

リアルタイム顔・視線計測システムの開発と 知的インタフェースへの応用

松本吉央[†] 怡土順一[†]
竹村憲太郎^{††} 小笠原司[†]

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、人間同士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良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、その計測システムの開発およびそれを応用して構築した知的なインタフェースについて述べる。

Development of Real-time Face and Gaze Measurement System and Its Application to Intelligent Interfaces

YOSHIO MATSUMOTO,[†] JUN'ICHI IDO,[†] KENTARO TAKEMURA^{††}
and TSUKASA OGASAWARA[†]

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.

1. はじめに

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に入力デバイスとして用いたり、機器の操作として用いたりするといった次世代インタフェースの研究がさかんに行われている。人の視線は他のポインティングデバイス（マウスなど）と比較して以下の利点を持つ。

- (1) 手を使うことなく、素早い移動が可能である。
- (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。さらに顔の動きは「うなずき」などのジェスチャを表現する手段としても用いられるうへ、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とる場合にも、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。

我々は、このようなヒューマンインタフェースの高度化を目的として、視線、顔向きをはじめとした様々な顔情報を計測するためのシステムを開発してきた。本稿では、2章で顔・視線計測の関連研究について述べ、次に3章では開発したリアルタイム顔・視線計測システムについて述べる。4章では顔計測に基づく行動認識の手法について述べる。5章では、それらを応

[†] 奈良先端科学技術大学院大学
Nara Institute of Science and Technology

^{††} 奈良工業高等専門学校
Nara National College of Technology

用して構築してきたヒューマンインタフェースに関するアプリケーションについて概説する。最後に6章でまとめと今後の課題について述べる。

2. 関連研究

従来より、視線計測装置としては頭部に小型カメラと赤外LEDを装着して眼球を撮影し、照射された赤外線の影響反射や強膜反射の情報から視線方向を検出するアクティブ方式のアイカメラ^{1),2)}が開発、商品化され、人間工学、心理学などの研究分野で広く用いられてきた。これらの手法では高精度な視線計測が可能であるものの、頭部に装着する装置にかなりの重量があったり、利用者に大きな拘束感、違和感を与えたりするため、人間の自然な状態での行動を計測することが難しいという欠点があった。据え置き型のものであるが、カメラの視野が狭い、キャリブレーションがずれるといった問題から頭部を固定した状況でしか使えないものが多い^{3),4)}。

また、頭部を動かすことのできるように改良された据え置き型のものとして、伴野ら⁵⁾のシステムでは、マーカ付きの眼鏡をかけ顔のトラッキング処理を単純化したうえで、視線方向の検出を行った。Ohnoらは顔全体の位置を計測する広角カメラと目を拡大して計測するパンチルトカメラを併用し、顔の動きに対応した視線計測装置を開発し⁶⁾、商品化もされている。しかしアクティブである点は変わらず、またサイズが大きく高価であることから普及はしていない。また、小型サーチコイルを内蔵したコンタクトレンズを装着する方法、眼球の電位差を計測する方法(EOG法)などの視線計測技術も開発されているが、ともに被験者の負担が大きいため一般のヒューマンインタフェースで利用するのは困難である。

視線以外の顔に関する情報としては、頭部の位置・姿勢(6自由度)を計測する装置として、磁界を発生するトランスミッタとレーザセンサを利用したシステムが一般的に用いられている。しかし、磁界を乱す金属の近くでは計測精度が低下するという問題がある。また機械的に頭部位置を計測するリンク式は拘束感が強い。頭部にマーカを取り付けてモーションキャプチャにより動きを計測することもできるが、非常に高価かつ大がかりなシステムになる。

一方、ヒューマンインタフェースに用いることを目的に、通常のカメラを用いて、非接触かつパッシブで顔の動きや視線方向を計測する手法も数多く研究されてきた。Azarbayejaniら⁷⁾、Zelinskyら⁸⁾は、単眼カメラで顔全体を写し、特徴点をトラッキングし、カルマン

フィルタを利用して顔の位置・姿勢を検出している。また、Active Appearance Model, Adaptive Appearance Model⁹⁾やGabor Wavelet Networks¹⁰⁾などを用いて非剛体として顔をトラッキングする手法も近年研究されているが、顔のモデリングが容易でなかったり、6自由度の動きを推定する処理が複雑でリアルタイム性に欠けるなどの問題点がある。ただし、このように顔全体を写すアプローチでは、顔の動きや視線方向以外にも、表情などに関する顔情報を計測するには適している。

我々は、3次元の視線方向、顔の位置・姿勢に加え、瞬き、眉や唇の動きなどの情報を、

- モーションキャプチャ技術のような定量的な情報として、
- 非接触、マーカレスかつ受動的(赤外線を投射しない)に、
- リアルタイム(ビデオレート)で、
- カメラとPCだけの小型かつ安価な構成により、計測できるシステムの開発を1998年にはじめた¹¹⁾。本システムの開発においてはリアルタイム・ステレオビジョンの技術を用い、顔全体をカメラで撮影して処理することを特徴とする。同様に顔情報の計測にステレオカメラを用いる研究としては岡ら¹²⁾、Morencyら¹³⁾の顔の動き計測があり、いずれもリアルタイム性に優れているが、この研究分野の中では少数派である。

3. 顔情報の計測システム

我々の開発した顔情報計測システムは、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ることができるものである¹⁴⁾。テンプレートマッチングにより顔特徴の3次元の動きを追跡し、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。ハードウェアとしては、図1に示すような2台のカメラとPCを用いる。カメラにはNTSCもしくはIEEE1394で外部同期がかかるものを、またPCのスペックとしてはPentiumIII 500 [MHz]以上のものを用いれば30 [Hz]での計測が可能である。さらに高速なPCとIEEE1394の高速カメラ(VidereDesign社 Mega-D)を用いると80 [Hz]での計測も可能である。

3.1 3次元顔モデル

顔の位置・姿勢計測に用いる顔モデルの例を図2に示す。このモデルは顔特徴(目、眉、口の両端などテンプレートマッチングしやすい特徴)のテンプレート



図 1 顔情報計測システム外観

Fig. 1 Overview of Face Measurement System.

