けるメガネやイヤホン，アクセサリ等のような装着型のデバイスによって各種機能が実現されることを想定しており，これをウェアラブル(装着型)端末コンセプトとして掲げている．このコンセプトでは，端末自体は情報の入出力のみを行い，各種データ処理・管理はモバイルネットワークに接続されたクラウド側で全て処理することを想定している．また，ウェアラブル端末には，スマートフォンなどのハンドヘルド型デバイスと異なり，手が塞がれない(=ハンズフリー)という特徴もある．ウェアラブル端末の中でもメガネ型の端末は，ヘッドマウントディスプレイ（HMD：Head Mounted Display）を備えることで，眼前で大画面の映像提示が可能という特徴があり，拡張現実（AR：Augmented Reality）アプリケーションなどへの応用が期待されている．

本論文では，メガネ型端末の特徴である“ハンズフリー”と“大画面での映像提示が可能”な 2 点を活かして，前述のテレビ電話の課題を解決するハンズフリービデオフォンを提案する．このゴールイメージを図 1 に示す．メガネ型端末を装着したユーザは自分の姿を撮影する(=自分撮り)ためのカメラを保持する必要なく，自分の姿を通話相手に送ることができ，また，通話相手の映像を HMD 越しに大画面で見ることができる．

2．メガネ型端末での自分撮り

メガネ型端末上に搭載されたカメラは，自分の視点に近い画像を撮影する外向きのものがほとんどであり，常時自分の視点画像を記録するライフログや，AR を利用したアプリケーションでの利用が主に想定されている[2]．一方，自分撮りを行うためには，メガネ型端末上に内向きのカメラを搭載して自分の姿(この場合，特に顔)を撮影する必要

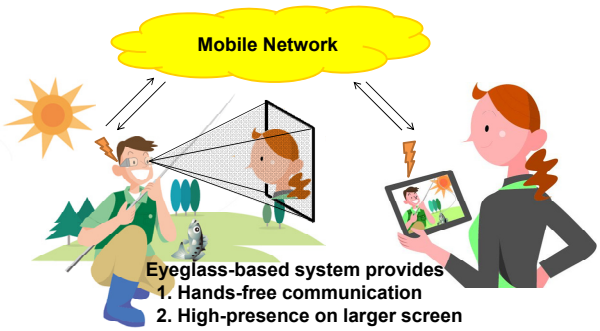


図 1. Goal of hands-free videophone

がある．しかしながら，顔とメガネ型端末上のカメラは非常に近い距離にあるため，目など，顔の一部分のみしか撮影することができない．装着者の視線を検出することでメガネ型端末の入力操作とする[3]場合にはこれで十分であるが，通話相手に自分の姿を送る本目的のためには顔の一部では無く，大部分を撮影する必要がある．メガネでは無く，ヘルメットにカメラを取り付け，顔から一定距離を離すことで顔の大部分を撮影するシステム[4]もあるが，常に自分の眼前にカメラがあることになり，日常での利用には不向きである．これを解決する方法として，凸面鏡と耳かけカメラを用いた手法[5]が提案されているが，これには見た目上の問題がある．メガネから凸面鏡が飛び出しているので，装着者自身の視界を邪魔するだけではなく，通常のメガネと異なる見た目となるため周囲の人へ違和感を与える結果となる．

