

表情伝達を目的としたアバターチャットシステム

森永 英文[†], 松本 吉央[†], 小笠原 司[†]

奈良先端科学技術大学院大学[†]

1. はじめに

近年、インターネットの急速な普及により、コミュニケーションの在り方が変化してきた。ネットワークを使ったコミュニケーションの向上が求められている。コミュニケーションにおいて、相手に与える印象の50%以上が表情からであるといわれており、コミュニケーションを向上させるには表情を伝達することが必要である。

しかし、インターネットの匿名性やプライバシーの問題等から顔を知られたくないといった問題からビデオチャットシステムは普及していない。そこで、ユーザの表情を仮想人物アバターに投影することでコミュニケーションを行うことができるアバターチャットシステムが開発されている。このようなシステムでは、アバターを動かすパラメータのみを転送すれば良いため、画像をそのまま転送するのに比べて非常に少ない転送量ですむという利点もある。

本研究では、従来の多くのシステム([1]など)のように、ユーザの表情を認識し、あらかじめ用意しておいたいくつかの表情にアバターを変化させるのではなく、顔のパーツを各々計測し、その結果をアバターに直接反映させることにより表情を伝達させるというアプローチをとる。

本稿では、表情が伝達できるチャットシステムの実現に向けて、ステレオカメラを用いてユーザの頭部・眼球・眉毛・口唇の各運動をそれぞれ計測し、計測された運動を3次元のモデルに投影し表情の再現を行うシステムについて述べる。

2. システム概要

Fig.1 に、開発したアバターチャットシステムの実験システムの外観を示す。本システムは、顔情報を計測するためのステレオカメラ、音声を取得・再生するためのマイク・スピーカー、これらを制御するためのPCで構成される。

本システムは、以下の5つの機能からなる。

- 1 顔向き・視線方向の計測
- 2 眉毛運動の計測
- 3 口唇運動の計測
- 4 アバターによる表情の再現
- 5 音声の伝達

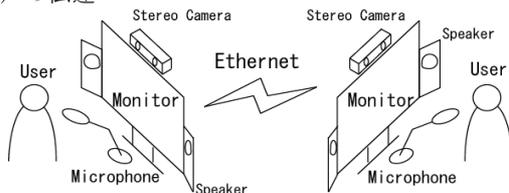


Fig.1: システム外観

2-1. 顔向き・視線方向の計測

我々は、これまでにステレオカメラからの画像を用いて、ユーザの顔向き、視線方向を実時間で検出するシステムを開発してきた[2]。このシステムは非接触であるため、ユーザに負担をかけることなく自然な動作を検出することができる。

2-2. 眉毛運動の計測

眉毛を、眉頭・眉山・眉尻をつないだ折れ線として定義する。ただし、眉尻はほとんど動かないことから眉頭・眉山を計測することで眉毛の運動の計測とする。

眉山と眉頭の画像は、あらかじめテンプレートとして登録しておき、左右のカメラ画像からテンプレートマッチングによりその位置を検出する。ステレオ画像上での位置を用い三角測量の原理から眉頭と眉山の3次元位置を求められる。

頭部を動かしながら眉毛を上下に動かす実験を行った結果を Fig.2 に示す。眉毛の上下運動に対して、安定して眉頭と眉山が計測できていることがわかる。

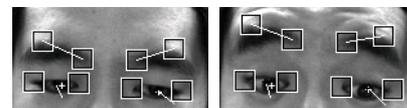


Fig.2: 眉毛の追跡結果

2-3. 口唇運動の計測

口唇運動の計測は、唇の外輪郭を計測することで行う。唇輪郭の検出手法には、トラッキングした唇の特徴点を補間する方法[3]や Snake を用いて検出する方法[4]などがある。前者の方法は、処理時間が短い精度が低く、後者の方法は精度は高いが処理時間がかかる。

そこで、本研究ではトラッキングする唇輪郭上の特徴点の数を増やし、唇輪郭を上端・下端・右端・左端をつないだ4つの曲線に分割し、各特徴点をテンプレートマッチングにより追跡を行う。特徴点の位置に曲線をフィッティングすることで輪郭検出を行う。通常のテンプレートマッチングでは、特徴の少ない唇輪郭上の点を安定して検出することは困難であるが、特徴的な唇の両端をまず見つけ、両端位置から探索範囲を細かく分割することで残りの特徴点の探索を可能にした。

頭部を動かしながら”あえいうえおあおん”と発音する実験を行った。実験では、追跡する特徴点の数を8~24個(4個刻み)で変化させた。また、特徴点のフィッティング手法として、(1)特徴点を直線でつないだ線形補間、(2)唇の輪郭曲線を2次関数と考え最小自乗によるフィッティング(曲線同士の両端を接続)、(3)3次のスプライン関数により補間を行う Cubic-Spline、(4)特徴点の数により曲線の次数を変化させるラグランジュ補間の4つ

“Avatar Chat System for Facial Expression Transfer”

[†]Hideaki Morinaga, Yoshio Matsumoto, Tsukasa Ogasawara
Nara Institute of Science and Technology

を試みた。

特徴点 16 個の場合の実験結果を Fig.3 に示す。Fig.3(a)を見ると、トラッキングが成功している場合は、どの補間方法を用いても輪郭形状に大差がない。しかし、Fig.3(b)のようにいくつかの特徴点のトラッキングが失敗している場合には、輪郭に誤差が生じる。実験の結果、特徴点 16 点・各曲線の端点を固定した 2 次関数によるフィッティング(3)が最も安定して輪郭を抽出できた。そこで、後述の表情の再現実験では、この手法を用いる。