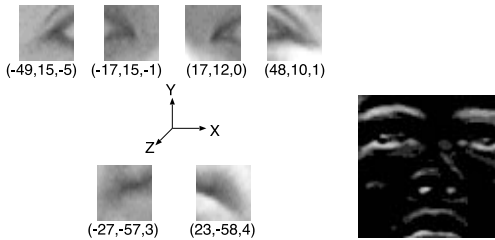


図 2 3次元顔位置・姿勢計測に用いるモデル

Fig. 2 Facial model used for 3D head pose estimation.

画像と3次元座標から構成されている。また、計測開始時に画面中から顔を発見するためには図2右に示すような顔全体のテンプレート画像を用いる。この画像は処理速度を上げるために低解像度にし、照明変化の影響を受けにくくするためにエッジ画像としてある。

なお、顔の3次元形状は個人ごとに微妙に異なるため、計測精度を上げるためには顔モデルを個人ごとに作成する必要がある。この登録を手動で行う場合、画面上の顔特徴をマウスでクリックしていく。その特徴の対応点を、テンプレートマッチングによりもう一方の画像上で探索することで、3次元座標が求まる。この登録作業は、1人につき1分以内に終わる程度の簡単なものである。

3.2 顔位置・姿勢の計測

前節で述べた3次元顔モデルと、時々刻々得られる3次元観測値を用いて、信頼性に基づく重み付けを考慮したモデルフィッティング処理を行うことにより、高速に頭部の位置・姿勢6自由度を求めることができる。各特徴の3次元観測値は、顔モデルに含まれるテンプレート画像を用いたマッチング(正規化相関)により求められる。フィッティングは以下の評価式に従って行う。

$$E = \sum_{i=0}^{N-1} w_i \|R x_i + t - y_i\|^2$$

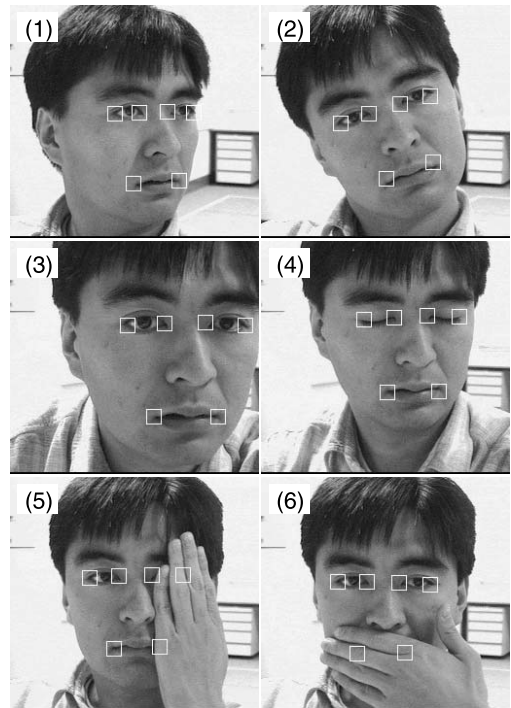


図 3 顔位置・姿勢計測の結果

Fig. 3 Result of 3D face tracking.

ここで、 N は特徴の数、 x_i は顔モデル中の特徴の位置ベクトル、 y_i は対応する特徴が観測された位置ベクトル、 w_i は観測の信頼性(相関演算の相関値を用いるため、 $0 \leq w_i \leq 1$ となる)、 t, R は並進ベクトルおよび回転行列、そして E はフィッティングエラーである。最も良いフィッティングを行うことは、 E を最小化する t, R を求める問題に帰着でき、本研究ではこれを最急降下法を用いて解いている。

上記の計算により3次元位置・姿勢を求めるのに最低限必要な特徴の数 N は、(一直線上に並んでいない)3点である。それ以上多い特徴を用いる場合は冗長になるので、計測はより安定する。6点程度を用いると、図3の(4)~(6)に示すように一部の特徴が隠れたり、表情により特徴の見え方が変化したりした場合でも計測が可能となる。計測精度は画像の解像度や写っている顔の大きさ、表情の変化などに依存するが、 320×240 のステレオ画像を用い、その中に顔全体が写っている場合、顔の並進・回転が 2 [mm], 2 [deg] 程度である。

画像処理を行うPCにPentium4 2.53 [GHz]を用いて、特徴の数が6点、各特徴領域の大きさが 16×16 画素、各特徴領域の探索範囲が 32×32 画素の場合、顔位置・姿勢の推定には約 10 [ms] かかる。また、顔を

見失った場合の復帰に必要となる全画面での顔発見処理は、顔全体のテンプレート画像を用いて約 70 [ms] かかる。

3.3 視線方向の計測

本研究では、視線方向を推定するために必要となる眼球のモデルについて、眼球は球形で、その中心が回転中心と一致すると仮定する。この単純化により、視線方向は頭部の位置・姿勢と、目の虹彩の位置との関係により決定される。さらに、眼球の半径と虹彩の半径を個人に依存しない定数と仮定する。実際に眼球の半径は、およそ 13 [mm]、虹彩の半径は 7 [mm] でこれらのパラメータはほとんど個人差がないことが知られている。これにより、個人ごとに設定しなければならないパラメータは「顔モデルに対する眼球の中心の相対位置」だけとなる。顔モデルに対する眼球の中心の相対位置は、左右の目でトラッキングに用いられている両端の特徴の midpoint から眼球へ向かう 3 次元ベクトルとして与えられ、「オフセットベクトル」と呼ばれる。このパラメータは、既知の方向を見たときの視線の向きが合うようにキャリブレーションされる。

