

顔・注視点計測を用いた知的コンピュータインタフェース

松本 吉央 竹村憲太郎 上田悦子 小笠原 司

奈良先端科学技術大学院大学 情報科学研究科 ロボティクス講座
〒630-0101 奈良県生駒市高山町 8916-5
Email: yoshio@is.naist.jp

あらまし 相手が何を見ているか、どのくらいの時間見ているか、どんな表情をしているか、など視線をはじめ顔から得られる情報は、人間どうしのコミュニケーションにおいてもノンバーバルな情報として非常に重要な役割を果たしている。これらのノンバーバルな情報を簡単かつ精度よく計測することができれば、ユーザの「意図」や「感情」といったものをコンピュータが理解することにつながる。これによりユーザの負担を軽減することができれば、「気が利く」知的なインタフェースを持つ機器が実現できるであろう。我々はこれまでに顔の動きや視線をリアルタイムに計測する技術を開発してきた。本稿では、その計測技術を応用し、知的なインタフェースを目指して構築しているシステムについて述べる。

キーワード ヒューマンインタフェース、顔トラッキング、視線検出、実時間画像処理

Intelligent Computer Interfaces Based on Face and Gaze Measurement

Yoshio Matsumoto, Kentaro Takemura, Etsuko Ueda and Tsukasa Ogasawara

Robotics Lab, Graduate School of Information Science, NAIST
8916-5 Takayamacho, Ikoma city, Nara 630-0101 Japan
Email: yoshio@is.naist.jp

Abstract The information obtained by observing a face of a person (e.g. what the person is looking at, how long, with what expression...) plays an important role as non-verbal information in human-human communication. If a computer system is capable of measuring such non-verbal information, it will be useful for the system to recognize the intention and the emotion of its user. This will reduce the burden of the user, which will eventually lead to realizing attentive interfaces. In this paper, novel intelligent computer interfaces are described, which are based on our face measurement system for facial motion and gaze direction.

keywords Human Interface, Face Tracking, Gaze Detection, Real-time Image Processing

1 ヒューマンインターフェースとしての顔情報

人間の視線の動きは、人間の意図や注意と関係が深い。この性質を利用して、視線をキーボードやマウスの代わりの入力デバイスとして用いたり、機器の操作として用いるといった次世代インターフェースの研究が盛んにされている。人の視線は他のポインティングデバイス（マウス等）と比較して以下の利点を持つ。

1. 手を使うことなくすばやく移動させることができる
2. 操作する対象を見るという行為は自然で直感的であり、疲労が少ない
3. コンピュータの画面上だけでなく、実世界にある対象を指すこともできる

また、人の顔の動きも視線に追従して動くことが多い。そのため、視線を正確に計測するのが困難な状況でも、顔の向きを計測することができればある程度視線の方向を推測することができる。

また顔の動きは「うなずき」などのジェスチャーを表現する手段としても用いられる上、表情もノンバーバルな感情の表現として用いられるなど、顔から表出される情報は人同士のコミュニケーションにおいて、非常に表現力に富んでいる。そのため、遠隔の人とコミュニケーションを取る場合にも、顔情報を計測することは有効な情報伝達手段になると考えられる。

本研究では、視線、顔の動き、眉や唇の動きを計測できるシステムを用いて、以下のインタフェースを構築する。

1. 視線を用いた機器操作の支援
2. 遠隔コミュニケーションの支援

これらは単一のユーザの顔情報の計測に基づいたシステムであるが、さらに本稿では複数人の注視情報を推定する試みについても述べる。

2 顔情報の計測システム

2.1 概要

顔の情報をヒューマンインタフェースに利用するには、非接触な計測装置が望ましく、装着感の強い頭部装着式のもの（Eye Camera による視線計測や Polhemus による頭部位置・姿勢計測など）は適していない。

本稿で用いる計測システムは、ステレオカメラを用い、非接触かつリアルタイムで顔の情報を計測す

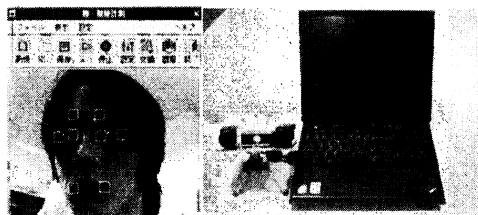


Figure 1 : Face Measurement System

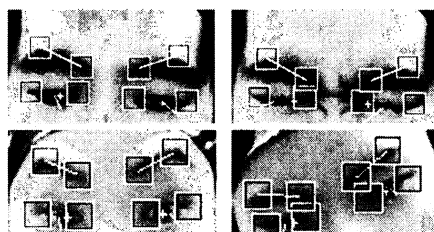


Figure 2 : Tracking of Eyebrows.

