テレビ電話　メガネ型端末　ウェアラブルデバイス

Technology Reports

# テレビ電話が可能なメガネ型端末の試作 ―ハンズフリービデオフォン―

**テレビ電話は対面に近いコミュニケーションを提供する重要な手段である．しかし，従来の携帯端末におけるテレビ電話では，端末のカメラを手に持ち，常に自分の姿を撮影しながら会話をしなければならない．これは，肉体疲労を伴い，視線が合わない状態での会話を強いられる場合が多く，テレビ電話の利用を妨げる要因となっている．そこで，手にカメラを持たずに，メガネ型端末をかけるだけで自分の姿を撮影・合成し，相手と会話が可能なハンズフリービデオフォンの開発をめざし，試作を行った．本稿ではその概要と今後の課題について解説する．**

先進技術研究所　木村 真治　堀越 力

## 1. まえがき

従来の携帯電話によるテレビ電話は，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つつ，通話相手の顔を画面上で見る利用スタイルである．しかしながら，長時間の利用による腕の疲れや，カメラの向きが変わることで自分の姿がカメラから外れてしまうこと，通話相手と視線が合わないことが多くある．また，端末の画面が小さく相手の表情が分かりづらい，端末を持っているため身振り手振りが伝えづらいなど，円滑なコミュニケーションを行ううえでの課題が多くあった．そこで，カメラを手で持つことなく，メガネ型端末を用いて自然なコミュニケーションを可能とするハンズフリービデオフォンの試作機を開発した．

昨今，Google™*1 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が，これらは，ユーザの眼前で映像提示を可能とするヘッドマウントディスプレイ（HMD：Head Mounted Display）*2と，自分の視線に近い映像を撮影する外向きのカメラを備えたものがほとんどであり，主に拡張現実（AR：Augmented Reality）*3を利用したアプリケーションを想定している[2]．一方，本システムでは自分自身を撮影する「自分撮り」の手段として，メガネ型端末を利用する．メガネ型端末には，超小型魚眼カメラが複数取り付けられ，自分の周囲だけでなく，自分のリアルな表情を撮影することが可能である．従来の研究において，ヘルメットに取り付けられたカメラなどを用いて自分自身の表情の動きを画像処理で検出して，CG で顔映像を再現する技術[3]がある（以下，このようにして再現された映像をCG アバター*4と呼ぶ）．CG アバターでは，大まかな動きの再現のみに留まり，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ったが，本システムではメガネ型端末上の魚眼カメラで顔を撮影し，その顔画像を用いることで，その微妙な表情の変化をそのまま再現することが可能となっている．

なお，ハンズフリービデオフォンは，スマートフォンに代わる将来の携帯電話としてドコモが提案しているウェアラブル（装着型）端末コンセプトを具現化した 1 つの例であり，CEATEC JAPAN 2012 に出展した内容である．



*1 **Google™**：Google, Inc.の商標または登録商標．
*2 **HMD**：頭部に装着するディスプレイ装置の総称．目のすぐ前における映像提示を可能とし，片目だけに映像を提示する単眼タイプと，両目に映像を提示する両眼タイプがある．メガネやゴーグルなどのレンズ部分に映像を提示するものが代表的である．
*3 **拡張現実（AR）**：現実世界を写した映像に，電子的な情報を実際にそこにあるかのように重ねて，ユーザに提示する技術．
*4 **アバター**：自分の分身として画面上に表示するキャラクター．

## 2. ウェアラブル端末コンセプト

ウェアラブル端末コンセプトは，将来の携帯電話の姿として提案しているものであり，ユーザが普段身に付けるメガネやイヤホン，アクセサリなどで携帯電話の機能を実現するコンセプトである．ウェアラブル端末自体は情報の入出力のみ行い，各種データ処理・管理は，すべてクラウド側で処理することを想定している．このコンセプトの中でも，**図 1**（a）に示すようなメガネ型の端末に特に注目している．メガネ型端末には　①ハンズフリー　②視界に合わせた情報提示が可能　③常時装用で情報へのアクセスが早い　といったメリットがある．これを活かすことで，新たなユーザ体験の提供や，従来の携帯電話の機能をより便利にすることが可能となる．例として，自分の視界そのものに対して情報が付加される直感的な AR（メリット②），両手を使ったジェスチャ入力（図 1（b），メリット①），メガネに搭載された生体センサによる常時健康管理（図 1（c），メリット③）などがある．この中で，ハンズフリーというメリットを活かし，自然なコミュニケーションを可能とするテレビ電話をコンセプト具現化の第一弾として提案した．