3 次元の視線方向を計測する手順を図 4 に示す。まず (1) に示すように、眼球中心の 3 次元位置（実際には見えていない）は、頭部の位置・姿勢の計測結果とオフセットベクトルを用いて決定される。眼球の位置が決まると、カメラ画像上での眼球の大きさ（これも実際には見えていない）は、眼球の形状を透視変換することで得られる（図 4 (2)）。また、虹彩の中心は二値化やハフ変換により求められる（図 4 (3)）。この処理は、すでに前節で得られている画像上の目の両端の間の小領域についてだけ行えばよいから、処理時間は

わずかである。

こうして求めた画像上での眼球の中心と半径、それに虹彩の中心の位置関係のみにより、視線の 3 次元方向が決まる。図 4 (4) は、この考えに基づいて視線の水平面上の方向 θ が決められる様子を示す。同様にして視線の垂直面上の方向も決められ、その 2 つのパラメータにより視線の 3 次元空間中での方向が表現される。本システムでの視線のモデリングは幾何学的に表現されておりシンプルである。しかし実際には、このようなモデリングで視線方向を計測するには、高い精度で頭部の位置・姿勢が求まっている必要がある。そのため、従来の研究ではこのようなアプローチをとっているものは少なく、頭の位置・姿勢の変動は微小であると仮定したうえで、虹彩の画像上での位置とモニタなどの平面上での視点の位置を直接対応付ける方法をとるものが多かった。

ステレオ画像中には合計で 4 つの目の画像があり、それぞれから 3 次元の視線方向を得ることができる。しかし、カメラは顔全体を写しているため、目のサイズ（横幅）は通常数十画素、虹彩の半径は数画素しかない。このような低い分解能では、左右の眼球の視線の輻輳角から交点を求め、注視点の奥行きを求めることは難しい。そこで、本システムでは奥行きを求めることはあきらめ、単純に 4 つの視線ベクトルの平均をとることでノイズの影響を減らし、1 つの視線ベクトルを求めることにした。

図 5 は、視線方向の計測実験の結果のスナップショットである。計測結果の 3 次元の視線ベクトルは、顔のトラッキング結果の画像上に重ねて表示されている。視線方向の計測誤差は顔の向きや画像の解像度によ

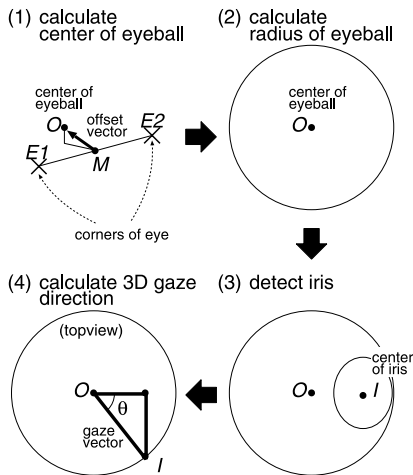


図 4 視線のモデリング

Fig. 4 Modeling of gaze direction.

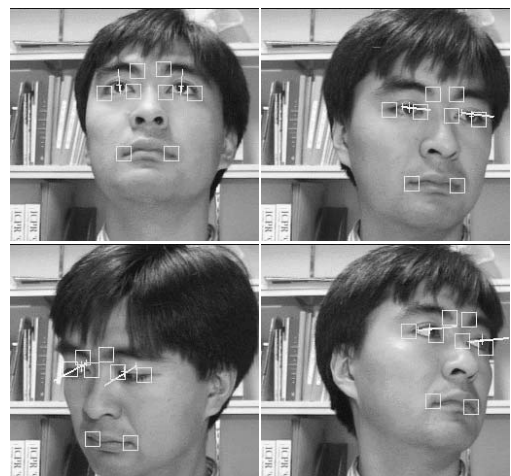


図 5 視線方向の認識結果

Fig. 5 Result of gaze measurement.

て変化するが、 320×240 画素の画像を用いて顔がカメラ正面方向を向いている場合、最大でも 5 [deg] 程度 (1 [m] 先にある 10 [cm] 間隔の注視対象を十分に識別できる程度) である。

3.4 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見なされていた。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。

図 6 は眉のトラッキング結果を示す。眉の中点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 (図 7)。眉の中点と同様に、(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ため、それでもテンプレートマッチングの失敗が起こりうる。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

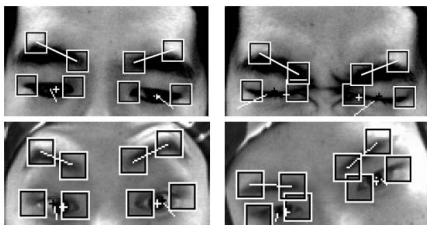


図 6 眉のトラッキング

Fig. 6 Tracking of eyebrows.

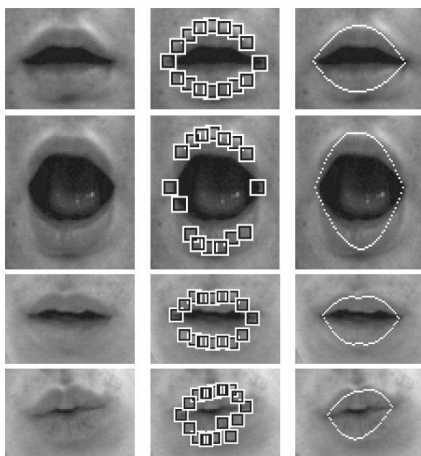


図 7 唇のトラッキング

Fig. 7 Tracking of lip.