これらを考慮して，自分撮りを可能とするメガネ型端末の設計において必要となる要件を以下の 4 つとした．

A)　通常のメガネに近い見た目
B)　装着者の視界を邪魔しない
C)　装着者の顔をできるだけ広く撮影する
D)　全周囲画像の取得を可能とする

要件 D)に関しては，メガネ型端末をハンズフリービデオフォン用の端末とするだけで無く，将来的にライフログや AR などの他のアプリケーションへの応用を可能とするために加えた要件である．この 4 つの要件を満たすために，本システムでは小型魚眼カメラをメガネのフレーム上に複数取り付けることとした．実際にメガネフレームのヒンジ部の内側（顔向き）及び，下部（下向き）に取り付けられた魚眼カメラで撮影された画像を図 2 に示す．一般に魚眼カメラは画角が 180° 程度あり，更に，非常に近い距離でも焦点が合う特徴を持っている．図 2 からも，顔に非常に近い距離にカメラがあるにも関わらず，顔の大部分が焦点ボケの無い状態で撮影されていることがわかる．複数の魚眼カメラを用いて，1 つの画像を合成する技術としては，自動車の周囲環境取得システム[6]が代表的である．これと同様に，本システムでは複数の魚眼カメラで自分の顔，手元，背景などを撮影し，それらの画像を合成することで，自分

†（株）エヌ・ティ・ティ・ドコモ　先進技術研究所

図 2. Captured images by fish-eye cameras mounted on eyeglasses

を向いているカメラがあたかも前方にあるような自分撮り画像を，ハンズフリーで生成することが可能になると考えた．

また，[4][5]による装着者自身の表情再現は，撮影画像から表情の動きを検出し，予め用意した装着者の CG 顔モデルに表情の動きをパラメータとして反映させる方法をとっている．しかしながら，パラメータとして動きを CG モデルに反映させる手法では，目・眉周辺のシワや微妙な表情の変化を再現することが困難である．テレビ電話におけるコミュニケーションでは，この微妙な表情変化を伝えることが重要な要素であり，CG モデルで動きを再現するだけでは不十分であると考えた．そこで，本システムでは，できるだけ実写画像を用いることで，写実的な自分撮り画像を合成する構成とした．

3. プロトタイプ

3.1 システム構成

前述のメガネ型端末設計の必要要件にもとづき，複数の魚眼カメラを搭載したメガネ型端末のプロトタイプを試作した．プロトタイプ自身と，プロトタイプを実際に装着した際に各カメラで撮影される画像を図 3 に示す．本プロトタイプには 7 個の魚眼カメラ(上・下・内向き×左右 ＋背景)，6 軸傾きセンサ，マイク，イヤホンが搭載されており，各モジュールはプロトタイプの後頭部にある USB ハブを介して PC に接続されている．図 3 から，内向きカメラで顔の大部分が，また，上下及び背景カメラによって装着者の全周囲画像が撮影できていることが分かる．搭載された魚眼カメラの仕様を表 1 に示す．この魚眼カメラは，プロトタイプに搭載可能な小型サイズでありながら，180°超の画角での撮影を可能としている．

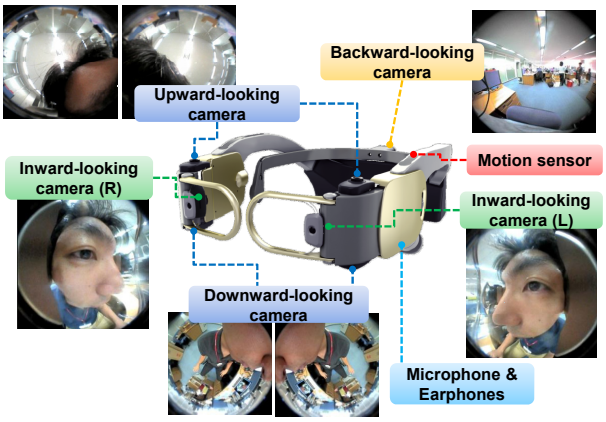


図 3. Prototype of eyeglass-based hands-free videophone system

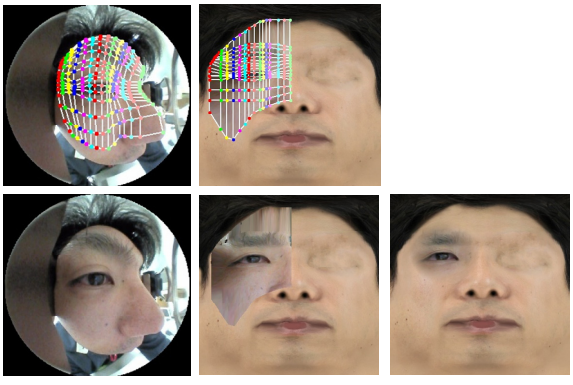
表 1. Specifications of fish-eye camera

Camera	
Size of CMOS [inch]	1/6.9
Resolution [pix] (Valid resolution)	1280 x 720 (720 x 720)
Compression	Motion JPEG
Fish-eye lens	
Diameter[mm]	7.2
Diagonal viewing angle [degree]	184.9
Projection method	Stereographic

