

応用論文

拡張現実感による顔表情操作を用いたビデオ通話方式

佐藤 翔悟*¹ 北原 格*² 大田 友一*²

A Video Chatting System by Using AR Deformation of Facial Expression

Shogo Sato*¹ Itaru Kitahara*² and Yuichi Ohta*²

Abstract --- Facial expression, which is one of the most important non-verbal communication media, makes communication smoother by compensating the missed message in language communication. However, some shy people are not as good at using facial expressions as they want. Such poor emotional expression by the conversation partner makes it difficult to read feelings correctly, and as the result, smooth communication is hindered. In order to solve this problem, this paper proposes a facial expression enhancement method that realizes the smooth communication with rich facial expression. To enhance facial expression, the facial shapes and textures are expressed as the parameters in parametric spaces reconstructed from the personal facial images naturally. In the parametric spaces of facial expression, the difference between two facial expressions can be handled as a multidimensional vector. By controlling the scale of the difference vector between input vector and the norm vector, the facial expression difference is enhanced without having to recognize the facial expression. Then we generate the texture, which enhanced the texture of expression by re-projecting into image space. Finally, we overlay the synthesized facial image of the conversation partner onto the face region in the video chatting sequence. We conduct on evaluations to confirm the effect of expression enhancement by our method using CG faces and a real video sequence.

Keywords: facial expression, face, point distribution models, augmented reality, video chat

1 はじめに

人と人とのコミュニケーションでは、感情や関心を相手に正確に伝えることが重要であるが、言語情報だけでは、それらを十分伝えられない場合がある。そのような場合、我々は、表情、しぐさ、声の抑揚といった非言語情報を用いることで、情報を補っている。Albert Mehrabian の実験[1]でも、非言語情報が感情伝達において大きな役割を担うことが示されており、特に表情は重要な伝達手段である。そのため、会話相手の表情表出が乏しい場合には、その感情や関心を正しく読み取ることが困難となり、結果として円滑なコミュニケーションが妨げられることが懸念される。

我々は、画像処理技術を用いて顔の表情を操作し、それを AR (Augmented Reality) 提示することにより、豊かな表情表出を伴った映像コミュニケーションメディアの実現を目的とした研究に取り組んでいる。

本研究で提案するシステムでは、様々な表情画像を用いた事前学習により、表情変化を表現するパラメトリッ

ク空間を構築し、入力された顔画像をその空間に射影して得られるパラメータを操作することで、表情操作を実現する。表情識別処理を行うことなく、表情の操作が可能であるため、微細な表情変化にも対応可能という特長を有する。撮影から提示までの処理をリアルタイム実行することにより、ビデオ通話システムへの応用を試みる。

2 関連研究

表情操作処理は、大きく二種類のアプローチに分類される。一つは、Fig. 1 (a)に示すように、入力顔画像から、怒り、笑い、悲しみといった表情変化の要因（感情）を認識し、その要因に対応した処理を顔画像に施す手法である[2][3]。しかし、これらの手法には、要因の正確な識別が困難な場合（例えば、感情表現が苦手で大きな表情を表出できないユーザの顔画像が入力された場合）、意図した表情操作処理が困難となる問題が発生する。

もう一方のアプローチとして、Fig. 1 (b)に示すように、顔の表情変化を表現するパラメトリック空間を構築し、入力顔画像をその空間中に射影した点と基準表情の画像を射影した点の差分ベクトルを操作することにより、表情を操作する手法がある。類似した試みとして、興水ら[4]は、顔の見え方を表現す

*1 筑波大学 大学院システム情報工学研究科

*2 筑波大学 システム情報系

*1 Department of Intelligent Interaction Technologies, Graduate School of Systems and Information Engineering, University of Tsukuba

*2 Faculty of Engineering, Information and Systems, University of Tsukuba

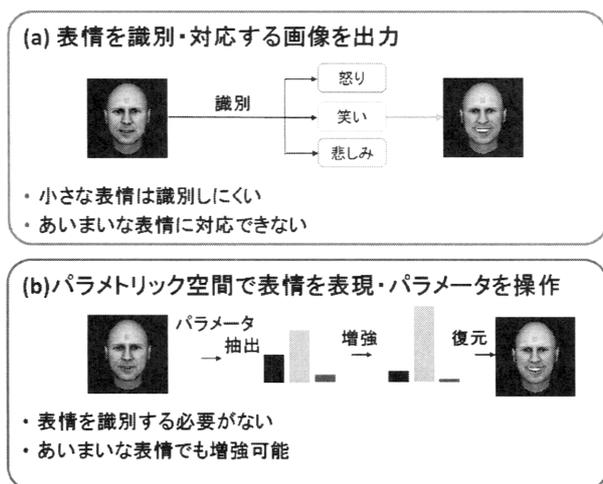


Fig. 1 表情操作のアプローチ

るパラメトリック空間において、平均顔と入力画像との差を誇張することにより、顔の特徴（個性）を増強するデフォルメ肖像画を生成している。このアプローチの場合、表情が僅かでも表出していれば、それを増強することが可能であるため、比較的表出の乏しい顔画像への適用が可能である。また、構築されるパラメトリック空間と、Woodworthら[5]によって提案された、愛・喜び・幸福・驚き・恐怖・苦痛・怒り・決意・嫌悪・軽蔑の感情カテゴリを一次的に配置した表情表現や、Schlosbergらによる、「快—不快」、「注意—拒否」、「緊張—睡眠」を軸とした表情表現の心理空間[6][7]は、データ構造が似ているため、その対応関係を解明すれば、認知心理学の知見の導入が可能であるという特長も有する。

本手法では、表情表現パラメトリック空間を介して表情の操作を行う。Calderら[8]は、画像上で観測される顔形状から各特徴点の位置ベクトル群を獲得し、入力画像と基準表情の位置ベクトルの差分を誇張操作することにより、顔の表情を増強する手法を提案し、相手の表情を理解するまでの時間が早くなることを示している。しかし、顔の形状を手動で取得しており、表情表出に伴う見え方（テクスチャ）の変化の操作についても検討がなされていない。