## 3. システム要件

メガネ型端末でテレビ電話を行うためには，メガネに搭載されたカメラで自分撮りをする必要がある．通常のテレビ電話や，ヘルメットに取り付けられたカメラ[3]では，顔からある程度（数十センチメートル）離すことで，自分自身の顔や上半身を広く撮影している．一方，メガネに搭載したカメラでは，顔までの距離が非常に近いため，顔の一部しか撮影できず，焦点を合わせることも難しい．つまり，メガネに取り付けられたカメラで自分の顔を撮影するためには，近距離で広い範囲をカバーし，かつ，焦点があった状態で映像を取得する必要がある．

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[4]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，周囲の人に違和感を与える結果となる．これを受けて，自分撮りを可能とするメガネ型端末の要件を以下の 4 つとした．

①通常のメガネに近い見た目
②装着者の視界を邪魔しない
③装着者の顔をできるだけ広い範囲撮影する
④全周囲画像を取得する（他アプリケーションへの応用を想定）

上記 4 つの要件を満たすために，本システムでは超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．一般に魚眼カメラは画角が 180° 程度あり，しかも，非常に近い距離でも焦点が合うという特徴を持っている．複数の魚眼カメラを用いて，1 つの映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[5]が代表的である．これと同様に，本システムでは複数の

(a)メガネ型端末の将来コンセプトモック

(b)コンセプトユースケース：バーチャルオフィス
レンズ越しにキーボードが表示され，指の動きをカメラで読み取ることで文字入力が可能．いつでもどこでも仮想的なオフィス環境を実現


(c)コンセプトユースケース：バイタルモニタリング
脈波や生体ガスを検知する生体センサをメガネに組み込むことで，常に装着者の健康管理をすることが可能

**図 1　メガネ型ウェアラブル端末コンセプト例**

魚眼カメラで自分の顔，手元，背景などを撮影し，映像を合成している．これにより，自分の前方に，あたかも自分の方を向いているカメラがあるかのような映像（以下，自分撮り映像）を，手に何も持たない状態で生成することが可能となると考えた．

## 4. プロトタイプ

前述のシステム要件に基づき，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．このプロトタイプと，各カメラで撮影された画像を**図 2** に示す．本プロトタイプには 7 個の魚眼カメラ（上中下向き×左右＋背景），傾きセンサ，マイク，イヤホンが搭載されている．また，魚眼カメラの仕様は**表 1** に示すとおりであり，メガネ型端末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．

### 4.1 自分撮り映像の生成

内向きカメラで撮影された顔画像は，正面ではなく左右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．このため，この画像の歪を補正し，正面から撮影したような画像になるような座標変換を行って，1 枚の正面顔画像を生成する．

ただし，図 2 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できない．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，当然ながら耳・首・髪の毛などについても撮影をすることはできない．そこで，装着者の正面顔写真からあらかじめベースとなる CG の顔，および，上半身モデルを作成しておき，そのベースモデルに貼り付けるテクスチャ*5 として，生成された正面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映像を用いているため，従来の CG アバターに比べて，よりリアルで豊かな表情再現が可能である．なお，ベースモデルと生成された正面顔写真では，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写真の色を補正し，境界部分が目立たないようなブレンド処理を行ったうえで，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 3** に示す．図 3 で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，および，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