3.2 自分撮りの顔画像生成

内向きカメラで撮影された顔画像は，図 3 に示すように，正面ではなく左右斜めから撮影された画像となっている．また，魚眼カメラは 180°超の画角で撮影するため，通常のカメラ画像に比べ，大きく歪んだ画像となっている．テレビ電話で必要となる顔画像を得るためには，魚眼画像の歪を補正し，正面から撮影したような画像に変換する必要がある．メガネ型という特性上，カメラは装着者の頭部の動きと連動して動くため，頭部を動かしても，装着者の顔部分は常にカメラ画像上の固定の位置に投影されるという特徴がある．そこで，魚眼カメラ画像を正面顔に変換する変換行列を求めた．この変換行列は，魚眼画像と正面顔画像との対応点を複数与えることで得られる．(図 4 上段(a)(b)参照)

ただし，図 4(a)を見て分かるように，内向きカメラでは口の周辺を撮影することは出来ない．上・下向きカメラの画像を見ても同様である．これは，レンズフレームにカメラを設置する位置関係上，顔の凹凸によって死角となってしまう部分である．また，耳・首・髪の毛などについても同様に死角となり，一部のみしか撮影できない．つまり，テレビ電話で必要となる正面顔画像を，魚眼カメラの画像群から完全に実写で再現することは不可能である．そこで，装着者の顔，及び，上半身の 3D-CG モデルをベースモデルとして予め作成しておき，そのベースモデルに貼り付けるテクスチャとして，魚眼画像の変換によって得られる正面顔画像を使用する手法をとった．表情を構成する要素として最も重要な目の周辺部分は実写画像をそのまま用いているため，CG 顔モデルに動きのみをパラメータとして反映させる手法に比べて，よりリアルで豊かな表情再現が可能である．なお，ベースモデルと，変換された正面顔画像は，



(a) Captured image (b) Mapped texture (c) Blended texture

図 4. Mapped real-captured texture onto base model

異なる環境で撮影されているため肌の輝度が異なる．このため，ベースモデルの肌色に合わせて正面顔画像の色を補正し，境界部分が目立たないようなブレンド処理を行なった上で，ベースモデルと正面顔画像を合成している．

この様子を図 4 下段に示す．図 4(b)(c)は顔の右半分が実際の画像を変換して合成した結果を，左半分がベースモデルそのものを示す．撮影画像(a) に対して歪補正，及び，正面変換を行った結果をベースモデルのテクスチャに合成した結果が(b)である．さらに，色補正とテクスチャとのブレンド処理を行い，違和感を低減させた結果が(c)となっている．同様の処理を，顔の左半分に対しても行い，合成されたテクスチャをリアルタイムでユーザのベースモデルの顔部分にマッピングすることで，自分撮りの上半身画像が取得できる．