3 Point Distribution Models を用いた パラメトリック表情空間の生成

本研究では、二種類の Point Distribution Models (PDM) [9]を用いて、顔の形状とテクスチャのパラメトリック表情空間を生成し、表情操作を実現する。表情をパラメトリック空間に射影することにより、ある表情をパラメトリック空間における一点として考えることができる。ある表情点からある表情点へのベクトルの長さを操作することにより、表情識別

処理を行うことなく、顔表情変化を操作することが可能となる。また、この手法を用いれば、基本的な五つの表情（驚き、喜び、恐怖、悲しみ、嫌悪）に属さない曖昧な表情の増強も可能である。

顔形状は、Constrained Local Models (CLM) [10]を用いて顔形状を推定し、パラメトリック空間を構築する。入力されたビデオ会話映像に CLM を適用してリアルタイムで顔形状のベクトルを推定し、事前に求めておいた基準表情の形状ベクトルを推定し、それを誇張するような操作を施すことで顔の形状の増強を行う。表情表出に伴うテクスチャの変化は、Active Appearance Models (AAM)[11]を用いてパラメトリック空間を構築し、表情操作を実現する。入力画像を AAM のパラメトリック空間に射影し、事前に求めておいた基準表情の射影点との間で与えられるベクトルを操作することで、顔のテクスチャ変化を操作する。

4 パラメトリック空間を用いた顔表情の操作

本研究では顔の形状とテクスチャの変化をリアルタイムで推定するために、PDM を用いて、入力画像中に写った顔をモデル化する。PDM は、学習画像群から抽出したランドマーク群に対して主成分分析を行い、式(1)に示すように、形状変化をモデル化した固有空間を生成する。

$$\mathbf{s} = \tilde{\mathbf{s}} + \Phi_{\mathbf{s}} \mathbf{b}_{\mathbf{s}} \quad (1)$$

ここで、 $\mathbf{s} = \{x_1, \dots, x_r\}^T$ は、 i 番目のランドマークの座標 $\mathbf{x}_i = \{x_i, y_i\}^T$ をランドマーク数 r まで一列に並べた形状ベクトルで、 $\tilde{\mathbf{s}}$ は平均形状のベクトル、 $\Phi_{\mathbf{s}}$ は固有ベクトル、 $\mathbf{b}_{\mathbf{s}}$ は固有空間のパラメータベクトルを表している。入力画像からある形状 \mathbf{s} が入力されたとき、 \mathbf{s} は式(1)によって固有空間に射影され、パラメータベクトル $\mathbf{b}_{\mathbf{s}}$ で表示される。

非剛体の物体追跡に PDM を適用した手法がいくつか提案されているが、これらは大きく分けて二種類のアプローチ、すなわち AAM と CLM に分類される。

AAM では、式(1)と同じように、テクスチャについても同様にモデル化し固有空間を生成する。

$$\mathbf{g} = \tilde{\mathbf{g}} + \Phi_{\mathbf{g}} \mathbf{b}_{\mathbf{g}} \quad (2)$$

ここで、 \mathbf{g} はテクスチャの輝度値を平均 0、分散 1 に正規化したものを一列に並べたテクスチャベクトルで、 $\tilde{\mathbf{g}}$ は平均テクスチャのベクトル、 $\Phi_{\mathbf{g}}$ はテクスチャの固有ベクトル、 $\mathbf{b}_{\mathbf{g}}$ は固有空間のパラメータベクトルを表している。

AAM では形状とテクスチャのパラメータ \mathbf{b}_s と \mathbf{b}_g を変化させることで、任意の形状・テクスチャを表現できる。そうして生成したテクスチャと入力画像のテクスチャとの誤差が最小になるように \mathbf{b}_s と \mathbf{b}_g を最適化することにより、入力画像に対応する形状とテクスチャのモデルを推定する。

一方、CLM では、局所的なパッチ情報のみをモデル化し、PDM により動作を拘束することで最適化を行い、入力画像に対応するモデルを推定する。本手法では、高速動作が可能で見え方の変化にロバストな CLM を用いて顔の形状推定と操作を行い、AAM を用いて顔テクスチャの記述と操作を実現する。

4.1 CLM による顔形状の操作

式(1)を用いて、無表情の顔画像を射影することで得られる顔のパラメータ(形状)ベクトルを $\mathbf{b}_{s,N}$ とする。式(3)に示すように、入力された顔のパラメータ(形状)ベクトル $\mathbf{b}_{s,I}$ と $\mathbf{b}_{s,N}$ との差分ベクトルを算出し、それをスケールリングすることにより、表情表出に伴う形状変化を操作する。ここで、 $\mathbf{b}_{s,E}$ は表情を操作した顔のパラメータベクトルで、 $k (>1)$ は増強倍率とする。

$$\mathbf{b}_{s,E} = \mathbf{b}_{s,N} + k(\mathbf{b}_{s,I} - \mathbf{b}_{s,N}) \quad (3)$$

4.2 AAM による顔テクスチャの増強

表情表出に伴うテクスチャ変化の増強には、AAM のパラメトリック空間を用いる。AAM は形状とテクスチャの変化を線形なパラメータとして表現するモデルで、任意の形状とテクスチャを再現することが可能である。AAM を用いれば、テクスチャと顔形状の変化を同時に追跡することもできるが、AAM は形状と明度を統合して扱い、顔全体のテクスチャを用いて誤差を評価するため、全体のテクスチャさえ十分に似ていれば、パーツの位置が間違っても評価が高くなってしまふ。そのため、細かな形状の推定には向いていない。そこで本研究では、表情表出に伴う形状変化とテクスチャの変化を分離し、AAM ではテクスチャ変化の増強だけを行うこととする。

学習段階では、様々な表情を撮影した正面顔画像か

ら顔領域を切り出したデータに主成分分析を適用し、表情変化に伴うテクスチャの変化を表現する固有空間を生成する。入力画像から、前述した CLM によって推定した形状情報を用いて顔領域を切り出し、正面から撮影した参照形状に変形した後、テクスチャ情報を一次元の画素ベクトルとして獲得する。照明変動を小さくするために、このベクトルを正規化した後、パラメトリック空間(固有空間)に射影する。