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。

4. 顔情報に基づく行動認識

4.1 ジェスチャの認識

本システムでは顔の動きも検出できるため、顔の動きによるジェスチャ表現を認識することも可能である¹⁵⁾。特に、ビデオレートでかつ高い精度で動き検出ができることから、従来のジェスチャ認識システムでは困難であった小さく速い (つまりオーバアクションでない自然な) ジェスチャも認識ができることが期待できる。現状のシステムでは、“Yes”, “No”, “ZoomIn/Out”, “Winking” などが認識できる。ジェスチャの認識には連続 DP マッチング¹⁶⁾ を用いた。図 8 (左) は “Yes”, “No” を行っているときの顔向き計測結果と CDP 値である。CDP 値が小さい部分 (網掛け部) がそれぞれのジェスチャを認識していることを示している。また図 8 (右) はジェスチャを認識しているときの処理結果である。

4.2 注視対象の認識

頭部位置と視線方向が同時に求められると、環境の配置があらかじめ分かっていたら視線の先の注視対象を認識することができる¹⁵⁾。頭部位置を p_h 、視線方向を g とすると、その判別式は以下の式で表現できる。

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

ここで p_{a_i} は i 番目の対象の位置、また L_i はそのサイズ (半径) である。すべての機器の形状は、扱いを簡単にするため球と見なしている。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。図 9 (上) に実験を行ったデスクトップの環境を示す。注視対象としては、電話、キーボード、マウス、モニター、窓などがある。それぞれの 3 次元座標はカメラを原点とした 3 次元座標 (単位は [cm]) で表現されている。図 9 (下) はこのようにして認識された注視対象である。

5. インタフェースへの応用

5.1 コンピュータインタフェース

5.1.1 ハンズフリーマウス

これまでも、視線をマウスカーソルの代用として用いる研究はあった¹⁷⁾⁻¹⁹⁾。我々は、視線よりも顔の向きの方が随意的に動かしやすいことに注目し、顔の向きによってマウスカーソルを移動させ、左右のウィ

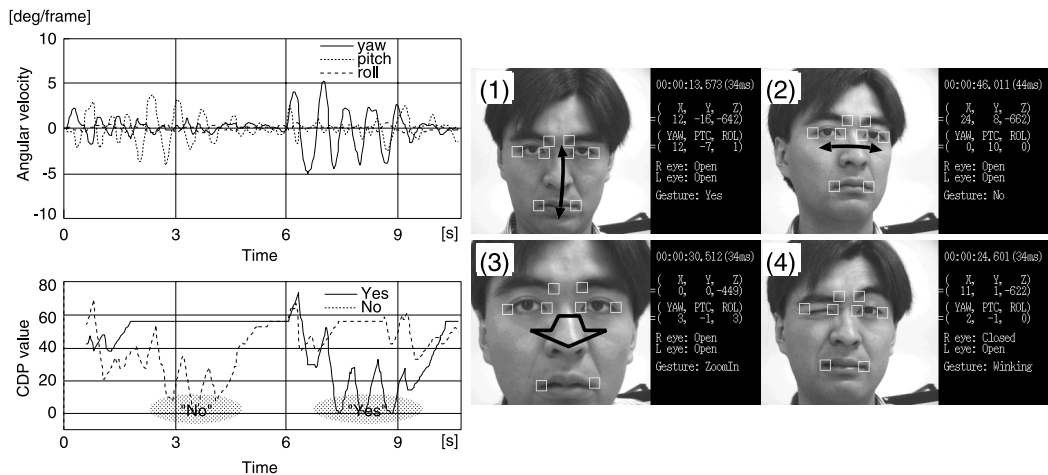


図 8 ジェスチャ認識結果：(1) Yes，(2) No，(3) 近づき，(4) ウィンク

Fig. 8 Experimental results of gesture recognition: (1) Yes, (2) No, (3) ZoomIn and (4) Winking gestures.

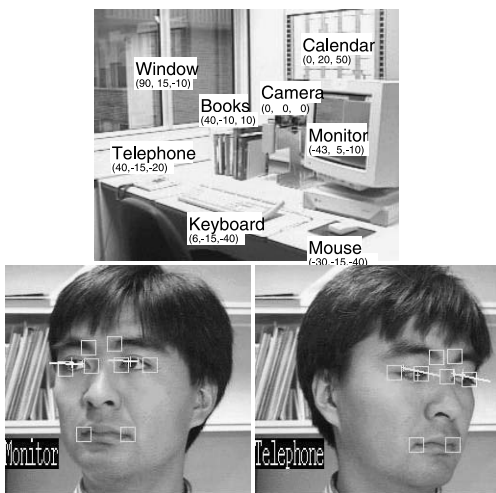


図 9 注視対象の認識結果

Fig. 9 Experimental results of gaze point detection.



図 10 頭部方向と瞬きを用了ハンズフリーマウス

Fig. 10 Hands-free mouse using head pose and winking.

リンクをマウスクリックの代用とすることで、手を用いずにコンピュータの画面上で任意のソフトウェアを使うことができるインタフェースを試作した。図 10 にそのインタフェースを用いてウェブブラウザを操作している様子を示す。

顔の向き(上下左右)は、2軸の単純なジョイスティック入力と同様、カーソルの速度に変換される。このときに5度程度の不感帯を設けたり、同じ向きの入力が続くと加速したりするなどの工夫を入れることにより、無意識な顔の動きに反応してしまうことが少なくなり、使い勝手は向上した。それでも小さなボタンを押したり、スクロールバーをつかんだりするには多少操作に

慣れる必要があり、またウィンクも連続して行うことは容易ではないため、現状では通常のマウス操作によるコンピュータ操作を代替できるレベルには達していない。

しかし、カーソルの動き特性を個人ごとにチューニングしたり、音声認識機能と組み合わせたりするなどの改良を行うことで、「マウスを手で操作することはできないが顔を動かすことは可能」というシチュエーションにおいて、コンピュータを利用するための実用的なハンズフリーインタフェースを構築できると考えている。

5.1.2 遠隔コミュニケーション

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がっている。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い、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チャットに，顔計測を応用したシステムを構築した²⁰⁾．システムの概要は，図 11 に示すとおり遠隔の 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．

図 12 は，計測した顔とそれを用いて表示したアバタ画像である．通信データ量としては，たとえば画像をそのまま (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には 18 [kbps] となり，通信量を大幅に削減できている．

評価実験として，あるユーザの顔をランダムに変化

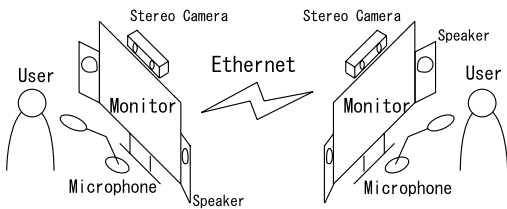


図 11 アバタチャットシステム

Fig. 11 System configuration for Avatar Chat.