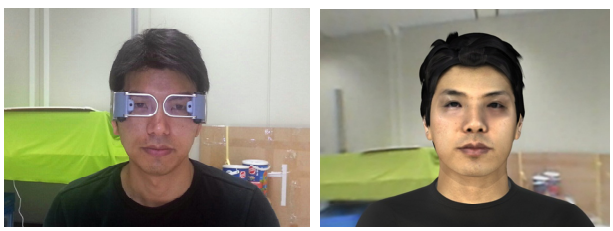
3.3 背景合成

テレビ電話では，相手に自分の姿を伝えるだけでなく，互いの場を共有することも重要な要素である．一般的なテレビ電話では，自分を撮る際に背景も含まれる構図となり，これが場の共有に繋がる．そこで，本システムでも背景画像との合成を検討した．ただし，装着者の前方フレーム上にある 6 個のカメラでは，装着者自身の頭部が遮ることで背景の撮影が困難であるため，図 3 に示すように，背景撮影用の魚眼カメラを後頭部に設置することで装着者の背景を取得した．実際のテレビ電話の構図に近づけるため，背景用魚眼カメラの画像はテレビ電話で通常用いられる一般的な画角の画像に平面展開した上で，**3.2** で生成された上半身画像と合成する．なお，6 軸センサによって得られた頭部の傾きを用いて，魚眼カメラ画像上で平面展開する位置を動的に変更することで，頭部を傾けた際でも背景画像が連動して傾かないようにする補正も行なっている．この合成結果を図 5 に示す．図 5(a)は，プロトタイプを実際に装着した様子であり，このプロトタイプで取得された各カメラ画像を元に合成した結果が図 5(b)となる．この図から，装着者の表情がリアルに再現されているだけで無く，装着者の背景も再現できていることが分かる．

3.4 動きのベースモデル反映

目周辺部の表情，及び，背景は実写画像をそのまま用いることができるが，装着者自身の口の動き，また，身振り・手振りについては，カメラ配置上，画像の取得が困難だったり，解像度が不十分であったりして，実写画像をそのまま用いることができない．そこで，ベースモデルを利用して，CG で動きを補完することを考えた．

口の動きについては，マイクで取得した発話音声から推定している．音声の周波数解析を行って得られるフォルマントから発話音声中の母音が認識できるため，口の動きを推定することができる[7]．この推定結果をベースモデル



(a) Appearance of wearer　　(b) Correspond image

図 5. Synthesized self-portrait image

(a) With head motion　　(b) With hand gesture

図 6. Reflected head and hand motion

に動きとして反映させることで，装着者の声に合わせて自分撮り画像の口が動くようになる．

また，プロトタイプに搭載された 6 軸センサの値から，装着者の頭部の動きを推定することができる．頭部の動きをベースモデルに反映させた結果を図 6(a)に示す．更に，図 3 中の下向きカメラ画像には，装着者の手が写っているため，この左右（ステレオ）の下向きカメラ画像内からユーザの手を示す肌色領域を検出・追従することで，手が 3 次元空間中でどの位置にあるかを特定することができる．手の 3 次元位置から，CG モデリングで一般的なインバースキネマティクスの考えを用いて腕全体の動きを推定してベースモデルに反映させることで，図 6(b)のように身振り・手振りも再現可能である．なお，現状のアルゴリズムでは，手全体の位置追従のみ可能であり，手や指の細かな動きは認識できていない．また，周囲の環境や照明環境によって，認識が不安定となる．これらの点については今後の課題としたい．

4. 今後の課題

メガネに搭載された複数の魚眼カメラ画像を元に，正面から撮影したような自分撮り画像を生成する手法について，基本原理を確認することができた．ここでは，より汎用的で，より高品質な自分撮り画像を生成してハンズフリービデオフォンを実現するにあたっての課題について述べる．

4.1 個人差への対応

3.2 で示したように，撮影された魚眼カメラ画像を正面顔画像に変換することは確認できた．ただし，当然ながら装着者の頭部全体の大きさや，顔の各部位の位置関係は個人毎にバラバラである．現状のシステムでは，この変換にあたって，個人毎に最適化した変換行列を装着時に求める必要があり，自動化は未だできていない．よって，今後はメガネ型端末を装着する際に，その人に合わせて上記変換行列を都度補正するような仕組みが必要となる．これには，画像内から顔の輪郭を抽出し，更に，各部位（例：目頭，目尻，眉尻等）の位置を自動で検出する必要がある．