提案手法では、入力ベクトルと基準表情の差をスケールリングすることで表情を操作する。Fig. 2のように、この基準表情を求めるために、同一種類の表情で表出強度の異なる2枚の顔表情画像をパラメトリック空間に射影し、それらを通る直線を求める。同様の処理を複数の表情について行い複数の直線を求め、それらの直線間の距離が最短となる点を基準表情のパラメータベクトル(表情変化の原点) $\mathbf{b}_{g,o}$ とする。式(4)に示すように、 $\mathbf{b}_{g,o}$ と入力画像の射影点 $\mathbf{b}_{g,I}$ を結ぶベクトルをスケールリングすることにより、表情操作処理を施した顔テクスチャのパラメータベクトルを得る。

$$\mathbf{b}_{g,E} = \mathbf{b}_{g,o} + k(\mathbf{b}_{g,I} - \mathbf{b}_{g,o}) \quad (4)$$

このパラメータベクトルを画像空間に逆射影することにより、表情操作によるテクスチャの変化を生成する。

最後に、このテクスチャを 4.1 節で獲得した形状変化情報を用いて、表情操作済みの形状に変形し、入力画像に重畳する。

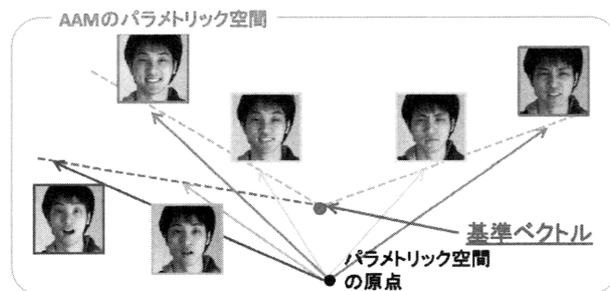


Fig. 2 基準表情のパラメータベクトルの生成

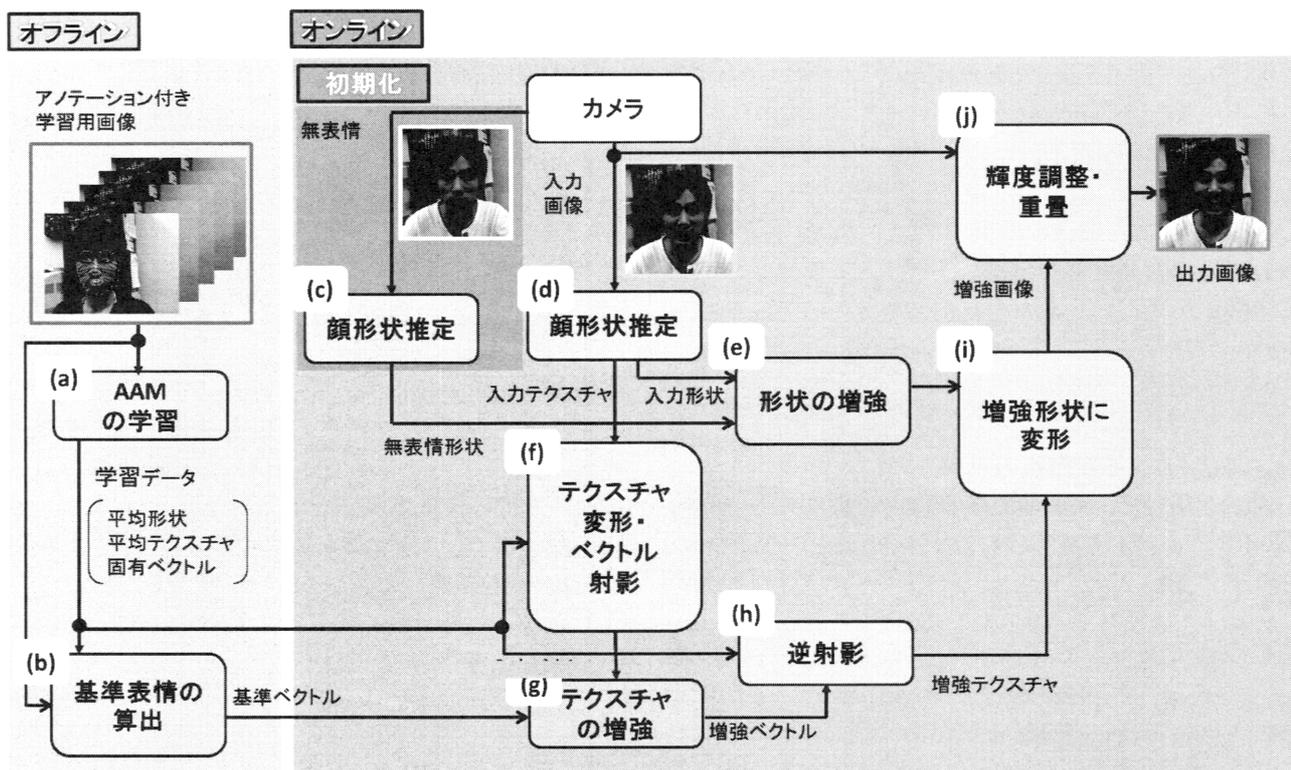


Fig. 3 提案手法の処理の流れ

5 顔表情増強機能付きビデオ通話システムの構築

$$A_i + t_i(A_i - B_i) \quad (5)$$

5.1 オフライン処理

顔の形状変化の推定に用いる CLM のパラメトリック空間は、一般的な顔画像データベース[12]-[14]から構築したもので十分な推定精度を実現可能であるが、顔の見え方(テクスチャ)の変化の推定に用いる AAM のパラメトリック空間を一般的なデータベースから構築すると、個々の顔つきの差異など、表情変化とは関係のない見え方の変化が含まれたパラメトリック空間が生成されてしまう。ビデオ会話では、Fig.3(a)に示すように、被写体となるユーザの特定が可能であることから、システム利用時に4種類の表情変化(喜び、怒り、悲しみ、驚き)を伴う顔画像を各2段階ずつと無表情の正面顔画像を撮影し、それらを用いて CLM と AAM のパラメトリック空間を構築する。