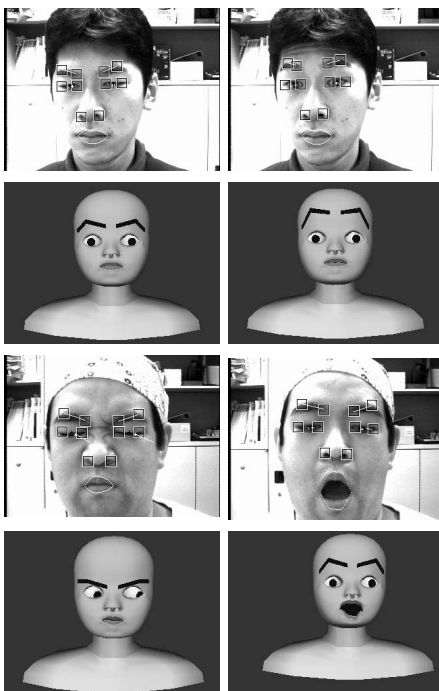


図 12 計測された顔と表示されるアバタ

Fig. 12 Measured face and animated avatar.

させ (無表情，喜び，怒り，驚きの 4 パターン)，その様子をアバタに投影した場合に，各被験者がどの程度認識できるかを調べた．その結果，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．

5.2 実世界ポインティングインタフェース

5.2.1 マルチディスプレイ環境のための PC 切替え

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²¹⁾．そのシステム構成を図 13 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出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 (図 14)．2 台の PC では OS としてそれぞれ Linux と WindowsXP を利用しており，PC 切替えプログラムの実装には VNC プロトコルを用いている．

評価実験として，頭部方向により切り替える (提案手法) とマウスの移動により切り替える (従来手法) の 2 つの方法を比較した．被験者 5 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

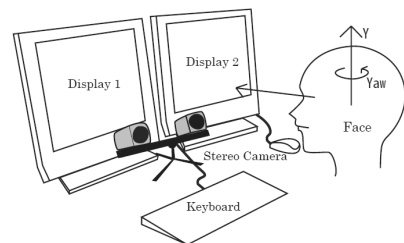


図 13 PC 切替えシステム

Fig. 13 Gaze-based PC Switching System.

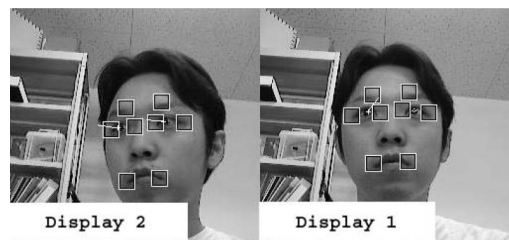


図 14 2 台の PC 利用時の認識結果

Fig. 14 Result of PC switching with two monitors.



図 15 視線を用いた機器操作

Fig. 15 Gaze-based Universal Remote Control.

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、すべての被験者において提案手法の方が達成時間が短かく(切替え 1 回あたり平均 0.38 [s]), 作業効率を向上させることができたのが確認された。

5.2.2 家電操作のためのユニバーサルリモコン

今日では家電製品の数が増え、何台ものリモコンから操作したい機器のリモコンを探すという状況も珍しくない。Shell らの開発した EyePliance²²⁾ 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(見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²¹⁾(図 15)。我々のアプローチでは、4.2 節で述べたように各機器の 3 次元配置とユーザの頭部位置・視線ベクトルから「見ている」ことが判別できるため、EyePliance のように各機器に視線検出用のカメラを取り付ける必要がない。また、ユーザの頭部位置の変化が小さい状況では、各機器とそこを見るとききの視線方向を直接対応付けてしまうことも可能である。

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。図 16 の点線内が機器操作のためのシステムである。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付け例を表 1 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

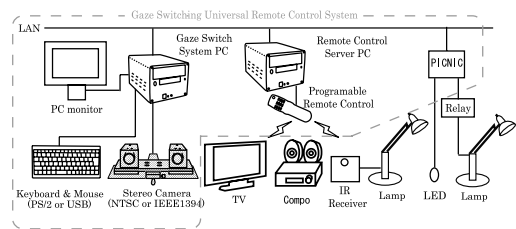


図 16 ユニバーサルリモコンのシステム構成

Fig. 16 Hardware configuration of Universal Remote System.

表 1 マウスボタンの割当て

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	-	モード切替え	モード切替え
右ボタン	-	消音	Play/Stop
ホイール	-	(Vol/Ch) ±	(Vol/Track) ±

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。

マウスのボタンだけで機器を操作しようとする、PC 画面上のマウスポインタの操作と区別し機器を選択するために「モード切替え」が必要になる。また機器が複数ある場合には、その中から操作対象を 1 つ選択する必要もある。これに対し、視線を用いた場合は実世界にある操作対象を自然に「見る」だけで、直感的に機器の選択が実現できる点が特徴である。

開発した視線インタフェースについて、機器ごとに別々のリモコンを用いて操作する場合と比較する評価実験を行った。5 人の被験者で、4 台の機器を対象にランダムに指示した機器の On/Off 操作を 10 回ずつ行った結果、1 回の機器操作あたり 0.47 [s] 早くなるという結果が得られ、有効性が確認された。

5.2.3 ドライバ支援

自動車のドライバの挙動を計測し、用いるインタフェースを構築した²³⁾。顔計測用カメラは、図 17(上)に示すように運転手前方に置かれた。ほかに、車外撮影用全方位カメラ、GPS、ジャイロセンサ、加速度センサなどが搭載されている。これらの車載計測システムの構築により、運転動作と頭部・視線情報の関係を記録・解析することが可能になった。図 17 は、全方位カメラによって撮影された道路環境の上に、視線方向と頭部方向をカーソルにより表示した例である。

また、計測した情報はオンラインでも利用可能であるので、ドライバの運転支援インタフェースにも用いることができる。たとえば、運転席の配置をあらかじめモデリングしておくことで、メータ類、ミラーなど



図 17 ドライバ計測カメラと視線の計測結果

Fig. 17 In-car camera system and experimental result of gaze measurement.