4.2 魚眼カメラの解像度

魚眼カメラは非常に広い画角で撮影できるという特徴があるが，その反面，1 画素あたりの空間解像度は減ってしまう．これは，画像中心から離れる程顕著であり，図 4(a)で，眉と髪の生え際の間の額部分は魚眼カメラ画像上では，少しの領域しか撮影できていないことが確認できる．一方，額部分は図 4(b)の正面テクスチャ画像では広い面積を占めており，元の魚眼カメラ画像に比べて，より多くの情報量が必要となる．この解像度の差によって，現状のプロトタイプでは，図 4(b)で額部分が間延びしたような画像となってまっている．将来的には，より高解像度の魚眼カメラを

用いることで，この問題を解決できると考えている．

4.3 HMD との組み合わせ

本プロトタイプでは自分撮り画像を生成する事に特化しており，HMD を搭載していない．つまり，現時点のプロトタイプでは通話相手を見るために別のディスプレイを用意する必要がある．今後は，プロトタイプに HMD を組み合わせることで，“ハンズフリー”かつ“大画面での映像提示”が可能なテレビ電話とすることで，真のハンズフリービデオフォンを実現させていく必要がある．

4.4 小型化・無線化

プロトタイプは，自分撮り画像生成の原理確認を目的として試作しており，メガネ型端末自身が通常利用に耐えうる小型なものになっていない．また，現状はメガネ型端末と有線で接続された PC 間での処理である．今後は HMD の組み合わせ等と合わせて，メガネ型端末自身の小型化や無線化を進めていく．また，最終的には通常のメガネと遜色ない見た目・重さにするためにも，合成処理はローカルな PC ではなく，無線ネットワークを通じてクラウド側で実施する構成としていき，図 1 に示したゴールイメージを実現していく必要がある．

5. あとがき

従来の携帯電話によるテレビ電話において，自然なコミュニケーションを妨げる各種要因を解決する手段として，ハンズフリーでのテレビ電話を実現するメガネ型端末のプロトタイプを提案・試作した．プロトタイプでは，メガネ型端末に搭載された複数の魚眼カメラ画像を組み合わせることで，実写画像を用いたリアルな表情を再現する自分撮り画像を生成することが可能となった．

今後は，現状のシステムにおける課題を解決して，生成される自分撮り画像の高品質化，及び，HMD の搭載によって，“ハンズフリー”かつ“大画面での映像提示”が可能な真のハンズフリービデオフォンを実現していく．また，メガネ型端末の小型化や無線化を行うことで，ハンズフリービデオフォンだけで無く，様々なアプリケーションを実現可能としていき，スマートフォンに変わる新たな携帯電話の世界を切り拓いていきたい．

参考文献

[1] Google Inc. Project Glass. http://www.google.com/glass/

[2] Kanade, T. and Hebert, M. First-Person Vision. Proc. IEEE, 100, 8 (2012), 2442-2453.

[3] Ye, Z., Li, Y., Fathi, A., Han, Y., Rozga, A., Abowd, G. D. and Rehg, J. M. Detecting eye contact using wearable eye-tracking glasses. Proc. the 2012 ACM Conference on Ubiquitous Computing, ACM Press (2012), 699-704.

[4] Jones, A., Fyffe, G., Xueming, Y., Wan-Chun, M., Busch, J., Ichikari, R., Bolas, M. and Debevec, P. Head-Mounted Photometric Stereo for Performance Capture. Proc. ACM SIGGRAPH 2010 Emerging Technologies, 14 (2010).

[5] Reddy, C. K., Stockman, G. C., Rolland, J. P. and Biocca, F. A. Mobile Face Capture for Virtual Face Videos. Proc. the 2004 Conference on Computer Vision and Pattern Recognition Workshop, 5 (2004), 77.

[6] Shimizu, S., Kawai, J. and Yamada, H. Wraparound View System for Motor Vehicles. Fujitsu scientific and technical journal, 46, 1 (2010), 95-102.

[7] Brand, M. Voice puppetry. Proc. the 26th annual conference on Computer graphics and interactive techniques, ACM Press (1999), 21-28.