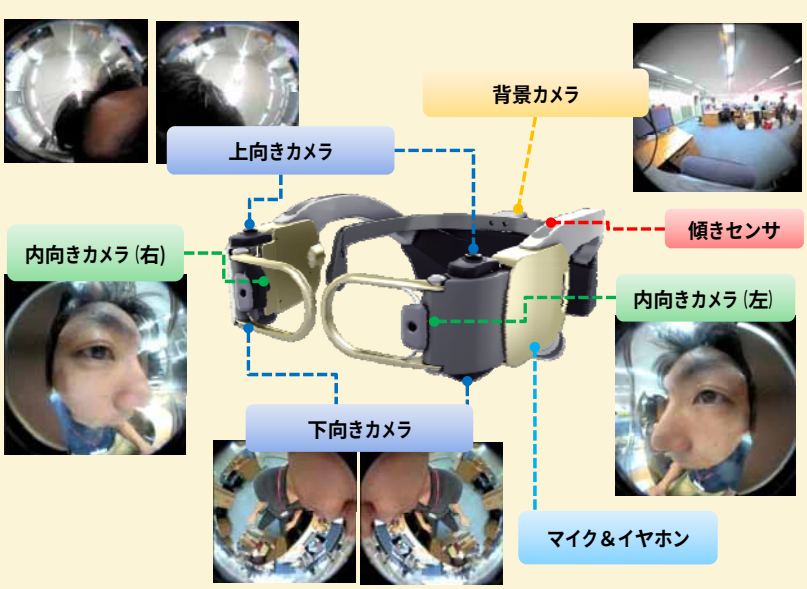
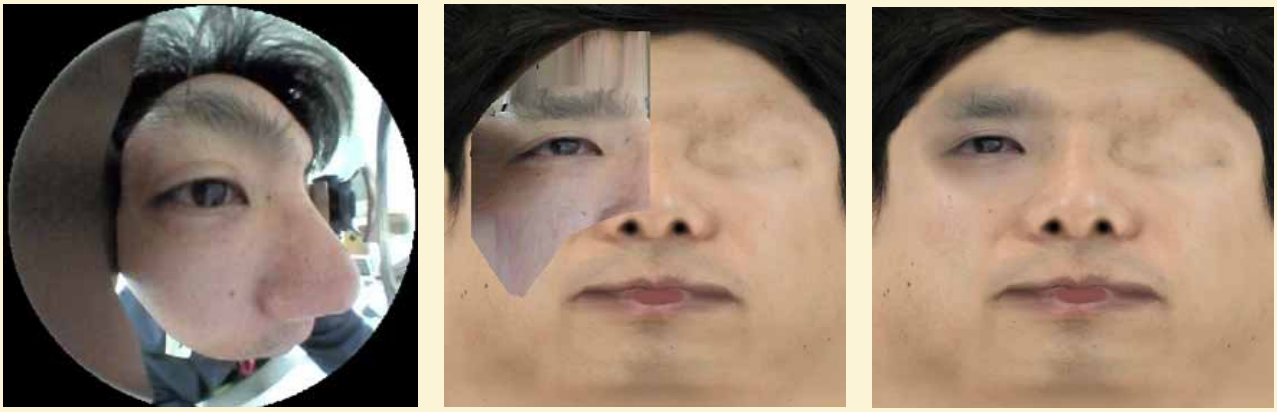



**図 2　プロトタイプと撮影画像**

**表 1　魚眼カメラの仕様**

| | |
|---|---|
| **カメラ** | |
| CMOS サイズ［inch］ | 1/6.9 |
| 解像度［pix］（有効解像度） | 1,280×720 (720×720) |
| 出力形式 | Motion JPEG |
| **魚眼レンズ** | |
| 直径［mm］ | 7.2 |
| 対角画角［degree］ | 184.9 |
| 射影方式 | 立体射影 |

*5 **テクスチャ**：3D 形式のデータにおいて，物体表面の質感を表現するために物体表面に貼り付ける画像のこと．



(a)撮影画像　(b)歪補正＆正面変換後　(c)色補正＆ブレンド後

**図 3　正面顔画像への変換**

ャをリアルタイムでユーザのベースCG モデルに貼り合わせることで，自分撮りの上半身映像が取得できる．

## 4.2　背景合成

テレビ電話では，相手に表情を伝えることも重要であるが，互いの場を共有することも重要な要素である．一般にテレビ電話では，自分を撮る際に背景も含まれる構図となる．そこで，本システムでも背景映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．そこで，図 2 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することを可能とした．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を一般的な画角をもつ画像に変換した上で，4.1 節で生成された上半身映像と合成する．この結果を**図 4** に示す．図 4（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映像を基に合成した結果が図 4（b）となる．

## 4.3　口の動きの反映

前述のように，装着者の口の周辺を撮影することは困難であり，画像処理によって口の動きを認識することはできない．そこで，口の動きについては，プロトタイプに搭載したマイクで取得した発話音声から口の動きを推定している．口の動きは発話音声中の母音に連動していることから，発話音声の周波数解析を行うことで得られる第一フォルマント*6，および，第二フォルマントの周波数帯から母音を認識し，結果的に口の動きを推定することができる[6]．この推定結果をベースの CG モデル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映像の口が動くようになる．

## 4.4　身振り・手振りの反映

プロトタイプには 6 軸の傾きセンサが搭載されている．このセンサの値から，カメラ画像からでは認識が難しい，装着者の頭部の動きを推定することができる．頭部の動きを反映させた結果を**図 5**（a）に示す．

また，図 2 中の下向きカメラ映像には，装着者の手が写っていることが分かる．この左右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，手がどのように動いているかを認識することができる．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．

# 5．今後の課題

メガネに搭載された複数の魚眼カメラ映像を基に，正面から撮影したような自分撮り映像を生成する手法について，基本原理を解説した．ここでは，より汎用的でより高品質な自分撮り映像を生成するにあたって