図 18 車内注視の計測結果

Fig. 18 Experimental result of in-car gaze point estimation.

の対象物を注視していることを検出することができる。図 18 は車内での注視対象認識の結果である。このうち、スピードメータを見た場合に、PC からスピードを読み上げてくれるという支援インタフェースを試作した。今後、注視方向や瞬きの履歴を用いれば前方不注意や居眠り防止の警告を行うなどのインタフェースも構築できるであろう。

なお、構築したシステムを用いて走行実験を行った結果、晴天時・昼間の走行では約 1 時間のうち 92% の区間で顔・視線の計測が可能であった。また、夜間はドライバーの顔情報を計測するのが困難になるが、近赤外対応の CCD カメラと、赤外線投光器を用いると、図 19 に示すように容易に計測できることが分かった。

5.3 ロボットインタフェース

5.3.1 視線操作型インテリジェント車いす

人は通常、行きたい方向を見る。このことを利用して、視線や顔の方向に基づいて走行を支援する車いす

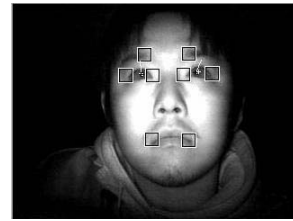


図 19 夜間の注視計測結果

Fig. 19 Experimental result of in-car gaze point estimation at night.

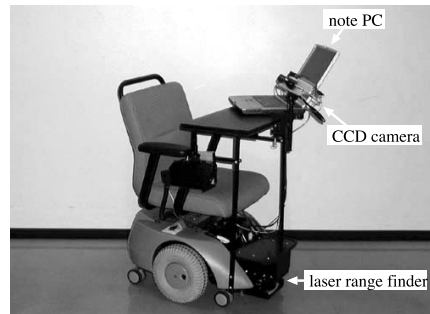


図 20 インテリジェント車いす外観

Fig. 20 Overview of Intelligent Wheelchair System.

システムを開発した^{24),25)}。図 20 に、開発したシステムの外観を示す。屋内用の電動車いす M-Smart (ミサワホーム製) をベースに作られている。搭乗者の前方に顔計測カメラが搭載され、また足もとにはレーザーレンジファインダ LMS 200 (SICK 製) を搭載し、ロボット前方 180 度の距離情報をスキャンすることで自己位置を認識したり障害物を検出したりすることができる。

この車いすには、顔・視線計測に基づく代表的な 2 つの走行モードがある。1 つは地図のない未知環境における走行のための“視線操縦モード”で、基本的にはユーザの顔向き方向に進む。レーザーレンジファインダを用いた障害物回避機能と組み合わせて用いられ、たとえば壁のポスタなどを見ているにもかかわらずそこに衝突することはなく、停止したり、壁に沿って滑らかに移動したりすることができる。

もう 1 つは、地図を持っている既知環境における“注視推定機能を持つ自動走行モード”で、目的地までの自動走行を基本としながら、搭乗者の注視状況により経路を変更/選択する。図 21 はこの走行モードでの実験の様子である。自己位置を認識し、その上に搭乗者の注視量の分布 (注視ヒストグラムと呼ぶ) が表示されている。地図上の特定の場所に注視が集中している場合に経路を変更してそこに近づいたり、また

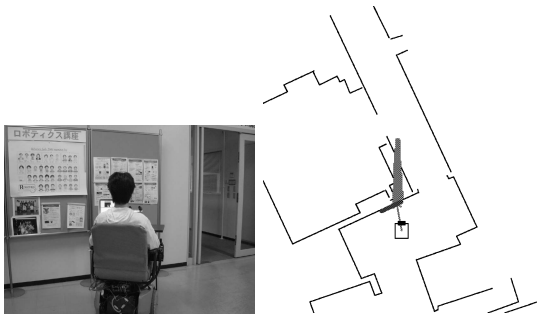


図 21 走行中の注意推定

Fig. 21 Estimated attention in experiment.

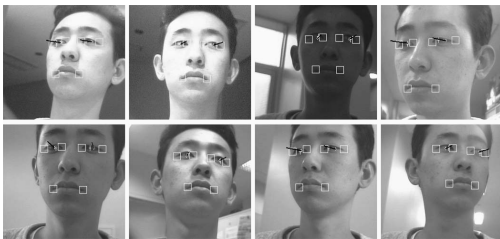


図 22 車いす搭乗時の顔計測の様子

Fig. 22 Face measurement on wheelchair.