ることができるものである [1]. テンプレートとして眉毛や目、唇の両端等を利用し、現在の計測結果に 3 次元の顔モデルをフィッティングさせることにより顔の位置・姿勢を計測し、また 3 次元ベクトルとして視線方向を計測することができる (Fig. 1 (左)). 計測精度は画像の解像度に依存するが、 320×240 のステレオ画像を用い、その中に顔が大きく写っている場合、顔の並進・回転が $2[\text{mm}]$, $2[\text{deg}]$ 、視線方向が $5[\text{deg}]$ 程度である。また現在は、Fig. 1 (右) に示すようにノート型 PC を用いた安価かつポータブルなシステムが実現されている。

2.2 眉・口唇の計測

従来の計測システムでは、顔を剛体として扱っており、追跡されている点（目尻、目頭、口の端点など）の位置の変化は誤差とみなされていた。しかし眉や口の変形は、表情や発話を認識するのに重要である。そこで、顔全体の位置・姿勢を計測したのちに、それらの位置のずれを計測する処理を加えた。

Fig. 2 は眉のトラッキング結果を示す。眉の midpoint のテンプレート画像は横方向に似たパターンが続くため、トラッキング位置が安定しない。そこで探索範囲は縦方向にのみを設定した。

また、口唇にも多数のテンプレートを設定し、その輪郭をトラッキングする (Fig. 3)。眉の midpoint と同様に、(口唇の両端以外の) テンプレートの探索範囲は縦方向のみである。しかし唇は変形が大きいので、

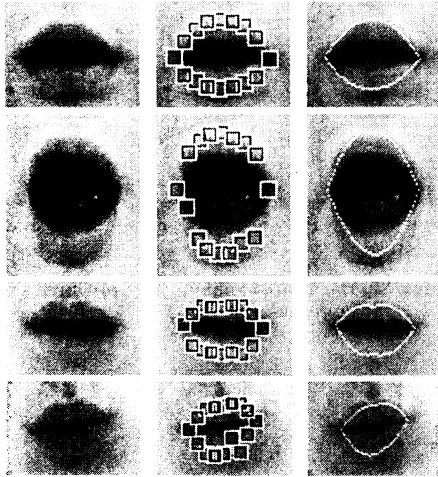


Figure 3 : Tracking of Lip.

それでもテンプレートマッチングの失敗が起きうる。いくつか失敗したテンプレートがある場合でも口唇の輪郭を安定して認識できるように、テンプレートマッチングの相関値を重みとして用いる最小自乗法により、唇の上下合計4本の2次曲線をフィッティングさせることにした。

3 視線を用いた機器操作の支援

3.1 PC切り替え

従来、複数台のPCを1組のインタフェース（キーボード、マウス）で操作する場合にはキーボード切り替え器や、モニタをまたいだマウスの移動が利用されている。しかし、タッチタイピングができるユーザにとっては、このような切り替え操作は作業の中断を生じさせる。モニタを2台並べたデスクトップ環

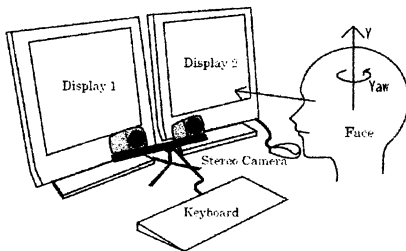


Figure 4 : Gaze-based PC switch system.

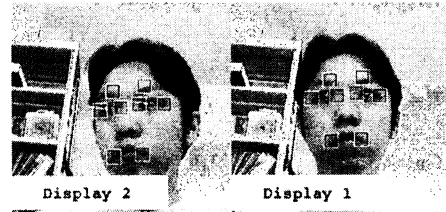


Figure 5 : Estimation of user's attention.