基準表情のパラメータベクトルを求めるために、構築した AAM のパラメトリック空間において、Fig.3(b)に示すように、4種類の表情変化によって定義される直線の交点を求める。各直線は必ずしも交わるとは限らないので、本手法では、各直線からの距離が最小となる点を以下の手順で求める。

パラメトリック空間の次元数を n とし、 m 種類の表情を使用すると考えると、 i 番目の直線は、直線上の2点を $A_i = \{a_{i1}, \dots, a_{in}\}^T$, $B_i = \{b_{i1}, \dots, b_{in}\}^T$, t_i を変数として、式(5)のように表せる。

このとき、直線の変数を $\mathbf{t} = \{t_1, \dots, t_m\}^T$, R^n 上の任意点を $\mathbf{p} = \{p_1, \dots, p_n\}^T$ として、式(7)を最小化する。

$$f(\mathbf{t}, \mathbf{p}) = \sum_{i=1}^m \sum_{j=1}^n \{a_{ij} + t_i(a_{ij} - b_{ij}) - p_j\}^2 \quad (6)$$

ここで、 $f(\mathbf{t}, \mathbf{p})$ は、 $t_i (i=1, \dots, m)$, $p_j (j=1, \dots, n)$ それぞれについての下に凸な2次関数であるので最小二乗法により \mathbf{t} , \mathbf{p} を求めることができる。

5.2 オンライン処理

Fig.3(d)に示すように、入力された動画像に対して、CLM を適用し、顔形状を追跡する。まず、ユーザを、無表情でカメラの前に立たせ、追跡用ランドマークの初期化を行う (Fig.3(c))。ここで得られた無表情の形状と入力画像の形状の差分ベクトルに対して、式(4)を用いた表情操作処理を行う (Fig.3(e))。表情変化の増強処理では、CLM によって追跡した顔形状情報を用いて顔領域を切り出し、テクスチャ情報として AAM のパラメトリック空間に射影する (Fig.3(f))。射影された入力画像のパラメータベクトルと、オフラインで求めておいた基準表情のパラメータベクトルの差分ベクトルに対して、式(4)のように操作処理を行う (Fig.3(g))。その後、テクスチャ空間

に逆射影を行い、表情を操作したテクスチャを得る (Fig.3(h)). そして、式(3)を用いて、無表情画像から獲得した形状ベクトルと入力画像の形状ベクトルの差分ベクトルを、操作し、表情を操作したテクスチャをマッピングする (Fig.3(i)). 最後に、入力画像に重畳するが、この際、カラーチャンネルごとに操作前・後のテクスチャの輝度値の平均値を算出し、両者が同等の値になるよう輝度値を補正する (Fig. 3(j)).

5.3 表情操作を用いたビデオ通話システムの実装

提案手法をビデオ通話システムに適用することにより、Fig. 4に示す、豊かな表情を提示可能な通話システムを実装する。Fig. 4では、左に表情増強を用いたビデオ通話の一例が示されており、右下に通話の様子を撮影した画像が示されている。



Fig.4 表情操作を用いたビデオ通話システム

Fig.5 に通話システムの処理を示す。通話システムは USB カメラとディスプレイ、PC で構成される。USB カメラによって撮影された顔画像がシステムに入力されると、提案システムを用いた表情操作が行われる。この処理結果を画面キャプチャソフトウェア[15]を用いて仮想的に撮影し、この画像をカメラからの映像として汎用的なビデオ通話ソフトウェア (例えば skype[16]) に入力する。ビデオ通話ソフトウェアの機能により、ネットワークを介して対話相手に伝送された画像は、対話相手のビデオ通話ソフトウェアを介してディスプレイ上に提示される。対

話相手の顔画像に対しても同様の処理を施すことにより、互いの表情が操作されたビデオ通話を実現することができる。

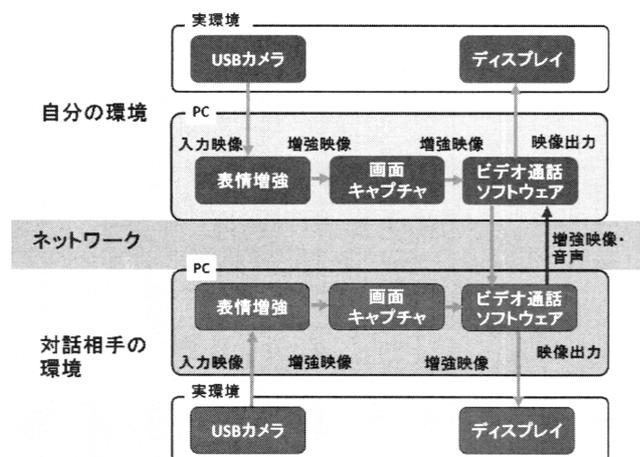


Fig.5 ビデオ通話システムの処理

6 実験

6.1 表情操作手法の評価

6.1.1 CG 顔画像を用いた画像評価実験

CG 顔画像データセットを用いて、表情が適切に操作されているかを確認する実験を行う。CG 顔画像データセットは、FaceGen Modeler[17]を用いて様々な顔表情を生成・取得した。顔形状の推定と増強操作は、CLMのC++のオープンソース[18]を用いて実装する。基準表情の導出には、喜び、怒り、悲しみ、驚きの各2段階(計8枚)を、FaceGen Modelerの表情変化スケールを0.5, 1.0として合成したものを使用する。AAMのパラメトリック空間の構築には、上述した8枚に無表情1枚を加えた9枚の顔画像を使用する。これらの画像サイズは、全て400画素×400画素とした。なお、表情が十分に表出されない状況下での提案手法の性能を評価するために、入力画像中の表情は比較的小さいものとなっている。