経路の分岐点においてはユーザが注意を払っている方向を選択したりすることができる。

この車いすシステムは、これまでに学外のイベントを含め様々な場所においてデモを行い、多くの被験者に体験してもらっている。その結果、屋内・屋外を問わず様々な環境で走行が可能であることを確認しており、顔情報計測における画像処理の安定性の高さを示している(図 22)。

5.3.2 マルチモーダル対話インタフェース

我々はこれまでに、人間型ロボットを用いた対話システムの研究を行ってきた²⁶⁾。これは、受付に設置したロボットが、音声認識によりユーザの発話を理解し、ジェスチャを交えながら質問に答えるという学内の受付案内システムである(図 23)。しかし、従来のシステムでは、背景雑音の状況によっては応答すべきでないときに間違った反応をしてしまうことが頻繁にあった。顔情報計測を用いると、ユーザの顔や視線、口の開閉を認識することができる。これにより、視線や顔が自分に向いていることを確認したら音声入力をスタートさせ、また口の動きが止まったら音声入力をストップさせ、背景雑音に反応しないようなシステムを実現した²⁷⁾。ロボットがユーザの視線を意識しながらコミュニケーションを行うという意味で、ロボットと人間のある種の“アイコンタクト”が実現できたと

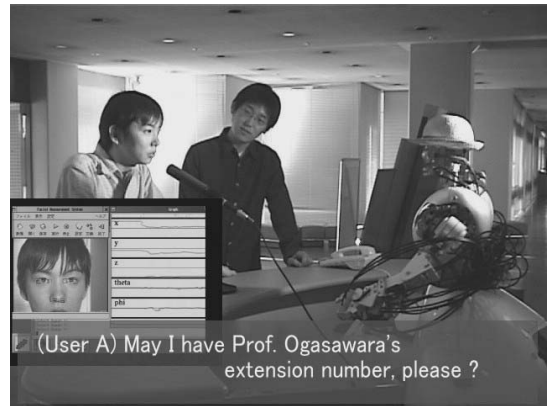


図 23 受付ロボット ASKA との対話の様子

Fig. 23 Eye-contact interaction with ASKA.

いえる。

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システムの開発と、それを用いて実現してきた様々なインタフェースについて述べた。開発したシステムの特徴は、

- 3次元的な顔向き、視線方向などの定量的な動きが計測可能、
 - 瞬き、眉や口の動きなども計測可能、
 - 非接触、マーカレスかつ受動的、
 - 実時間かつ実環境で動き、照明変動に強い、
 - カメラと PC だけの小型かつ安価なシステム、
- という特徴がある。これらは、
- 正規化相関演算により特徴領域をトラッキング、
 - ステレオ視により直接 3 次元情報を得ることで、フィッティングの問題を簡単化、
 - 顔の位置・姿勢が推定された後は、限られたごく小領域だけ処理を行い視線方向などの情報を得る、
- という工夫により実現できたと考えている。本システムにより、ユーザにいったいの負担をかけず自然な状態での実験が可能となるため、乳幼児から高齢者までの計測が可能となり、自然な人間の行動計測とそれに基づくヒューマンインタフェースの構築が可能になった。

我々は、究極的には人の顔を見て“その人の注意、意図、感情などを推定して気の利いた反応ができる”機器やロボットを実現したいと考えているが、現状では単に“ある対象を見ているかどうか”、“どの方向を見ているか”、あるいは“どんな顔の変形が起きているか”を計測してインタフェースに用いているにすぎない。今後はこのような表層的な情報から、人の内部状

態を推定するための“人”に関する研究が必要になっていくであろう。本稿で述べた計測システムは、心理学、人間工学の分野でも利用可能であるので、このような研究にもチャレンジしたい。

謝辞 本研究の一部は、奈良先端科学技術大学院大学 21 世紀 COE プログラム「ユビキタス統合メディアコンピューティング」、科学研究費補助金・基盤研究(B)「注視情報の伝達に基づくコミュニケーション支援空間の構築(課題番号 16300058)」, NEDO 次世代ロボット実用化プロジェクト(プロトタイプ開発支援事業)「ヒューマノイドロボットのためのインタラクティブミドルウェアの開発」、IPA 未踏ソフトウェア創造事業「ヒューマン・モデリングのための顔情報計測ソフトウェアの開発」、JST 戦略的基礎研究推進事業(CREST)「高度メディア社会の生活情報技術」の支援によるものである。