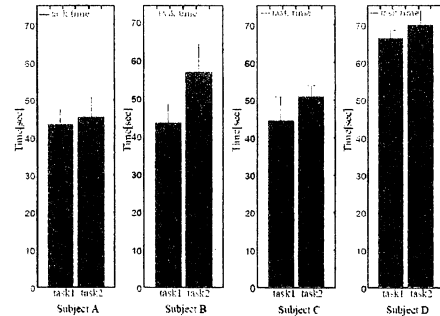


Figure 6 : Evaluation of PC switch system.

境は近年一般的になりつつあるが、そのような環境で顔情報を用いてPCを自動的に切り替えるシステムを構築した [2]。そのシステム構成を Fig. 4 に示す。予備実験の結果、2台並んだモニタをそれぞれ見たときに頭部方向に十分な差がでることが分かったため、操作している対象の推定は、単純に頭部の水平方向の回転量から行う (Fig. 5)。2台のPCではOSとしてそれぞれLinuxとWindowsXPが走っており、PC切り替えプログラムの実装にはVNCプロトコルを用いている。

Fig. 6 は、構築したシステムの評価実験の結果である。被験者は以下の2つの方法で操作対象のPCの切り替えを行う。

1. Task1: 頭部方向によりPCを切り替える (提案手法)
2. Task2: マウスの移動によりPCを切り替える (従来手法)

被験者4人（ある程度タッチタイピングができる）には、それぞれ10ワードずつ交互に2台のPCに入力してもらおうというタスクを5回ずつ行ってもらい時間を計測した。その結果、全ての被験者において提案手法の方が達成時間が短かく、作業効率を向上できたことが確認された。

3.2 ユニバーサルリモコンへの拡張

今日では家電製品の数が増え、リビングに何台ものリモコンが置かれていて、操作したい機器のリモコンをいちいち探すという状況も珍しくない。SONYのGaze-Link[3]やHITACHIのAir-real[4]は、リモコンを向けたコンピュータや機器を操作できるというインタフェースを実現している。またShellらのEyePliance[5]では「見ているかどうか」に基づき操作する家電製品のインタフェース（みていないテレビは消えるなど）を構築している。我々も同様に、見ているものが操作対象になる（What you look at is what you control）というコンセプトのインタフェースを、ユーザの視線情報に基づき構築した[2]。

2台のPCを切り替える場合と違い、Fig. 7に示すように対象物の位置が複雑である場合には、顔向きだけでは正確に操作対象を識別することは難しい。そこで、ユーザの頭部位置 P_h および視線方向 g を用いて以下の式により交差判定を行い注視している操作対象を識別する。

$$\left\| \frac{(P_{a_i} - P_h) \cdot g}{\|g\|^2} g + P_h - P_{a_i} \right\| \leq L_i$$

ここで P_{a_i} は i 番目の機器の位置、また L_i はそのサイズ（半径）である。全ての機器の形状は、扱いを簡単にするため球とみなす。この式が満たされている場合は視線ベクトルと i 番目の機器が交差していることになる。

システムの実装に際しては、各機器が情報家電のようにネットワーク経由で操作できれば、視線を認識しているPCから操作指令を送るだけで機器を操作できることになる。しかしそのような機器はまだ一般的でないため、ここでは各機器をPCから制御できるように改造した。改造方法には以下の2通りがある。

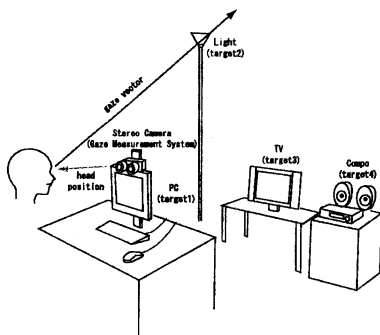


Figure 7 : Gaze-based Universal Remote Control.

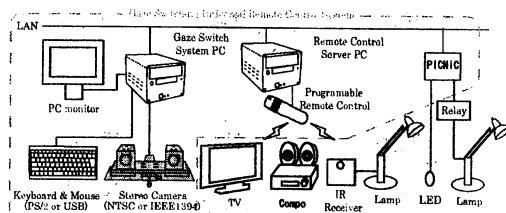


Figure 8 : Hardware Configuration of Universal Remote System.



Figure 9 : Experiment on Implemented System.