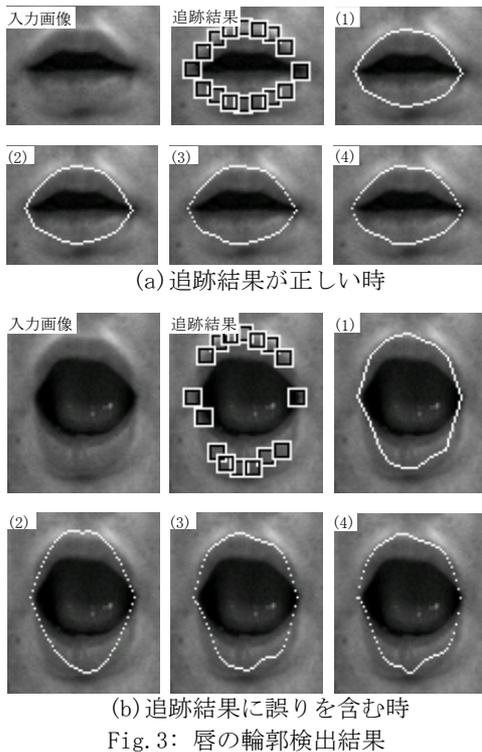


Fig. 3: 唇の輪郭検出結果

2-4. アバターによる表情の再現

あらかじめ、口を閉じ、正面を向いた時の計測結果をアバターの初期状態として与えておく。

顔向きと視線方向の計測結果は、頭部・眼球にそのまま反映させ、眉毛の運動・口唇運動は、初期状態からの変位量を計算し、眉毛や唇を構成するポリゴンの頂点データに反映することで再現させる。顔の情報計測部から送られてくる、頭部の姿勢や向きなどアバターを動かすのに必要なパラメータは、16bit の整数 44 個で表すことができ、これらをビデオレートで転送するのに必要な通信速度は以下ようになる。

$$44 \times 16(\text{bit}) \times 30(\text{Hz}) = 21.1(\text{kbps})$$

これは、80×60 のサイズの画像を転送する時の約 1/50 に相当する。

2-5. 音声伝達

音声伝達部では、音声をマイクから取り込み、一般的な電話の音質である 8kHz、モノラル、8bit でサンプリングを行う。サンプリングされたデータは、Socket 通信により転送され、受信側のスピーカーから再生される。

3. 表情の再現実験

本システムを用いた表情の再現実験の様子を Fig.4 に示す。Fig.4 の各図は、左が計測結果、右がアバターによる表情の再現である。Fig.4(a)は、アバターに与えた初期状態であり、Fig.4(b)は、口を開いた時、Fig.4(c)は、眉毛を上げた時の結果である。各 3 つの状態の計測結果がアバターに正しく反映されているのがわかる。

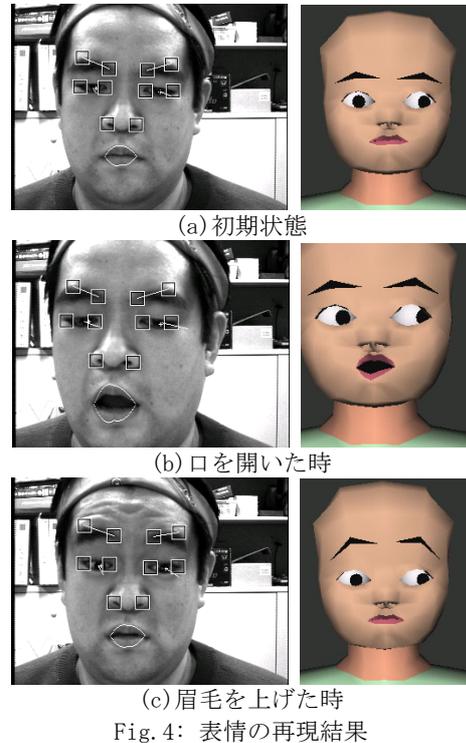


Fig. 4: 表情の再現結果

4. おわりに

表情が伝達できるチャットシステムの実現に向けて、ユーザの頭部・眼球・眉毛・口唇の各運動をそれぞれ計測し、計測された運動を 3 次元のモデルに投影し表情の再現を行うシステムについて述べた。実験結果より、本システムを用いて、顔の各情報を計測する事ができ、アバターに反映できることを確認した。

さらに計測できる特徴を増やすとともに、今後は開発したアバターチャットシステムの評価を行っていく。

参考文献

- [1] Kohtaro Ohba, Takehito Tsukada, Tetsuo Kotoku and Kazuo Tanie: Facial Expression Space for Smooth Tele-Communications, FG'98, pp.378-383, 1998.
- [2] Y.Matsumoto, T.Ogasawara, A.Zelinsky: Behavior Recognition Based on Head Pose and Gaze Direction Measurement, IROS2000, pp.2127-2132, 2000.
- [3] Ying-li Tian, Takeo Kanade, Jeffrey F.Cohn: Robust Lip Tracking by Combining Shape, Color and Motion, ACCV2000, pp.1040-1045, 2000.
- [4] Patrice Delmas, Nicolas Eveno, Marc Lievin: Towards Robust Lip Tracking, ICPR2002, pp.528-531, 2002.