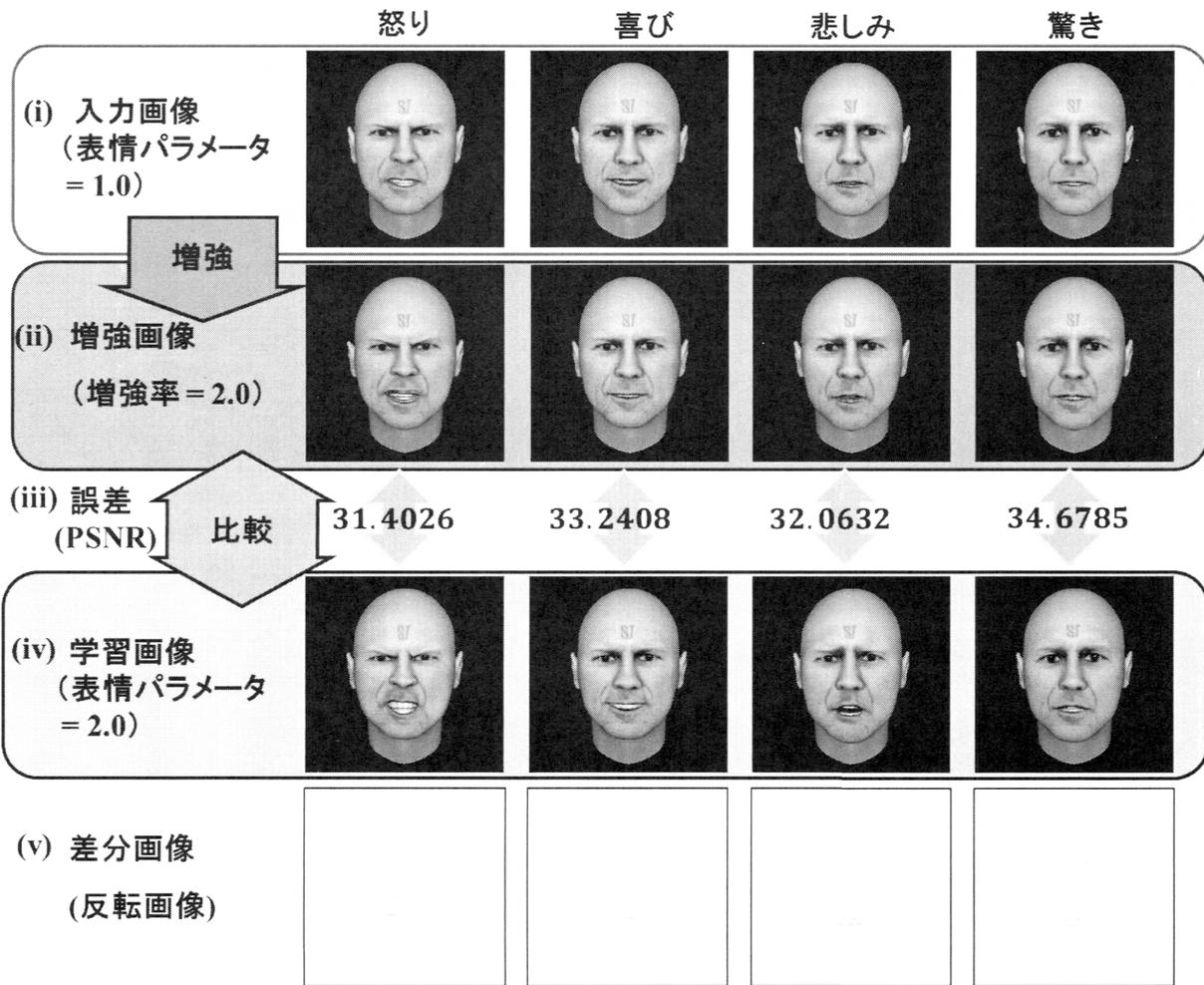


Fig.6 CGにおける表情増強の結果

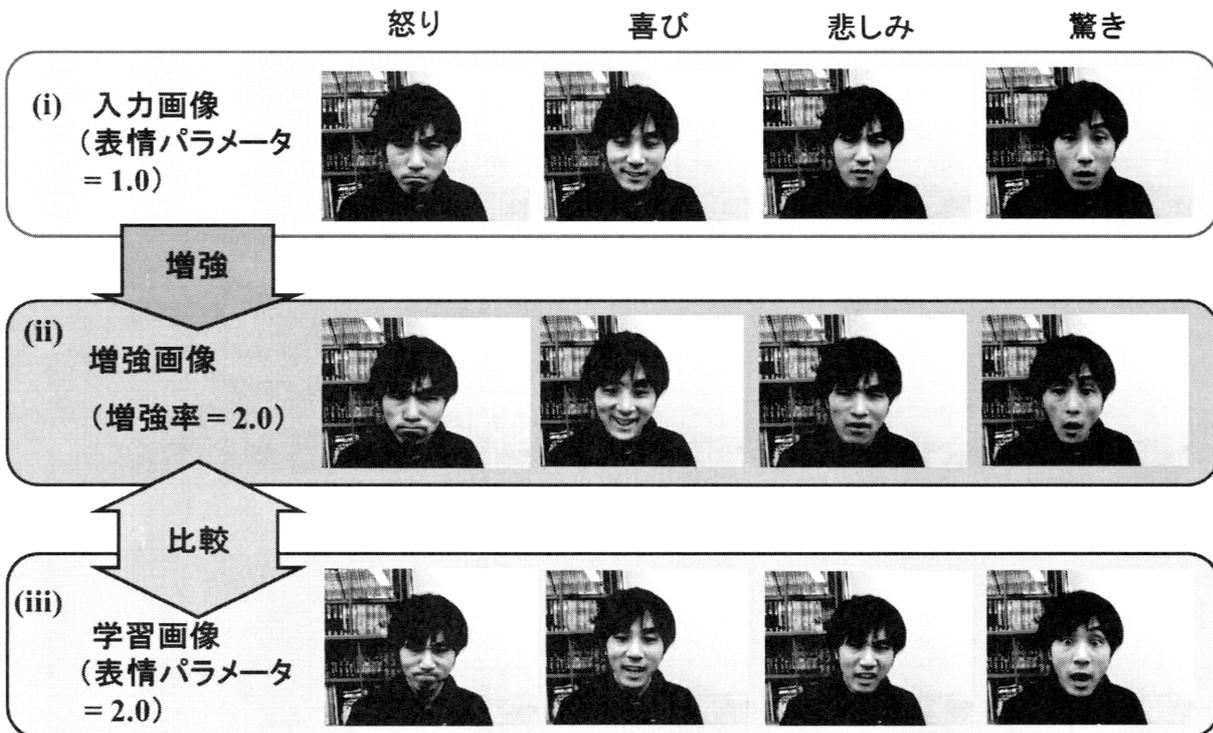


Fig.7 映像中の顔画像に対する増強結果

Table 1. 被験者1~12に対する提示

	1	2	3	4	5	6	7	8	9	10	11	12
Subject1	A1a	A1b	A1c	A1d	B2a	B2b	B2c	B2d	C3a	C3b	C3c	C3d
Subject2	D2b	D2c	D2d	D2a	A3b	A3c	A3d	A3a	B1b	B1c	B1d	B1a
Subject3	C3c	C3d	C3a	C3b	D1c	D1d	D1a	D1b	A2c	A2d	A2a	A2b
Subject4	B1d	B1a	B1b	B1c	C2d	C2a	C2b	C2c	D3d	D3a	D3b	D3c
Subject5	A2a	A2b	A2c	A2d	B3a	B3b	B3c	B3d	C1a	C1b	C1c	C1d
Subject6	D3b	D3c	D3d	D3a	A1b	A1c	A1d	A1a	B2b	B2c	B2d	B2a
Subject7	C1c	C1d	C1a	C1b	D2c	D2d	D2a	D2b	A3c	A3d	A3a	A3b
Subject8	B2d	B2a	B2b	B2c	C3d	C3a	C3b	C3c	D1d	D1a	D1b	D1c
Subject9	A3a	A3b	A3c	A3d	B1a	B1b	B1c	B1d	C2a	C2b	C2c	C2d
Subject10	D1b	D1c	D1d	D1a	A2b	A2c	A2d	A2a	B3b	B3c	B3d	B3a
Subject11	C2c	C2d	C2a	C2b	D3c	D3d	D3a	D3b	A1c	A1d	A1a	A1b
Subject12	B3d	B3a	B3b	B3c	C1d	C1a	C1b	C1c	D2d	D2a	D2b	D2c

“A”~“D” は映像中の人物の種類を表す。

“1”~“3” は増強倍率を表す (1: そのままの映像, 2: 増強倍率 $k=1.5$, 3: 増強倍率 $k=2$)。

“a”~“c”は、表情の種類を表す。

Q1. 発話者はどのような表情を表出していたように感じましたか?
 a. 怒り b. 喜び c. 悲しみ d. 驚き e. 無感情 f. その他()

Q2. 発話者の感情は伝わりやすく感じましたか?