1. リモコンが付属している機器は、学習リモコンのクロスサム2+USB(スギヤマエレクトロン)を経由してPCから操作信号を送る。
2. そうでないものは、PICNIC(PIC Network Interface Card)のI/Oポートに接続したリレーを経由して操作対象の機器のスイッチ等を操作する。

Fig. 8の点線内が機器操作のためのシステムである。また、現在の操作対象機器をユーザに明示することが有効であると考え、各機器にはLEDを取り付け、ユーザがそちらを見ていると認識されている場合に点灯するようにした。

各機器の操作方法としては、PC作業との親和性や、PCへの信号の取り込みやすさを考慮し、マウスを利用することにした。また、ライト、テレビ、CDプレーヤを操作する場合のマウスのボタンへの機能の割り付け例を以下に示す。ここで各機器のOn/Offは共通とし、多機能で操作が複雑な（＝リモコンのボタンが多い）機器についてはその一部だけを実装している。

開発したインタフェースの評価を行った結果をFig. 10に示す。ここでは、ユーザの前方に配置された4台のライトを、指示された通りに点灯させるというものである。この実験では、4台の個別のリモコンを用いたもの、学習リモコン（操作する前に操作対

Table 1 : Assignment of Mouse Buttons

	ライト	テレビ	CD プレーヤ
左ボタン	On/Off	On/Off	On/Off
中ボタン	-	モード切替	モード切替
右ボタン	-	消音	Play/Stop
ホイール	-	(Vol/Ch) ±	(Vol/Track) ±

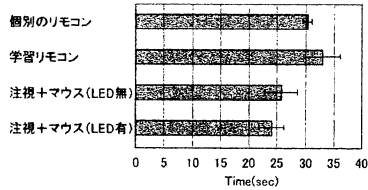


Figure 10 : Result of Evaluation Experiment.

象によってモードを切り替える), 注視とマウス操作によるもの(機器に取り付けられたLEDが, ユーザの注視に反応して点灯する場合としない場合の2通り)の4通りの操作を行った. 各操作ではそれぞれ10回ずつライトを点灯させた. その結果, 提案手法(注視とマウス操作とLED点灯を併用するもの)が最も操作時間が短くなり, 手法の有効性が示された.

4 遠隔コミュニケーションの支援

近年, 遠隔コミュニケーションの1つとしてビデオチャットが広まっている. ビデオチャットは相手の表情が分かり円滑なコミュニケーションが可能である反面,

1. 自分の顔や背景の様子が見られてしま, プライバシーや匿名性を守れない
2. 通信量が多くテキストや音声に比べて大きな通信帯域が必要

という問題点もある. そこで, 実画像でなくCG画像を用いてチャットするアバターチャットに, 顔計測を

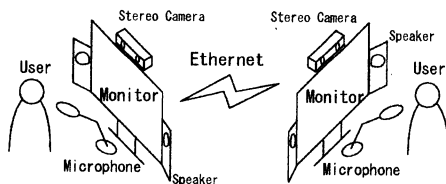


Figure 11 : System Configuration for Avatar Chat.

応用したシステムを構築した [6]. システムの概要は, Fig.11 に示す通り遠隔の2台のPCがネットワークで接続されており, それぞれの端末に顔計測用のステレオカメラが接続されている.

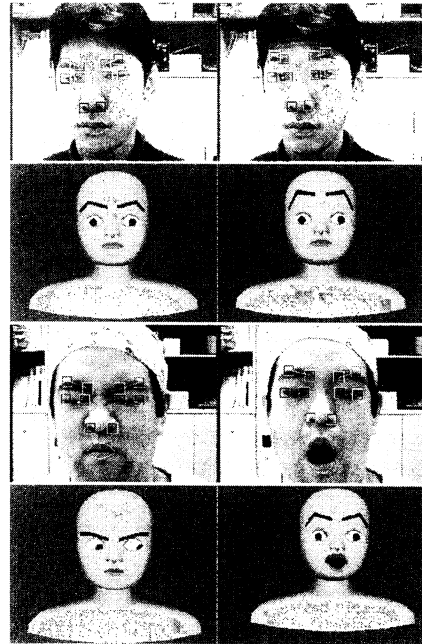


Figure 12 : Measured Face and Animated Avatar

Fig.12は, 計測した顔とそれを用いて表示したアバター画像である. 通信データ量としては, 画像をそのまま (80×60 [pixel] のグレースケール画像を 30[Hz] で) 送る場合に 1.2[Mbps] であるのに対し, 計測された顔情報のパラメータだけ送る場合は 18[kbps] となり, 大幅に通信量を削減できている.