参 考 文 献

- 1) 中野倫明, 坂口靖雄, 山本 新: 屋外環境における視線検出の一方式, 電子情報通信学会論文誌(D-II), Vol.J79-D-II, No.9, pp.1638-1642 (1996).
- 2) Clarke, A.H.: Current trends in eye movement measurement techniques, *Visual Attention and Cognition*, Zangemeister, W.H., Stiehl, H.S. and Freksa, C. (Eds.), pp.347-364, Elsevier Science B.V. (1996).
- 3) 松田圭司, 永見武司, 山根 茂: 高速楕円近似による汎用視線位置計測システム, 第 10 回生体工学シンポジウム論文集, pp.401-404 (1995).
- 4) Morimoto, C., Koons, D., Amir, A. and Flickner, M.: BlueEyes: Pupil Detection and Tracking Using Multiple Light Sources, *Image and Vision Computing*, Vol.18, No.4, pp.331-334 (2000).
- 5) 伴野 明, 岸野文郎, 小林幸雄: 瞳孔の抽出処理と頭部の動きを許容する視線検出装置の試作, 電子情報通信学会論文誌(D-II), Vol.J76-D-II, No.3, pp.636-646 (1993).
- 6) Ohno, N.T.: A Free-head, Simple Calibration, Gaze Tracking System That Enables Gaze-Based Interaction, *Proc. symposium on ETRA 2004: Eye Tracking Research & Applications Symposium*, pp.115-122 (2004).
- 7) Azarbayejani, A., Starner, T., Horowitz, B. and Pentland, A.: Visually controlled graphics, *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol.15, No.6, pp.602-605 (1993).
- 8) Zelinsky, A. and Heinzmann, J.: Real-time Visual Recognition of Facial Gestures for Human Computer Interaction, *Proc. 2nd Int. Conf. on Automatic Face and Gesture Recognition*, pp.351-356 (1996).
- 9) Davoine, F.F.: Head and Facial Animation Tracking Using Appearance-Adaptive Models and Particle Filters, *Real-Time Vision for Human Computer Interaction*, Kisacanin, G., Pavlovic, V. and Huang, T.S. (Eds.), chapter 8, pp.121-140, Springer Verlag (2005).
- 10) Kruger, V., Happe, A. and Sommer, G.: Affine Real-Time Face Tracking using Gabor Wavelet Networks, *Proc. Int. Conf. on Pattern Recognition*, pp.127-130 (2000).
- 11) Matsumoto, Y. and Zelinsky, A.: An Algorithm for Real-time Stereo Vision Implementation of Head Pose and Gaze Direction Measurement, *Proc. 4th Int. Conf. on Automatic Face and Gesture Recognition (FG'2000)*, pp.499-505 (2000).
- 12) 岡 兼司, 佐藤洋一, 中西泰人, 小池英樹: 適応的拡散制御を伴うパーティクルフィルタを用いた頭部姿勢推定システム, 電子情報通信学会論文誌(D-II), Vol.J88-D-II, No.8, pp.1601-1613 (2005).
- 13) Morency, L.-P. and Darrell, T.: Stereo Tracking using ICP and Normal Flow Constraint, *Proc. International Conference on Pattern Recognition* (2002).
- 14) Matsumoto, Y., Ido, J., Takemura, K., Koeda, M. and Ogasawara, T.: Portable Facial Information Measurement System and Its Application to Human Modeling and Human Interfaces, *Proc. 6th Int. Conf. on Automatic Face and Gesture Recognition*, pp.475-480 (2004).
- 15) Matsumoto, Y., Ogasawara, T. and Zelinsk, A.: Behavior Recognition Based on Head Pose and Gaze Direction Measurement, *Proc. 2000 IEEE/RSJ Intl. Conf. on Intelligent Robots and Systems (IROS'2000)*, pp.2127-2132 (2000).
- 16) 高橋勝彦, 関 進, 小島 浩, 岡 隆一: ジェスチャ動画像のスポッティング認識, 電子情報通信学会論文誌(D-II), Vol.J77-D-II, No.8, pp.1552-1561 (1994).
- 17) Hansen, J.P., Andersen, W. and Roed, P.: Eye-Gaze Control of Multimedia Systems, *Symbiosis of Human and Artifact*, Etal, Y., (Ed.), Elsevier Science B.V. (1995).
- 18) Jacob, R.J.K.: What You Look At Is What You Get: Eye Movement-based Interaction Techniques, *Proc. ACM CHI'90*, pp.11-18 (1990).
- 19) 高木啓伸: 視線を用いたユーザインターフェース開発環境 EyeTk, 情報処理学会 HI・SLP 合同研究報告会 HI65-15 SLP 10-15 (1996).
- 20) 森永英文, 松本吉央, 小笠原司: 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).

- 21) Takemura, K., Minamide, H., Matsumoto, Y. and Ogasawara, T.: What You Look at Is What You Control: A Universal Remote Control Based on Gaze Measurement Technology, *Proc. 1st IEEE Technical Exhibition Based Conference on Robotics and Automation* (2004).
- 22) Shell, J.S., Vertegaal, R. and Skaburskis, A.W.: EyePliances: Attention-Seeking Devices that Respond to Visual Attention, *CHI'03 extended abstracts on Human factors in computing systems*, pp.770-771 (2003).
- 23) 竹村憲太郎, 松本吉央, 小笠原司: ドライブ行動解析のための非侵襲な注視計測システムの開発, 日本機械学会論文集 C 編, Vol.71, No.702, pp.165-170 (2005).
- 24) 松本吉央, 後藤健志, 谷 裕章, 怡土順一, 紙弘和, 小笠原司: 搭乗型ガイドロボット WATSON の開発, 第 8 回ロボティクスシンポジウム予稿集 (2003).
- 25) 足立佳久, 恒成裕行, 松本吉央, 小笠原司: ユーザーの注意測定に基づく搭乗型移動ロボットの走行支援システム, ヒューマンインタフェース学会論文誌, Vol.6, No.3, pp.265-274 (2004).
- 26) Nishimura, R., Uchida, T., Lee, A., Saruwatari, H., Shikano, K. and Matsumoto, Y.: ASKA: Receptionist Robot System with Speech Dialogue System, *Proc. 2002 IEEE/RSJ Int. Conf. on Intelligent Robots and Systems*, pp.1314-1319 (2002).
- 27) Ido, J., Myouga, Y., Matsumoto, Y. and Ogasawara, T.: Interaction of Receptionist ASKA Using Vision and Speech Information, *Proc. International Conference on Multisensor Fusion and Integration for Intelligent Systems*, pp.335-340 (2003).

(平成 18 年 1 月 27 日受付)

(平成 18 年 7 月 21 日採録)

(担当編集委員 杉本 晃宏)



松本 吉央

1998 年東京大学大学院工学系研究科博士課程修了。1998 年オーストラリア国立大学情報工学研究科研究員。1999 年奈良先端科学技術大学院大学情報科学研究科助手。2002 年同助教授。博士(工学)。ロボットビジョン, 移動ロボット, ヒューマンインタフェースに関する研究に従事。



怡土 順一

2006 年 3 月奈良先端科学技術大学院大学情報科学研究科博士課程単位認定退学。2006 年 4 月より奈良先端科学技術大学院大学情報科学研究科研究員。



竹村憲太郎

2006 年 3 月奈良先端科学技術大学院大学情報科学研究科博士課程修了。2004 年ユビキタス統合メディアコンピューティング COE 奨励研究員。2006 年 4 月より奈良工業高等専門学校電気工学科助手。博士(工学)。



小笠原 司(正会員)

1983 年 3 月東京大学大学院工学系研究科博士課程修了。1983 年通商産業省工業技術院電子技術総合研究所入所。1993 年カールスルーエ大学客員研究員。1994 年電子技術総合研究所知能システム部行動知能研究室長。1998 年 4 月より奈良先端科学技術大学院大学情報科学研究科教授。工学博士。