 0. そう思わない 1. あまりそう思わない 2. どちらでもない 3. ややそう思う 4. そう思う

Q3. 発話者の表情に違和感を覚えましたか?
 0. そう思わない 1. あまりそう思わない 2. どちらでもない 3. ややそう思う 4. そう思う

Fig.8 主観評価実験で用いた質問

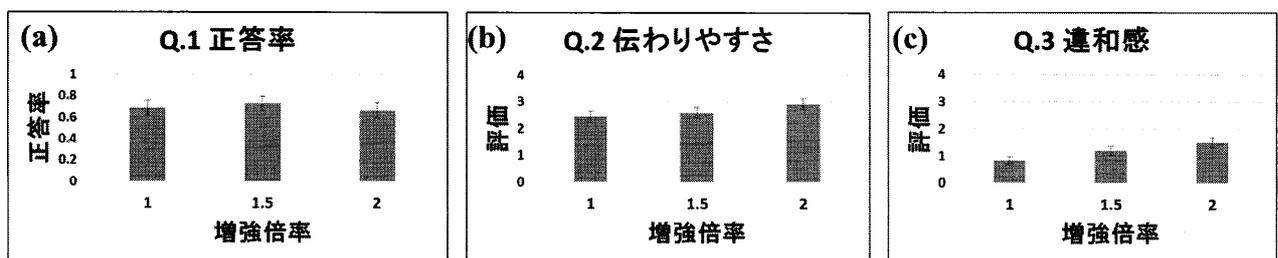


Fig.9 実写画像を用いた表情操作に関する主観評価実験の結果

Fig.6 (i)に、表情変化スケールが1.0のCG画像群を示す。この画像に対し提案手法を用いて操作(2倍に増強)した結果、画像群(ii)が生成された。FaceGen Modelerでは、各表情操作に伴うCGモデルの変形(頂点の移動)が、予めモデル化されているため、表情の表

出具合を「n倍に増強」のように線形操作可能である。本実験では、その特性に着目しFaceGen Modelerで生成した画像を真値データとする。生成画像群(ii)と、表情変化スケール2.0のCG画像群(iv)との差分画像を(v)に示す。(iii)は、差分画像中の顔領域において算出し

た Peak Signal to Noise Ratio (PSNR) の値である。

PSNR は、画像圧縮などの画像同士の類似度を定量的に評価するために使用され、原画像と加工画像の差異が小さいほど値が大きくなる特性がある。どの表情変化においても、PSNR は大きな値を維持していることから、生成画質が高いことが確認できる。(v)差分画像において差異を生じている領域に注目すると、各表情変化に特徴的な部位に差分が出ている。このことから、生成画像における見え方の変化が適切な箇所が発生していることがわかる。

6.1.2 実写画像を用いた表情操作実験

あるビデオシーケンス中の、表情変化の乏しい画像に対し、提案手法(表情の増強処理)を施し、豊かな表情を有するシーケンスを生成する。この際、表情変化の基準表情のパラメータベクトルは、被写体に喜び、怒り、悲しみ、驚きの表情を各2段階表出するよう指示しながら撮影した8枚の顔画像に基づいて設定する。AAMのパラメトリック空間は、上述した8枚の画像に無表情1枚を加えた9枚の顔画像を用いて構築する。実験に用いた画像サイズは、全て640画素×480画素である。画像の入力から、表情操作、映像提示までの処理は、Intel Core-i7 processor を搭載した計算機を用いて行い、平均動作速度は毎秒15フレームであった。