評価実験として, あるユーザの顔をランダムに変化させその様子をアバターに投影した場合に, 各被験者がどの程度認識できるかを調べた. 認識させる表情は以下の4表情である. この実験では約80%の表情が正しく認識されており, 遠隔コミュニケーションへの顔情報の応用が可能であることが分かった.

Table 2 : Result of Expression Recognition

	認識された表情			
	無表情	喜び	怒り	驚き
提示: 無表情	10	1	2	2
提示: 喜び	3	3	4	0
提示: 怒り	1	0	9	0
提示: 驚き	0	0	0	15

5 複数人による「場の注意」の推定

これまでに述べたシステムは、それぞれ単一のユーザの顔情報を計測し、インタフェースに利用していた。これに対して同時に複数のユーザの顔情報を計測し利用する研究も現在行っている [7]。

複数人（あるいは集団）のいる環境で注意の集まる場所を、我々は「場の注意」と呼んでいる。場の注意は、日常生活の中でもミーティング、学校での授業、ゲームなど知的共同作業中では重要なコミュニケーションの要素である。場の注意を計測できれば、

- その場の臨場感のある画像を、ビデオに記録したり、遠隔地からの参加メンバーへ送るための自動カメラ撮影
- メンバーの注意状態や集中度を記録する（例えば授業の）評価システム

などへ応用可能であろう。

Fig.13 は、4 人の顔を同時計測するために試作したカメラシステムである。このシステムで計測された各自の顔向き情報を統合して、最小自乗法により最も注意の集中していると思われる点を推定した結果を Fig. 14 に示す。図中にはカメラの配置、各自の顔向き方向と同時に「場の注意」の位置およびその大きさを表す楕円（“Person2” の下にある楕円（実際には楕円体））が示されている。楕円の大きさは、全員の注意が 1 点に集中しているほど小さくなる。この推定結果をオンラインで用いる遠隔ミーティング支援システムを現在開発中である。



Figure 13 : Measurement System of Group Attention.

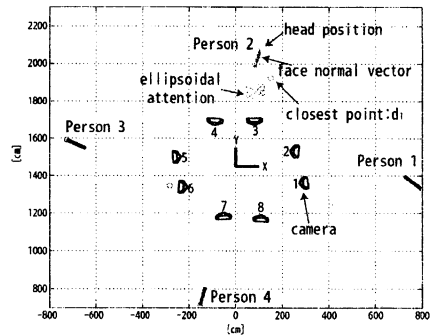


Figure 14 : Estimated Group Attention.

6 おわりに

本稿では、人間の視線や頭部の動き、表情などの情報をリアルタイムに計測することで可能になったインタフェースについて述べた。

今後は、複数人の注意推定を用いて、日常生活や遠隔コミュニケーションの支援を行うシステムの研究を行っていく予定である。また、本稿で述べたコンピュータインタフェースへの応用だけでなく、認知心理学、人間工学、ロボットインタフェース、アミューズメント等、様々な分野へ顔情報の計測技術を応用していく予定である。

References

- [1] Yoshio Matsumoto et al.: "Portable Facial Information Measurement System and Its Application to Human Modeling and Human Interfaces," The 6th International Conference on Automatic Face and Gesture Recognition (FG2004), pp.475-480, 2004.
- [2] K.Takemura, et al.: "What You Look at Is What You Control: A Universal Remote Control Based on Gaze Measurement Technology," 1st IEEE Technical Exhibition Based Conference on Robotics and Automation, 2004.
- [3] 綾塚ほか: "実世界指向ユーザインタフェースにおける「見ているものに接続する」というメタファ", 情報処理学会論文誌, Vol.42, No.6, pp.1330-1337, 2001.
- [4] <http://hhil.hitachi.co.jp/products/airreal.html>
- [5] J.Shell, et al.: "EyePliances and EyeReason: Using Attention to Drive Interactions with Ubiquitous Appliances," Extended Abstracts of UIST 2003
- [6] 森永英文ほか, "表情伝達を目的としたアバターチャットシステム", 情報処理学会 第 66 回全国大会講演論文集, Vol.4, pp.239-240, 2004.
- [7] 竹村憲太郎ほか, "複数人の視線計測に基づく「場の注意」の推定", 情報処理学会 ヒューマンインタフェース研究会研究報告 HI-110, pp25-30, 2004.