---

*6 **フォルマント**：発話音声のスペクトルを観察することで得られる，時間的に移動しているピークのこと．周波数の低いピークから，第一フォルマント，第二フォルマント…　と呼び，第一フォルマントと第二フォルマントの周波数帯から，発話音声中の母音を認識することが可能である．

(a)プロトタイプを装着した様子

(b)合成された自分撮り映像

**図4　自分撮り映像**

(a)顔の傾きを反映した様子

(b)手の追従結果を反映した様子

**図5　動きを反映した自分撮り画像**

の課題について述べる．

## 5.1　個人差への対応

4.1 節で示したように，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ごとにバラバラである．現状のシステムでは，この変換にあたって，個人ごと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，さらに，各部位（例：目頭，目尻，眉尻など）の位置を自動で検出する必要がある．

## 5.2　魚眼カメラの解像度不足

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 画素当りの空間解像度*7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 3（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 3（b）の正面画

*7 **空間解像度**：デジタルカメラにおいて，1 画素に投影される，実空間中の広さを示す指標．例えば，同じ大きさの物体を異なるカメラで撮影した場合に，その物体が撮影画像中で，より多くの画素数を占めるカメラの方が，空間解像度が高いカメラといえる．なお，空間分解能とも呼ぶ．

像では広い面積を占めており，元の魚眼カメラ映像に比べて，より多くの情報量が必要となる．よって，現状のプロトタイプでは，図 3（b）で額部分が間延びしたような映像となっている．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．

### 5.3 HMD との組合せ

テレビ電話を実現するためには，当然ながら通話相手の姿が見える必要がある．前述のようにメガネ型端末としては HMD を搭載したものが多く開発されているが，本プロトタイプは現時点で自分撮り映像を生成する方に特化しており，HMD を搭載していない．よってプロトタイプに HMD を組み合わせることで，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．

### 5.4 カメラの配置

今回試作したプロトタイプは，メガネフレーム前方に 6 個，背面に 1 個の計 7 個のカメラを配置することで，自分の顔画像だけではなく，全周囲画像も取得することができている．しかし，例えば AR に関するアプリケーションを実現するには，前方を向いたカメラのほうが使いやすいと考えられる．また，テレビ電話の利用には上向きカメラを必要としない．つまり，用途によって使用されるカメラの位置や個数は異なる．この解決には 2 つの方法が考えられる．1 つは使うアプリケーションによってカメラの位置を移動できるようなハードウェア構成とすること．もう 1 つは，カメラの位置を気にせずに使える全周囲画像合成を内部で行い，その全周囲画像の中からアプリケーションごとに必要な部分だけ切り出して使うことである．

### 5.5 小型化・無線化

現状のプロトタイプは，自分撮り映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はなっていない．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．今回，自分撮り映像生成の原理確認は実施できたため，HMD の組合せなどと合わせて，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必要がある．

## 6. 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，CG アバターによる表現とは一線を画した自分撮り映像を生成することが可能となった．今後は，現状の試作機における課題を解決して，生成される自分撮り映像の高品質化，および，HMD の搭載による真のハンズフリービデオフォンを実現していく．

メガネ型端末は 2 章で述べたメリットを有し，従来の携帯電話では実現できなかったユーザ体験を提供する可能性を秘めた，新たなプラットフォームとして今後の普及が期待される．ハンズフリービデオフォンの具現化はこの一歩に過ぎず，小型化や無線化によって応用範囲を拡げていくことで，スマートフォンに代わる，新たな携帯電話の世界を切り拓いていきたい．

### 文 献


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

[2] T. Kanade and M. Hebert : “First-Person Vision,” Proc. of IEEE, Vol. 100, No.8, pp.2442-2453, Aug. 2012.

[3] A. Jones, G. Fyffe, Y. Xueming, M. Wan-Chun, J. Busch, R. Ichikari, M. Bolas and P. Debevec : “Head-Mounted Photometric Stereo for Performance Capture,” Proc. of ACM SIGGRAPH 2010 Emerging Technologies, 2010.

[4] C. K. Reddy, G. C. Stockman, J. P. Rolland and F. A. Biocca : “Mobile Face Capture for Virtual Face Videos,” Proc. of the 2004 Conference on Computer Vision and Pattern Recognition Workshop, pp.77-83, Jun. 2004.

[5] S. Shimizu, J. Kawai and H. Yamada : “Wraparound View System for Motor Vehicles,” Fujitsu scientific and technical journal, Vol.46, No.1, pp.95-102, Jun. 2010.

[6] M. Brand : “Voice puppetry,” Proc. of the 26th annual conference on Computer graphics and interactive techniques, pp.21-28, 1999.