Fig.7 に実験結果を示す。提案手法を用いて入力画像(i)に写った顔の表情を2倍に増強した結果が(ii)生成画像である。(ii)の生成画像と、実際に表情が大きく表出した画像(iii)とを比較すると、類似した表情変化が再現されていることが確認できる。なお本実験では、ビデオ会話での利用を想定し、被写体を完全に固定せず、自由な姿勢をとらせた状態で増強処理の比較実験を実施している。

6.2 表情操作画像を用いた主観評価実験

6.2.1 表情増強に対する評価

提案手法による表情増強により、表情の伝達がどの程度円滑に行えるようになるかを検証するため、23-25歳の男性12名を被験者として主観評価実験を実施する。実験で観察する動画は、23-25歳の男性4名を撮影した動画画像に対し、操作倍率を1.0, 1.5, 2.0の3種類設定して生成したものをを用いる。

各被験者には、Table1 に示すように、無作為に抽出

した12パターンの動画画像を提示し、Fig.8に示す質問に回答してもらう。Q1では、表出していると感じた表情を答え、Q2では感情が伝わりやすく感じたか、Q3では発話者の表情に違和感を覚えたかどうかを0から4の5段階で評価するよう指示した。

各質問に対する回答結果をFig.9に示す。Fig.9(a)は、各々の増強倍率に対して、正しい表情を認識できた確率の平均値と標準偏差を示している。Fig.9(b), (c)は、Q2, Q3に対する評価値の平均値と標準偏差を示している。増強倍率1.0と2.0の場合に対する質問結果について、有意水準0.05でMann-WhitneyのU検定を行った。Q1の有意確率は0.699と有意差は認められなかったが、Q2の有意確率は0.119となり有意差が認められた。このことから、増強倍率が上がるにつれ表情の伝わりやすさが向上することが確認された。

一方Q3では、有意確率0.012で有意差が認められ、表情が大きくなるに伴い違和感も大きくなることが確認された。「非現実的な顔の形になっている場合は違和感が増す。」というコメントが被験者からあり、実際の顔の表情変化では起こりえない見え方の変化が、違和感の原因と考えられる。その一つの要因として、映像中の現実の顔とCGモデルによる仮想顔が幾何学的に不整合となることがあげられる。これまでも顔の位置姿勢および形状変化を実時間で計測・推定し、顔のCGモデルを重畳提示する研究が行われているが[19][20]、表情を増強させた顔のCGモデルの重畳においては、幾何学的不整合の問題がより顕著に現れている。本提案手法では、顔CGモデルの全ての頂点に対して、同じ条件で変形処理を施しているが、表情変化の印象度と違和感への影響(例えば顔全体でのパーツの位置関係)のバランスを考慮し、頂点毎に変形処理を設定することで、より有効な表情増強処理の実現が可能であると考えている。

本実験の結果では、表情の伝わりやすさが向上しているにも関わらず正解率が低下している。パラメトリック空間を介して、表情認識を行わずに表情を増強するため、基準表情と入力画像の射影点間のベクトルの方向が、学習時に与えた方向と異なる場合、このような問題が現れやすいが、学習に用いる表情変化のバリエーションを増やすなどの対応により、軽減可能であると考えている。

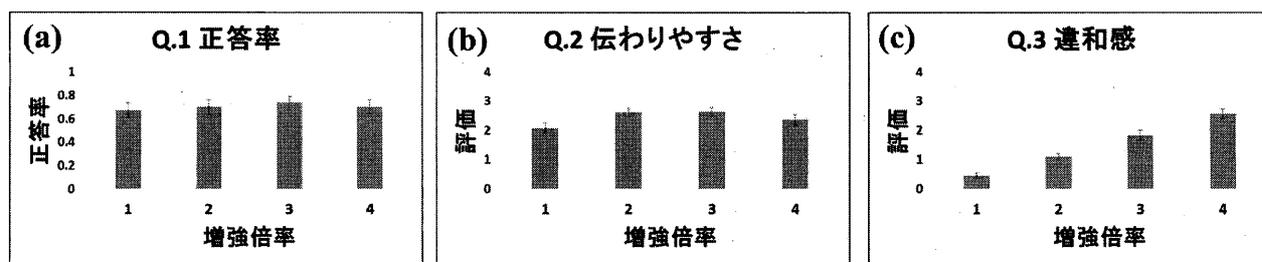


Fig.10 表情表出の乏しい顔画像における増強の効果に関する主観評価実験の結果

6.2.2 表情表出の乏しい顔画像における増強の効果

6.2.1 節の実験で表情の正答率の低かった2種類の人物の映像を、提案手法を用いてより大きく表情を操作(増強倍率を2, 3, 4倍に設定)した場合、正答率がどのように変化するかを調査する。被験者は23-25歳の男性16名であり、質問項目は、6.2.1 節と同様のものを使用する。被験者は、ある被写体の表情4種類と倍率4種類の動画を無作為な順序で観察し、Fig. 8の質問に回答する。

評価値の平均と標準偏差をFig.10に示す。増強倍率が大きくなるにつれ(c)の違和感の評価値が高くなっている一方で、(a)の正答率、(b)の伝わりやすさについては、増強倍率を大きくしていくと正答率、評価値共に増加傾向にあるものの、増強倍率が4倍に達すると、評価値が低下している。表情表出の乏しい顔画像に対して顔画像の増強処理を施す場合、顔形状の変化が小さいため、パラメトリック空間における基準表情と入力表情間の表情変化ベクトルの推定が不安定になる。そのようなベクトルに対して、比較的大きな倍率による表情増強処理を行うと、実験で意図していなかった表情が表出し、性能低下を招いたことが考えられる。また、大きく首を左右に振るなど入力画像中の顔の角度が大きい場合、パラメトリック空間におけるパラメータベクトルの大部分が姿勢変動の記述に割り当てられてしまうため、表情変化の記述性能が低下するという問題も考えられる。

7 おわりに

本稿では、画像処理技術を用いて顔の表情を操作し、それをAR提示することにより、豊かな表情表出を伴った映像コミュニケーションメディアの実現を目的とした研究について紹介した。様々な表情画像を用いた事前学習により、表情変化を実現するパラメトリック空間を構築し、入力された顔画像をその空間に射影して得られるパラメータを操作することで、表情操作を実現する手法を提案し、表情操作機能を有するビデオ通話システムを実装した。

CG画像データセットを用いた画質評価実験により、

表情変化の小さな顔画像から、大きな表情の画像の合成に成功したことを確認した。また、主観評価実験により、表情が乏しく伝わりにくい表情画像を増強することが、被験者の表情認識を助けることを確認した。

参考文献

- [1] Mehrabian A. "Silent messages." Wadsworth, Belmont, California, (1981)
- [2] Nakaguchi T., Tsumura N., Takase K., Makino T., Okaguchi S., Usuba R., Ojima N., Miyake Y. "Color-Enhanced Emotion," ACM SIGGRAPH 2005, Emerging Technologies, (2005)
- [3] Te-Feng Su, Chih-Hsueh Duan, Shu-Fan Wang, Yu-Tzu Lee, Shang-Hong Lai, "Automatic Facial Expression Exaggeration System with Parallelized Implementation on a Multi-Core Embedded Computing Platform", Journal of Signal Processing Systems, Volume 75 Issue 2, pp 155-168, 2014
- [4] Murakami K., Koshimizu H., Nakayama A., Fukumura T. "Facial Caricaturing Based on Visual Illusion--A mechanism to evaluate caricature in PICASSO system--." IEICE TRANSACTIONS on Information and Systems, (1993), 76.4: 470-478.
- [5] Woodworth R. S., Schlosberg H. "Experimental psychology." Oxford and IBH Publishing, (1954).
- [6] Schlosberg H. "A scale for the judgment of facial expressions." Journal of experimental psychology, (1941), 29.6: 497.
- [7] Schlosberg H. "Three dimensions of emotion. Psychological review." (1954), 61.2: 81.
- [8] Calder A. J., Young A. W., Rowland D., Perrett D. I. "Computer-enhanced emotion in facial expressions." Proceedings of the Royal Society of London. Series B: Biological Sciences, (1997), 264.1383: 919-925.
- [9] Cootes T. F., Taylor C. J. "Active shape models--'smart snakes'." BMVC. (1992): 266-275.
- [10] Cristinacce D., Cootes T. F. "Feature Detection and Tracking with Constrained Local Models." BMVC. (2006): 929-938.
- [11] Cootes T. F., Edwards G. J., Taylor C. J. "A Comparative Evaluation of Active Appearance Model Algorithms." BMVC. (1998). 680-689.
- [12] Gross R., Matthews I., Cohn J. F., Kanade T., Baker S. "The CMU Multiple Pose, Illumination and Expression (Mul-tiPIE) Database." Technical report, Robotics Institute, Carnegie Mellon University, (2007).

- [13] Pantic M., Alstar V., Rademaker R., Maat L. "Web-based database for facial expression analysis." Proc. IEEE Int'l Conf. on Multimedia and Expo, (2005).
- [14] Alstar V., Pantic M. "Induced Disgust, Happiness and Surprise: an Addition to the MMI Facial Expression Database." Proceedings of the International Language Resources and Evaluation Conference, (2010): 65-70.
- [15] ManyCam LLC "ManyCam free virtual webcam effects software" <http://www.manycam.com/2014/3>
- [16] Microsoft "Skype - Free internet calls and online cheap calls to phones and mobiles" <http://www.skype.com/en/2014/3>
- [17] Singular Inversions "FaceGen Modeller: 3D Generator", <http://www.facegen.com/modeller.htm> 2014/3
- [18] Saragih J. M., Lucey S., Cohn J. F. "Face alignment through sub-space constrained mean-shifts." Computer Vision, IEEE 12th International Conference, (2009): 1034-1041.
- [19] 竹村雅幸, 大田友一, "協調型複合現実空間のための顔映像生成による視線情報の復元", 日本バーチャルリアリティ学会論文誌, Vol10, No3, pp.313-321, Sep. 2005.
- [20] K Dale, K Sunkavalli, MK Johnson, D Vlastic, W Matusik, H Pfister, "Video face replacement", Proceedings of the 2011 SIGGRAPH Asia Conference, Volume.30 Issue.6, Article No.130, 2011

2008-2009 年度日本バーチャルリアリティ学会論文章受賞.
パターン認識国際連盟フェロー, 情報処理学会フェロー.
第 21 回パターン認識国際会議(ICPR2012)ゼネラルチェア.

(2014 年 3 月 10 日受付)

[著者紹介]



佐藤 翔悟 (非会員)

筑波大学大学院システム情報工学研究科
博士前期課程 2014 年 3 月修了。在学中、画
像情報研究室に在籍し、表情増強の研究に
従事。



北原 格 (正会員)

1996 年筑波大学大学院理工学研究科修了。
同年シャープ (株) 入社。2000 年筑波大学
先端学際領域研究センター助手。2003 年
ATR 研究員。2005 年筑波大学大学院システ
ム情報工学研究科講師, 2008 年同准教授, 2001 年電子情報
通信学会学術奨励賞。2003 年 IEEE VR2003 Honorable
Mention, 2009 年度本会論文賞など受賞, 博士(工学)。



大田 友一 (正会員)

1977 年筑波大学大学院博士課程了。京都
大学情報工学科助手, 筑波大学電子。情報
工学系講師, カーネギーメロン大学計算機
会科学科客員研究員, 筑波大学電子・情報
工学系助教授を経て, 1992 年同教授 2004 年同大学院シ
ステム情報工学研究科教授。工学博士, コンピュータビジョ
ン, 視覚情報メディア, 複合現実感の研究に従事。